

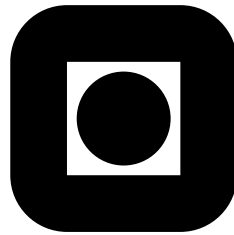
NORGES TEKNISK-NATURVITENSKAPELIGE
UNIVERSITET

Localized/Shrinkage Kriging Predictors

by

Zeytu Gashaw Asfaw and Henning Omre

PREPRINT
STATISTICS NO. 1/2014



NORWEGIAN UNIVERSITY OF SCIENCE AND
TECHNOLOGY
TRONDHEIM, NORWAY

This report has URL

<http://www.math.ntnu.no/preprint/statistics/2014/S1-2014.pdf>

Zeytu Gashaw Asfaw has homepage: <http://www.math.ntnu.no/~zeytugas>

E-mail: zeytugas@math.ntnu.no

Address: Department of Mathematical Sciences, NTNU, N-7491 Trondheim,
Norway.

Localized/Shrinkage Kriging Predictors

Zeytu Gashaw Asfaw/zeytugas@math.ntnu.no
Henning Omre/omre@math.ntnu.no
Department of Mathematical Sciences
Norwegian University of Science and Technology(NTNU)
P.O.Box 7491, Trondheim, Norway

May 28, 2014

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 appears with a shrinkage effect and is coined a localized/shrinkage kriging predictor. The predictor is compared to traditional localized kriging predictors in a case study on observations of annual cumulated precipitation. A crossvalidation criterion is used in the comparison. The shrinkage predictor appears as uniformly preferable to the traditional kriging predictors. A simulation study on prediction in non-stationary Gaussian random fields is conducted. The results from this study confirms that the shrinkage predictor is favorable to the traditional ones. Moreover, the crossvalidation criterion is found to be suitable for selection of predictor. Lastly, the shrinkage predictor appears as particularly robust towards spatially varying expectation functions.

1 Introduction

Consider a set of exact observations from a continuous regionalized variable. Focus is on prediction of the regionalized variable in an unobserved location with associated prediction variance. One option is to use traditional kriging prediction, see Journel and Huijbregts (1978) and Chiles and Delfiner (1999). If one assume a model with spatially constant expectation and variance with a shift-invariant spatial correlation function, then a global, ordinary kriging predictor will be a natural choice. This model assumption may be tested statistically, see Fuentes (2005).

A more flexible and robust spatial predictor can be defined by applying the ordinary kriging predictor locally. This entails using only observations in a specified finite neighborhood around the location of the variable to be predicted. This approach is termed local neighborhood kriging, see Chiles and Delfiner (1999), and it robustifies the predictor with respect to deviations from the assumption of spatially constant expectation and variance. Moreover, local neighborhood predictors can give huge computational gains in large scale problems.

The major challenge in using localized predictors is to specify the size of the neighborhoods, or the set of neighboring observations involved. Classical statistical trade-offs between bias and variance in the local predictor must be made. The spatial correlation structure may provide a screening effect by the neighboring observations, see Stein (2002), and this effect may be used to justify localization. The localized predictors can cause artifacts in the predicted regionalized variable as discontinuities when extreme observations are included or excluded in the neighborhood as it is shifted, see Gribov and Krivoruchko (2004).

The objective of this study is to improve the flexibility and robustness of the spatial predictor. We define a spatial model as a Gaussian random field with spatially varying expectation and variance. The spatial correlation function is shift invariant and known. Under these model assumptions the local neighborhood kriging predictor require expectation and variance to be assessed in the prediction and observation locations. In traditional kriging approaches this inference is made by a sliding neighborhood maximum-likelihood estimator.

We define a new localized predictor inspired by the empirical Bayes approach discussed in Efron and Morris (1973). We phrase the inference of the spatial expectation and variance in a Bayesian setting along the lines of Røislien and Omre (2006). The conjugate prior models are assessed empirically from the global set of observations. The resulting local kriging predictor appear with shrinkage caused by the global prior model. We term the predictor as localized/shrinkage kriging.

The report is organized as follows : Section 2 contain a list of notation. In Section 3 general random field models are defined and discussed. Model parameter inference in these random field models are discussed in Section 4, while Section 5 contain definitions of the localized predictors. In Section 6 a presentation of the evaluation criteria is included. Section 7 contain a demonstration and evaluation of the predictors on a couple of real data examples, while Section 8 presents the empirical simulation study on Gaussian random fields. Lastly, Section 9 contains the conclusions from the study.

2 Notation

The following notation is used:

$\mathcal{L}_{\mathbf{D}}$	grid over \mathbf{D}
n	number of grid nodes in \mathbf{D}
\mathcal{L}_o	locations of sampled observations
n_o	number of sampled observation
k	number of closest observation
\mathbf{r}	vector of values in grid $\mathcal{L}_{\mathbf{D}}$
\mathbf{r}_o	vector of values of observations in \mathcal{L}_o
r_+	value at location x_+
\mathbf{r}_{+o}^k	vector of values in k-closest observations to location x_+
i_n	unit $[n \times 1]$ vector
I_n	unit diagonal $[n \times n]$ matrix
μ_r	expected value
σ_r^2	variance value
Σ_{rr}	covariance matrix
Ω_r	correlation matrix
Γ_r	diagonal standard deviation matrix

H selection matrix
 ν degree of freedom for hierarichical representation

3 Random field models

A random field (RF) is a generalization of a stochastic process, taking referenes on some topological space. Due to the complexity of natural phenomena and the actual problem, several kind of random fields are defined, among them Gaussian RF, Poisson RF and Markov RF. The former is of concern in the current study. We consider Gaussian RF $\{r(x); x \in \mathbf{D} \subset \mathfrak{R}\}$, where x is a reference location running over the domain \mathbf{D} as a subset of \mathfrak{R}^m , with $r(x)$ being the random variable of interest.

3.1 General Gaussian Random Field Model

A Gaussian RF is defined by the Gaussian probability density functions. The Gaussian RF is a preferable model for continuous, or almost continuous, spatial variables due to its simplicity in inferences and analytical tractability.

The definition of a Gaussian RF is:

A RF $\{r(x); x \in \mathbf{D} \subset \mathfrak{R}^m\}$ is denoted a Gaussian RF if

$$\mathbf{r} = \begin{bmatrix} r(x_1) \\ \vdots \\ r(x_n) \end{bmatrix} \sim N_n(\boldsymbol{\mu}_r, \Sigma_{rr}) \quad (1)$$

for \forall conf $(x_1, \dots, x_n) \in \mathbf{D}^n, \forall n \geq 1$

and the corresponding pdf can be written as:

$$f(\mathbf{r}) = (2\pi)^{-\frac{n}{2}} |\Sigma_{rr}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{r} - \boldsymbol{\mu}_r)^T \Sigma_{rr}^{-1} (\mathbf{r} - \boldsymbol{\mu}_r) \right\},$$

where

$$\boldsymbol{\mu}_r = \begin{bmatrix} \mu_r(x_1) \\ \vdots \\ \mu_r(x_n) \end{bmatrix},$$

and

$$\Sigma_{rr} = \Gamma_r \Omega_{rr} \Gamma_r^T,$$

with

$$\Gamma_r = \begin{bmatrix} \sigma_r(x_1) & 0 & \dots & 0 \\ 0 & \sigma_r(x_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_r(x_n) \end{bmatrix}$$

$$\Omega_{rr} = \begin{bmatrix} \rho_r(x_1, x_1) & \rho_r(x_1, x_2) & \dots & \rho_r(x_1, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ \rho_r(x_n, x_1) & \rho_r(x_n, x_2) & \dots & \rho_r(x_n, x_n) \end{bmatrix}$$

The model parameters for Gaussian RFs are:

$\{\mu_r(x) = E[r(x)]; \quad x \in \mathbf{D}\}$ - spatial expectation field.

$\{\sigma_r^2(x) = Var[r(x)]; \quad x \in \mathbf{D}\}$ - spatial variance field.

$\{\rho_r(x', x'') = Corr[r(x'), r(x'')]; \quad x', x'' \in \mathbf{D}^2\}$ - spatial correlation field.

$\{\phi_r(x', x'') = Cov[r(x'), r(x'')] = \sigma_r(x')\sigma_r(x'')\rho_r(x', x''); \quad x', x'' \in \mathbf{D}^2\}$

- spatial covariance field.

Hence, the model parametrization for a Gaussian RF is: $\{\mu_r(x), \sigma_r^2(x); x \in \mathbf{D}\}$ and $\{\rho_r(x', x''); x', x'' \in \mathbf{D}^2\}$. The requirements for the model parameters are:

- $\sigma_r(x) \geq 0$ for $x \in \mathbf{D}$.
- $-1 \leq \rho_r(x', x'') \leq 1$ for $x', x'' \in \mathbf{D}^2$
- Ω_{rr} - non-negative definite matrix.

Thus, a correlation field must be a positive semi-definite field to ensure that Ω_{rr} is valid. A correlation field $\rho_r(x', x'')$ is called positive semi-definite if the associated quadratic form is non-negative, then:

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \rho_r(x_i, x_j) \geq 0$$

$$\forall \text{conf}(x_1, \dots, x_n) \in \mathbf{D}^n, \forall n > 1, \forall \boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)^T \in \Re^n$$

If in addition the quadratic form $\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \rho_r(x_i, x_j) = \mathbf{0}$ only for $\boldsymbol{\alpha} = \mathbf{0i}_n$, then the correlation field $\rho_r(x', x'')$ is called positive definite. Expectation, variance and correlation fields determine all stochastic properties of a Gaussian RF.

Consider a regular grid over \mathbf{D} , denote it $\mathcal{L}_{\mathbf{D}}$ and let the number of grid nodes be n . Define the discretized Gaussian RF $\mathbf{r} = \{r(x); x \in \mathcal{L}_{\mathbf{D}}\}$.

Further let the expectation $[n \times 1]$ vector $\boldsymbol{\mu}_r = \{\mu_r(x); x \in \mathcal{L}_{\mathbf{D}}\}$, the standard deviation $[n \times n]$ matrix Γ_r be diagonal with elements $\{\sigma_r(x); x \in \mathcal{L}_{\mathbf{D}}\}$ and the correlation $[n \times n]$ matrix Ω_{rr} have elements $\{\rho_r(x', x''); x', x'' \in \mathcal{L}_{\mathbf{D}} \times \mathcal{L}_{\mathbf{D}}\}$. Moreover, the covariance $[n \times n]$ matrix is $\Sigma_{rr} = \Gamma_r \Omega_{rr} \Gamma_r^T$.

Let \mathbf{r}_o be a $[n_o \times 1]$ vector of observations which occur at grid locations hence at a subset of $\mathcal{L}_{\mathbf{D}}$,

$$\mathbf{r}_o = \begin{bmatrix} r(x_{o1}) \\ r(x_{o2}) \\ \vdots \\ r(x_{on_o}) \end{bmatrix} = \begin{bmatrix} r_{o1} \\ r_{o2} \\ \vdots \\ r_{on_o} \end{bmatrix}$$

Hence, $[\mathbf{r}_o | \mathbf{r}] = H\mathbf{r}$, where H is a binary selection $[n_o \times n]$ matrix having one on all sampled location and zero in unsampled location.

Consider the combined vector of values at grid nodes and observations:

$$\begin{bmatrix} \mathbf{r} \\ \mathbf{r}_o \end{bmatrix} \sim N_{n+n_o} \left[\begin{bmatrix} \boldsymbol{\mu}_r \\ H\boldsymbol{\mu}_r \end{bmatrix}, \begin{bmatrix} \Sigma_{rr} & \Sigma_{rr}H \\ H\Sigma_{rr} & H\Sigma_{rr}H^T \end{bmatrix} \right]$$

Hence, the conditional Gaussian RF is

$$[\mathbf{r} \mid \mathbf{r}_o] \sim N_n(\boldsymbol{\mu}_{\mathbf{r}|\mathbf{r}_o}, \Sigma_{\mathbf{r}\mathbf{r}|\mathbf{r}_o}). \quad (2)$$

where

$$\boldsymbol{\mu}_{\mathbf{r}|\mathbf{r}_o} = \mathbb{E}[\mathbf{r} \mid \mathbf{r}_o] = \boldsymbol{\mu}_r + \Sigma_{rr} H^T [H \Sigma_{rr} H^T]^{-1} (\mathbf{r}_o - H \boldsymbol{\mu}_r).$$

$$\Sigma_{\mathbf{r}\mathbf{r}|\mathbf{r}_o} = \text{Var}[\mathbf{r} \mid \mathbf{r}_o] = \Sigma_{rr} - \Sigma_{rr} H^T [H \Sigma_{rr} H^T]^{-1} H \Sigma_{rr}$$

Consider prediction of the value in an arbitrary location $x_+ \in \mathbf{D}$, and denote it $r_+ = r(x_+)$. The available observations are \mathbf{r}_o .

$$\begin{bmatrix} r_+ \\ \mathbf{r}_o \end{bmatrix} \sim N_{1+n_o} \left[\begin{bmatrix} \mu_+ \\ H \boldsymbol{\mu}_r \end{bmatrix}, \begin{bmatrix} \sigma_+^2 & \sigma_+ \omega_{o+}^T H \Gamma_r H^T \\ \sigma_+ H \Gamma_r H \omega_{o+} & H \Gamma_r \Omega_{rr} \Gamma_r H^T \end{bmatrix} \right]$$

$$= N_{1+n_o} \left[\begin{bmatrix} \mu_+ \\ \boldsymbol{\mu}_o \end{bmatrix}, \begin{bmatrix} \sigma_+^2 & \sigma_+ \omega_{o+}^T \Gamma_o \\ \sigma_+ \Gamma_o \omega_{o+} & \Gamma_o \Omega_{oo} \Gamma_o^T \end{bmatrix} \right]$$

where $\mu_+ = \mu(x_+)$, $\sigma_+^2 = \sigma^2(x_+)$ and

$$\boldsymbol{\mu}_o = \begin{bmatrix} \mu(x_{o1}) \\ \vdots \\ \mu(x_{on_o}) \end{bmatrix} = H \boldsymbol{\mu}_r$$

$$\omega_{o+} = \begin{bmatrix} \rho(x_{o1}, x_+) \\ \vdots \\ \rho(x_{on_o}, x_+) \end{bmatrix}$$

$$\Gamma_o = \begin{bmatrix} \sigma(x_{o1}) & 0 & \dots & 0 \\ 0 & \sigma(x_{o2}) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma(x_{on_o}) \end{bmatrix} = H\Gamma_r H^T$$

$$\Omega_{oo} = \begin{bmatrix} \rho(x_{o1}, x_{o1}) & \rho(x_{o1}, x_{o2}) & \dots & \rho(x_{o1}, x_{on_o}) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(x_{on_o}, x_{o1}) & \rho(x_{on_o}, x_{o2}) & \dots & \rho(x_{on_o}, x_{on_o}) \end{bmatrix} = H\Omega_{rr}H^T$$

and let $\Sigma_{oo} = \Gamma_o\Omega_{oo}\Gamma_o^T$.

The conditional Gaussian random variable in location x_+ given the observation vector \mathbf{r}_o is:

$$[r_+ | \mathbf{r}_o] \sim N_1 \left[\mu_{+|o}, \sigma_{+|o}^2 \right]$$

with

$$\begin{aligned} \mu_{+|o} &= \mu_+ + \sigma_+ \omega_{o+}^T \Gamma_o [\Gamma_o \Omega_{oo} \Gamma_o]^{-1} [\mathbf{r}_o - \boldsymbol{\mu}_o] \\ &= \mu_+ + K_G [\mathbf{r}_o - \boldsymbol{\mu}_o] \\ \sigma_{+|o}^2 &= \sigma_+^2 - \sigma_+ \omega_{o+}^T \Gamma_o [\Gamma_o \Omega_{oo} \Gamma_o]^{-1} \Gamma_o \omega_{o+} \sigma_+ \\ &= \sigma_+^2 [1 - \omega_{o+}^T \Gamma_o [\Gamma_o \Omega_{oo} \Gamma_o]^{-1} \Gamma_o \omega_{o+}] \\ &= \sigma_+^2 - K_G \Gamma_o \Omega_{oo} \Gamma_o K_G^T \end{aligned}$$

where $K_G^T = \sigma_+ \omega_{o+}^T \Gamma_o [\Gamma_o \Omega_{oo} \Gamma_o]^{-1}$ is a weight $[n_o \times 1]$ vector.

Note that the weight vector, termed generalized kriging weights, are functions of the variances in all the locations involved in the prediction and the correlation structure.

Simulation of a Gaussian RF

Consider a conditional discretized Gaussian RF represented by $[\mathbf{r} | \mathbf{r}_o]$ with parameters as in Equation (1). We may want to generate a sample of $[\mathbf{r} | \mathbf{r}_o]$ from the conditional model. This simulated surface can be generated as follows:

1. Cholesky factorization of $\Sigma_{rr|r_o} = LL^T$ where L and L^T are lower/upper triangular $[n \times n]$ matrices.
2. Sample $\mathbf{z} \sim N_n(\mathbf{0}, I_n)$
3. Compute $\mathbf{v} = L\mathbf{z}$
4. Compute $[\mathbf{r} | \mathbf{r}_o] = \boldsymbol{\mu}_{r|r_o} + \mathbf{v}$
5. Return $[\mathbf{r} | \mathbf{r}_o]$

3.2 Stationary Gaussian RF Model

Stationarity is a property of a regionalized variable that has shift invariant statistical properties within the area of interest. Stationary random field is a random field whose joint probability distribution does not change when shifted in location. Prediction in stationary Gaussian RF is termed either simple or ordinary kriging dependent on whether the expectation is known or must be estimated.

A RF $\{r(x); x \in \mathbf{D}\}$ is defined to be stationary if it satisfies:

$$\begin{aligned} & \{\mu_r(x) = E[r(x)] = \mu_r; \quad x \in \mathbf{D}\} \\ & \{\sigma_r^2(x) = Var[r(x)] = \sigma_r^2; \quad x \in \mathbf{D}\} \\ & \{\rho_r(x', x'') = Corr[r(x'), r(x'')] = \rho_r(x' - x''); \quad x', x'' \in \mathbf{D}^2\} \\ & \{\phi_r(x', x'') = Cov[r(x'), r(x'')] = \sigma_r^2 \rho_r(x' - x''); \quad x', x'' \in \mathbf{D}^2\} \end{aligned}$$

A Gaussian RF having stationary model parameters is said to be stationary Gaussian RF. The predictor in $r_+ = r(x_+)$, in a stationary Gaussian RF is defined by the expression:

$$\begin{bmatrix} r_+ \\ \mathbf{r}_o \end{bmatrix} \sim N_{1+n_o} \left[\begin{bmatrix} \mu_r \\ \mu_r \mathbf{i}_{n_o} \end{bmatrix}, \begin{bmatrix} \sigma_r^2 & \sigma_r^2 \omega_{o+}^T \\ \omega_{o+} \sigma_r^2 & \sigma_r^2 \Omega_{oo} \end{bmatrix} \right]$$

hence,

$$\begin{aligned}
[r_+ \mid \mathbf{r}_o] &\sim N_1 \left[\mu_{r|o}, \sigma_{r|o}^2 \right] \\
\mu_{r|o} &= \mu_r + \sigma_r^2 \omega_{o+}^T \left[\sigma_r^2 \Omega_{oo} \right]^{-1} [\mathbf{r}_o - \mu_r i_{n_o}] \\
&= \mu_r + K_S [\mathbf{r}_o - \mu_r i_{n_o}] \\
\sigma_{r|o}^2 &= \sigma_r^2 - \sigma_r^2 \omega_{o+}^T \left[\sigma_r^2 \Omega_{oo} \right]^{-1} \omega_{o+} \sigma_r^2 \\
&= \sigma_r^2 [1 - K_S \Omega_{oo} K_S^T]
\end{aligned}$$

where $K_S^T = \omega_{o+}^T \Omega_{oo}^{-1}$, is a weight $[n_o \times 1]$ vector. Note that the weight vector, termed stationary (simple) kriging weights, is dependent on the correlation structure only, not on the variance.

3.3 Hierarchical Stationary Gaussian Random Field

For a stationary Gaussian RF to be fully specified, the model parameters μ_r , σ_r^2 and $\rho_r(\cdot)$ need to be known. In the hierarchical representation we let μ_r and σ_r^2 be represented by random variables m and s^2 while $\rho_r(\cdot)$ is considered to be known. By conditioning on $[m, s^2]$, the Gaussian RF is fully specified.

Consider the discrete representation of stationary Gaussian RF. Let \mathbf{r} conditional on the random parameters $[m, s^2]$ be distributed as

$$[\mathbf{r} \mid m, s^2] \sim N_n(m i_n, s^2 \Omega_{rr})$$

where, m and s^2 are univariate random variables and Ω_{rr} is a known positive definite correlation $[n \times n]$ matrix.

Moreover, assume the following prior model for $[m, s^2]$:

$$[m \mid s^2] \sim N_1(\mu_m, \tau_m s^2) \tag{3}$$

$$s^2 \sim IG(\xi_s, \gamma_s) \quad (4)$$

with $\mu_m \in \mathfrak{R}$, $\tau_m \in \mathfrak{R}_+$ and $IG(\xi_s, \gamma_s)$ representing the inverse gamma pdf

$$f(s^2) = \frac{1}{\Gamma(\xi_s)} \gamma_s^{\xi_s} [s^2]^{-(\xi_s+1)} \exp\{-\gamma_s [s^2]^{-1}\}, \quad s^2 > 0.$$

where $\Gamma(x)$ is the gamma function, $\xi_s \in \mathfrak{R}_+$ is a shape parameter and $\gamma_s \in \mathfrak{R}_+$ is a scale parameter.

Hence, it can be demonstrated, see Røislien and Omre (2006) that:

$$\mathbf{r} \sim T_n(\mu_m i_n, \Sigma_{rr}, \nu)$$

represent a T-dist RF defined by the multivariate T-distribution:

$$f(\mathbf{r}) = \frac{\Gamma(\frac{\nu+n}{2})}{\Gamma(\frac{\nu}{2})(\nu\pi)^{\frac{n}{2}}} |\Sigma_{rr}|^{-\frac{1}{2}} \left[1 + \frac{1}{\nu} [\mathbf{r} - \mu_m i_n]^T \Sigma_{rr}^{-1} [\mathbf{r} - \mu_m i_n] \right]^{-\frac{\nu+n}{2}}$$

where $\nu \in \mathfrak{R}_+$ is the degrees of freedom which is defined by $\nu = 2\xi_s$. This definition specifies a spherical-symmetric pdf centered at $\mu_m i_n$ with Ω_{rr} controlling scale and multivariate dependence, while ν controls the tail behavior (Mardia et al., 1979).

Multivariate Gaussian and Cauchy distributions are special cases of multivariate T-distributions. That is,

$$T_n(\mu_m i_n, \Sigma_{rr}, \nu) \xrightarrow{\nu \rightarrow \infty} N_n(\mu_m i_n, \Sigma_{rr}) - \text{Gaussian distribution.}$$

$$T_n(\mu_m i_n, \Sigma_{rr}, 1) = C_n(\mu_m i_n, \Sigma_{rr}) - \text{Cauchy distribution.}$$

This hierarchical representation can be interpreted in a Bayesian setting with $[m, s^2]$ being random hyperparameters.

Consider a set of observations \mathbf{r}_o as previously defined. The posterior model for the model parameters $[m, s^2 \mid \mathbf{r}_o]$ can then be determined. From the definition of hierarchical stationary Gaussian RF one has:

$$[\mathbf{r}_o \mid m, s^2] \sim N_{n_o}(m i_{n_o}, s^2 \Omega_{oo})$$

and using the prior model for $[m \mid s^2]$, in Equation (3), one obtains:

$$\begin{bmatrix} \mathbf{r}_o \\ m \end{bmatrix} \mid s^2 \sim N_{n_o+1} \left[\begin{bmatrix} \mu_m i_{n_o} \\ \mu_m \end{bmatrix}, \begin{bmatrix} \tau_m s^2 i_{n_o} i_{n_o}^T + s^2 \Omega_{oo} & \tau_m s^2 i_{n_o} \\ \tau_m s^2 i_{n_o}^T & \tau_m s^2 \end{bmatrix} \right]$$

Consequently,

$$[m \mid \mathbf{r}_o, s^2] \sim N_1 \left[\mu_{m|o}, \sigma_{m|o}^2 \right]$$

with

$$\begin{aligned} \mu_{m|o} &= \mu_m + \tau_m s^2 i_{n_o}^T [\tau_m s^2 i_{n_o} i_{n_o}^T + s^2 \Omega_{oo}]^{-1} (\mathbf{r}_o - \mu_m i_{n_o}) \\ &= \mu_m + \tau_m i_{n_o}^T [\tau_m i_{n_o} i_{n_o}^T + \Omega_{oo}]^{-1} (\mathbf{r}_o - \mu_m i_{n_o}). \\ \sigma_{m|o}^2 &= \tau_m s^2 - \tau_m s^2 i_{n_o}^T [\tau_m s^2 i_{n_o} i_{n_o}^T + s^2 \Omega_{oo}]^{-1} \tau_m s^2 i_{n_o} \end{aligned}$$

Note that $\mu_{m|o} = E[m \mid \mathbf{r}_o, s^2] = E[m \mid \mathbf{r}_o]$ and hence independent of s^2 . Similarly, the marginal pdf of $[\mathbf{r}_o \mid s^2]$ is,

$$[\mathbf{r}_o \mid s^2] \sim N_{n_o} [\mu_m i_{n_o}, \tau_m s^2 i_{n_o} i_{n_o}^T + s^2 \Omega_{oo}]$$

and using the prior model for s^2 , in Equation (4), one obtains, see Appendix A:

$$[s^2 \mid \mathbf{r}_o] \sim IG(\xi_{s|o}, \gamma_{s|o})$$

with

$$\begin{aligned} \xi_{s|o} &= \xi_s + \frac{n_o}{2} \\ \gamma_{s|o} &= \gamma_s + \frac{1}{2} \left[(\mathbf{r}_o - \mu_m i_{n_o})^T [\Omega_{oo} + \tau_m i_{n_o} i_{n_o}^T]^{-1} (\mathbf{r}_o - \mu_m i_{n_o}) \right] \end{aligned}$$

Note that from the characteristics of the inverse Gamma distribution we have:

$$\begin{aligned} \mu_{s|o} &= E[s^2 \mid \mathbf{r}_o] = \frac{\gamma_{s|o}}{\xi_{s|o} - 1}, \quad \xi_{s|o} > 1 \\ \sigma_{s|o}^2 &= Var[s^2 \mid \mathbf{r}_o] = \frac{\gamma_{s|o}}{(\xi_{s|o} - 1)^2 (\xi_{s|o} - 2)}, \quad \xi_{s|o} > 2 \end{aligned}$$

4 Model Parameter Inference

Both the stationary Gaussian RF model, see Section 3.2, and the hierarchical, stationary Gaussian RF model, see Section 3.3, depends on a set of model parameters. These model parameters must be assessed from the available observations \mathbf{r}_o in order to make the respective models operable.

In the study we also use localized estimators for the model parameters. Consider location $x_+ \in \mathbf{D}$ and define the $[k \times 1]$ vector of observations:

$$\mathbf{r}_{+o}^k = G_+^k \mathbf{r}_o$$

where G_+^k is a binary $[k \times n_o]$ matrix which selects the k observations located closest to x_+ . The selection may also include some symmetry criteria.

4.1 Stationary Gaussian RF model

The actual set of model parameters are $[\mu_r, \sigma_r^2, \rho_r(\tau)]$. We consider the spatial correlation function $\rho_r(\tau)$ to be known, hence the expected value μ_r and variance value σ_r^2 must be assessed from \mathbf{r}_o .

We choose to use a maximum likelihood criterion in the assessment, and the log-likelihood function is:

$$\begin{aligned} l(\mu_r, \sigma_r^2; \mathbf{r}_o) &= -\frac{n_o}{2} \log(2\pi) - \frac{n_o}{2} \log(\sigma_r^2) \\ &\quad - \frac{1}{2} \log |\Omega_{oo}| - \frac{1}{2} [\sigma_r^2]^{-1} [(\mathbf{r}^o - \mu_r i_{n_o})^T \Omega_{oo}^{-1} (\mathbf{r}_o - \mu_r i_{n_o})] \end{aligned}$$

Hence the maximum likelihood estimates are:

$$\begin{aligned} \hat{\mu}_r &= [i_{n_o}^T \Omega_{oo}^{-1} \mathbf{r}_o] [i_{n_o}^T \Omega_{oo}^{-1} i_{n_o}]^{-1} \\ \hat{\sigma}_r^2 &= \frac{1}{n_o} (\mathbf{r}_o - \hat{\mu}_r i_{n_o})^T \Omega_{oo}^{-1} (\mathbf{r}_o - \hat{\mu}_r i_{n_o}) \end{aligned}$$

The corresponding localized estimators of $[\mu_r, \sigma_r^2]$ centered at location $x_+ \in \mathbf{D}$ based on $\mathbf{r}_{+o}^k = G_+^k \mathbf{r}_o$ are:

$$\hat{\mu}_+^k = [i_k^T [G_+^k \Omega_{oo} [G_+^k]^T]^{-1} G_+^k \mathbf{r}_o] [i_k^T [G_+^k \Omega_{oo} [G_+^k]^T]^{-1} i_k]^{-1}$$

$$\hat{\sigma}_+^{k2} = \frac{1}{k} (G_+^k \mathbf{r}_o - \hat{\mu}_+^k i_k)^T [G_+^k \Omega_{oo} [G_+^k]^T]^{-1} (G_+^k \mathbf{r}_o - \hat{\mu}_+^k i_k)$$

Define also the expectation $[n_o \times 1]$ vector centered at the observation locations:

$$\hat{\boldsymbol{\mu}}_o^k = \begin{bmatrix} \hat{\mu}_1^k \\ \vdots \\ \hat{\mu}_{n_o}^k \end{bmatrix},$$

and the corresponding standard deviation diagonal $[n_o \times n_o]$ matrix

$$\hat{\Gamma}_o^k = \begin{bmatrix} \hat{\sigma}_1^k & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \hat{\sigma}_{n_o}^k \end{bmatrix}$$

4.2 Hierarchical, stationary Gaussian RF model

The actual set of model parameters is $[\mu_m, \tau_m, \xi_s, \gamma_s, \rho_r(\tau)]$. We consider the spatial correlation function $\rho_r(\tau)$ to be known, hence the prior model parameters for expectation $[\mu_m, \tau_m]$ and for variance $[\xi_s, \gamma_s]$ must be assessed from \mathbf{r}_o .

We choose to make this assessment in an empirical Bayes setting based on the observations \mathbf{r}_o . The k -closest localization is used to define a set of localizations centered at the observations over the domain \mathbf{D} . This set is considered to be a super-population from which the k -closest prior model is assessed.

The estimates for the Gaussian prior model parameters for expectation are:

$$\begin{aligned} \hat{\mu}_m^k &= \frac{1}{n_o} i_{n_o}^T \hat{\boldsymbol{\mu}}_o^k \\ \hat{\sigma}_m^{k2} &= \frac{1}{n_o} [\hat{\boldsymbol{\mu}}_o^k - \hat{\mu}_m^k i_{n_o}]^T [\hat{\boldsymbol{\mu}}_o^k - \hat{\mu}_m^k i_{n_o}] \\ \hat{\sigma}_{r|\cdot}^{k2} &= \frac{1}{n_o} Tr[\hat{\Gamma}_o^k] \\ \hat{\tau}_m^k &= \frac{\hat{\sigma}_m^{k2}}{\hat{\sigma}_{r|\cdot}^{k2}} \end{aligned}$$

The corresponding localized estimators for the posterior expectation m centered at location $x_+ \in \mathbf{D}$ based on observations $\mathbf{r}_{+o}^k = G_+^k \mathbf{r}_o$ is:

$$\begin{aligned}\hat{m}_+^k &= E[m \mid s^2, G_+^k \mathbf{r}_o] \\ &= \hat{\mu}_m^k + \hat{\tau}_m^k i_k^T [\hat{\tau}_m^k i_k i_k^T + [G_+^k \Omega_{oo} [G_+^k]^T]^{-1}]^{-1} [G_+^k \mathbf{r}_o - \hat{\mu}_m^k i_k]\end{aligned}$$

which is independent of s^2 .

Define also the expectation $[n_o \times 1]$ vector centered at the observation locations:

$$\hat{\mathbf{m}}_o^k = \begin{bmatrix} \hat{m}_1^k \\ \vdots \\ \hat{m}_{n_o}^k \end{bmatrix},$$

The estimates for the inverse gamma prior model parameters for variance are more complicated. Note first that the prior expectation and variance for $\xi_s > 2$ are :

$$\begin{aligned}\mu_s &= E[s^2] = \frac{\gamma_s}{\xi_s - 1} \\ \sigma_s^2 &= Var[s^2] = \frac{\gamma_s^2}{[\xi_s - 1]^2 [\xi_s - 2]}\end{aligned}$$

Consequently,

$$\begin{aligned}\xi_s &= \frac{\mu_s^2}{\sigma_s^2} + 2 \\ \gamma_s &= \mu_s \left[\frac{\mu_s^2}{\sigma_s^2} + 1 \right]\end{aligned}$$

Define the $[n_o \times 1]$ vector defined for a k -neighborhood

$$\mathbf{s}^2 = \begin{bmatrix} (r_{o1} - \hat{\mu}_m^k)^2 \\ \vdots \\ (r_{on_o} - \hat{\mu}_m^k)^2 \end{bmatrix}$$

The two first moments are estimated by:

$$\hat{\mu}_s = \frac{1}{n_o} i_{n_o}^T \mathbf{s}^2$$

$$\hat{\sigma}_s^2 = \frac{1}{n_o} [\mathbf{s}^2 - \hat{\mu}_s i_{n_o}]^T [\mathbf{s}^2 - \hat{\mu}_s i_{n_o}]$$

The prior model estimates $\hat{\xi}_s$ and $\hat{\gamma}_s$ are obtained by inserting $\hat{\mu}_s$ and $\hat{\sigma}_s^2$ into the expressions above.

The corresponding localized estimator for the posterior variance s^2 centered at location $x_+ \in \mathbf{D}$ based on observations $\mathbf{r}_{+o}^k = G_+^k \mathbf{r}_o$ is:

$$\begin{aligned} \hat{s}_+^{k2} &= E[s^2 \mid G_+^k \mathbf{r}_o] \\ &= \frac{\hat{\gamma}_{s|o}}{\hat{\xi}_{s|o} - 1} \\ &= \frac{\hat{\gamma}_s + \frac{1}{2} \left[(G_+^k \mathbf{r}_o - \hat{\mu}_m^k i_k)^T \left[(G_+^k \Omega_{oo} [G_+^k]^T) + \hat{\tau}_m^k i_k i_k^T \right]^{-1} (G_+^k \mathbf{r}_o - \hat{\mu}_m^k i_k) \right]}{\hat{\xi}_s + \frac{k}{2} - 1} \end{aligned}$$

Define also the diagonal standard deviation $[n_o \times n_o]$ matrix centered at the observation locations

$$\hat{S}_o^k = \begin{bmatrix} \hat{s}_1^k & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \hat{s}_{n_o}^k \end{bmatrix}$$

5 Prediction Models

The objective of the study is to define improved spatial predictors, and we consider localized predictors which only utilizes observations in a neighborhood of the location in focus for prediction. Two model types are defined: localized/stationary [Loc/Stat] model and localized/non-stationary [Loc/Non-stat] model. For each model type we consider a traditional [Trad] predictor and a shrinkage [Shr] predictor.

Focus is on predicting $r(x_+) = r_+$ in arbitrary location $x_+ \in \mathbf{D}$. The prediction is based on the observation $[n_o \times 1]$ vector $\mathbf{r}_o = [r(x_{o1}), \dots, r(x_{on_o})] = [r_{o1}, \dots, r_{on_o}]$. Define also the binary, selection $[k \times n_o]$ matrix G_+^k which selects the k closest observations to location x_+ . Note that G_+^k may also include some symmetry criteria.

5.1 Localized/Stationary Model

The predictor is based on the stationary Gaussian RF model, see Section 3.2, with the model parameters assessed in two different localized ways.

5.1.1 Traditional Predictor

The Loc/Stat/Trad predictor of r_+ with associated prediction variance is defined as:

$$\begin{aligned}\hat{r}_{STP+}^k &= \hat{\mu}_+^k + [G_+^k \omega_{o+}]^T [G_+^k \Omega_{oo} [G_+^k]^T]^{-1} [G_+^k \mathbf{r}_o - \hat{\mu}_+^k i_k] \\ \hat{\sigma}_{STP+}^{k2} &= \hat{\sigma}_+^{k2} [1 - [G_+^k \omega_{o+}]^T [G_+^k \Omega_{oo} [G_+^k]^T]^{-1} G_+^k \omega_{o+}]\end{aligned}$$

with the parameter estimators defined in Section 4.1.

The expectation μ_+ and the variance σ_+^2 are estimated by maximum likelihood in a neighborhood of x_+ , hence the predictor appears like a localized ordinary kriging predictor. This corresponds to the traditional approach to localized spatial interpolation, see Chiles and Delfiner (1999). The challenge is to define the size of the neighborhood to obtain a suitable bias-variance trade-off. The neighborhood must be small to adopt to possible spatially varying expectation /variance functions and large to contain enough observations to provide stable estimates.

5.1.2 Shrinkage Predictor

The Loc/Stat/Shr predictor of r_+ with associated prediction variance is defined as:

$$\begin{aligned}\hat{r}_{SSP+}^k &= \hat{m}_+^k + [G_+^k \omega_{o+}]^T [G_+^k \Omega_{oo} [G_+^k]^T]^{-1} [G_+^k \mathbf{r}_o - \hat{m}_+^k i_k] \\ \hat{\sigma}_{SSP+}^{k2} &= \hat{s}_+^{k2} [1 - [G_+^k \omega_{o+}]^T [G_+^k \Omega_{oo} [G_+^k]^T]^{-1} G_+^k \omega_{o+}]\end{aligned}$$

with the parameters estimators defined in Section 4.2.

The expectation m_+ and the variance s_+^2 are estimated in a Bayesian setting as the posterior expectations given the observations in a neighborhood of x_+ . Hence the prior model acts like a regularizer when estimating the local expectation and variance. The prior models for m_+ and s_+^2 are assessed from the available observations in an empirical Bayesian setting. This makes it possible to use smaller neighborhoods which hopefully provides predictors with less bias.

5.2 Localized/Non-stationary Model

The predictor is based on the general Gaussian RF model, see Section 3.1, with the model parameters assessed in two different localized ways.

5.2.1 Traditional Predictor

The Loc/Non-stat/Trad predictor of r_+ with associated prediction variance is defined as:

$$\begin{aligned}\hat{r}_{NTP+}^k &= \hat{\mu}_+^k + \hat{\sigma}_+^k [G_+^k \hat{\Gamma}_o^k \omega_{o+}]^T \left[G_+^k \hat{\Gamma}_o^k \Omega_{oo} \hat{\Gamma}_o^k [G_+^k]^T \right]^{-1} G_+^k [\mathbf{r}_o - \hat{\boldsymbol{\mu}}_o^k] \\ \hat{\sigma}_{NTP+}^{k2} &= \hat{\sigma}_+^{k2} \left[1 - [G_+^k \hat{\Gamma}_o^k \omega_{o+}]^T \left[G_+^k \hat{\Gamma}_o^k \Omega_{oo} \hat{\Gamma}_o^k [G_+^k]^T \right]^{-1} G_+^k \hat{\Gamma}_o^k \omega_{o+} \right]\end{aligned}$$

with the parameter estimators defined in Section 4.1.

The expectation and variance is locally and uniquely estimated for each observation according to the General Gaussian RF model in Section 3.1. The number of parameter estimates is $2(n_o + 1)$, expectation and variance for x_+ and all observation locations. Hence the predictor is very sensitive to the estimate precision, which favors large neighborhoods. Large neighborhoods will however introduce larger bias in the predictor, which is unfavorable.

5.2.2 Shrinkage Predictor

The Loc/Non-stat/Shr predictor of r_+ with associated prediction variance is defined as:

$$\begin{aligned}\hat{r}_{NSP+}^k &= \hat{m}_+^k + \hat{s}_+^k [G_+^k \hat{S}_o^k \omega_{o+}]^T \left[G_+^k \hat{S}_o^k \Omega_{oo} \hat{S}_o^k [G_+^k]^T \right]^{-1} G_+^k [\mathbf{r}_o - \hat{\boldsymbol{m}}_o^k] \\ \hat{\sigma}_{NSP+}^{k2} &= \hat{s}_+^{k2} \left[1 - [G_+^k \hat{S}_o^k \omega_{o+}]^T \left[G_+^k \hat{S}_o^k \Omega_{oo} \hat{S}_o^k [G_+^k]^T \right]^{-1} G_+^k \hat{S}_o^k \omega_{o+} \right]\end{aligned}$$

with the parameter estimators defined in Section 4.2.

The expectation and variance for both x_+ and all observation locations are assessed in a Bayesian setting as conditional expectations given the observations in a neighborhood. Hence the empirical prior model for expectation

and variance act as regularizers in the inference. Note that this regularization will influence each individual kriging weight under this model. Hence this can be seen as a truly shrinkage kriging predictor. This regularized approach makes it possible to use smaller neighborhoods which hopefully entails less biased predictions.

5.3 CrossValidation Calibrated Predictor

Consider a specific predictor with associated prediction variance in arbitrary location $x_+ \in \mathbf{D}$:

$$\begin{aligned}\hat{r}_{\wedge+} &= \hat{\mu}_+ = \hat{E}[r(x_+) \mid \mathbf{r}_o] \\ \hat{\sigma}_{\wedge+}^2 &= \hat{Var}[r(x_+) \mid \mathbf{r}_o]\end{aligned}$$

based on the observations $\mathbf{r}_o = [r(x_{o1}), \dots, r(x_{on_o})]$.

Define the crossvalidation predictions:

$$\begin{aligned}\hat{r}_{\wedge oi} &= \hat{E}[r(x_{oi}) \mid \mathbf{r}_{o(-i)}]; \quad i = 1, \dots, n_o \\ \hat{\sigma}_{\wedge oi}^2 &= \hat{Var}[r(x_{oi}) \mid \mathbf{r}_{o(-i)}]\end{aligned}$$

where $\mathbf{r}_{o(-i)}$ entails \mathbf{r}_o with observation i removed.

The corresponding normalized crossvalidation errors are:

$$\Delta_{\wedge i} = \frac{r(x_{oi}) - \hat{r}_{\wedge oi}}{\hat{\sigma}_{\wedge oi}}; \quad i = 1, \dots, n_o$$

and define the two first moments:

$$\begin{aligned}\Delta_{\wedge} &= \frac{1}{n_o} \sum_{i=1}^{n_o} \Delta_{\wedge i} \\ \kappa_{\wedge}^2 &= \frac{1}{n_o} \sum_{i=1}^{n_o} [\Delta_{\wedge i} - \Delta_{\wedge}]^2\end{aligned}$$

Note that both Δ_{\wedge} and κ_{\wedge}^2 can be calculated, and that for a reliable predictor we want $\Delta_{\wedge} \approx 0$ and $\kappa_{\wedge}^2 \approx 1$. Actually, we can adjust the predictor with associated prediction variance such that the global normalized crossvalidation statistics are exactly 0 and 1.

Define the crossvalidation calibrated (CVC) predictor with associated prediction variance at an arbitrary location $x_+ \in \mathbf{D}$:

$$\tilde{r}_{\wedge+} = \hat{r}_{\wedge+} + \hat{\sigma}_{\wedge+} \Delta_{\wedge}$$

$$\tilde{\sigma}_{\wedge+}^2 = \kappa_{\wedge}^2 \hat{\sigma}_{\wedge+}^2$$

where the associated normalized crossvalidation errors will have the two first moments equal to 0 and 1 respectively.

Hence we are able to construct a predictor which always reproduce the favored values for the global crossvalidation statistics. The CVC predictor can be seen as a globally centered and scale corrected version of the original predictor. These calibrations are particularly beneficial for localized predictors which often lack global references.

6 Evaluation Criteria

The test criteria are based on the CVC predictors with associated prediction variances, see Section 5.3:

$$\tilde{r}_{\wedge+} = \tilde{E}[r(x_+) | \mathbf{r}_o]$$

$$\tilde{\sigma}_{\wedge+}^2 = \tilde{Var}[r(x_+) | \mathbf{r}_o]$$

The corresponding crossvalidation predictions are:

$$\tilde{r}_{\wedge oi} = \tilde{E}[r(x_{oi}) | \mathbf{r}_{o(-i)}]; \quad i = 1, \dots, n_o$$

$$\tilde{\sigma}_{\wedge oi}^2 = \tilde{Var}[r(x_{oi}) | \mathbf{r}_{o(-i)}]$$

The CVC predictions will be globally centered and scaled with respect to the normalized cross-validation errors. This means however that large deviations in predictions may be compensated by a large estimated prediction variance.

A prediction criterion which favors precise predictions is mean squared cross-validation error non-normalized:

$$PMSE = \frac{1}{n_o} \sum_{i=1}^{n_o} [r(x_{oi}) - \tilde{r}_{\wedge oi}]^2$$

We favor small values for this prediction criterion of course.

The CVC prediction variances are also globally scaled. This scaling does not ensure close agreement between large observed deviation in predictions and large estimated prediction variances, however. This agreement is indicated by the normalized crossvalidation errors being close to one, not only unity in average.

A prediction variance criterion which favors agreement between observed prediction deviations and estimated prediction variances is :

$$VMSE = \frac{1}{n_o} \sum_{i=1}^{n_o} \left[\left[\frac{r(x_{oi}) - \hat{r}_{\wedge oi}}{\hat{\sigma}_{\wedge oi}} \right]^2 - 1 \right]^2$$

We favor small values for this prediction variance criterion of course.

7 Case Studies

Two case studies are presented: US precipitation and Gamma-log data. The former consists of observations in a number of locations in the US. The latter contain observations very densely located along a vertical subsurface profile.

7.1 US precipitation

We consider a data set of yearly accumulated precipitation in locations in an area in the US, see Figure 1. The study area contains 1001 locations with observations. The data is a subset of a much larger spatio-temporal data base Data (2014), see also Johns et al (2003), and we use data from 1997 in an subarea in the south-east US.

By inspecting the data in Figure 1.b there appears to be a slight increase in values in the south-east direction, but the observation density is very high. The empirical spatial correlation function is displayed in Figure 2, and we fit a generalized exponential correlation function:

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

which represents a fairly smooth precipitation surface. This correlation function is used throughout the study.

We aim at demonstrating and evaluating the predictors defined in Section 5. The evaluation criteria presented in Section 6 are used.

The Global ordinary kriging predictor, using all 1001 observations under a stationary model with unknown expectation/variance, is used as reference. In Figure 3 the results from the corresponding CVC predictor is presented. The crossvalidation predictions in the observation locations and the associated prediction standard deviations are displayed. These predictions results from predictions based on the global data set with the observation in the actual location removed. We observe that the predictions appear with similar patterns as the observations in Figure 1, and that the standard deviations are fairly constant across the area. The standard deviations are only dependent on the location configuration of the observations used in the predictor, not on the actually observed values. This explains the somewhat higher values along the boundary since the location configurations are unfavourably,

asymmetric along the boundary. In Figure 4 the corresponding normalized crossvalidation errors and the resulting histogram are displayed. Recall that the use of the CVC predictor ensures that these errors are globally centered to zero and scaled to one. From the figure we observe some larger errors in areas with high-value predictions and that the histogram appears as somewhat peaked with heavy tails. These effects may indicate that the variance of the observations does vary across the area.

We study localized kriging predictors with a $k = 10$ neighborhood in some detail, see Figure 5 through 13. In Figure 5 the CVC predictions from the Loc/Stat/Trad predictor defined in Section 5.1.1 are displayed. The format of the figure correspond to Figure 3. The global predictions in Figure 3.a and the Loc/Stat/Trad predictions in Figure 5.a look fairly similar, which is not surprising since there is a large number of observations and a fairly smooth spatial correlation model is used. The standard deviations for global and Loc/Stat/Trad predictors do differ significantly, though, see Figure 3.b and 5.b respectively. The former does only depend on location configuration while the latter also depend on the locally estimated variance. One concern is, however, that estimated variances based on $k = 10$ observations may be unstable. We observe that there is larger dispersion in Figure 5.b than in Figure 3.b. The resulting normalized crossvalidation errors in Figure 6 for the Loc/Stat/Trad predictor appears as homogeneous across the area. More so than the corresponding errors for the Global predictor in Figure 4. By comparing the histograms from the Loc/Stat/Trad and Global predictors in Figure 6.b and Figure 4.b respectively, we observe that the former has much lighter tails.

The Loc/Stat/Shr predictor is defined in Section 5.1.2 and we study a $k = 10$ neighborhood version. The empirical Bayes approach used in the predictor include prior models on the expectation and variance inferred from the global data set. These prior models are displayed in Figure 13. The results from the CVC Loc/Stat/Shr predictor are presented in Figure 7. These results are compared to the corresponding results for the Global and Loc/Stat/Trad predictors presented in Figure 3 and 5, respectively. The crossvalidation predictions appear as fairly similar for all predictors, while the standard deviations differs significantly. The Loc/Stat/Shr results seem to lie in between the results for the two other predictors. The former predictor does actually shrink the localized estimates towards the global ones, hence this results are

not surprising. The normalized crossvalidation error results are displayed in Figure 8. The results do not deviate much from the corresponding results in Figure 6. Note, however, that the histogram of the errors in Figure 8.b are very symmetrical with very light tails.

The Loc/Non-stat/Trad predictor is defined in Section 5.2.1. This predictor has weighting of the observations that depends on the spatially varying variance estimates. The results from the CVC Loc/Non-stat/Trad predictor with $k = 10$ are displayed in Figure 9 and 10. The crossvalidation predictions do not deviate much from the other predictors. The standard deviations are even more dispersed than the ones obtained from the Loc/Stat/Trad predictor in Figure 5. The normalized crossvalidation errors from the Loc/Non-stat/Trad predictor in Figure 10 have very little dispersion, hence the histogram is very compact almost without tails.

The Loc/Non-stat/Shr predictor is defined in Section 5.2.2 and it is based on the empirical Bayes approach and uses the prior models in Figure 13. The CVC prediction results are presented in Figure 11 and 12, and the crossvalidation predictions are similar to the other predictors. The prediction standard deviations appear as a shrunk version of to the ones for the Loc/Non-stat/Trad predictor in Figure 9. By inspecting the normalized crossvalidation errors, in Figure 12, they appear with little dispersion, and the histogram is compact with fairly light tails. Recall that for the Loc/Non-stat/Shr predictor shrinkage is enforced on the observation weighting itself.

In Table 1, the evaluation criteria defined in Section 6 are displayed for the various predictors. The results for the Global predictor, corresponding to Loc/Stat/Trad $k = 1000$, are presented in the leftmost column. The mean normalized error (MNE) and mean squared normalized error (MSNE) represents statistics of the normalized crossvalidation errors prior to crossvalidation calibration (CVC). We observe that the errors are correctly centered at zero, but somewhat over-dispersed. The evaluation criteria prediction mean square error (PMSE) and variance mean square error (VMSE) are based on the CVC predictors, see Section 5.3. These CVC predictors are globally centered to zero and scaled to unity. The PMSE criterion represents prediction quality while the VMSE represents prediction variance quality, and for both criteria small values are favored. We consider the prediction criteria PMSE as more important than the prediction variance criterion VMSE. The criteria

have no absolute scale and are only suitable for comparison between various predictors.

In Table 1, column two, the evaluation criteria for the Loc/Stat/Trad $k = 10$ predictor are displayed. The non-calibrated centering appear as very good, while the variance is too large. The former follows from the localized predictors all being unbiased, while the latter is caused by the localized prediction variances lacking a global reference. The criteria PMSE and VMSE are based on the corresponding CVC predictors which are globally calibrated. The prediction quality, represented by PMSE, appears to be slightly better for the Loc/Stat/Trad predictor than for the Global predictor. The prediction variance quality, represented by VMSE, however, appears as significantly better for the former than for the latter. Hence the Loc/Stat/Trad $k = 10$ predictor dominates the Global predictor in this study.

In Table 1, column three, the evaluation criteria for the Loc/Stat/Shr $k = 10$ predictor are presented. From the MNE and MSNE values we observe good centering and over dispersion in the non-calibrated normalized crossvalidation errors. The PMSE and VMSE values based on the corresponding CVC predictor, are very encouraging. The prediction quality appears as slightly better than for the Loc/Stat/Trad and Global predictors, while the prediction variance quality seems to be significantly better than for the two other predictors. Hence the localized/shrinkage kriging predictor appears to dominate the global and localized kriging predictors in this study.

Table 1, column four and five, contain results from the Loc/Non-stat/Trad $k = 10$ and Loc/Non-stat/Shr $k = 10$ predictors, respectively. Both these predictors are well centered, and are highly over-dispersed prior to global crossvalidation calibration. The criteria PMSE and VMSE based on the corresponding CVC predictors give some mixed signals. The PMSE representing prediction quality appear as poorer than for the global and localized, stationary predictors. This lack of precision in the predictions may be explained by the large number of model parameters that are implicitly estimated. Recall that the observation weights are based on locally estimated variances. The quality of the prediction variances, represented by PMSE, appear as very favourable compared to the other predictors. It is somewhat surprising that the Loc/Non-stat/Trad predictor performs better than the Loc/Non-stat/Shr one for prediction variance assessment for $k = 10$, how-

ever.

To summarize, localized predictors appear favorably to the global predictor. This effect is most likely caused by lack of stationarity in both expectation and variance of the phenomenon under study. For a phenomenon that is less smooth and with more sparse observations the localized predictors are expected to be even more favorable. The localized, shrinkage predictors stabilize the corresponding localized, traditional predictors and provide very encouraging prediction results in the study. The previous results are all based on a localization with a $k = 10$ neighborhood. In Table 2 through 4 corresponding results for $k = 4, 8, 16$ respectively, are presented.

In Table 2, results are displayed from the Global predictor, Loc/Stat/Trad $k = 1000$, and the other predictors with $k = 4$. By Comparing prediction quality PMSE and prediction variance quality VMSE we observe that such a small neighborhood provide very unstable local estimates of expectation and variance. The corresponding predictors have poor performance compared to the Global predictor. Note, however, the improvements in prediction variance quality VMSE by using shrinkage.

Table 3 contain results from the $k = 8$ neighborhoods. The results are very similar to ones for $k = 10$, in Table 1. If we compare each predictor for different k -neighborhoods, we observe that precision quality PMSE is best for $k = 8$ while precision variance quality VMSE is best for $k = 10$. There appears to be some kind of trade-off between the criteria PMSE and VMSE.

Table 4 presents results for $k = 16$ neighborhoods. The localized, shrinkage predictors perform very favorably for this case, particularly the Loc/Stat/Shr predictor which has high prediction quality PMSE and very favorable prediction variance quality VMSE. This may indicate the neighborhood somewhat larger than $k = 10$ should be used. In fact, one may optimize the size of the neighborhood with respect to a loss criterion combining PMSE and VMSE. One may also perform predictor selection along these lines.

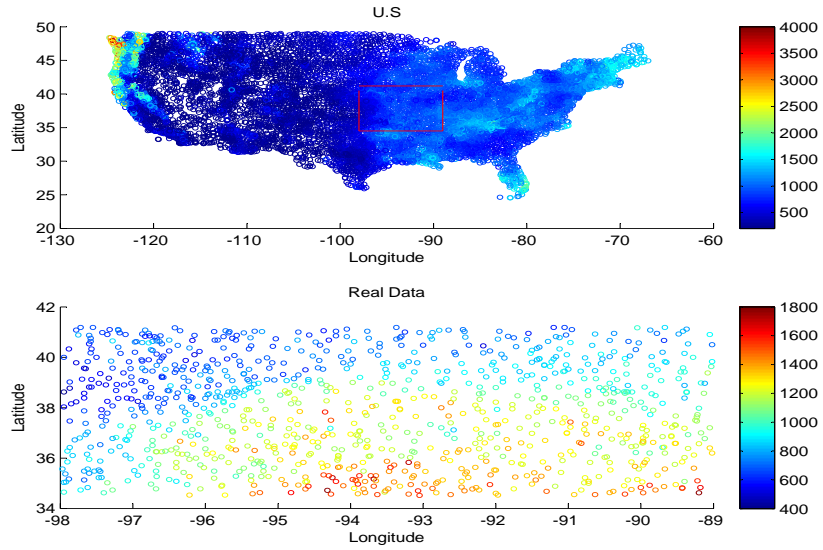


Figure 1: Annual accumulated precipitation observation in the US and sub-area studied.

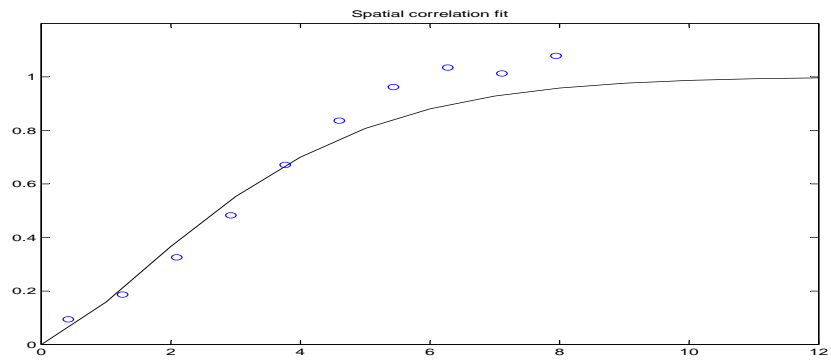


Figure 2: Spatial correlation function with estimated values.

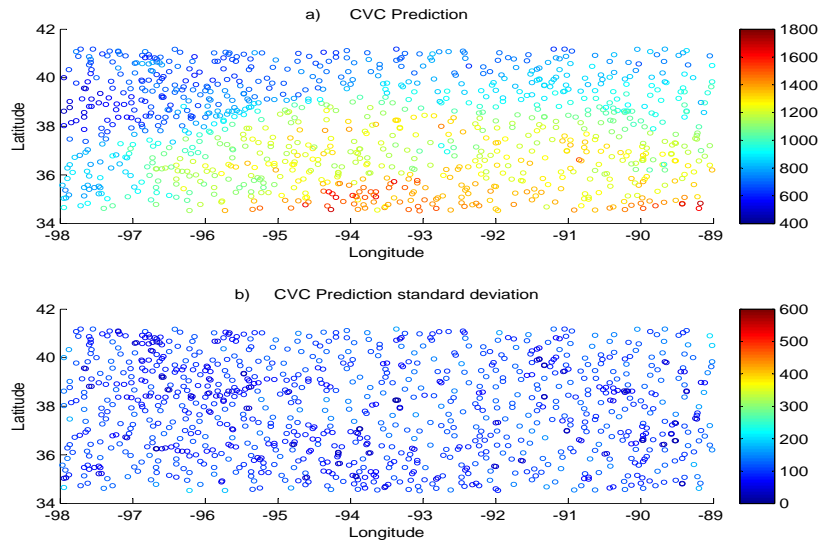


Figure 3: Global CVC predictor-ordinary kriging.

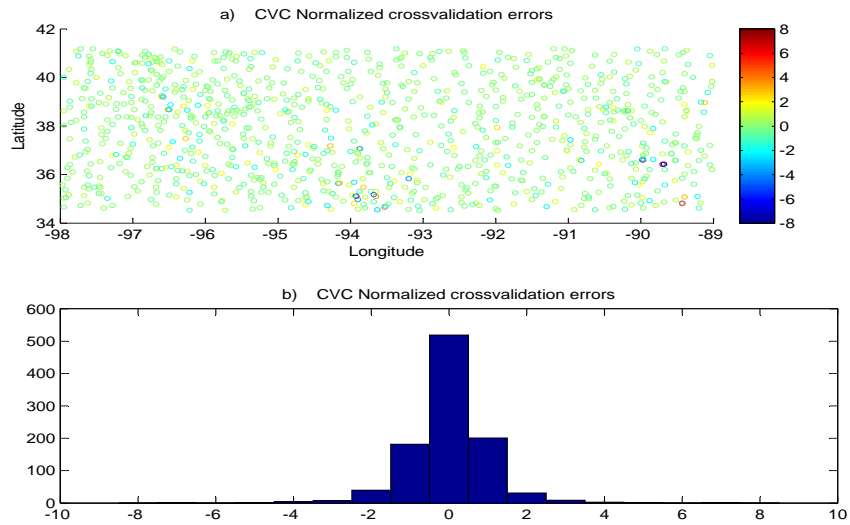


Figure 4: Global crossvalidation errors.

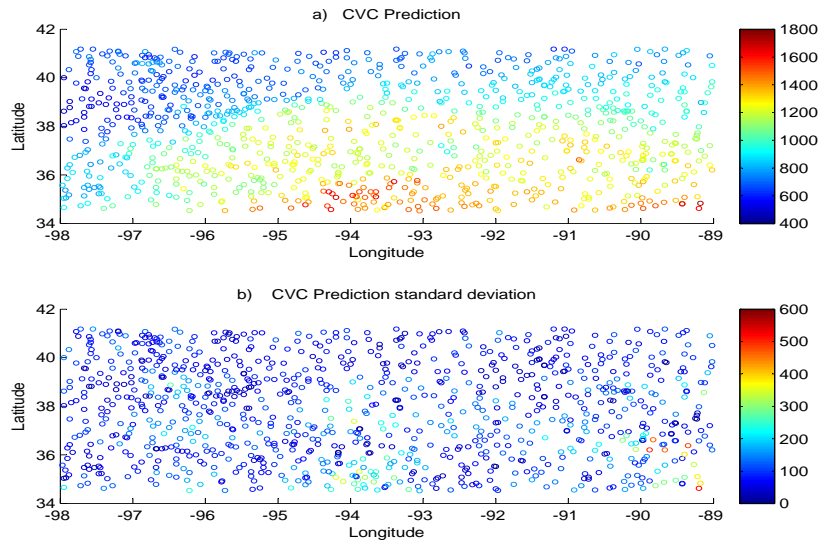


Figure 5: Loc/Stat/Trad/10 CVC predictor.

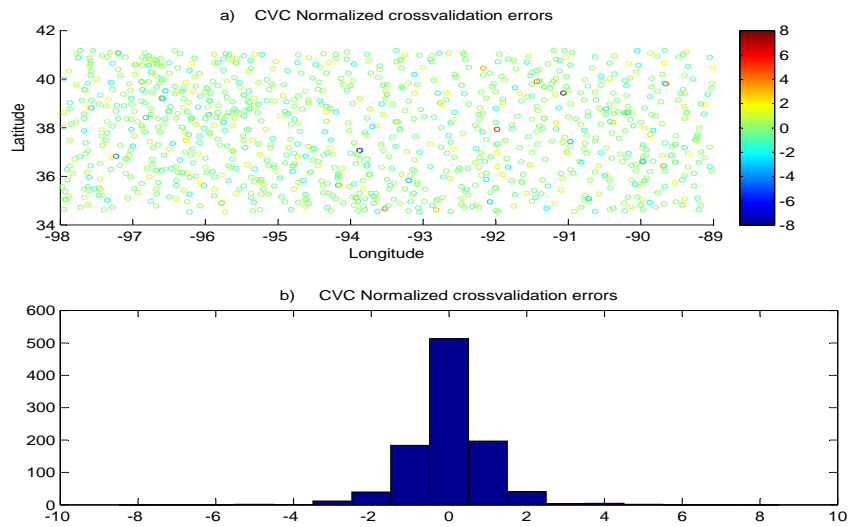


Figure 6: Loc/Stat/Trad/10 crossvalidation errors.

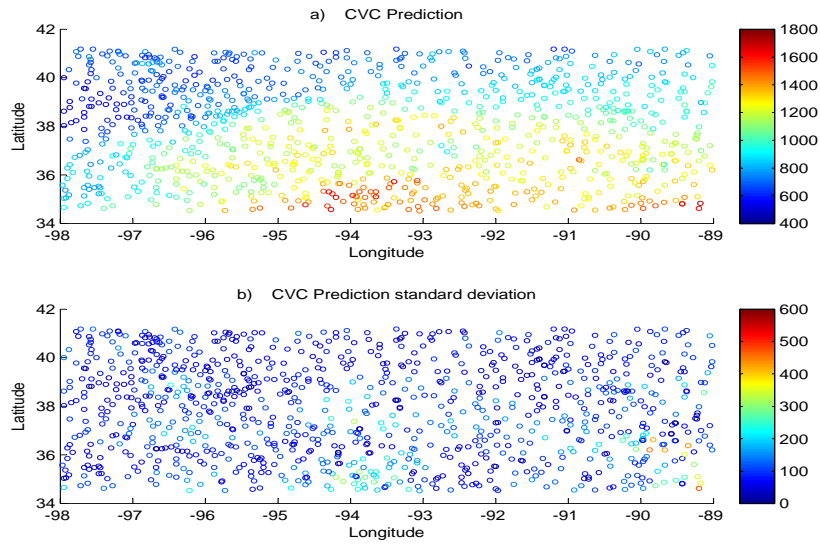


Figure 7: Loc/Stat/Shr/10 CVC predictor.

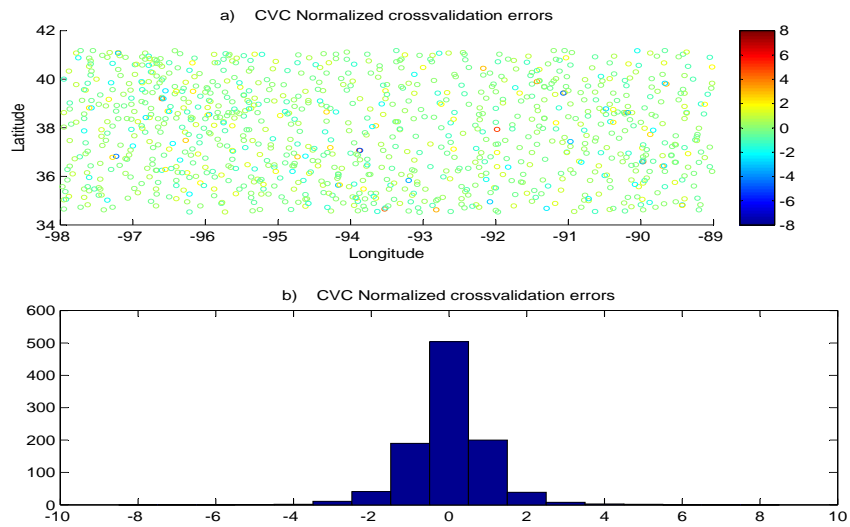


Figure 8: Loc/Stat/Shr/10 crossvalidation errors.

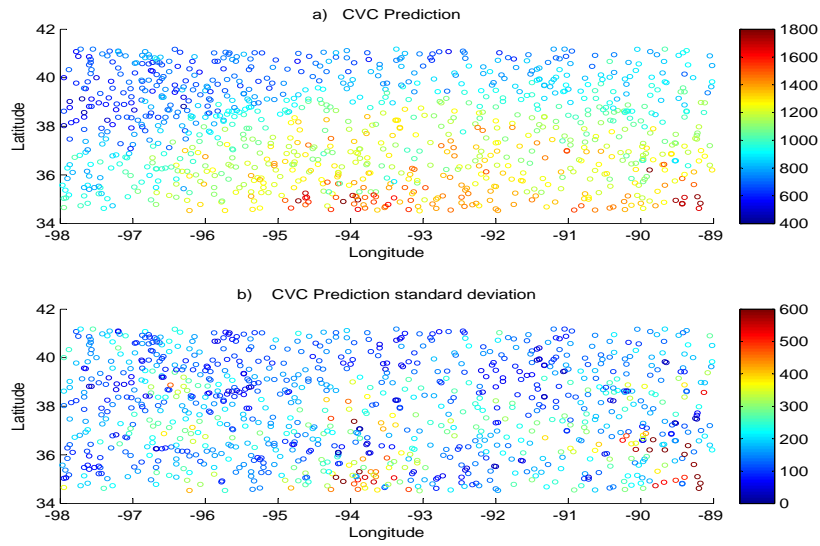


Figure 9: Loc/Non-stat/Trad/10 CVC predictor.

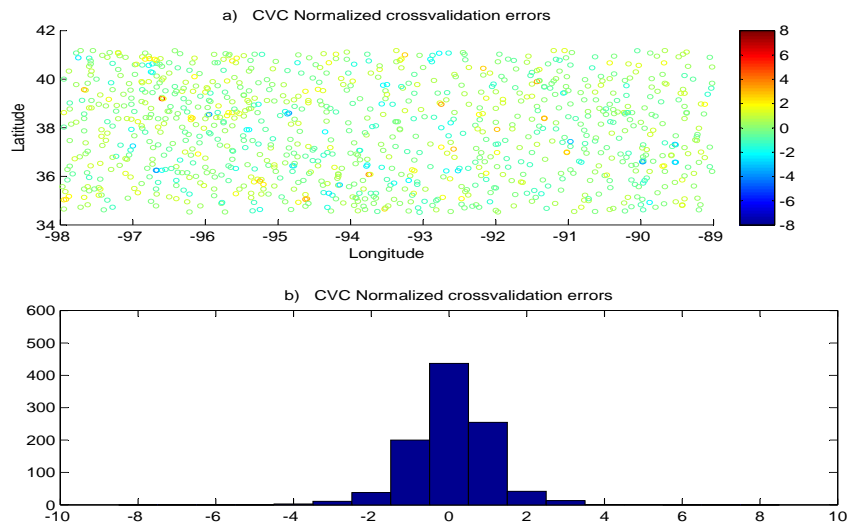


Figure 10: Loc/Non-stat/Trad/10 crossvalidation errors.

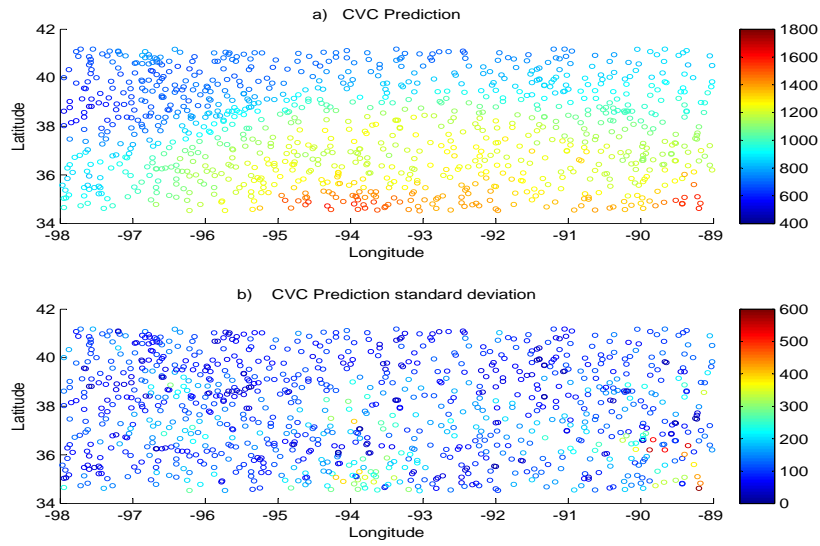


Figure 11: Loc/Non-stat/Shr/10 CVC predictor.

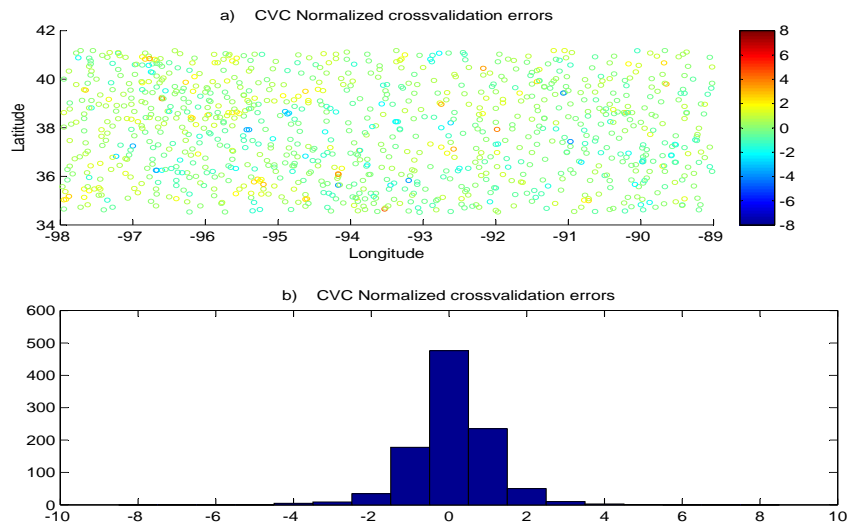


Figure 12: Loc/Non-stat/Shr/10 crossvalidation errors.

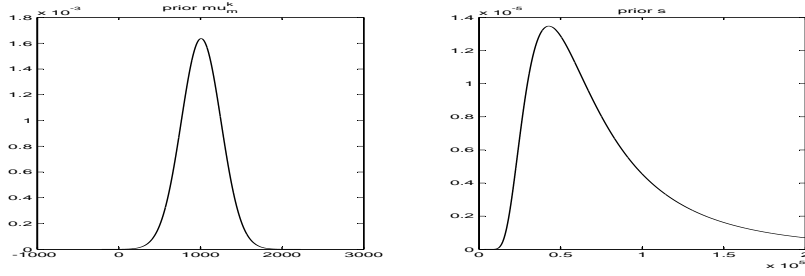


Figure 13: Priors model for expectation and variance.

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

Table 1: Precipitation crossvalidation: Mean normalized error (MNE), Mean square normalized error (MSNE), Prediction mean squared error (PMSE) and Variance mean squared error (VMSE).

Model	Localized/Stationary			Localized/Non-stationary	
	Traditional		Shrinkage	Traditional	Shrinkage
Test D.	$k = 1000$	$k = 4$	$k = 4$	$k = 4$	$k = 4$
MNE	1.0148e-17	2.3291e-18	-1.8677e-16	1.7080e-17	5.6565e-17
MSNE	1.5399	12.4782	6.8602	14.9119	7.0732
PMSE	6.8758e + 03	7.2351e + 03	7.1938e + 03	1.0271e + 04	6.2025e + 03
VMSE	9.8749	76.5237	14.7642	39.6900	10.3703

Table 2: Precipitation crossvalidation: Mean normalized error (MNE), Mean square normalized error (MSNE), Prediction mean squared error (PMSE) and Variance mean squared error (VMSE).

Model	Localized/Stationary			Localized/Non-stationary	
	Traditional		Shrinkage	Traditional	Shrinkage
Test D.	$k = 1000$	$k = 8$	$k = 8$	$k = 8$	$k = 8$
MNE	1.0148e-17	-1.5639e-17	3.9207e-17	-1.2422e-16	-1.7258e-16
MSNE	1.5399	3.6322	3.9626	8.7274	5.3399
PMSE	$6.8758e + 03$	$6.8421e + 03$	$6.8275e + 03$	$1.8969e + 04$	$8.3440e + 03$
VMSE	9.8749	8.0073	5.9935	4.2336	4.5284

Table 3: Precipitation crossvalidation: Mean normalized error (MNE), Mean square normalized error (MSNE), Prediction mean squared error (PMSE) and Variance mean squared error (VMSE).

Model	Localized/Stationary			Localized/Non-stationary	
	Traditional		Shrinkage	Traditional	Shrinkage
Test D.	$k = 1000$	$k = 16$	$k = 16$	$k = 16$	$k = 16$
MNE	1.0148e-17	1.7191e-17	1.6304e-17	-5.1907e-17	1.8367e-16
MSNE	1.5399	2.2864	2.5668	9.4840	4.9365
PMSE	$6.8758e + 03$	$6.8581e + 03$	$6.8549e + 03$	$2.5527e + 04$	$1.1038e + 04$
VMSE	9.8749	5.8588	4.8849	4.3886	4.0764

Table 4: Precipitation crossvalidation: Mean normalized error (MNE), Mean square normalized error (MSNE), Prediction mean squared error (PMSE) and Variance mean squared error (VMSE).

7.2 Gamma-log Data

In Figure 14 a Gamma ray data set from a vertical subsurface well is displayed. The data set locations are numbered as $[1, 2, \dots, 600]$. We split the data in an observations set $[1, 30, \dots, 570, 600]$ with $n_o = 21$, and a control set containing the 579 remaining data. The two sets are presented in Figure 14. Contrary to the US precipitation study, we have control data here, while we only operate in one dimension with a limited number of observations.

We use the observation set of size $n_o = 21$ in a cross-validation study and we also do prediction into the locations of the control data set. The spatial correlation function

$$\rho_r(\tau) = \exp\left\{-\frac{1}{30}\tau^{1.5}\right\}; \tau \geq 0$$

is used throughout the study. The predictors in Section 5 are evaluated by the evaluation criteria in Section 6.

In Figure 15 results from both the Global predictor and the localized predictor are displayed. The predictions look fairly similar, while the prediction variances differ. The localized predictors have prediction variances that vary with location. In the shrinkage predictors empirical prior models for expectation and variance are used, and these prior models are displayed in Figure 16. The former being Gaussian and the latter Inverse-Gamma. Note how the prediction variances in the shrinkage predictor shrink the localized ones towards the global ones.

Table 5 contains values of the evaluation criteria for the four predictors for different values of $k = 4, 8, 12$ by using crossvalidation within the observation data set of size $n_o = 21$. Note that all normalized crossvalidation errors are reasonably centered, MNE close to zero, but under-dispersed, MSNE greater than unity. The CVC predictions that normalize with respect to this under-dispersion provide the base for the prediction criterion PMSE and the prediction variance criterion VMSE. Both are favored to be small. Note that the shrinkage predictors make significantly better predictions than their traditional counterparts, ie smaller PMSE. For the prediction variance criterion VMSE we observe the same picture. The shrinkage predictors appears with smallest VMSE. The best shrinkage predictors have very small neighborhoods ± 2 which indicates fast changing model characteristics in the Gamma ray data.

Table 6 contain values of the evaluation criteria based on the differences between the control set and the different CVC predictors. The prediction criterion PMSE is almost equal for all predictors, while the prediction variance criterion VMSE appears as favourable for the traditional predictors. The latter constitutes a surprising result. Note, however, that these results are based on a very limited number of observations.

To summarize, if crossvalidation within the $n_o = 21$ observations is used in the calculations of the criteria, then we obtain results similar to the ones on the US precipitation data. The shrinkage predictors are clearly favorable

to their traditional counterparts. If, however, the criteria is based on a control set of data, shrinkage predictors do not appear as favourable. We do not understand why, but it may just be by chance in a limited dataset.

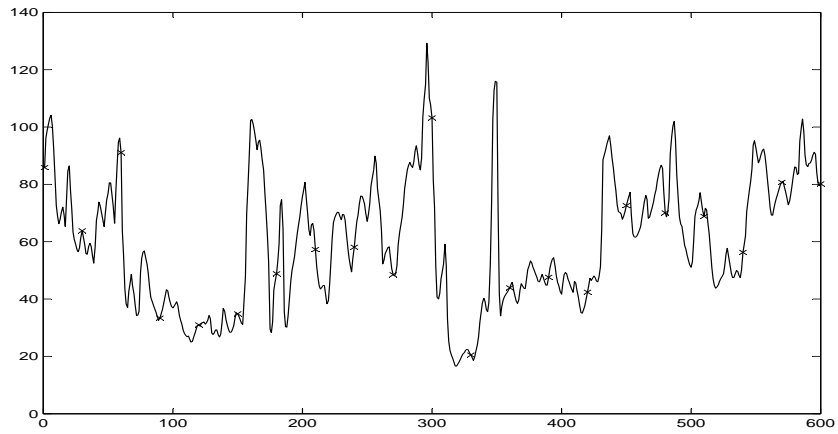


Figure 14: Gamma ray observations-with observations (*) and control values.

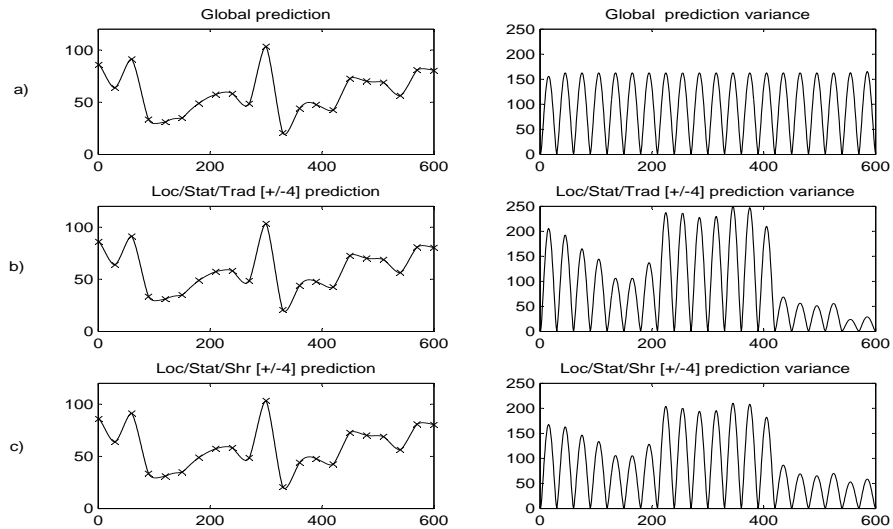


Figure 15: Gamma ray predictions and prediction variances.

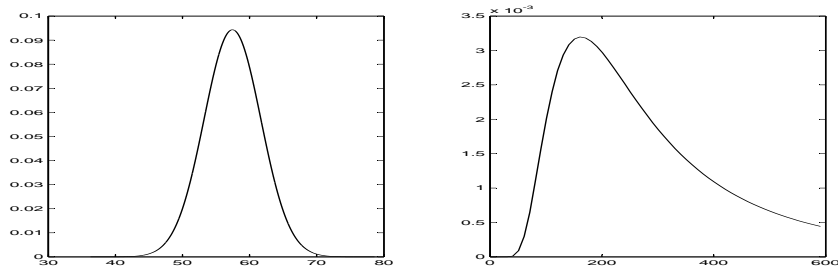


Figure 16: Prior model for expectation and variance.

Model	Localized/Stationary						Localized/Non-stationary					
	Traditional			Shrinkage			Traditional			Shrinkage		
Test D.	± 6	± 4	± 2	± 6	± 4	± 2	± 6	± 4	± 2	± 6	± 4	± 2
MNE	0.0	0.2	0.1	0.1	0.0	0.0	0.1	0.0	0.1	0.1	0.1	0.1
MSNE	3.2	4.5	3.0	2.3	2.5	1.9	4.1	6.0	4.2	2.5	2.7	1.8
PMSE	639.7	705.2	703.3	589.6	594.1	561.1	921.5	1318.4	1247.7	662.6	704.3	595.7
VMSE	5.3	6.7	4.5	4.3	5.0	3.5	3.8	4.3	2.8	3.5	3.8	2.5

Table 5: Gamma ray crossvalidation: Mean normalized error (MNE), Mean Squared normalized error (MSNE) Prediction Mean squared error (PMSE) and Variance Mean Squared error (VMSE).

Model	Localized/Stationary						Localized/Non-stationary					
	Traditional			Shrinkage			Traditional			Shrinkage		
Test D.	± 6	± 4	± 2	± 6	± 4	± 2	± 6	± 4	± 2	± 6	± 4	± 2
PMSE	336.71	331.40	330.40	340.98	335.22	333.86	346.16	340.86	335.60	345.93	342.59	358.97
VMSE	9.6033	4.9232	72.3516	19.4386	15.2300	50.1785	5.7387	3.1133	42.3435	16.4979	13.6487	57.3204

Table 6: Gamma ray control values: Prediction Mean squared error (PMSE) and Variance Mean Squared error (VMSE) .

8 Empirical Study

The prediction models defined in the previous section can be evaluated based on one realization only. We define a test design and several evaluation criteria and make the evaluation for a variety of Gaussian random field models for which the exact solutions are analytically obtainable.

8.1 Test Design and Criteria

We define a variety of 1D Gaussian RF: $\{r(x); x \in \mathbf{D} \in \mathfrak{R}^1\}$, discretised as $\mathbf{r} = \{r(x); x \in \mathcal{L}_{\mathbf{D}}\}$ where $\mathcal{L}_{\mathbf{D}} = \{1, 2, \dots, 199, 200\}$ and hence $n = 200$. The observations are obtained as $\mathbf{r}_o = \{r(x); x \in \mathcal{L}_o\}$ where $\mathcal{L}_o = \{1, 10, \dots, 190, 200\}$ and hence $n_o = 21$. We use the CVC predictors defined in Section 5 based on \mathbf{r}_o to obtain the predictions $\tilde{\mathbf{r}} = [\tilde{r}_1, \dots, \tilde{r}_{200}]^T$ and the associated prediction variances $\tilde{\boldsymbol{\sigma}}^2 = [\tilde{\sigma}_1^2, \dots, \tilde{\sigma}_{200}^2]^T$.

Note that all the Gaussian RF are analytically tractable when the model parameters are known, hence the optimal predictions $\mathbf{r}^* = [r_1^*, \dots, r_{200}^*]^T$ and associated prediction variances $\boldsymbol{\sigma}^{*2} = [\sigma_1^{*2}, \dots, \sigma_{200}^{*2}]^T$ are available.

The evaluation for a single realization is based on both comparison with the correct predictions and prediction variances and cross-validation in the 15 centrally located observations.

The evaluation criteria in the comparison with correct results are:

$$PMSC = \frac{1}{126} \sum_{\substack{i=30 \\ i \notin \mathcal{L}_o}}^{170} [\tilde{r}_i - r_i^*]^2$$

$$VMSC = \frac{1}{126} \sum_{\substack{i=30 \\ i \notin \mathcal{L}_o}}^{170} \left[\frac{\sigma_i^{*2}}{\tilde{\sigma}_i^2} - 1 \right]^2$$

We randomize over the model by averaging over 1000 realizations to obtain APMSC and AVMSC. Note that this is the ultimate criteria for goodness for the predictor and prediction variances.

The evaluation criteria in the cross-validation are PMSE and VMSE as in previous sections, but in addition we randomize over the model by averaging over 1000 realizations to obtain the criteria APMSE and AVMSE. These are the criteria we need to use when only one set of observations are available.

This study is targeted at identifying a good spatial predictor and suitable number of neighborhood observations involved in the predictor. Thus, $k = \pm 2, \pm 4, \pm 6$ observations located closest to prediction location x_+ is considered.

8.2 Test Cases

We generate four different test cases with varying expectation and variance fields but with one identical correlation field defined as:

$$\text{Corr}[r(x'), r(x'')] = \rho_r(x', x'') = \exp\left\{-\frac{1}{5}|x' - x''|^{1.5}\right\}$$

Case I-Test

This case defines a regular stationary Gaussian random field with constant expectation and variance. We use $\mu_r(x) = 10$ and $\sigma_r^2(x) = 20$, see Figure 17.a. In Figure 17.b, the optimal prediction and prediction variance are analytically obtained from the correct model for one realization of the field. The prediction and prediction variance in Figure 17.c are obtained from a stationary model with globally estimated model parameters. These results correspond to ordinary kriging. Figure 17.d displays Loc/Stat/Trad/ ± 4 predictions made according to CVC predictors defined in Section 5, while Figure 17.e displays the results from the corresponding Loc/Stat/Shr/ ± 4 CVC predictor.

Figure 18 displays the prior models for expectation and variance for one realization in the Loc/Shr predictors defined in Section 5. The parametric prior models for expectation and variance are inferred in an empirical Bayes framework as defined in Section 5.

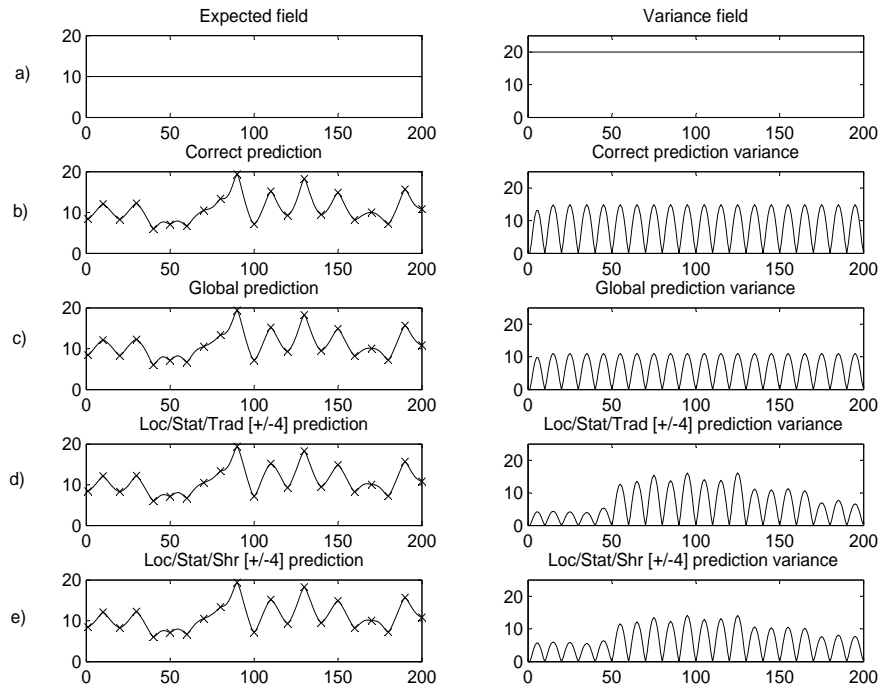


Figure 17: Case I - Predictions and prediction variances for one realization.

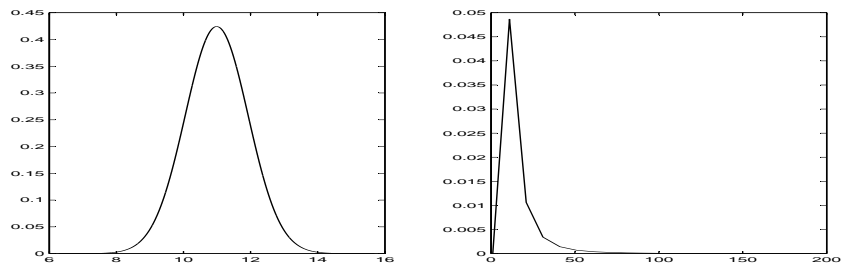


Figure 18: Case I - Prior model for expectation and variance for one realization.

The values of evaluation criteria are listed in Table 7. In Figure 19, histograms of the deviations between one realization and the corresponding predictors, for all the evaluated predictors are displayed.

Model	Localized/Stationary						Localized/Non-stationary					
	Traditional			Shrinkage			Traditional			Shrinkage		
Test D.	± 6	± 4	± 2	± 6	± 4	± 2	± 6	± 4	± 2	± 6	± 4	± 2
APMSC	0.2666	0.3484	1.3569	0.2027	0.2181	0.2696	0.3809	0.6220	2.2520	0.2318	0.2860	0.5083
AVMSC	0.4139	0.6684	43.3676	0.4149	0.5338	1.1251	0.4004	0.6353	37.1616	0.4099	0.5245	1.1335
APMSE	21.6177	22.5767	25.0502	18.5324	18.1406	16.3947	22.1637	23.4164	29.8467	18.6158	18.2514	16.5170
AVMSE	1.8692	1.9430	2.6963	1.6603	1.5907	1.3328	1.8588	1.9239	2.3750	1.6532	1.5778	1.2846

Table 7: Case I Deviation from correct predictions and crossvalidation: Average prediction mean squared correct (APMSC), Average variance mean squared correct (AVMSC), Average prediction mean squared error (APMSE) and Average variance mean squared error (AVMSE).

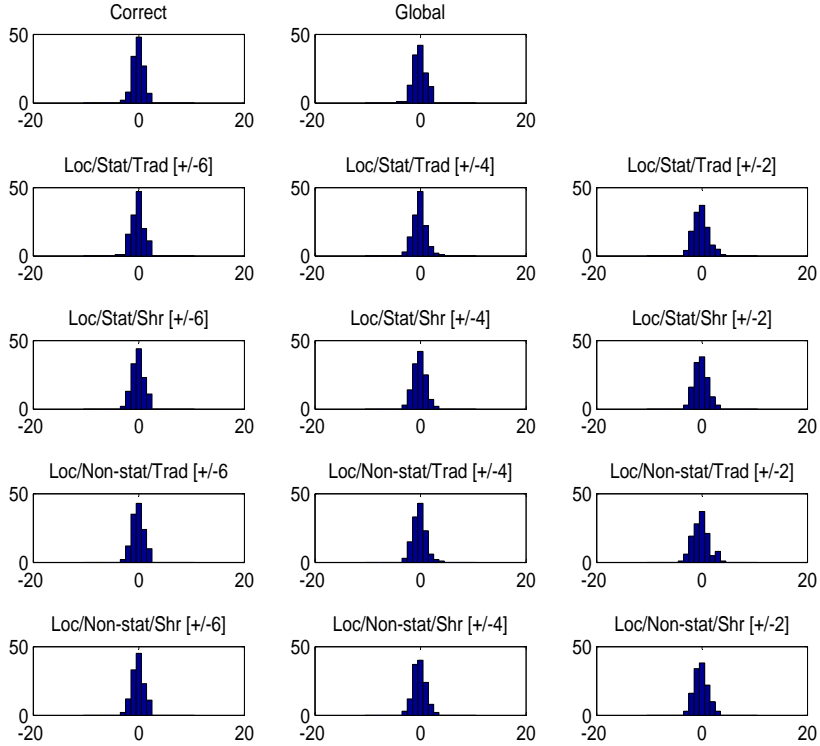


Figure 19: Case I - Histogram for normalized error for one realization.

Case I-Discussion

In Figure 17, the predictions from all CVC predictors appear fairly similar. This is not surprising since the model is stationary in both expectation and variance and the correlation function is known. The prediction variances, however, vary considerably between the predictors. The Global predictor appears with a somewhat under-estimated variance, but otherwise identical to the correct prediction variance. The localized predictors appear with locally varying variances. Note that the shrinkage predictor has variances that are in between the localized traditional and the global traditional one, and hence justifies its shrinkage label. The shrinkage is caused by the prior models on expectation and variance, see Figure 18, which are inferred in an empirical Bayesian setting.

Table 7 contain the values for the evaluation criteria APMSC/AVMSC and APMSE/AVMSE, defined at the begining of Section 8. The two former are

based on comparisons with the correct predictions which are analytically obtainable while the two later relates to crossvalidation. Recall that we favor all four criteria to be as small as possible. For each pair of criteria, the P-criterion is most important since it reflects prediction quality while the V-criterion reflects prediction variance quality. The results for the ultimate criteria APMSC/AVMSC are not surprising, all predictors improve with increasing neighborhood and the shrinkage predictors are almost consistently favorable to the corresponding traditional ones. For a model with stationarity in both expectation and variance, global predictors are optimal of course which favors large neighborhoods. Moreover, shrinkage does reduce local estimation variability which obviously improve the localized predictors. This effect is clearly observable for small neighborhoods, where shrinkage provides dramatic improvements in localized predictors.

In Table 7, also the corresponding crossvalidation criteria APMSE/AVMSE are listed, which we must rely on with only one set of observations available. Note that also these criteria consistently favor the shrinkage predictors relative to their traditional counterparts. The best shrinkage predictor with respect to the APMSE/AVMSE criterions appear with the smallest neighborhood, however. This is unfortunate, and it may be caused by overfitting to the observations.

The normalized error histograms for one realization for all CVC predictors are displayed in Figure 19. The histograms are fairly similar, but for small neighborhoods, ± 2 , we can observe that the errors are somewhat regularized by the shrinkage effect.

Case II-Test

This case defines a non-stationary Gaussian random field with varying expectation and variance. We use $\boldsymbol{\mu}_r(x) = 10 \sin(\pi \frac{x}{200})$ and $\boldsymbol{\sigma}_r^2(x) = 20 \sin(\pi \frac{x}{200})$, see Figure 20.a. In Figure 20.b, the optimal prediction and prediction variance are analytically obtained from the correct model. The prediction and prediction variance in Figure 20.c are obtained from a stationary model with globally estimated model parameters. These results correspond to ordinary kriging. Figure 20.d displays Loc/Stat/Trad/ ± 4 predictions made according to CVC predictors defined in Section 5, while Figure 20.e displays the results from the corresponding Loc/Stat/Shr/ ± 4 CVC predictor.

Figure 21 displays the prior models for Loc/Shr predictors, and these models capture the gross variability over the domain. Table 8 and Figure 22 correspond to Table 7 and Figure 19, for Case I.

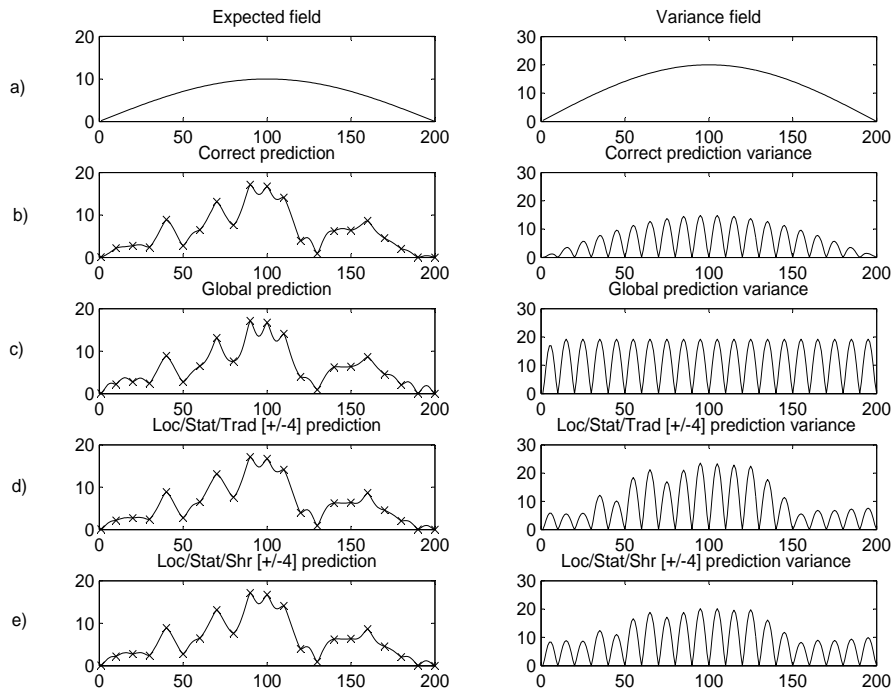


Figure 20: Case II - Predictions and prediction variances for one realization.

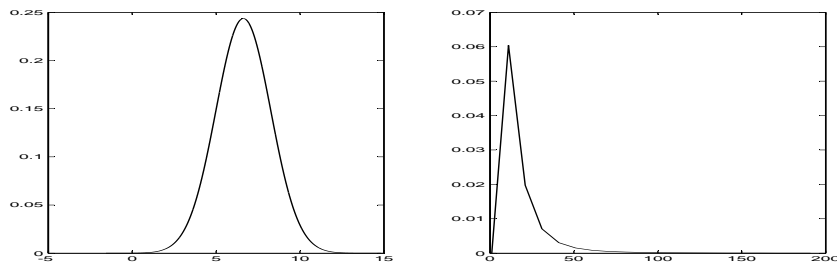


Figure 21: Case II - Prior model for expectation and variance for one realization.

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

Table 8: Case II Deviation from correct predictions and crossvalidation: Average prediction mean squared correct (APMSC), Average variance mean squared correct (AVMSC), Average prediction mean squared error (APMSE) and Average variance mean squared error (AVMSE).

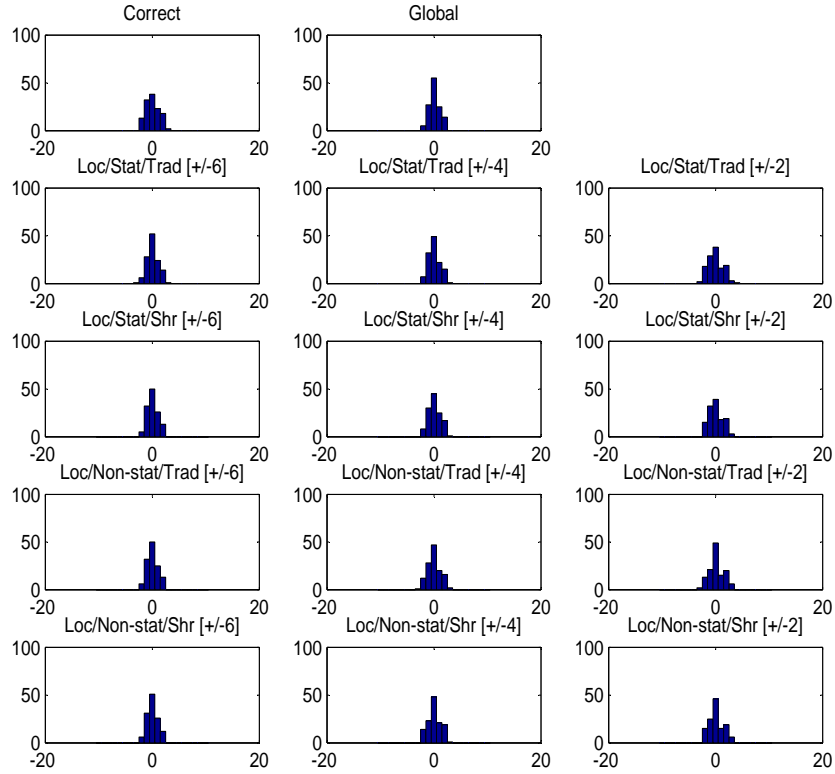


Figure 22: Case II - Histogram for normalized error for one realization.

Case II-Discussion

In Figure 20, the predictions vary considerably. The global predictor is based on a stationary model and predictions between observations are clearly biased towards the global average of the observations. The localized predictors are fairly similar and do not deviate much from the correct prediction. The prediction variances for the global predictor reflect the stationarity assumptions made and only capture the localization configuration of the observations. The prediction variances from the localized predictors capture much more of the non-stationarity of the model. Note also that the shrinkage predictor provides prediction variances in between the local traditional and global traditional predictors. The shrinkage is caused by the prior models displayed in Figure 21.

Table 8 contain the values of the evaluation criteria. Consider first the APMSC for Loc/Stat/Trad CVC predictors with varying neighborhoods. These are prediction quality relative to the correct predictions. Observe the pattern, best for ± 4 and poorer for ± 6 and ± 2 . We observe a bias/variance crossing point, since too large neighborhood provides biased estimates of the non-stationary expectation and variance, while too small neighborhood provides large estimation variance due to few observations. Consider now the APMSC criterion for corresponding shrinkage CVC predictor, Loc/Stat/Shr. This predictor regularizes the estimates of the expectation and variances and appear as clearly favorable for small neighborhoods. The improvements are so large that the bias/variance crossing point is moved to ± 2 for the shrinkage predictor. The prediction variance quality is reflected by AVMSC relative to the correct predictions. The stability of variance estimates are of course very poor, and use of shrinkage predictors has a dramatic positive effect for small neighborhood predictors. A joint assessment based on both prediction and prediction variance quality, would probably make us choose the Loc/Stat/Shr/ ± 4 CVC predictor among the stationary predictors.

Note, however, that for models which are non-stationary in both expectation and variance, localized, non-stationary predictors have the potential of being better than stationary ones. By inspecting the APMSC/AVMSC criteria for Loc/Non-stat/Trad and Loc/Non-stat/Shr predictors for varying neighborhoods, we observe that they are less favorable than the stationary ones. We believe this is caused by the need to estimate a large number of

parameters.

In practice, with one set of observations we need to select a predictor based on the crossvalidation criteria APMSE/AVMSE. Based on these criteria the shrinkage predictors appear as consistently favorable to the corresponding traditional ones. As they actually are. Moreover, we would select a stationary, shrinkage predictor prior to a non-stationary one. However, we would select the Loc/Stat/Shr ± 2 predictor since the prediction variance criterion appears as under-estimated for the crossvalidation criterion. Hence we select a too small neighborhood in the predictor.

In Figure 22, normalized error histograms for one realization are displayed. Observe the favorable shape of the global histogram, but this is caused by normalization based on severely over-estimated prediction variances. Note also that the histograms for the correct predictor and Loc/Stat/Shr with ± 4 and ± 2 are very similar. Lastly, a small regularization effect on the histograms can be seen from shrinkage.

Case III-Test

This case defines a non-stationary Gaussian random field with varying expectation and constant variance. We use $\mu_r(x) = 10 \sin(\pi \frac{x}{200})$ and $\sigma_r^2(x) = 20$, see Figure 23.a. In Figure 23.b, the optimal prediction and prediction variance are analytically obtained from the correct model. The prediction and prediction variance in Figure 23.c are obtained from a stationary model with globally estimated model parameters. These results correspond to ordinary kriging. Figure 23.d displays Loc/Stat/Trad/ ± 4 predictions made according to CVC predictors defined in Section 5, while Figure 23.e displays the results from the corresponding Loc/Stat/Shr ± 4 CVC predictor.

Figure 24 displays the prior models for Loc/Shr predictors, which represents the gross variability over the domain. Table 9 and Figure 25 correspond to Table 7 and 8, and Figure 19 and 22 for Case I and II, respectively.

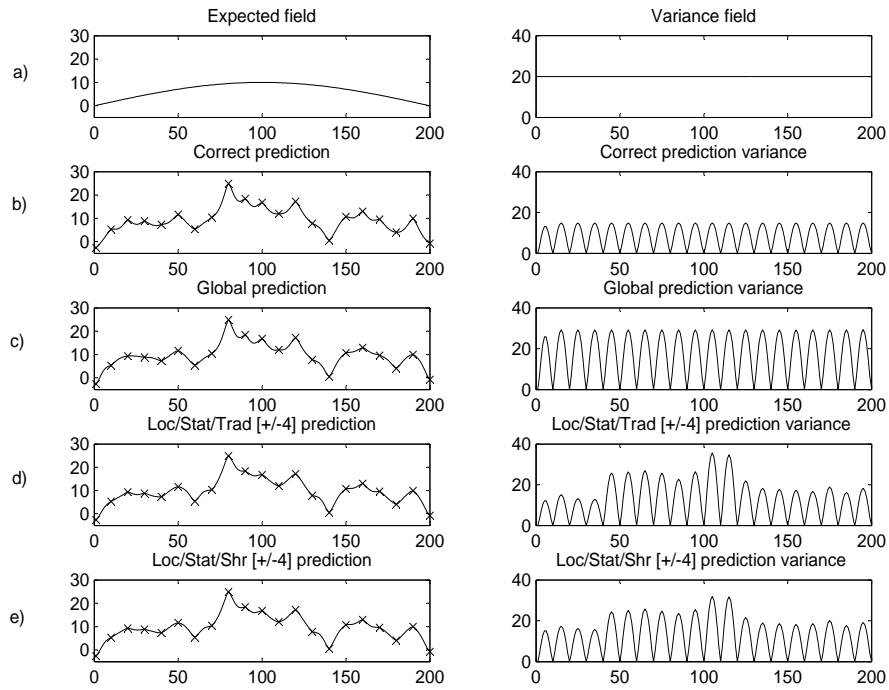


Figure 23: Case III - predictions and prediction variances for one realization.

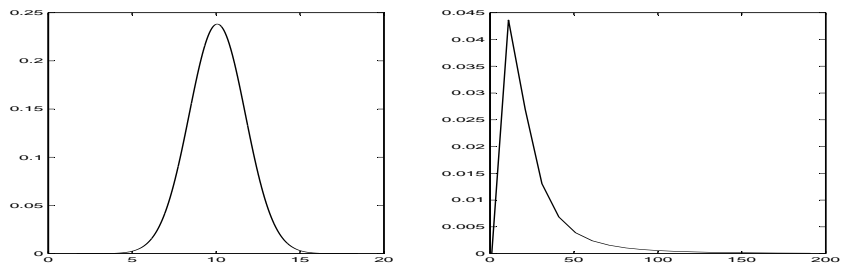


Figure 24: Case III - Prior model for expectation and variance for one realization.

Model	Localized/Stationary						Localized/Non-stationary					
	Traditional			Shrinkage			Traditional			Shrinkage		
Test D.	± 6	± 4	± 2	± 6	± 4	± 2	± 6	± 4	± 2	± 6	± 4	± 2
APMSC	0.4339	0.4027	1.5498	0.4628	0.3354	0.3152	0.5800	0.7993	3.0182	0.5014	0.4956	0.7195
AVMSC	0.5783	1.0568	687.1633	0.3919	0.5517	1.0074	0.5566	0.9922	590.5700	0.3847	0.5302	0.9891
APMSE	22.5294	22.3758	24.8330	20.6543	19.2866	17.0429	23.0463	23.1767	29.0106	20.7936	19.5368	17.3665
AVMSE	1.8404	1.9649	3.3623	1.6693	1.6335	1.3537	1.8366	1.9477	2.5569	1.6691	1.6236	1.3147

Table 9: Case III Deviation from correct predictions and crossvalidation: Average prediction mean squared correct (APMSC), Average variance mean squared correct (AVMSC), Average prediction mean squared error (APMSE) and Average variance mean squared error (AVMSE).

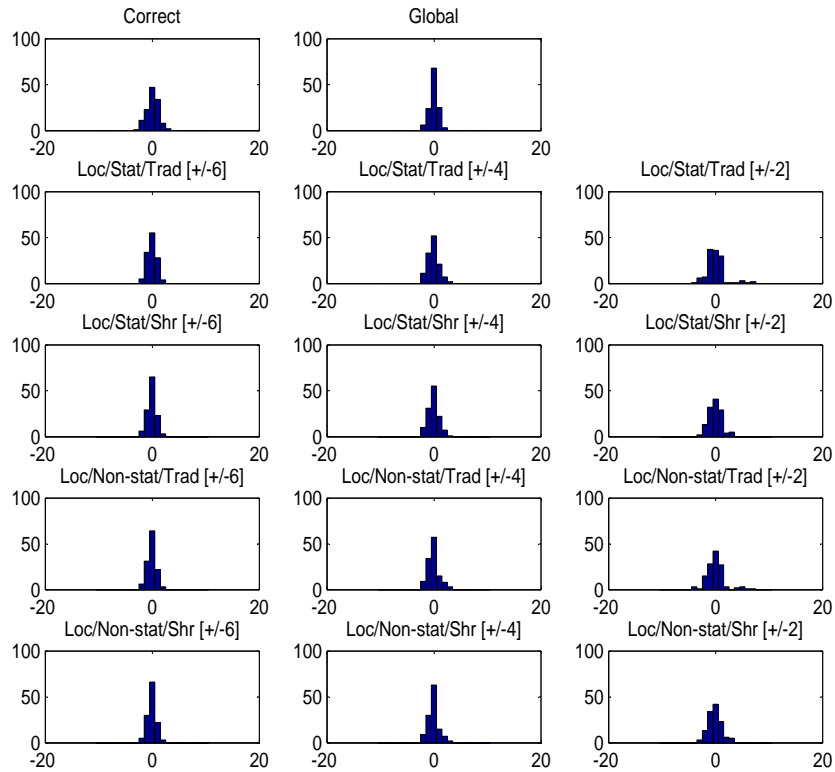


Figure 25: Case III - Histogram for normalized error for one realization.

Case III-Discussion

In Figure 23, the predictions vary somewhat, since the global predictions are biased towards the global average of the observations. The prediction variances are different because they merge variance and expectation curvature. Note the upwards bias in the variance of the global stationary model. The localized predictors appear with localized variance estimates and the shrinkage predictor is somewhat closer to the global prediction variance. The prior models in Figure 24 causes this shrinkage.

Table 9, contain the results for the evaluation criteria APMSC/AVMSC and they exposes patterns very similar to the results from the model with non-stationarity in both expectation and variance, see Figure 8. The stationary, traditional predictor makes bias/variance trade-offs in the predictor, while the stationary, shrinkage improve on the predictor by regularization of the estimation and hence reducing variance. This shrinkage make small neighborhood predictors more favorable. The prediction variances are more unstable, and we probably end up recommending the Loc/Stat/Shr/ ± 4 CVC predictor. The non-stationary predictors are less favorable due to their dependence on a large number of parameters.

In practice, with only one set of observations available, we must select a predictor based on the crossvalidation criteria APMSE/AVMSE. We would correctly favor shrinkage predictors for traditional ones and stationary predictors for non-stationary ones. However, we would select a too small neighborhood, as Loc/Stat/Shr ± 2 would be the favorable predictor.

In Figure 25, the normalized error histograms for one realization for all the CVC predictors are presented. Note in particular the localized predictors with a ± 2 neighborhood. The histograms for the shrinkage predictors are clearly regularized compared to the histograms for the traditional ones.

Case IV-Test

This case defines a Gaussian random field with constant expectation and varying variance. We use $\mu_r(x) = 10$ and $\sigma_r^2(x) = 20 \sin(\pi \frac{x}{200})$, see Figure 26.a. In Figure 26.b, the optimal prediction and prediction variance are analytically obtained from the correct model. The prediction and prediction

variance in Figure 26.c are obtained from a stationary model with globally estimated model parameters. These results correspond to ordinary kriging. Figure 26.d displays Loc/Stat/Trad/ ± 4 predictions are made according to CVC predictors defined in Section 5, while Figure 26.e displays the results from the corresponding Loc/Stat/Shr/ ± 4 CVC predictor.

Figure 27 displays the prior models for Loc/Shr predictors, and these models capture the gross variability over the domain. Table 10 and Figure 28 correspond to Table 7, 8 and 9, and Figure 19, 22 and 25, for Case I, II and III, respectively.

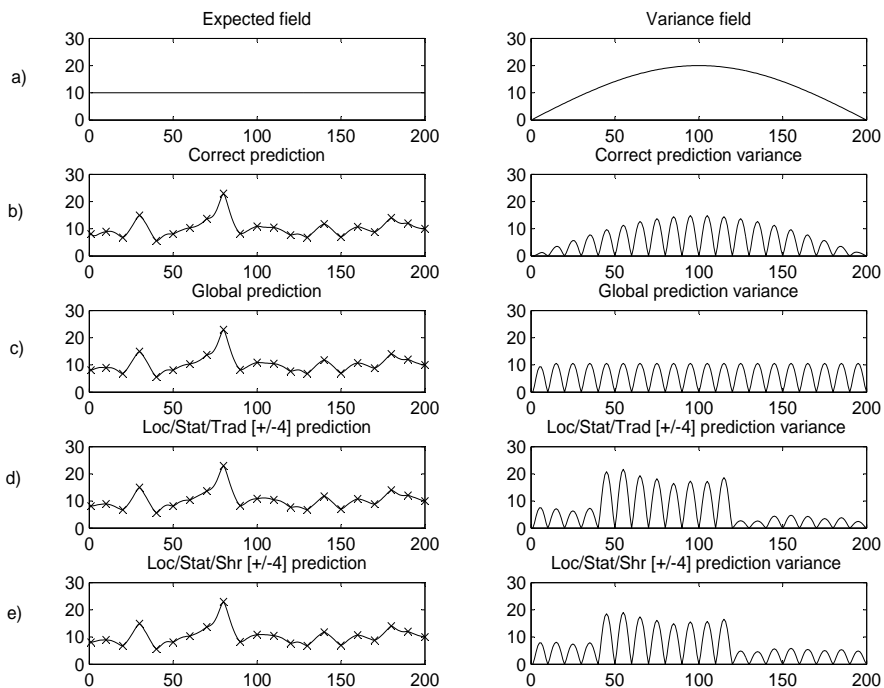


Figure 26: Case IV - Predictions and prediction variances for one realization.

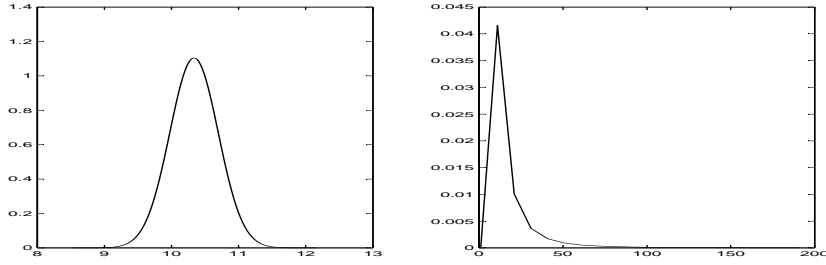


Figure 27: Case IV - Prior model for expectation and variance for one realization.

Model	Localized/Stationary						Localized/Non-stationary					
	Traditional			Shrinkage			Traditional			Shrinkage		
Test D.	± 6	± 4	± 2	± 6	± 4	± 2	± 6	± 4	± 2	± 6	± 4	± 2
APMSC	0.1891	0.2810	1.0623	0.1357	0.1510	0.2023	0.2576	0.4462	1.6301	0.1510	0.1886	0.4097
AVMSC	0.3821	0.7524	33.0924	0.4556	0.6696	1.9722	0.3689	0.7104	26.8692	0.4506	0.6606	1.9826
APMSE	16.9439	17.5722	19.1127	14.5025	14.1418	12.7449	17.3596	18.2110	22.9114	14.5563	14.2143	12.9016
AVMSE	1.8534	1.8717	2.7330	1.6950	1.5993	1.3174	1.8438	1.8514	2.3343	1.6881	1.5868	1.2709

Table 10: Case IV Deviation from correct predictions and crossvalidation: Average prediction mean squared correct (APMSC), Average variance mean squared correct (AVMSC), Average prediction mean squared error (APMSE) and Average variance mean squared error (AVMSE).

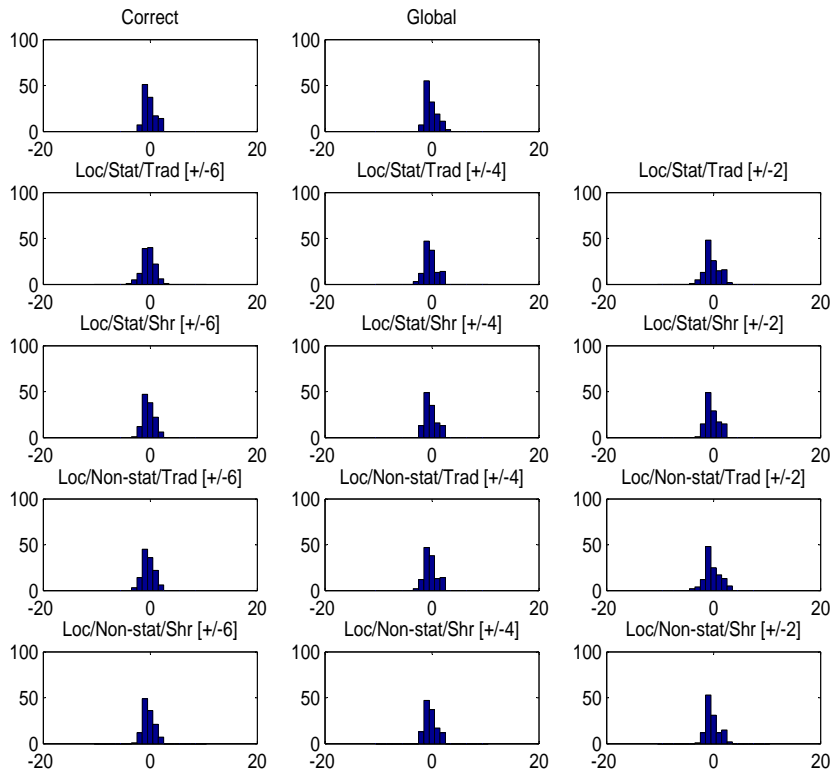


Figure 28: Case IV - Histogram for normalized error for one realization.

Case IV-Discussion

In Figure 26, all predictions appear as very similar. The prediction variances are of course very different since the global predictor estimate one global variance. The shrinkage predictor regularizes the prediction variances of the traditional predictor. The prior models are presented in Figure 27.

Table 10 contain the values of the evaluation criteria APMSC/AVMSC. Large neighborhoods are clearly favorable for all CVC predictors both for prediction and prediction variance criteria. Moreover, shrinkage predictors are clearly favorable to their traditional counterparts, and stationary are favorable to non-stationary. If we select predictor based on the crossvalidation criteria APMSE/AVMSE, we would select $\text{Loc/Stat/Shr}/\pm 2$, hence correct predictor type but with too small neighborhood. All histograms in Figure 28 are

very similar.

Summary

For a model with stationarity in both expectation and variance the global stationary predictor is of course favorable. Only the global expectation and variance must be estimated. If, however, localized predictors are required, for example to reduce computational demands, then localized shrinkage CVC predictors are clearly favorable.

For models with non-stationary expectation and variances across the domain of study, localized predictors will often appear as clearly favorable to global ones. Moreover, localized, shrinkage CVC predictors seem to be favorable to localized, traditional ones since it regularizes the expectation and variance estimates and hence can operate with larger localization hence smaller neighborhoods. The critical factor appears to be non-stationarity in expectation. If the expectation vary across the domain of study, global predictors can be severely biased towards the average of the observations and the variance estimates are upwards biased by the interplay of expectation and variance. Localized, shrinkage CVC predictors appear as robust towards non-stationary expectations. Non-stationary variances across the domain of study appears to have less influence on the predictors.

In practice, with only one set of observations available, we must rely on crossvalidation criteria in the selection of optimal CVC predictors. We will correctly select localized, stationary, shrinkage predictors, but we will tend to select too small neighborhoods.

9 Conclusion

Spatial prediction is usually made under Gaussian assumptions by kriging. In order to robustify the predictor towards lack of spatial stationarity, localized kriging, which includes only observations in a neighborhood, is frequently used. Bias/variance trade-off must be made in order to specify the local neighborhood. We introduce a shrinkage version of kriging in a Bayesian setting, where estimates in a local neighborhood are regularized by a global prior model. This prior model is assessed in an empirical Bayes tradition from the complete, global set of observations.

Two versions of the shrinkage kriging predictor are defined. One stationary version perform regularization only of the model parameter estimates, while the other non-stationary version make regularized estimates of both parameter estimates and kriging weights. Further we define crossvalidation calibrated (CVC) predictors which empirically calibrates the predictor for centering and scale.

The localized/shrinkage CVC predictors are compared to localized/traditional kriging CVC predictors in an empirical simulation study. The experimental design include 1D Gaussian random fields with varying expectation and variance trends. Deviation measures between the predictors specified above and the correct predictions and prediction variances which are analytically assessible under the model specification are compared. The major conclusions are:

- localized/shrinkage kriging appears to be almost uniformly superior to the corresponding localized/traditional kriging - for relatively small neighborhoods.
- localized/shrinkage kriging with a regularizer only on model parameter estimates appears superior to the version that also regularizes the kriging weight. In the latter a large number of parameters need to be inferred and uncertainty related to this estimation detoriate the predictor.
- cases with both non-stationary expectation and variance make the localized/shrinkage predictor more favorable to the localized/traditional

kriging predictors. Large curvature in the expected trend seems to be the important factor.

- optimal size of the neighborhood is not specifically studied since only one correlation function and one observation design is evaluated. The optimal neighborhood shall normally increase with more smoothness in expectation and variance trends and longer range in the correlation function.

A set of criteria for evaluating the localized CVC predictors based on cross-validation are developed. This set of criteria can be computed from the set of observations only. The characteristics of this criteria set are explored in the empirical study, and the major conclusions are:

- the crossvalidation criteria can be used to identify the favorable predictor type most frequently being localized, stationary, shrinkage predictors.
- the optimal neighborhood of the favorable predictor seems to be underestimated, hence too small neighborhoods are selected.

We evaluated the localized predictors on two different real datasets: Observations of annually cumulated precipitation in locations in an sub-area of the US, and Gamma ray recordings along a vertical well through the subsurface.

The findings are:

- in the precipitation study, 1001 observations are available, and cross-validation based evaluation criteria are used. Localized CVC predictors are found to be clearly favorable to the global ordinary kriging predictor. Since a model non-stationary in both expectation and variance appears as most representative for the observation, this conclusion make sense. The localized, shrinkage predictors are uniformly favorable to the corresponding localized traditional ones for the neighborhoods being studies. Moreover, the stationary, shrinkage predictors appear as favorable to the non-stationary shrinkages ones. Lastly, a neighborhood of about 10 observations are found to be suitable.
- in the Gamma ray study, one observation set of size 21 and a control set of 579 are used. By using the crossvalidation based criteria, we conclude

that the localized, stationary, shrinkage predictors are clearly favorable to the other predictors. The evaluation results from the control set of data, can not confirm this conclusion, though. We have no clear explanation for this result.

We conclude that whenever a localized predictor is preferred - either due to non-stationaries in the expectation and variance fields, or due to need for computational efficiency - one should use a localized, stationary, shrinkage CVC predictor. Since this predictor is also reliable for stationary models, the recommendation may as well be to always use localized, stationary, shrinkage CVC predictors whenever there are a fair number of observations.

References

- [1] Chiles, J.P and Delfiner, J.P; 1999: *Geostatistics: Modeling spatial uncertainty (wiley series in probability and statistics)*, Wiley, New York, 695p.
- [2] Data; 2014: <http://www.image.ucar.edu/GSP/Data/US.monthly.met/>
- [3] Efron, B. and Morris, C.; 1973: "Stein's estimation rule and its competitors-An empirical bayes approach", *Journal of the American Statistical Association*, Vol. 68, No. 341, pp. 117-130.
- [4] Fuentes M. A.; 2005: "formal test for nonstationarity of spatial stochastic processes", *Journal of Multivariate Analysis*, 96, pp. 30-55.
- [5] Gribov, A. and Krivoruchko, K. ; 2004: "Geostatistical Mapping with Continuous Moving Neighborhood", Published in *Mathematical Geology*, Volume 36, Number 2.
- [6] Johns, C. , Nychka, D., Kittel, T. , and Daly, C.; 2003: "Infilling sparse records of spatial fields", *Journal of the American Statistical Association*, 98(464), pp. 796-806.
- [7] Journel, A.G and Huijbregts, C.J; 1978: *Mining Geostatistics*, Academic Press Inc, London.
- [8] Mardia, K. V., Kent, J. T., and Bibby, J. M.; 1979: *Multivariate analysis*, Academic press, London, 518p.
- [9] Stein, M. L. ; 2002: "The Screening Effect in Kriging", *The Annals of Statistics*, Vol. 30, No. 1, pp. 298-323
- [10] Røislien, J. and Omre, H.; 2006: "T-distributed random fields: A parametric model for heavy-tailed well-log data," *Mathematical Geosciences*, 38(7), pp. 821-849.

Appendices

A Posterior pdf

The conditional pdf for $[s^2 \mid \mathbf{r}_o]$ is:

$$\begin{aligned}
 f(s^2 \mid \mathbf{r}_o) &= \text{const} \times f(\mathbf{r}_o \mid s^2) f(s^2) \\
 &= \text{const} \times [2\pi]^{-\frac{n_o}{2}} |s^2 [\Omega_{oo} + \tau_m i_{n_o} i_{n_o}^T]|^{-\frac{1}{2}} \\
 &\quad \times \exp \left[-\frac{1}{2} \left[(\mathbf{r}_o - \mu_m i_{n_o})^T [s^2 [\Omega_{oo} + \tau_m i_{n_o} i_{n_o}^T]]^{-1} (\mathbf{r}_o - \mu_m i_{n_o}) \right] \right] \\
 &\quad \times \frac{\gamma_s^{\xi_s}}{\Gamma(\xi_s)} [s^2]^{-(\xi_s+1)} \exp [-\gamma_s [s^2]^{-1}] \\
 &= \text{const}' \times [s^2]^{-(\xi_s + \frac{n_o}{2} + 1)} \\
 &\quad \times \exp \left\{ - \left[\gamma_s + \frac{1}{2} \left[(\mathbf{r}_o - \mu_m i_{n_o})^T [\Omega_{oo} + \tau_m i_{n_o} i_{n_o}^T]^{-1} (\mathbf{r}_o - \mu_m i_{n_o}) \right] \right] [s^2]^{-1} \right\} \\
 &= IG(\xi_{s|o}, \gamma_{s|o})
 \end{aligned}$$

with

$$\xi_{s|o} = \xi_s + \frac{n_o}{2}$$

$$\gamma_{s|o} = \gamma_s + \frac{1}{2} \left[(\mathbf{r}_o - \mu_m i_{n_o})^T [\Omega_{oo} + \tau_m i_{n_o} i_{n_o}^T]^{-1} (\mathbf{r}_o - \mu_m i_{n_o}) \right]$$