

# Localized/Shrinkage Kriging Predictors

Zeytu Gashaw Asfaw<sup>1</sup> · Henning Omre<sup>2</sup>

Received: 31 January 2015 / Accepted: 9 November 2015 / Published online: 10 December 2015  
© The Author(s) 2015. This article is published with open access at Springerlink.com

**Abstract** The objective of the study is to improve the robustness and flexibility of spatial kriging predictors with respect to deviations from spatial stationarity assumptions. A predictor based on a non-stationary Gaussian random field is defined. The model parameters are inferred in an empirical Bayesian setting, using observations in a local neighborhood and a prior model assessed from the global set of observations. The localized predictor exhibits a shrinkage effect and is termed a localized/shrinkage kriging predictor. The predictor is compared to traditional localized kriging predictors in a case study on observations of annual accumulated precipitation. A cross-validation criterion is used in the comparison. The shrinkage predictor appears as clearly preferable to the traditional kriging predictors. A simulation study on prediction in non-stationary Gaussian random fields is conducted. The results from this study confirm that the shrinkage predictor is preferable to the traditional one. Moreover, the cross-validation criterion is found to be suitable for selection of the local neighborhood in the predictor. Lastly, the computational demands of localized predictors are very modest, hence these localized/shrinkage predictors are suitable for large scale spatial prediction problems.

**Keywords** Spatial statistics · Gaussian random fields · Bayesian inference · Conjugate models

---

✉ Zeytu Gashaw Asfaw  
zeytugashaw@yahoo.com

Henning Omre  
omre@math.ntnu.no

<sup>1</sup> School of Mathematical and Statistical Sciences, Hawassa University, Hawassa, Ethiopia

<sup>2</sup> Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway

## 1 Introduction

Spatial prediction is required in many applications, and examples can be found in natural resource mapping, meteorology and image analysis. Consider a regionalized variable  $\{r(x); x \in \mathbf{D} \subset \mathfrak{R}^m\}$  where  $r(x) \in \mathfrak{R}^1$  is the variable of interest and  $x$  is a reference variable in the domain  $\mathbf{D}$ . The challenge is to predict the regionalized variable from a set of observations  $\mathbf{r}_o = [r(x_1), \dots, r(x_n)]; x_1, \dots, x_n \in \mathbf{D}$ . In the current study, the predictors are defined in a probabilistic setting and associated predictor uncertainties can also be obtained.

The classical probabilistic approach to spatial prediction is kriging (Journel and Huijbregts 1978; Chiles and Delfiner 2009). The traditional ordinary kriging predictor is based on a stationary model for the regionalized variable, with spatially constant expectation and variance, and a translation invariant spatial correlation function. The localized predictors, local neighborhood kriging (Chiles and Delfiner 2009) can be defined to robustify the predictor with respect to deviations from the stationarity assumptions. The major challenge in using localized predictors is to define the size of the local neighborhood where a bias/variance trade-off must be made. For some spatial correlation structures, a screening effect is provided by the observations closest to the prediction location (Stein 2002), and this effect may robustify the localization. A weighted localized approach is also defined (Anderes and Stein 2011) and it is demonstrated to be useful for non-stationary random fields.

Recent developments in computer and sensor technology have provided enormous spatio-temporal sets of observations (Johns et al. 2003). Computational demands in spatial prediction have been a critical factor and considerable research is devoted to numerical algorithms for large sparse correlation matrices. Localized predictors may provide an alternative solution to reduce these computational demands.

In the current study, spatial predictors which are robust with respect to deviations from assumptions about spatial stationarity are defined. These predictors are based on a Gaussian random field with spatially varying expectation and variance. The spatial correlation function is shift invariant and known. The expectation and variance are assessed locally by a sliding window approach. In traditional local neighborhood kriging, this assessment is done by some maximum likelihood procedure. The new feature of the study is that these local assessments are done by shrinkage estimators in an empirical Bayes setting (Efron and Morris 1973). The hierarchical, Gaussian random field model (Røislien and Omre 2006) is used locally and the hyper-parameters are assessed from the global set of observations. Since the predictor is locally defined it is extremely computationally efficient. The resulting spatial predictor is termed localized/shrinkage kriging.

In the two first sections, various Gaussian random field models for the regionalized variable  $\{r(x); x \in \mathbf{D}\}$  are presented with associated model parameter estimators. In the following section, five spatial predictors are defined, one global and four localized. Two of these predictors feature shrinkage. The next section contains an empirical evaluation of these predictors. The evaluation criteria used in the comparison are defined, and the five predictors are compared on a real set of annual accumulated precipitation data and in a synthetic simulation study. In the last section, the conclusions are

forwarded. The paper summarizes the major findings in an extended study (Asfaw 2014).

## 2 Predictor Models

The spatial predictors are based on probabilistic models for the regionalized variable  $\{r(x); x \in \mathbf{D}\}$ . In traditional kriging prediction  $\{r(x); x \in \mathbf{D}\}$  is associated with a stationary Gaussian random field (Chiles and Delfiner 2009). This assumption entails that

$$\begin{aligned} E[r(x)] &= \mu; \quad \forall x \in \mathbf{D} \\ \text{Var}[r(x)] &= \sigma^2; \quad \forall x \in \mathbf{D} \\ \text{Corr}[r(x'), r(x'')] &= \rho(x' - x''); \quad \forall x', x'' \in \mathbf{D} \end{aligned} \tag{1}$$

where the expected level  $\mu \in \mathfrak{R}^1$ , the variance level  $\sigma^2 \in \mathfrak{R}^1_+$  and the spatial correlation function  $\rho(x' - x'')$  is positive definite. Note that for these model assumptions, the random field is shift invariant, and this property is extensively used to make inference about the model parameters  $[\mu, \sigma^2, \rho(\cdot)]$  from the set of observations  $\mathbf{r}_o$ . This traditional kriging model may be extended to have an expectation surface  $\mu(x) = \sum_{l=1}^L \alpha_l g_l(x)$  where  $\{g_l(x); x \in \mathbf{D}\}; l = 1, \dots, L$  are known basis surfaces while  $\alpha = (\alpha_1, \dots, \alpha_L)$  are unknown coefficients. This model corresponds to a spatial regression model, and the shift invariance property is lost, which complicates model parameter inference. Note that for given correlation function  $\rho(\cdot)$ , maximum likelihood estimates based on  $\mathbf{r}_o$  are analytically assessable for the other model parameters, under both these model assumptions.

In the current study, it is assumed that  $\{r(x); x \in \mathbf{D}\}$  is associated with a general Gaussian random field, which entails that

$$\begin{aligned} E[r(x)] &= \mu(x); \quad \forall x \in \mathbf{D} \\ \text{Var}[r(x)] &= \sigma^2(x); \quad \forall x \in \mathbf{D} \\ \text{Corr}[r(x'), r(x'')] &= \rho(x', x''); \quad \forall x', x'' \in \mathbf{D} \end{aligned} \tag{2}$$

with  $\mu(x) \in \mathfrak{R}^1$ ,  $\sigma^2(x) \in \mathfrak{R}^1_+$  and  $\rho(x', x'')$  being positive definite. There is obviously a lack of translation invariance under these assumptions. One can expect that inference of the spatial model parameters  $\{\mu(x); x \in \mathbf{D}\}$ ,  $\{\sigma^2(x); x \in \mathbf{D}\}$  and  $\{\rho(x', x''); \forall x', x'' \in \mathbf{D}\}$  based on  $\mathbf{r}_o$  is complicated. If the correlation function  $\rho(\cdot, \cdot)$  is fixed, one may use localized estimators in a kernel spirit to assess the spatial model parameters  $\mu(\cdot)$  and  $\sigma^2(\cdot)$ . When selecting the size of the localization, a bias/variance trade-off must be made large local neighborhoods introduce bias in the estimators due to smoothing while small neighborhoods introduce instability in the estimators due to censoring of observations. In the current study, this bias/variance trade-off is addressed by defining shrinkage estimators in an empirical Bayes setting.

To define the shrinkage estimators, a stationary, hierarchical Gaussian random field model is introduced for a local neighbourhood  $\mathbf{D}_+$  around an arbitrary  $x_+ \in \mathbf{D}$ , i.e

for  $\{r(x); x \in \mathbf{D}_+\}$ . In this model the expected level  $m$  and the variance level  $s^2$  are considered to be random variables. Moreover,  $\{[r(x)|m, s^2]; x \in \mathbf{D}_+\}$  is defined to be a stationary Gaussian random field, hence

$$\begin{aligned} E[r(x)|m, s^2] &= m; \quad \forall x \in \mathbf{D}_+ \\ \text{Var}[r(x)|m, s^2] &= s^2; \quad \forall x \in \mathbf{D}_+ \\ \text{Corr}[r(x'), r(x'')|m, s^2] &= \rho(x' - x''); \quad \forall x', x'' \in \mathbf{D}_+. \end{aligned} \tag{3}$$

It is assumed that  $\rho(\cdot)$  is known, and that the model parameters  $[m|s^2]$  and  $s^2$  have prior models which are Gaussian and Inverse Gamma respectively, with

$$\begin{aligned} E[m|s^2] &= \mu_m \\ \text{Var}[m|s^2] &= \tau_m s^2 \\ E[s^2] &= [\xi_s - 1]^{-1} \gamma_s \\ \text{Var}[s^2] &= [(\xi_s - 1)^2 (\xi_s - 2)]^{-1} \gamma_s^2 \end{aligned} \tag{4}$$

where the model parameters are  $\mu_m \in \mathfrak{R}^1$ ,  $\tau_m \in \mathfrak{R}^1_+$ ,  $\xi_s \in 2 + \mathfrak{R}^1_+$  and  $\gamma_s \in \mathfrak{R}^1_+$ . The prior model on  $[m, s^2]$  is a conjugate model for the stationary Gaussian random field, and the marginal random field  $\{r(x); x \in \mathbf{D}_+\}$  will be a t-distributed random field and analytically tractable (Røislien and Omre 2006).

The estimators for the model parameters  $\mu(x_+)$  and  $\sigma^2(x_+)$  at an arbitrary location  $x_+ \in \mathbf{D}$  are localized to observations in  $\mathbf{D}_+$  and with shrinkage according to the localized hierarchical Gaussian model. These estimators will of course depend on the parameters of the prior model  $[\mu_m, \tau_m, \xi_s, \gamma_s]$ , which are assessed in an empirical Bayesian spirit from a set of local neighborhood  $\mathbf{D}_+$  covering  $\mathbf{D}$ .

Recent research on non-stationary Gaussian random fields (Higdon 1998), is usually based on models of the following form

$$\begin{aligned} E[r(x)] &= \mu; \quad \forall x \in \mathbf{D} \\ \text{Cov}[r(x'), r(x'')] &= \kappa(x', x''); \quad \forall x', x'' \in \mathbf{D} \end{aligned}$$

with stationary, shift invariant expectation  $\mu$  and non-stationary spatial covariance function  $\kappa(\cdot, \cdot)$ . The latter must be a positive definite function which complicates the model parametrization. The major challenge is, however, to make inference of the model parameters based on one set of observations from one realization of the random field. The current model defined in Eq. (2), can be cast in the framework above

$$\begin{aligned} E[r(x)] &= \mu; \quad \forall x \in \mathbf{D} \\ \text{Cov}[r(x'), r(x'')] &= \kappa(x', x'') \\ &= \sigma(x')\sigma(x'')\rho(x', x'') + (\mu(x') - \mu)(\mu(x'') - \mu) \quad \forall x', x'' \in \mathbf{D} \end{aligned}$$

with global centering value  $\mu$  and gross covariance function  $\kappa(\cdot, \cdot)$ . The latter will be a positive definite function whenever  $\rho(\cdot, \cdot)$  is a positive definite function, for arbitrary functions  $\mu(\cdot) \in \mathfrak{R}^1$  and  $\sigma(\cdot) \in \mathfrak{R}^1_+$  on  $\mathbf{D}$ . Note, in particular that for stationary,

translation invariant  $\rho(x', x'') = \rho(x' - x'')$  the corresponding  $\kappa(\cdot, \cdot)$  will still be non-stationary. One may consider the model in Eq. (2), to be a flexible parametrization of  $\kappa(\cdot, \cdot)$  in a non-stationary Gaussian random field. In the current study, localized, robust estimators for  $\mu(x)$  and  $\sigma(x)$  at arbitrary  $x \in \mathbf{D}$  are presented, given a stationary correlation function  $\rho(\cdot)$ , based on one set of observations from one realization. Consequently, inference of the non-stationary  $\kappa(\cdot, \cdot)$  under the parametrization discussed above can be made.

These developments are also valid for non-stationary  $\rho(\cdot, \cdot)$ , but this extension introduces new challenges in assessing  $\rho(\cdot, \cdot)$ . These challenges are not addressed in the current study.

### 3 Inference of Model Parameters

To use the probabilistic models for  $\{r(x); x \in \mathbf{D}\}$  in spatial prediction, the model parameters must be assessed from the set of observations represented by the  $n$ -vector  $\mathbf{r}_o = [r(x_1), \dots, r(x_n)]^T = [r_1, \dots, r_n]^T$ .

Throughout the study, the spatial correlation function  $\{\rho(x', x'') = \rho(x' - x''); \forall x', x'' \in \mathbf{D}\}$  is assumed to be known, with associated inter-observation correlation  $[n \times n]$ -matrix  $\Omega_{oo}$ . This correlation model must of course be inferred in one way or the other, but in the current study the uncertainty related to this is not accounted for. The other model parameters will be estimated conditional on this correlation function, and hence capture the remaining spatial structure in the observations.

For a stationary Gaussian random field, given the correlation function, the model parameters can be assessed by a maximum likelihood estimator

$$\begin{aligned} \hat{\mu} &= [\mathbf{i}_n^T \Omega_{oo}^{-1} \mathbf{i}_n]^{-1} [\mathbf{i}_n^T \Omega_{oo}^{-1} \mathbf{r}_o] \\ \hat{\sigma}^2 &= \frac{1}{n} [\mathbf{r}_o - \hat{\mu} \mathbf{i}_n]^T \Omega_{oo}^{-1} [\mathbf{r}_o - \hat{\mu} \mathbf{i}_n] \end{aligned} \tag{5}$$

where  $\mathbf{i}_n$  is the  $n$ -vector  $[1, \dots, 1]^T$ .

For a general Gaussian random field, given the correlation function, the assessments of the spatial model parameters  $\{\mu(x); x \in \mathbf{D}\}$  and  $\{\sigma^2(x); x \in \mathbf{D}\}$ , are more complicated. To define localized estimators, consider an arbitrary location  $x_+ \in \mathbf{D}$  and parameterize the localization by the  $k$  observations in  $\mathbf{r}_o$  localized closest to  $x_+$ , and denote this as  $k$ -localization. Define a binary-selection  $[k \times n]$ -matrix  $G_+^k$  such that  $G_+^k \mathbf{r}_o$  is a  $k$ -vector containing the  $k$  observations in the  $k$ -localization of  $x_+$ . Note that  $G_+^k$  can also be extended to account for favorable configurations of observations around  $x_+$ . The  $k$ -localized maximum likelihood estimators for the model parameters are

$$\begin{aligned} \hat{\mu}_+^k &= \hat{\mu}^k(x_+) = [\mathbf{i}_k^T [G_+^k \Omega_{oo} [G_+^k]^T]^{-1} G_+^k \mathbf{r}_o] [\mathbf{i}_k^T [G_+^k \Omega_{oo} [G_+^k]^T]^{-1} \mathbf{i}_k]^{-1} \\ \hat{\sigma}_+^{k2} &= \hat{\sigma}^{k2}(x_+) = \frac{1}{k} [G_+^k \mathbf{r}_o - \hat{\mu}_+^k \mathbf{i}_k]^T [G_+^k \Omega_{oo} [G_+^k]^T]^{-1} (G_+^k \mathbf{r}_o - \hat{\mu}_+^k \mathbf{i}_k). \end{aligned} \tag{6}$$

When  $x_+$  coincide with the observation locations  $[x_1, \dots, x_n]$  this produces estimators for the observation expectation  $n$ -vector, and the diagonal standard deviation  $[n \times n]$ -matrix

$$\hat{\boldsymbol{\mu}}_o^k = [\hat{\mu}_1^k, \dots, \hat{\mu}_n^k]^T$$

$$\hat{\Gamma}_o^k = \begin{bmatrix} \hat{\sigma}_1^k & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \hat{\sigma}_n^k \end{bmatrix}. \tag{7}$$

For a stationary, hierarchical Gaussian random field, with given correlation function, inference of the model parameters of the prior model  $[\mu_m, \tau_m, \xi_s, \gamma_s]$  is required. This assessment is made in an empirical Bayes setting, by considering the localized estimates in the observation locations,  $[\hat{\mu}_i^k, \hat{\sigma}_i^k]; i = 1, \dots, n$ , to be a super-population of the  $k$ -localized estimate in an arbitrary location. Moment estimators are used, based on the moment expressions defined in Eq. (4). Then natural estimators for the parameters of the prior model are dependent on  $k$ , and given by

$$\hat{\mu}_m^k = \frac{1}{n} \mathbf{i}_n^T \hat{\boldsymbol{\mu}}_o^k$$

$$\hat{\tau}_m^k = [\hat{\sigma}_m^{k2}]^{-1} \hat{\sigma}_m^{k2} \tag{8}$$

where

$$\hat{\sigma}_m^{k2} = \frac{1}{n} Tr[[\hat{\Gamma}_o^k]^2]$$

$$\hat{\sigma}_m^{k2} = \frac{1}{n} [\hat{\boldsymbol{\mu}}_o^k - \hat{\mu}_m^k \mathbf{i}_n]^T [\hat{\boldsymbol{\mu}}_o^k - \hat{\mu}_m^k \mathbf{i}_n]$$

and

$$\hat{\xi}_s^k = [\hat{\sigma}_s^{k2}]^{-1} [\hat{\mu}_s^{k2}] + 2$$

$$\hat{\gamma}_s^k = \hat{\mu}_s^k [[\hat{\sigma}_s^{k2}]^{-1} [\hat{\mu}_s^{k2}] + 1] \tag{9}$$

where

$$\hat{\mu}_s^k = \frac{1}{n} \mathbf{i}_n^T \mathbf{s}^2$$

$$\hat{\sigma}_s^{k2} = \frac{1}{n} [\mathbf{s}^2 - \hat{\mu}_s^k \mathbf{i}_n]^T [\mathbf{s}^2 - \hat{\mu}_s^k \mathbf{i}_n]$$

$$\mathbf{s}^2 = [(r_1 - \hat{\mu}_m^k)^2, \dots, (r_n - \hat{\mu}_m^k)^2]^T.$$

Estimators for all model parameters are now defined based on the observations  $\mathbf{r}_o$ . Hence all probabilistic models for the regionalized variable  $\{r(x); x \in \mathbf{D}\}$  are fully specified. The focus is on spatial prediction however, and in the following section spatial predictors are specified under the various model assumptions. The estimators for the model parameters can be inserted to obtain operable predictors.

### 4 Spatial Predictors

The focus of this study is on spatial prediction in a random field  $\{r(x); x \in \mathbf{D}\}$  based on a set of observations represented in a  $n$ -vector  $\mathbf{r}_o$ . Consider an arbitrary location  $x_+ \in \mathbf{D}$  with value  $r(x_+) = r_+$ . The challenge is to provide a reliable predictor for  $r_+$  based on  $\mathbf{r}_o$ . By using a squared error loss the predictor is  $\hat{r}_+ = \widehat{E}[r_+|r_o]$ , with associated estimated predictor variance  $\hat{\sigma}_+^2 = \widehat{\text{Var}}[r_+|r_o]$ . Note that a predictor for the entire regionalized variable  $\{r(x); x \in \mathbf{D}\}$  can be obtained by letting  $x_+$  run over the domain  $\mathbf{D}$ .

Recall that the correlation function  $\{\rho(x' - x''); x', x'' \in \mathbf{D}\}$  is assumed known and that the inter observation correlation  $[n \times n]$ -matrix is denoted  $\Omega_{oo}$ , while the observation to  $x_+$  correlation  $n$ -vector is denoted  $\omega_{o+}$ . Recall also that the localization operator  $G_+^k$  is defined such that  $G_+^k \mathbf{r}_o$  is an observation  $k$ -vector which contain the  $k$  observations located closest to  $x_+$ .

#### 4.1 Glob/Stat/Trad Predictor

This predictor is global and based on a stationary Gaussian random field model with traditional parameter estimates. It corresponds to the frequently used global ordinary kriging predictor, and it is defined by

$$[r|\mathbf{r}_o] \sim \text{Gauss}[\mu_{+|o}, \sigma_{+|o}^2]$$

with

$$\begin{aligned} \mu_{+|o} &= \mu + \omega_{o+}^T \Omega_{oo}^{-1} [\mathbf{r}_o - \mu \mathbf{i}_n] \\ \sigma_{+|o}^2 &= \sigma^2 [1 - \omega_{o+}^T \Omega_{oo}^{-1} \omega_{o+}]. \end{aligned} \tag{10}$$

Note that the predictor is independent of the variance  $\sigma^2$  while the prediction variance is independent of the observed values  $\mathbf{r}_o$ . The latter is only dependent on the location configuration of  $\mathbf{r}_o$ . These are well-known characteristics of kriging.

The Glob/Stat/Trad predictor with associated predictor variance is defined as

$$\begin{aligned} \hat{r}_{\text{GST}} &= \hat{\mu}_{+|o} \\ \hat{\sigma}_{\text{GST}}^2 &= \hat{\sigma}_{+|o}^2 \end{aligned}$$

which are defined by Eq. (10) with the estimates in Eq. (5) inserted.

#### 4.2 Loc/Stat/Trad Predictor

This predictor is  $k$ -localized and based on a stationary Gaussian random field model with traditional parameter estimators. It corresponds to a localized ordinary kriging predictor which is frequently used in practice, and is defined by

$$[r_+ | G_+^k \mathbf{r}_o] \sim \text{Gauss}[\mu_{+|o}^k, \sigma_{+|o}^{k2}]$$

with

$$\begin{aligned} \mu_{+|o}^k &= \mu_+^k + [G_+^k \omega_{o+}]^T [G_+^k \Omega_{oo} [G_+^k]^T]^{-1} [G_+^k \mathbf{r}_o - \mu_+^k \mathbf{i}_k] \\ \sigma_{+|o}^{k2} &= \sigma_+^{k2} [1 - [G_+^k \omega_{o+}]^T [G_+^k \Omega_{oo} [G_+^k]^T]^{-1} G_+^k \omega_{o+}]. \end{aligned} \tag{11}$$

This predictor is locally independent on  $\sigma^2$  with predictor variance locally independent of the observed values  $\mathbf{r}_o$ . Since both  $\mu_+^k$  and  $\sigma_+^{k2}$  will vary across the field, the predictor and predictor variance will vary across the field as well.

The Loc/Stat/Trad predictor with associated predictor variance is defined as

$$\begin{aligned} \hat{r}_{\text{LST}}^k &= \hat{\mu}_{+|o}^k \\ \hat{\sigma}_{\text{LST}}^{k2} &= \hat{\sigma}_{+|o}^{k2} \end{aligned}$$

which are defined by Eq. (11) with the estimates in Eq. (6) inserted.

### 4.3 Loc/Stat/Shr Predictor

This predictor is  $k$ -localized and based on a stationary Gaussian random field model with shrinkage parameter estimators defined in an empirical Bayes setting. The predictor is termed the stationary localized/shrinkage kriging predictor and constitutes a new predictor in the study

$$[r_+ | m_+^k, s_+^{k2}, G_+^k \mathbf{r}_o] \sim \text{Gauss}[\mu_{+|o}^k, \sigma_{+|o}^{k2}]$$

with

$$\begin{aligned} \mu_{+|o}^k &= m_+^k + [G_+^k \omega_{o+}]^T [G_+^k \Omega_{oo} [G_+^k]^T]^{-1} [G_+^k \mathbf{r}_o - m_+^k \mathbf{i}_k] \\ \sigma_{+|o}^{k2} &= s_+^{k2} [1 - [G_+^k \omega_{o+}]^T [G_+^k \Omega_{oo} [G_+^k]^T]^{-1} G_+^k \omega_{o+}] \end{aligned} \tag{12}$$

where the posterior expectations for the hyper-parameters are (“Appendix”)

$$\begin{aligned} m_+^k &= E[m | s^2, G_+^k \mathbf{r}_o] \\ &= \mu_m^k + \tau_m^k s^T [\tau_m^k \mathbf{i}_k \mathbf{i}_k^T + [G_+^k \Omega_{oo} [G_+^k]^T]^{-1} [G_+^k \mathbf{r}_o - \mu_m^k \mathbf{i}_k] \\ s_+^{k2} &= E[s_+^{k2} | G_+^k \mathbf{r}_o] = \left[ \xi_s^k + \frac{k}{2} - 1 \right]^{-1} \\ &\times \left[ \gamma_s^k + \frac{1}{2} \left[ [G_+^k \mathbf{r}_o - \mu_m^k \mathbf{i}_k]^T [[G_+^k \Omega_{oo} [G_+^k]^T] + \tau_m^k \mathbf{i}_k \mathbf{i}_k^T]^{-1} [G_+^k \mathbf{r}_o - \mu_m^k \mathbf{i}_k] \right] \right]. \end{aligned} \tag{13}$$

Equation (12) follows from the definition of a stationary hierarchical Gaussian random field conditioned on the model parameters and on  $G_+^k \mathbf{r}_o$ . Equation (13) follows from



the prior model of  $[m, s^2]$  being conjugate, hence the posterior models are analytically accessible, and so are their expectations. Note in particular that  $m_+^k = E[m|G_+^k \mathbf{r}_o]$  hence independent of  $s_+^{k2}$ . Note also that the predictor is independent of the variance while the predictor variance is dependent on the actual observed values.

The predictor exhibits shrinkage behaviour through the estimators for shift and scaling parameters  $m_+^k$  and  $s_+^{k2}$ . The actual observation weights are independent of  $m_+^k$  and  $s_+^{k2}$ .

The Loc/Stat/Shr predictor with associate predictor variance is defined as

$$\begin{aligned} \hat{r}_{\text{LSS}}^k &= \hat{\mu}_{+|o}^k \\ \hat{\sigma}_{\text{LSS}}^{k2} &= \hat{\sigma}_{+|o}^{k2} \end{aligned}$$

which are defined by Eqs. (12) and (13) with the estimators in Eqs. (8) and (9) inserted.

### 4.4 Loc/Non-stat/Trad Predictor

This predictor is  $k$ -localized and based on a general Gaussian random field model with traditional parameter estimators. This predictor is surprisingly seldom used, and it is defined as

$$[r_+ | G_+^k \mathbf{r}_o] \sim \text{Gauss}[\mu_{+|o}^k, \sigma_{+|o}^{k2}]$$

with

$$\begin{aligned} \mu_{+|o}^k &= \mu_+^k + \sigma_+^k [G_+^k \Gamma_o^k \omega_{o+}]^T [G_+^k \Gamma_o^k \Omega_{oo} \Gamma_o^k [G_+^k]^T]^{-1} [G_+^k [\mathbf{r}_o - \boldsymbol{\mu}_o^k]] \\ \sigma_{+|o}^{k2} &= \sigma_+^{k2} [1 - [G_+^k \Gamma_o^k \omega_{o+}]^T [G_+^k \Gamma_o^k \Omega_{oo} \Gamma_o^k [G_+^k]^T]^{-1} G_+^k \Gamma_o^k \omega_{o+}] \end{aligned} \tag{14}$$

where

$$\begin{aligned} \boldsymbol{\mu}_o^k &= [\mu_1^k, \dots, \mu_n^k]^T \\ \Gamma_o^k &= \begin{bmatrix} \sigma_1^k & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_n^k \end{bmatrix}. \end{aligned}$$

Note that this predictor is dependent on the variance variability across the field while the predictor variance is independent of the observation values.

The Loc/Non-stat/Trad predictor with associated predictor variance are defined as

$$\begin{aligned} \hat{r}_{\text{LNT}}^k &= \hat{\mu}_{+|o}^k \\ \hat{\sigma}_{\text{LNT}}^{k2} &= \hat{\sigma}_{+|o}^{k2} \end{aligned}$$

which are defined by Eq. (14) with the estimates in Eqs. (6) and (7) inserted.

### 4.5 Loc/Non-stat/Shr Predictor

This predictor is  $k$ -localized and based on a non-stationary Gaussian random field model with shrinkage parameter estimators defined in an empirical Bayesian setting. This predictor is termed the non-stationary localized/shrinkage kriging predictor and constitutes another new predictor in this study

$$[r_+ | m_+^k, \mathbf{m}_o^k, s_+^{k2}, S_o^k, G_+^k \mathbf{r}_o] \sim \text{Gauss}[\mu_{+|o}^k, \sigma_{+|o}^{k2}]$$

with

$$\begin{aligned} \mu_{+|o}^k &= m_+^k + s_+^k [G_+^k S_o^k \omega_{o+}]^T [G_+^k S_o^k \Omega_{oo} S_o^k [G_+^k]^T]^{-1} [G_+^k [\mathbf{r}_o - \mathbf{m}_o^k]] \\ \sigma_{+|o}^{k2} &= s_+^{k2} [1 - [G_+^k S_o^k \omega_{o+}]^T [G_+^k S_o^k \Omega_{oo} S_o^k [G_+^k]^T]^{-1} G_+^k S_o^k \omega_{o+}] \end{aligned} \tag{15}$$

where  $m_+^k$  and  $s_+^{k2}$  are defined as in Eq. (13) while

$$\begin{aligned} \mathbf{m}_o^k &= [m_1^k, \dots, m_n^k]^T \\ S_o^k &= \begin{bmatrix} s_1^k & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & s_n^k \end{bmatrix} \end{aligned}$$

are defined from the  $x_+$ -centered  $m_+$  and  $s_+^{k2}$  shifted to the observation locations  $[x_1, \dots, x_n]$ .

The predictor exhibits shrinkage through the estimators of  $m_+^k, s_+^{k2}$  and  $m_i^k, s_i^{k2}; i = 1, \dots, n$ . Hence both shift and scaling, as well as observation weights are influenced by the shrinkage effect. This is a full shrinkage predictor.

The Loc/Non-stat/Shr predictor with associated predictor variance are defined as

$$\begin{aligned} \hat{r}_{\text{LNS}}^k &= \hat{\mu}_{+|o}^k \\ \hat{\sigma}_{\text{LNS}}^{k2} &= \hat{\sigma}_{+|o}^{k2} \end{aligned}$$

which are defined by Eq. (15) with the estimates in Eqs. (8) and (9) inserted.

### 4.6 Cross-Validation Calibrated (CVC) Predictors

Five predictors are defined, one global and four localized. One challenge with localized predictors is lack of global anchoring. The variance estimates are localized and coupled with the expectation estimates which may cause incorrect scaling of the prediction variances. To account for this cross-validation calibrated (CVC) predictors are introduced to provide a global calibration.

Consider an arbitrary predictor in an arbitrary location  $x_+ \in \mathbf{D}, \hat{r}_+ = \hat{E}[r_+ | \mathbf{r}_o]$ . The predictor may be global or localized, and it is based on the observations  $\mathbf{r}_o =$

$[r_1, \dots, r_n]^T$  in locations  $[x_1, \dots, x_n]$ . The associated prediction variance is  $\hat{\sigma}_+^2 = \widehat{\text{Var}}[r_+ | \mathbf{r}_o]$ .

Define the cross-validation predictors with associated predictor variances in the observation locations  $[x_1, \dots, x_n]$

$$\begin{aligned} \hat{r}_i &= \widehat{E}[r_i | \mathbf{r}_{o(-i)}]; \quad i = 1, \dots, n \\ \hat{\sigma}_i^2 &= \widehat{\text{Var}}[r_i | \mathbf{r}_{o(-i)}]; \quad i = 1, \dots, n \end{aligned}$$

where  $\mathbf{r}_{o(-i)}$  represents observations  $\mathbf{r}_o$  with observations  $r_i$  removed.

The normalized cross-validation errors are defined as

$$e_i = [\hat{\sigma}_i]^{-1}[r_i - \hat{r}_i]; \quad i = 1, \dots, n.$$

Under a fully specified model, these errors will be centered at zero and scaled to unity. Consider the estimators

$$\begin{aligned} \hat{\mu}_e &= \frac{1}{n} \sum_{i=1}^n e_i \\ \hat{\sigma}_e^2 &= \frac{1}{n} \sum_{i=1}^n [e_i - \hat{\mu}_e]^2 \end{aligned}$$

so the mean normalized error (MNE)  $\hat{\mu}_e$ , and mean square normalized error (MSNE)  $\hat{\sigma}_e^2$ , should be close to zero and unity respectively.

The CVC predictor and CVC prediction variance are defined in an arbitrary location  $x_+ \in \mathbf{D}$  by

$$\begin{aligned} \tilde{r}_+ &= \hat{r}_+ + \hat{\sigma}_+ \hat{\mu}_e \\ \tilde{\sigma}_+^2 &= \hat{\sigma}_e^2 \hat{\sigma}_+^2. \end{aligned} \tag{16}$$

Note that the corresponding normalized cross-validation errors will have  $\hat{\mu}_e$  and  $\hat{\sigma}_e^2$  which are identical to zero and unity, respectively. These CVC predictors with associated CVC prediction variances will be used in the following study.

## 5 Empirical Evaluation

It is difficult to compare the various predictors analytically, hence an empirical evaluation is conducted. First suitable evaluation criteria are defined, thereafter two studies are described and the results are summarized.

### 5.1 Evaluation Criteria

Several spatial CVC predictors with associated CVC prediction variances are defined. Recall that all predictors have normalized cross-validation errors centered at zero and scaled to unity.

The precision of the CVC predictors  $\tilde{r}_+$  measured by mean square cross-validation error

$$\text{PMSE} = \frac{1}{n} \sum_{i=1}^n [r_i - \tilde{r}_i]^2 \quad (17)$$

may vary, however. In the CVC predictor the scale of the normalized cross-validation error is identical to unity but large deviations may be reduced by large prediction variances in this measure. The PMSE value is used as measure for precision in the CVC predictor  $\tilde{r}_+$  and small values of PMSE are preferable, of course.

The precision of the CVC prediction variances  $\tilde{\sigma}_+^2$  is indicated by the dependence between cross-validation squared errors  $[r_i - \tilde{r}_i]^2$  and corresponding prediction variances  $\tilde{\sigma}_i^2$ . Since these are variance estimates the mean square measure is defined as

$$\text{VMSE} = \frac{1}{n} \sum_{i=1}^n \left[ \frac{[r_i - \tilde{r}_i]^2}{\tilde{\sigma}_i^2} - 1 \right]^2. \quad (18)$$

Recall that normalized cross-validation squared error in this expression is centered exactly at unity. The VMSE value is used as measure for precision in the CVC prediction variance  $\tilde{\sigma}_+^2$  and small values of VMSE are preferable.

By comparing values of PMSE and VMSE for different CVC predictors one can evaluate the relative quality of the predictors. Normally, the criterion PMSE is considered more important than the criterion VMSE.

## 5.2 Empirical Studies

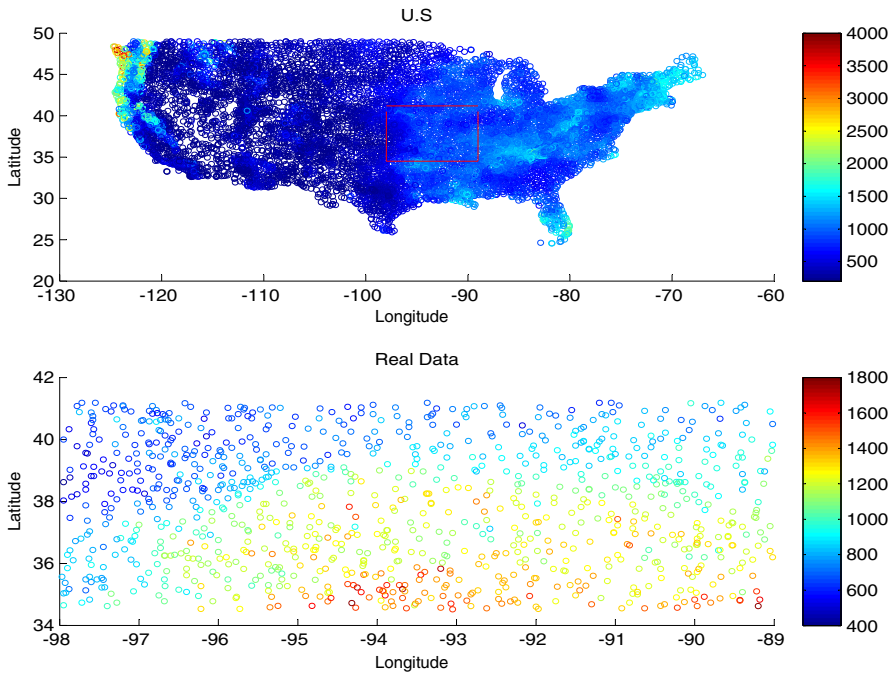
Two cases are presented. The first is based on a set of observations of yearly accumulated precipitation and the second on a simulation study of a general Gaussian random field.

### 5.2.1 US Precipitation Study

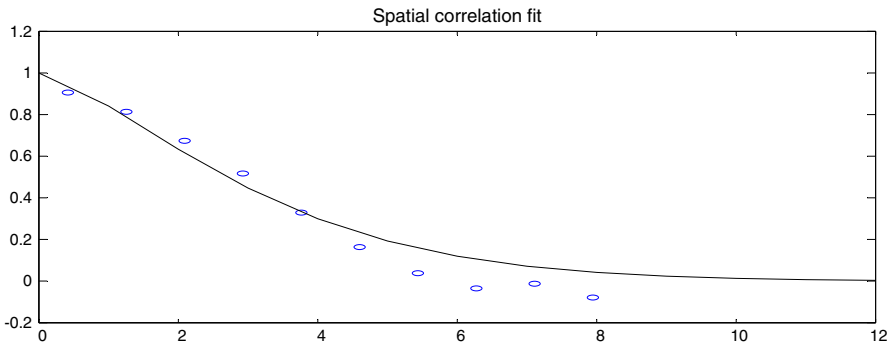
The data set consists of observations of accumulated precipitation from 1997 in 1001 locations in an area of the US (Fig. 1). The observations are a subset of a much larger data set (Johns et al. 2003).

By inspecting the observations in Fig. 1, one sees relatively dense, uniform coverage of observations with a slight south-eastern trend in the values. The translation invariant correlation function  $\rho(\cdot)$  is inferred from correlation values computed at regular shifts assuming translation invariant expectations and variances. These correlations (Fig. 2) constitutes some sort of spatial average values. The following model is fitted

$$\rho(\tau) = \exp \left\{ - \left[ \frac{\tau}{3.5} \right]^{1.4} \right\}; \quad \tau \geq 0$$



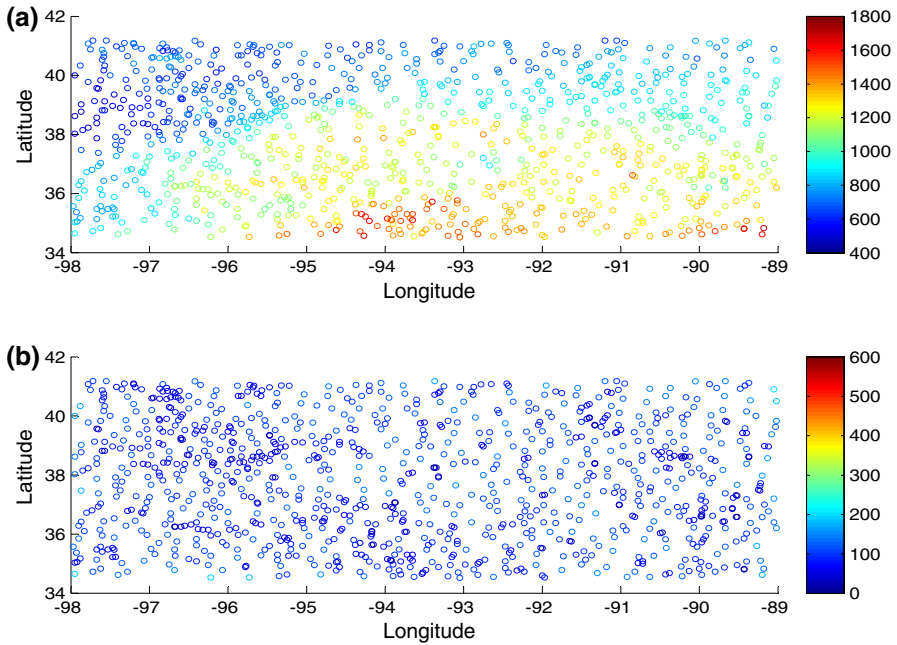
**Fig. 1** The 1997 accumulated precipitation observations in the US with sub-area to be studied (*top*). The 1,001 observations used in the study (*bottom*)



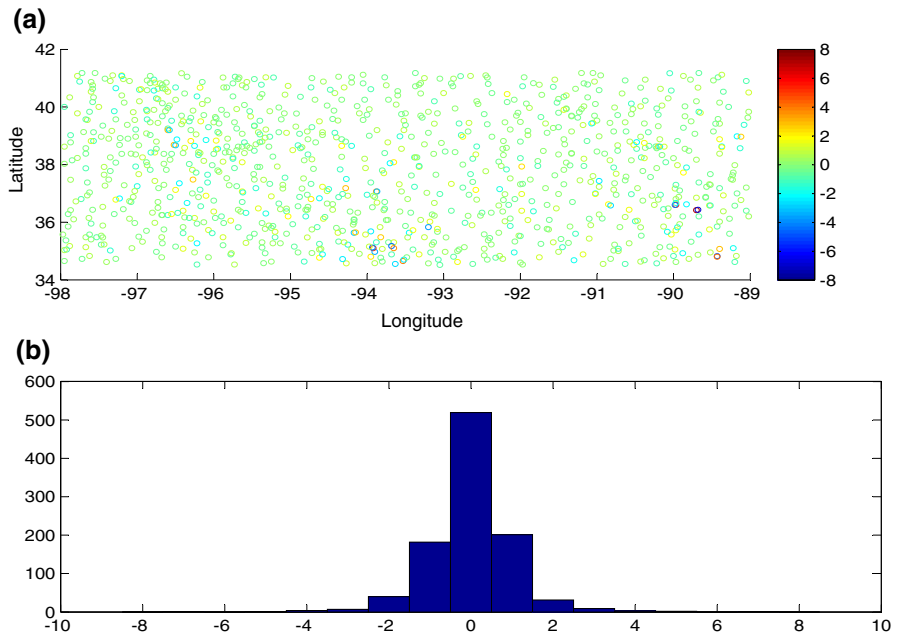
**Fig. 2** Spatial correlation function used with estimated correlation values

with  $\tau = |x' - x''|$ . This spatial correlation model is used throughout the study and the estimates of the model parameters  $\mu(\cdot)$  and  $\sigma^2(\cdot)$  are obtained conditional on this correlation model. The estimators and the localized predictors require that  $k$ , the number of observations in the neighborhood, is specified. After a small preliminary study, the value  $k = 10$  is chosen.

The comparison is made between the five alternative CVC predictors defined in Sect. 4 using the evaluation criteria defined in Sect. 5. The results from the evaluation are displayed in Figs. 3, 4, 5, 6, 7, 8, 9 and Table 1.



**Fig. 3** Glob/Stat/Trad CVC predictor-ordinary kriging **a** Predictions, **b** Prediction standard deviations



**Fig. 4** Glob/Stat/Trad cross-validation errors **a** Normalized cross-validation errors, **b** Normalized cross-validation error histogram

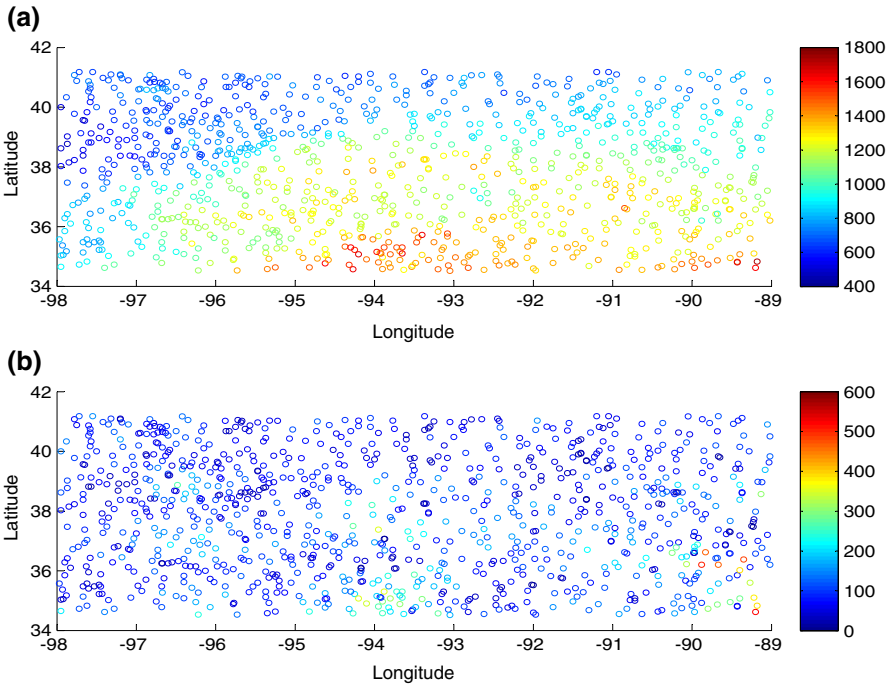


Fig. 5 Loc/Stat/Trad  $k = 10$  CVC predictor **a** Predictions, **b** Prediction standard deviations

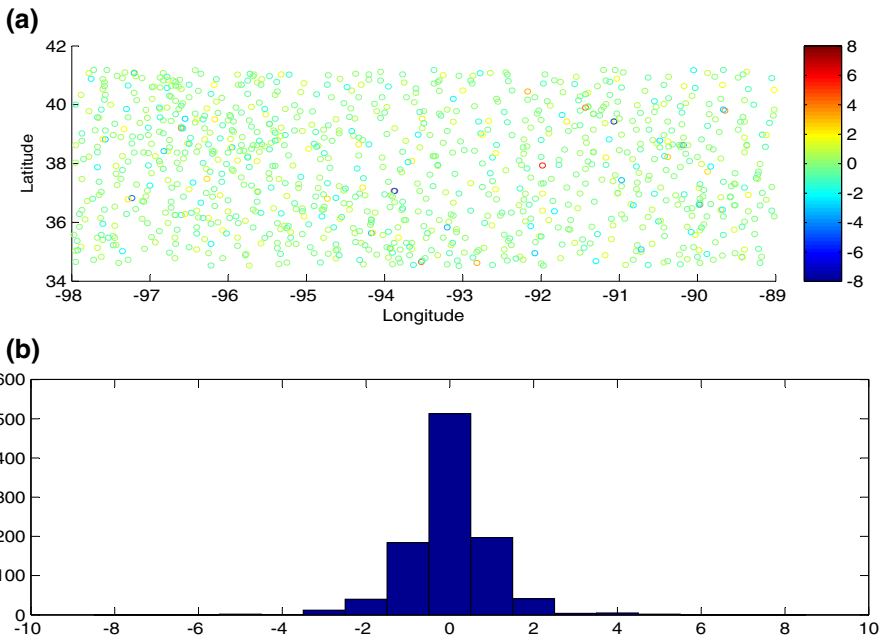


Fig. 6 Loc/Stat/Trad  $k = 10$  cross-validation errors **a** Normalized cross-validation errors, **b** Normalized cross-validation error histogram

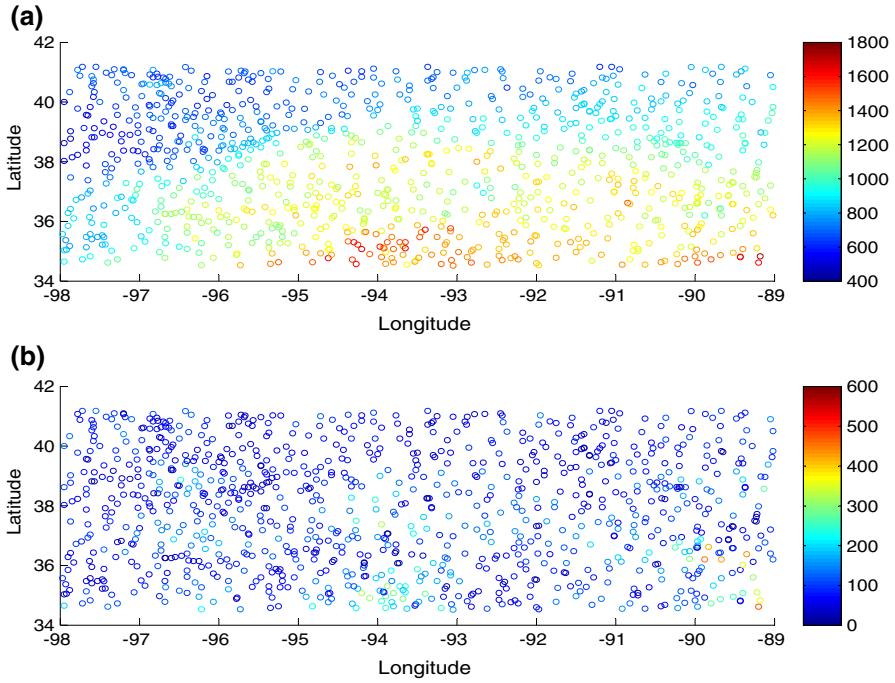


Fig. 7 Loc/Stat/Shr  $k = 10$  CVC predictor **a** Predictions, **b** Prediction standard deviations

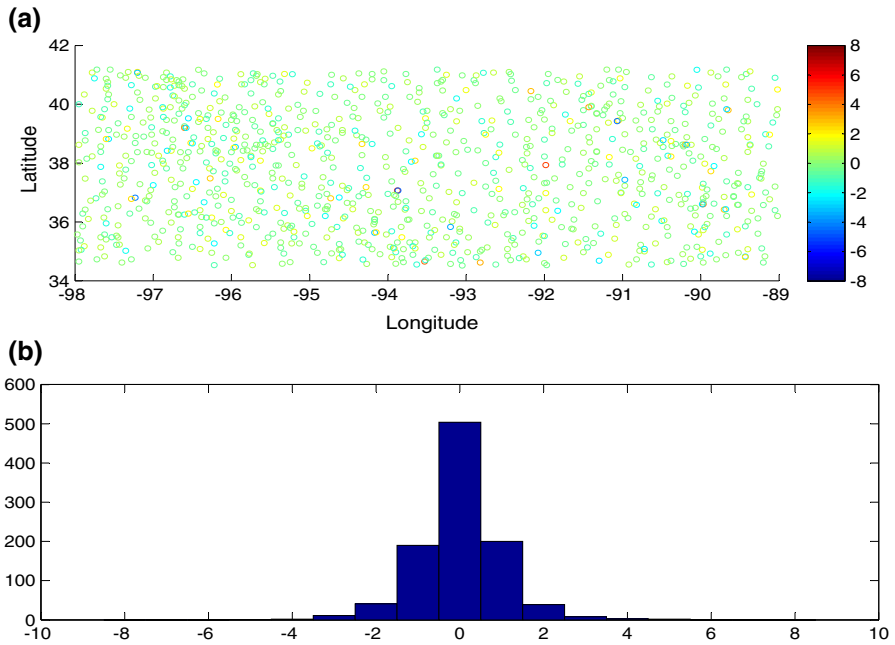
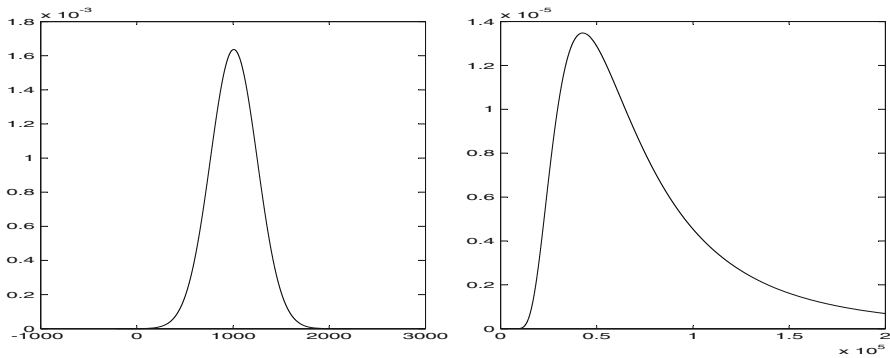


Fig. 8 Loc/Stat/Shr  $k = 10$  cross-validation errors **a** Normalized cross-validation errors, **b** Normalized cross-validation error histogram





**Fig. 9** US precipitation study. Priors model for expectation and variance

**Table 1** US precipitation study

Model	Localized/stationary			Localized/non-stationary	
	Traditional		Shrinkage	Traditional	Shrinkage
Test D	$k = 1000$	$k = 10$	$k = 10$	$k = 10$	$k = 10$
MNE	1.0148e-17	-5.3792e-18	5.9892e-17	1.5084e-16	-1.1579e-16
MSNE	1.5399	2.9740	3.3928	9.2487	5.1702
PMSE	6.8758e+03	6.8745e+03	6.8654e+03	2.2181e+04	9.3555e+03
VMSE	9.8749	5.9746	5.2027	4.0787	4.3475

Evaluation criteria: *MNE* mean normalized error, *MSNE* mean square normalized error, *PMSE* prediction mean squared error, *VMSE* variance mean squared error

The results from the Glob/Stat/Trad CVC predictor are displayed in Figs. 3, 4 and Table 1. Figure 3 displays the actual cross-validation predictions and cross-validation prediction standard deviations in the observations locations. The predictions can be compared to the actual observations in Fig. 1, and no dramatic deviations are seen since the observation coverage is dense. Figure 4 displays the normalized cross-validation errors both spatially and as a histogram. Note that the histogram is centered at zero and scaled to unity since the CVC predictor is used. The locations of large errors seem to fall in the south-eastern corner where the trend effect is largest. The values of the evaluation criteria are listed in Table 1, first column ( $k = 1000$ ). The two first lines MNE and MSNE contain the values of  $\hat{\mu}_e$  and  $\hat{\sigma}_e^2$  respectively, hence the empirical moments of the normalized cross-validation errors of the non-calibrated predictor. The predictor appears as well centered but with downward biased prediction variances. The two next lines contain the values of the evaluation criteria PMSE and VMSE, for the CVC predictor. The former criteria is related to prediction precision while the latter is related to precision in the prediction variance. These criteria provide the basis for comparison of the various CVC predictors.

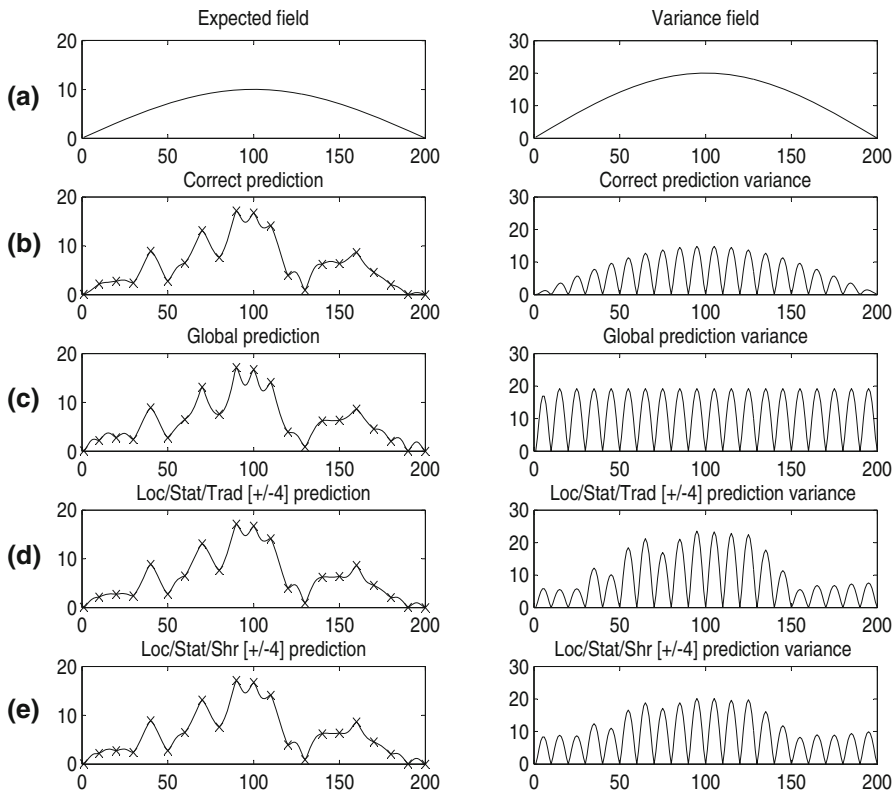
The results from the Loc/Stat/Trad  $k = 10$  CVC predictor are displayed in Figs. 5, 6 and Table 1. The formats are identical to the ones discussed in the previous paragraph. The cross-validation predictions and prediction standard deviations in Fig. 5 are very

similar to the results for the global predictor in Fig. 3. The normalized cross-validation errors in Fig. 6 deviate noticeable from the results for the global predictor in Fig. 4, since the large errors tend to be more uniformly located in the area and the histogram has somewhat lighter tails. From Table 1 one sees that the non-calibrated predictor is well centered but with downward biased prediction variances. The evaluation criteria PMSE and VMSE of the Loc/Stat/Trad  $k = 10$  CVC predictor have values that are favorable compared to the global CVC predictor. It is mildly favorable for prediction and clearly favorable for prediction variance.

The results from the Loc/Stat/Shr  $k = 10$  CVC predictor are displayed in Figs. 7, 8, 9 and Table 1, in similar formats as above. The predictor relies on a set of hyper-parameters that defines the prior model for localized expectation and variance. These prior models are assessed in an empirical Bayesian spirit from the complete set of observations. For the current predictor with  $k = 10$  we obtain the prior model displayed in Fig. 9. The cross-validation predictions and cross-validation standard deviations in Fig. 7 appear as very similar to the results for the other CVC predictors in Figs. 3 and 5. The normalized cross-validation errors in Fig. 8 appear as similar to the ones for the traditional localized predictor in Fig. 6. Note, however, that the histograms are different in the sense that the histogram of the shrinkage predictor has lighter tails than the traditional one. This is very much in the shrinkage spirit, since extreme predictions, often caused by unstable model parameter estimates, are dampened towards the center of the model. From Table 1, one observe that the non-calibrated predictor is well centered but with downward biased prediction variances. The evaluation criteria PMSE and VMSE for the Loc/Stat/Shr  $k = 10$  CVC predictor are both favorable compared to both the traditional global and localized predictors. Minor improvement in prediction precision is obtained, while the precision in the prediction variance is clearly improved.

The results from the Loc/Non-stat/Trad  $k = 10$  CVC predictor and the Loc/Non-stat/Shr  $k = 10$  CVC predictor is summarized in Table 1. By comparing the values of the evaluation criteria PMSE and VMSE to the other predictors, we conclude that the precision in prediction is clearly poorer. Note, however, the improvement in precision for the prediction variance. These results may indicate that there is a trade-off in the precision of the prediction and prediction variance.

To summarize, it can be concluded that the localized predictors are clearly preferable to the global one, both in prediction precision and particularly in the precision of prediction variance. The localized models are robust with respect to deviations from assumptions of global stationary, and this robustness improves the localized predictors. Localized stationary predictors are preferable to the localized non-stationary ones, since the precisions in prediction are clearly better. The precision in the prediction variance can, however, be improved by non-stationary predictors. In the non-stationary models one must estimate the expectation and variance at each observation location, which introduces additional uncertainty in the model. This uncertainty dominates the advantage of using a more general model. Among localized, stationary predictors, the shrinkage predictor is clearly favorable to the traditional one. The prediction precision is slightly better, while the precision in prediction variance is clearly favorable. For localized models, one needs to make bias/variance trade-offs when selecting the localization. Using a regularizer representing the global variability in the parameter



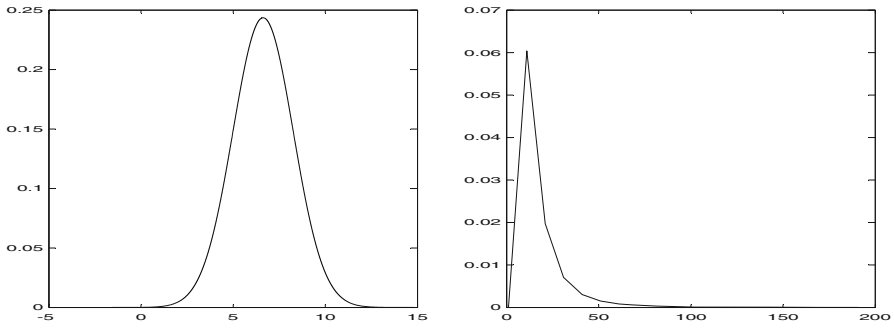
**Fig. 10** General Gaussian random field. Expectation and variance field (a), predictions and prediction variances for one realization (b–e)

estimators provide more stable estimates. It is not surprising that the effect is largest for prediction variance, since traditional variance estimators are notoriously unstable.

In the current study, localization with  $k = 10$  is used. In the extended study (Asfaw and Omre 2014), the sensitivity to choice of  $k$  value is evaluated. If  $k$  is considerably reduced, to for example  $k = 4$ , the localized predictor are poorer than the global one. This deterioration is probably caused by overfitting to the observations. Results for  $k$  in the range 8–16 are consistent with the results in the current study with localized shrinkage predictors clearly preferable. By increasing  $k$  the localized and global predictors will eventually coincide.

### 5.2.2 Simulation Study

A simulation study is conducted on a general Gaussian random field model  $\{r(x); x \in \mathbf{D} \subset \mathbb{R}^1\}$  with  $\mathbf{D} = [1, 200]$  discretized to a grid  $\mathcal{L}_{\mathbf{D}} = \{1, \dots, 200\}$ . The expectation and variance fields used in the study are non-stationary with similar shape (Fig. 10a). The known correlation function is  $\rho(\tau) = \exp\{-0.2\tau^{1.5}\}$  with  $\tau = |x'' - x'|$ . One realization is generated from this random field model and the realizations at locations  $\mathcal{L}_0 = \{1, 10, \dots, 190, 200\}$  are used as observations, hence  $n = 21$ .



**Fig. 11** General Gaussian random field. Prior model for expectation and variance for one realization

With known model parameters, the correct predictions and prediction variances in locations  $\mathcal{L}_D$  are analytically assessable (Fig. 10b). The predictor reproduces the observations and the prediction variances reflect the non-stationarity. The various CVC predictors Glob/Stat/Trad, Loc/Stat/Trad and Loc/Stat/Shr are based on localization  $k = \pm 4$  whenever relevant. The corresponding results are displayed in Fig. 10c–e. Note that the global predictor has regression towards the observation average and prediction variance accounting only for the observation configuration. The localized predictors are fairly similar, capturing the local variability in the observations. The shrinkage results appear as somewhat damped relative to the traditional ones. This dampening is caused by the prior models for local expectation and variance which are assessed in an empirical Bayesian spirit. The prior model for this realization is displayed in Fig. 11. Predictions based on the non-stationary models are also evaluated but these results are not displayed in Fig. 10.

To summarize, the global predictor, which corresponds to classical ordinary kriging, appears as highly unreliable since the underlying model is clearly non-stationary. This predictor will not be further evaluated. The results from the localized predictors are difficult to distinguish by visual inspection.

The evaluation criteria discussed in Sect. 5 can be revised to characterize the deviations between the CVC predictions and prediction variances and the correct ones displayed in Fig. 10a. One thousand realizations are generated and the evaluation criteria to obtain APMSC and AVMSC are averaged for prediction and prediction variance respectively, see Table 2. This procedure is repeated for varying localizations  $k$ . From Table 2, one observes that for localized, stationary predictors the shrinkage-versions are preferable to the traditional ones for both criteria and for all  $k$ , except for one extreme case with large  $k$ . For the localized, non-stationary predictors the shrinkage-versions are uniformly preferable to the traditional ones.

One typical feature is observed for localized/stationary predictors for the criterion APMSC, characterizing precision in prediction, for varying localizations  $k$ . The traditional predictor makes bias/variance trade-offs, resulting in poor performance for large  $k$  due to bias and also poor performance for small  $k$  due to instability in the parameter estimates. Localization at  $k = \pm 4$  provides a favorable traditional predictor. The shrinkage predictor stabilizes the variance estimates by shrinkage and performs well for smaller localizations with  $k = \pm 2$ . The Loc/Stat/Shr  $k = \pm 4$  CVC predictor

**Table 2** General Gaussian random field

Model	Localized/stationary						Localized/non-stationary					
	Traditional			Shrinkage			Traditional			Shrinkage		
	±6	±4	±2	±6	±4	±2	±6	±4	±2	±6	±4	±2
Test D	0.3665	0.3357	1.0100	0.3968	0.2838	0.2563	0.4815	0.6669	2.1658	0.4259	0.4305	0.6264
APMSC	0.5287	1.0939	159.3760	0.3927	0.5944	1.2491	0.5108	1.0369	138.6639	0.3860	0.5723	1.2327
AVMSC	18.2536	17.8311	19.8851	16.9502	15.6486	13.8070	18.6699	18.4643	23.7843	17.0732	15.8705	14.1365
APMSE	1.8936	2.0277	2.9888	1.7271	1.7306	1.4011	1.8908	2.0086	2.4491	1.7277	1.7187	1.3483

Evaluation criteria: *APMSC* average prediction mean squared correct, *AVMSC* average variance mean squared correct, *APMSE* average prediction mean squared error, *AVMSE* average variance mean squared error

appears to be the preferable one, since the APMSC criterion is seen as more important than the AVMSC one.

The evaluation criteria discussed above require the correct predictors to be available, which is not the case in real studies. Therefore, the cross-validation criteria discussed in Sect. 5, which are always available are also computed. One thousand realizations are averaged over to obtain APMSE and AVMSE, and the resulting values are listed in Table 2. For these criteria, the shrinkage versions are uniformly preferable to the traditional ones for all cases. The Loc/Stat/Shr  $k = \pm 2$  CVC predictor appears to be superior based on these criteria as well.

To summarize, the cross-validation criteria appear as representative for the exact criteria based on the correct predictions. The former criteria can be computed in real studies with only one set of observations available. The Loc/Stat/Shr CVC predictors are identified as preferable to the other predictors, although the best localization  $k$  appears to be somewhat underestimated by cross-validation.

In the current simulation study, only one expectation and one variance function are used. In the extended study (Asfaw 2014), many other cases are considered. The results are consistent with the ones presented here, and it is demonstrated that the traditional predictors are particularly sensitive to deviations from stationarity in expectation, which also influences the variance estimates. Lastly, the synthetic simulation case is Gaussian so no outliers and no heavy-tailed distributions are involved. In spite of this, the shrinkage predictors are found to be preferable to the traditional ones. In the presence of outliers and heavy-tailed models, one can expect that the shrinkage predictors perform even more favorably.

## 6 Conclusions

Two versions of localized, shrinkage CVC predictors are defined, one based on local stationarity and one with local non-stationarity. The shrinkage is defined in an empirical Bayes setting while the cross-validation calibration (CVC) ensures correct global scaling of the predictor variance. The introduction of spatial shrinkage predictors constitutes the new feature of this study, and are termed localized/shrinkage kriging predictors.

The localized/shrinkage kriging predictors are compared to traditional kriging predictors, both global and localized, in a study on real precipitation data and in a synthetic simulation study. Two cross-validation-based criteria are used in the comparison. The localized/shrinkage kriging predictors are found to be clearly favorable to traditional kriging predictors on the real data set of yearly accumulated precipitation. The synthetic study is based on a Gaussian random field with spatially varying expectation and variance which makes local predictors suitable. The localized/shrinkage kriging predictors emerge as clearly favorable to traditional localized kriging predictors also in this simulation study. The shrinkage predictors based on local stationarity seems to be the superior models.

Our recommendation is to use localized, shrinkage kriging predictors, based on a local stationarity model, for spatial prediction whenever deviation from stationarity in the observations is suspected. Even for a stationary Gaussian model localized,

shrinkage kriging predictors can be preferable to global ordinary kriging, if the focus is on computational demands.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

### Appendix: Hierarchical, Gaussian Model

Consider a stationary hierarchical, Gaussian random field  $\{r(x); x \in \mathbf{D} \subset \mathfrak{R}^m\}$  and a set of observations  $\mathbf{r}_o = [r(x_1), \dots, r(x_n)]$ ,  $x_1, \dots, x_n \in \mathbf{D}$ .

The pdf's are

$$\begin{aligned} [\mathbf{r}_o|m, s^2] &\sim f[\mathbf{r}_o|m, s^2] \\ &= \text{Gauss}_n[m\mathbf{i}_n, s^2\Omega_{oo}] \\ &= [2\pi]^{-\frac{n}{2}} |s^2\Omega_{oo}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} [\mathbf{r}_o - m\mathbf{i}_n]^T [s^2\Omega_{oo}]^{-1} [\mathbf{r}_o - m\mathbf{i}_n]\right\} \\ [m|s^2] &\sim f[m|s^2] = \text{Gauss}_1[\mu_m, \tau_m s^2] \\ &= [2\pi]^{-\frac{1}{2}} [\tau_m s^2]^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} [\tau_m s^2]^{-1} [m - \mu_m]^2\right\} \\ [s^2] &\sim f[s^2] = \text{InvGam}_1[\xi_s, \gamma_s] \\ &= [\Gamma(\xi_s)]^{-1} \gamma_s^{\xi_s} [s^2]^{-[\xi_s+1]} \exp\{-\gamma_s [s^2]^{-1}\}. \end{aligned}$$

Note that

$$\begin{aligned} E[s^2] &= [\xi - 1]^{-1} \gamma_s; \quad \xi_s > 1, \gamma_s > 0 \\ \text{Var}[s^2] &= \left[ (\xi - 1)^2 (\xi - 2) \right]^{-1} \gamma_s^2; \quad \xi_s > 2, \gamma_s > 0. \end{aligned}$$

From the pdfs above

$$\begin{aligned} [m|s^2, \mathbf{r}_o] &\sim f[m|s^2, \mathbf{r}_o] = \frac{f[\mathbf{r}_o|m, s^2] f[m|s^2]}{\int f[\mathbf{r}_o|m, s^2] f[m|s^2] dm} \\ &= \text{Gauss}_1(\mu_{m|o}, \sigma_{m|o}^2) \\ \mu_{m|o} &= \mu_m + \tau_m \mathbf{i}_n^T [\tau_m \mathbf{i}_n \mathbf{i}_n^T + \Omega_{oo}]^{-1} [\mathbf{r}_o - \mu_m \mathbf{i}_n] \\ \sigma_{m|o}^2 &= \tau_m s^2 - \tau_m s^2 \mathbf{i}_n^T [\tau_m s^2 \mathbf{i}_n \mathbf{i}_n^T + s^2 \Omega_{oo}]^{-1} [\tau_m s^2 \mathbf{i}_n] \\ [s^2|\mathbf{r}_o] &\sim f[s^2|\mathbf{r}_o] = \frac{\int f[\mathbf{r}_o|m, s^2] f[m|s^2] f[s^2] dm}{\int \int f[\mathbf{r}_o|m, s^2] f[m|s^2] f[s^2] dm ds^2} \\ &= \text{InvGam}_1(\xi_{s|o}, \gamma_{s|o}) \end{aligned}$$

$$\begin{aligned}\xi_{s|o} &= \xi_s + \frac{n}{2} \\ \gamma_{s|o} &= \gamma_s + \frac{1}{2} [[\mathbf{r}_o - \mu_m \mathbf{i}_n]^T [\tau_m \mathbf{i}_n \mathbf{i}_n^T + \Omega_{oo}]^{-1} [\mathbf{r}_o - \mu_m \mathbf{i}_n]].\end{aligned}$$

## References

- Anderes EB, Stein ML (2011) Local likelihood estimation for nonstationary random fields. *J Multivar Anal* 102(3):506–520
- Asfaw ZG (2014) Inference and Prediction in Non-stationary Stochastic Models: Survival Analysis and Kriging Interpolation. Dissertation NTNU
- Chiles JP, Delfiner P (2009) Geostatistics: modeling spatial uncertainty. Wiley, New York
- Efron B, Morris C (1973) Stein's estimation rule and its competitors—an empirical Bayes approach. *J Am Stat Assoc* 68(341):117–130
- Higdon D (1998) A process-convolution approach to modelling temperatures in the North Atlantic Ocean. *Environ Ecol Stat* 5(2):173–190
- Johns CJ, Nychka D, Kittel TGF, Daly C (2003) Infilling sparse records of spatial fields. *J Am Stat Assoc* 98(464):796–806
- Journel AG, Huijbregts CJ (1978) Mining geostatistics. Academic press, New York
- Røislien J, Omre H (2006) T-distributed random fields: a parametric model for heavy-tailed well-log data. *Math Geol* 38(7):821–849
- Stein ML (2002) The screening effect in kriging. *Ann Stat* 30(1):298–323