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

July 23, 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 accumulated precipitation. A crossvalidation 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 confirms that the shrinkage predictor is favorable to the traditional one. Moreover, the crossvalidation criterion is found to be suitable for selection of predictor. Lastly, the computational demands of localized predictors are very modest, hence the localized/shrinkage predictors are suitable for large scale spatial prediction problems.
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 \( \{ r(x); x \in D \subset \mathbb{R}^m \} \) where \( r(x) \in \mathbb{R}^1 \) is the variable of interest while \( x \) is a reference variable in the domain \( D \). The challenge is to predict the regionalized variable from a set of observations \( \mathbf{r}_o = [r(x_1), \ldots, r(x_n)]; x_1, \ldots, x_n \in D \). We define the predictors in a probabilistic setting and can also obtain associated predictor uncertainties.

The classical probabilistic approach to spatial prediction is kriging, see Journel and Huijbregts (1978) and Chiles and Delfiner (1999). The traditional ordinary kriging predictor is based on a stationary model for the regionalized variable, with spatially constant expectation and variance, and a shift invariant spatial correlation function. Localized predictors, local neighborhood kriging, see Chiles and Delfiner (1999) 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, see Stein (2002), and this effect may robustify the localization. In Anderes and Stein (2011) a weighted localized approach is defined and demonstrated to be useful for non-stationary random fields.

Recent developments in computer and sensor technology have provided enormous spatio-temporal sets of observations, see 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 we define spatial predictors which are robust with respect to deviations from assumptions about spatial stationary. We define a Gaussian random field with spatially varying expectation and variance. The spatial correlation function is shift invariant and known. We assess the expectation and variance locally by a sliding window approach. In traditional local neighborhood kriging this assessment is done by some maximum likeli-
hood procedure. The new feature of the study is that these local assessments are done by shrinkage estimators in an empirical Bayes setting, inspired by the study of Efron and Morris (1973). The hierarchical, Gaussian random field model discussed in 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. We denote the spatial predictor localized/shrinkage kriging.

In the two first sections, various Gaussian random field models for the regionalized variable \( \{ r(x); x \in D \} \) 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 appear with shrinkage. The next section defines the evaluation criteria used in the comparison of the predictors. In the next to last section the five predictors are compared on a real set of annual accumulated precipitation data and in a synthetic simulation study. Thereafter, in the last section, the conclusions are forwarded. The paper summarises the major findings in an extended study, reported in Asfaw and Omre (2014).

2 Predictor Models

The spatial predictors will be based on probabilistic models for the regionalized variable \( \{ r(x); x \in D \} \).

In traditional kriging prediction \( \{ r(x); x \in D \} \) is associated with a stationary Gaussian random field, see Chiles and Delfiner (1999). This assumption entails that:

\[
\begin{align*}
E[r(x)] &= \mu; \quad \forall x \in D \\
Var[r(x)] &= \sigma^2; \quad \forall x \in D \\
Corr[r(x'), r(x'')] &= \rho(x' - x''); \quad \forall x', x'' \in D
\end{align*}
\]

where expected level \( \mu \in \mathbb{R}^1 \), variance level \( \sigma^2 \in \mathbb{R}_+^1 \) and 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 \( [\mu, \sigma^2, \rho(.)] \) from the set of observations \( r_o \). This traditional kriging model may be extended to have an
expectation surface \( \mu(x) = \sum_{l=1}^{L} \alpha_l g_l(x) \) where \( \{g_l(x); x \in D\} \); \( l = 1, ..., L \) are known basis surfaces while \( \alpha = (\alpha_1, ..., \alpha_L) \) 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 \( \rho(.,.) \), maximum likelihood estimates based on \( r_o \) is analytically assessable for the other model parameters, under both these model assumptions.

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

\[
\begin{align*}
E[r(x)] &= \mu(x); \quad \forall x \in D \\
Var[r(x)] &= \sigma^2(x); \quad \forall x \in D \\
Corr[r(x'), r(x'')] &= \rho(x' - x''); \quad \forall x', x'' \in D
\end{align*}
\]

with \( \mu \in \mathbb{R}^1 \), \( \sigma^2 \in \mathbb{R}^1_+ \) and \( \rho(x', x'') \) being positive definite. There is obviously lack of shift invariance under these assumptions. We can expect that inference of the spatial model parameters \( \{\mu(x); x \in D\}, \{\sigma^2(x); x \in D\} \) and \( \{\rho(x', x''); \quad \forall x', x'' \in D\} \), based on \( r_o \) is complicated. If we fix the correlation function \( \rho(.,.) \) we may use localized estimators in a kernel spirit to assess the spatial model parameters \( \mu(.) \) and \( \sigma^2(.) \). When selecting the size of localization we face a bias/variance trade-off. Large local neighborhood introduce bias in the estimators due to smoothing while small neighborhoods introduce instability in the estimators due to sensoring of observations. In the current study we adress this bias/variance trade-off by defining shrinkage estimators in an empirical Bayes setting.

In order to define the shrinkage estimators we introduce a stationary, hierarchical Gaussian random field model for a local neighborhood \( D_+ \) around an arbitrary \( x_+ \in D \), i.e for \( \{r(x); x \in D_+\} \). In this model the expected level \( m \) and the variance level \( s^2 \) are considered to be random variables. Moreover, \( \{[r(x) | m, s^2]; x \in D_+\} \) is defined to be a stationary Gaussian random field, hence:

\[
\begin{align*}
E[r(x) | m, s^2] &= m; \quad \forall x \in D_+ \\
Var[r(x) | m, s^2] &= s^2; \quad \forall x \in D_+ \\
Corr[r(x'), r(x'') | m, s^2] &= \rho(x' - x''); \quad \forall x', x'' \in D_+
\end{align*}
\]
By assuming that $\rho(\cdot)$ is known, and that the model parameters $[m \mid s^2]$ and $s^2$ have prior models which are Gaussian and Inverse Gamma respectively, with:

$$
E\left[ m \mid s^2 \right] = \mu_m \\
Var\left[ m \mid s^2 \right] = \tau_m s^2 \\
E\left[ s^2 \right] = [\xi_s - 1]^{-1} \gamma_s \\
Var\left[ s^2 \right] = [(\xi_s - 1)^2(\xi_s - 2)]^{-1} \gamma_s^2
$$

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

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

### 3 Inference of Model Parameters

In order to use the probabilistic models for $\{r(x); x \in D\}$ to spatial prediction the model parameters must be assessed from the set of observations represented by the $[n \times 1]$-vector $r_o = [r(x_1), ..., r(x_n)]^T = [r_1, ..., r_n]^T$.

Throughout the study we assume the spatial correlation function $\{\rho(x', x''); \forall x', x'' \in D\}$ to be known, with associated inter-observation correlation $[n \times n]$-matrix $\Omega_{oo}$. This correlation model must of course also be inferred in one way or the other, but in our study we do not account for the uncertainty related to this.

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

$$
\hat{\mu} = [i_n^T \Omega_{oo}^{-1} i_n]^{-1} i_n^T \Omega_{oo}^{-1} r_o
$$
\[ \hat{\sigma}^2 = \frac{1}{n} [r_o - \hat{\mu}_o]^T \Omega_{oo}^{-1} [r_o - \hat{\mu}_o] \]

For a general Gaussian random field, given the correlation function, the assessments of the spatial model parameters \( \{\mu(x); x \in D\} \) and \( \{\sigma^2(x); x \in D\} \), are more complicated. We prescribe localized estimators. Consider an arbitrary location \( x_+ \in D \) and parameterize the localization by the \( k \) observations in \( r_o \) localized closest to \( x_+ \), and denote it as \( k \)-localization. Define a binary-selection \( [k \times n] \)-matrix \( G^k_+ \) such that \( G^k_+[r_o] \) is a \( [k \times 1] \)-vector containing the \( k \) observations in the \( k \)-localization of \( x_+ \). Note that \( G^k_+ \) easily can be extended to also account for favorable configurations of observations around \( x_+ \). The \( k \)-localized maximum likelihood estimators for the model parameters are:

\[
\hat{\mu}_+^k = \hat{\mu}^k(x_+) = [i^T_k [G^k_+ \Omega_{oo}[G^k_+]^{-1} G^k_+ r_o][i^T_k [G^k_+ \Omega_{oo}[G^k_+]^{-1} i^T_k]^{-1} \]
\]

\[
\hat{\sigma}^2_k = \hat{\sigma}^{k2}(x_+) = \frac{1}{k} [G^k_+ r_o - \hat{\mu}_+^k i^T_k]^T [G^k_+ \Omega_{oo}[G^k_+]^{-1} (G^k_+ r_o - \hat{\mu}_+^k i^T_k)]^{-1} \]

By letting \( x_+ \) coincide with the observation locations \( [x_1, ..., x_n] \) we obtain the estimators for observation expectation \( [n \times 1] \)-vector and for diagonal standard deviation \( [n \times n] \)-matrix respectively.

\[
\hat{\mu}_o^k = [\hat{\mu}_1^k, ..., \hat{\mu}_n^k]^T
\]

\[
\hat{\Gamma}_o^k = \begin{bmatrix}
\hat{\sigma}_1^k & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & \hat{\sigma}_n^k
\end{bmatrix}
\]

For a stationary, hierarchical Gaussian random field, given the correlation function, we need to assess the model parameters of the prior model \( \{\mu_m, \tau_m, \xi_s, \gamma_s\} \). We choose to make this assessment in an empirical Bayes setting, by considering the localized estimates in the observation locations, \( [\hat{\mu}_i^k, \hat{\sigma}_i^{k2}]; i = 1, ..., n \), to be a super population of the \( k \)-localized estimate in an arbitrary location. Then natural estimators for the parameters of the prior model are dependent on \( k \):

\[
\hat{\mu}_m^k = \frac{1}{n} i^T_k \hat{\mu}_o^k
\]

\[
\hat{\tau}_m^k = [\hat{\sigma}_i^{k2}]^{-1} \hat{\sigma}_m^{k2}
\]

where
\begin{align*}
\hat{\sigma}_{k}^2 &= \frac{1}{n} Tr[\hat{\Gamma}_k] \\
\hat{\sigma}_{m}^2 &= \frac{1}{n} [\hat{\mu}_{m} - \hat{\mu}_{m} i_n]^T [\hat{\mu}_{m} - \hat{\mu}_{m} i_n]
\end{align*}

and

\begin{align*}
\hat{\xi}_s &= [\hat{\sigma}_s^2]^{-1} - [\hat{\mu}_s^2] + 2 \\
\hat{\gamma}_s &= \hat{\mu}_s \left[ ([\hat{\sigma}_s^2]^{-1} - [\hat{\mu}_s^2]) + 1 \right]
\end{align*} \tag{9}

where

\begin{align*}
\hat{\mu}_s &= \frac{1}{n} I_n^T s^2 \\
\hat{\sigma}_s^2 &= \frac{1}{n} [s^2 - \hat{\mu}_s^2 i_n]^T [s^2 - \hat{\mu}_s^2 i_n] \\
s^2 &= [(r_1 - \hat{\mu}_m^k)^2, ..., (r_n - \hat{\mu}_m^k)^2]^T
\end{align*}

We have now defined estimators for all model parameters based on the observations \( r_o \). Hence all probabilistic models for the regionalized variable \( \{r(x); x \in D\} \) are fully specified. Focus is on spatial prediction however, and in the following Section we define spatial predictors under the various model assumptions and insert the estimators for the model parameters to obtain operable predictors.

4 Spatial Predictors

Focus of the study is on spatial prediction in a random field \( \{r(x); x \in D\} \) based on a set of observations represented in a \([n \times 1]\)-vector \( r_o \). Consider an arbitrary location \( x_+ \in D \) with value \( r(x_+) = r_+ \). The challenge is to provide a reliable predictor for \( r_+ \) based on \( r_o \). We use a squared error loss, hence the predictor is \( \hat{r}_+ = \hat{E}[r_+ | r_o] \), with associated estimated predictor variance \( \hat{\sigma}_+^2 = \hat{V}ar[r_+ | r_o] \). Note that a predictor for the entire regionalized variable \( \{r(x); x \in D\} \) can be obtained by letting \( x_+ \) run over the domain \( D \).

Recall that the correlation function \( \{\rho(x' - x''); x', x'' \in D\} \) is assumed known and that the inter observation correlation \([n \times n]\)-matrix is denoted \( \Omega_{oo} \), while the observation to \( x_+ \) correlation \([n \times 1]\)-vector is denoted \( \omega_{o+} \). Recall also that the localization operator \( G^k_+ \) is defined such that \( G^k_+ r_o \) is a observation
[\mathbf{k} \times 1]-vector which contain the \(k\) observations located closest to \(x_+\).

**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:

\[
[r \mid \mathbf{r}_o] \sim Gauss[\mu_{+\mid o}, \sigma^2_{+\mid o}]
\]

with

\[
\mu_{+\mid o} = \mu + \omega^T \Omega_{oo}^{-1} \left[ \mathbf{r}_o - \mathbf{\mu}_n \right]
\]

\[
\sigma^2_{+\mid o} = \sigma^2 \left[ 1 - \omega^T \Omega_{oo}^{-1} \omega_o \right]
\]

Note that the predictor is independent of the variance \(\sigma^2\) while the prediction variance is independent of the observed values \(\mathbf{r}_o\) only the location configuration. These are well recognized characteristics of kriging.

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

\[
\hat{r}_{GST} = \hat{\mu}_{+\mid o}
\]

\[
\hat{\sigma}^2_{GST} = \hat{\sigma}^2_{+\mid o}
\]

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

**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 it is defined by:

\[
[r_+ \mid G^k \mathbf{r}_o] \sim Gauss[\mu_{+\mid o}^k, \sigma^k_{+\mid o}]
\]

with

\[
\mu_{+\mid o}^k = \mu^k + \left[ G^k_{+\mid o} \Omega_{oo} [G^k_{+\mid o}]^T \right]^{-1} \left[ G^k_{+\mid o} \mathbf{r}_o - \mu^k \mathbf{i}_k \right]
\]

\[
\sigma^k_{+\mid o} = \sigma^2 \left[ 1 - [G^k_{+\mid o}]^T [G^k_{+\mid o} \Omega_{oo} [G^k_{+\mid o}]^T]^{-1} G^k_{+\mid o} \right]
\]
Also this predictor is locally independent on $\sigma^2$ with predictor variance locally independent on the observed values $r_o$. Since both $\mu_+^k$ and $\sigma_+^{k^2}$ will vary across the field, the predictor and predictor variance will vary across the field.

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

$$\hat{r}^k_{LST} = \hat{\mu}_+^k | o$$
$$\hat{\sigma}^{k^2}_{LST} = \hat{\sigma}_+^{k^2} | o$$

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

**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 denoted the stationary localized/shrinkage kriging predictor and it constitutes one new predictor in the study:

$$[r_+ | m^k_+, s^{k^2}_+, G^k_+ r_o] \sim Gauss[\mu^k_+ | o, \sigma^{k^2}_+ | o]$$

with

$$\mu^k_+ = m^k_+ + [G^k_+ \omega_o+] T [G^k_+ \Omega_{oo} [G^k_+] T]^{-1} [G^k_+ r_o - m^k_+ i_k]$$

$$\sigma^{k^2}_+ = s^{k^2}_+ [1 - [G^k_+ \omega_o+] T [G^k_+ \Omega_{oo} [G^k_+] T]^{-1} G^k_+ \omega_o+]$$

where the posterior expectations for the hyper-parameters are, see Appendix A:

$$m^k_+ = E[m | s^2, G^k_+ r_o]$$

$$= \mu^k_+ + \tau^k_m i_k [\tau^k_m i_k i_k^T + [G^k_+ \Omega_{oo} [G^k_+] T]^{-1} [G^k_+ r_o - \mu^k_m i_k]$$

$$s^{k^2}_+ = E[s^{k^2}_+ | G^k_+ r_o]$$

$$= \left[ \xi^k_s + \frac{k}{2} - 1 \right]^{-1}$$

$$\times \left[ \gamma^k_s + \frac{1}{2} [G^k_+ r_o - \mu^k_m i_k] T \left[ [G^k_+ \Omega_{oo} [G^k_+] T] + \tau^k_m i_k i_k^T \right]^{-1} [G^k_+ r_o - \mu^k_m i_k] \right]$$

9
Expression (12) follows from the definition of a stationary hierarchical Gaussian random field conditioned on the model parameters and on $G^k_+ r_o$. Expression (13) follows from the prior model of $[m, s^2]$ being conjugate, hence the posterior models are analytically assessible, and so are their expectations. Note in particular that $m^k_+ = E[m \mid G^k_+ r_o]$ 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 appears with shrinkage through the estimators for shift and scaling parameters $m^k_+$ and $s^{k2}_+$. The actual observations weights are independent of $m^k_+$ and $s^{k2}_+$.

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

$$
\hat{r}^k_{LSS} = \hat{\mu}^k_{+|o}
$$

$$
\hat{\sigma}^{k2}_{LSS} = \hat{\sigma}^{k2}_{+|o}
$$

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

**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:

$$
[r_+ \mid G^k_+ r_o] \sim Gauss[\mu^k_{+|o}, \sigma^{k2}_{+|o}]
$$

with

$$
\mu^k_{+|o} = \mu^k_+ + \sigma^k_+ [G^k_+ \Gamma^k_o \omega_o^+]^T [G^k_+ \Gamma^k_o \Omega_o \Gamma^k_o [G^k_+]^T]^{-1} [G^k_+ [r_o - \mu^k_o]]
$$

$$
\sigma^{k2}_{+|o} = \sigma^{k2}_+ [1 - [G^k_+ \Gamma^k_o \omega_o^+]^T [G^k_+ \Gamma^k_o \Omega_o \Gamma^k_o [G^k_+]^T]^{-1} G^k_+ \Gamma^k_o \omega_o^+]^{-1}
$$

where

$$
\mu^k_o = [\mu^k_1, \ldots, \mu^k_n]^T
$$

$$
\Gamma^k_o = \begin{bmatrix}
\sigma^k_1 & \ldots & 0 \\
\vdots & \ddots & \vdots \\
0 & \ldots & \sigma^k_n
\end{bmatrix}
$$
Note that this predictor is dependent of 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:

\[ \hat{r}_{LNT} = \hat{\mu}_{+|o} \]

\[ \hat{\sigma}_{LNT}^2 = \hat{\sigma}_{+|o}^2 \]

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

**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. The predictor is denoted the non-stationary localized/shrinkage kriging predictor and it constitutes another new predictor in the study:

\[
[r_+ | m_k^+, m_o^k, s_{+o}^k, S_k^o, r_o] \sim \text{Gauss}[\mu_{+|o}, \sigma_{+|o}^2]
\]

with

\[
\mu_{+|o} = m_k^+ + s_k^+ [G_{+}^k S_o^k \omega_{+o}^k]^{T} [G_{+}^k S_o^k \Omega S_o^k[G_{+}^k]^{T} ]^{-1} [G_{+}^k [r_o - m_o^k]]
\]

\[
\sigma_{+|o}^2 = s_{+o}^2 \left[ 1 - [G_{+}^k S_o^k \omega_{+o}^k]^{T} [G_{+}^k S_o^k \Omega S_o^k[G_{+}^k]^{T} ]^{-1} G_{+}^k S_o^k \omega_{+o}^k \right]
\]

where \( m_k^+ \) and \( s_{+o}^k \) are defined as in Expression (13) while

\[
m_o^k = [m_1^k, ..., m_n^k]^T
\]

\[
S_o^k = \begin{bmatrix}
s_1^k & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & s_n^k
\end{bmatrix}
\]

are defined from the \( x_+ \)-centred \( m_+ \) and \( s_{+o}^k \) shifted to the observation locations \([x_1, ..., x_n]\).
The predictor appears with shrinkage through the estimators of \( m_{k+}^k, s_{k+}^{k^2} \) and \( m_i^k, s_i^{k^2}, \ i = 1, \ldots, n \). 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:

\[
\hat{r}_{LNS}^k = \hat{\mu}_{i+|o}^k
\]

\[
\hat{\sigma}_{LNS}^{k^2} = \hat{\sigma}_{i+|o}^{k^2}
\]

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

**Crossvalidation Calibrated (CVC) Predictors**

We have defined five predictors, 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 wrong scaling of the prediction variances. We introduce crossvalidation calibrated (CVC) predictors to provide a global calibration.

Consider an arbitrary predictor in an arbitrary location \( x_+ \in D, \hat{r}_+ = \hat{E}[r_+ | r_o] \). The predictor may be global or localized, and it is based on the observations \( r_o = [r_1, \ldots, r_n] \) in locations \( [x_1, \ldots, x_n] \). The associated prediction variance is \( \hat{\sigma}_+^2 = \hat{V}ar[r_+ | r_o] \).

Define the crossvalidation predictors with associated predictor variances in the observation locations \( [x_1, \ldots, x_n] \):

\[
\hat{r}_i = \hat{E}[r_i | r_o(-i)] \quad ; \quad i = 1, \ldots, n
\]

\[
\hat{\sigma}_i^2 = \hat{V}ar[r_i | r_o(-i)] \quad ; \quad i = 1, \ldots, n
\]

where \( r_o(-i) \) represents observations \( r_o \) with observations \( r_i \) removed.

The normalized crossvalidation errors are defined as:

\[
e_i = [\hat{\sigma}_i]^{-1}[r_i - \hat{r}_i] \quad ; \quad i = 1, \ldots, n
\]
Under a fully specified model, these errors will be centred at zero and be scaled to unity. Consider the estimators:

\[ \hat{\mu}_e = \frac{1}{n} \sum_{i=1}^{n} e_i \]

\[ \hat{\sigma}_e^2 = \frac{1}{n} \sum_{i=1}^{n} [e_i - \hat{\mu}_e]^2 \]

and the \( \hat{\mu}_e \) and \( \hat{\sigma}_e^2 \) should be close to zero and unity, respectively.

We define the CVC predictor and CVC prediction variance in an arbitrary location \( x_+ \in D \) by:

\[ \tilde{r}_+ = \tilde{r}_+ + \tilde{\sigma}_+ \hat{\mu}_e \]

\[ \tilde{\sigma}_+^2 = \hat{\sigma}_e^2 \hat{\sigma}_+^2 \]

Note that the corresponding normalized crossvalidation 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 study.

5 Evaluation Criteria

We have defined several spatial CVC predictors with associated CVC prediction variances. We need to select one based on the set of observations \( r_o \) only. Recall that all predictors have normalized crossvalidation errors centred at zero and scaled to unity.

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

\[ PMSE = \frac{1}{n} \sum_{i=1}^{n} [r_i - \tilde{r}_i]^2 \]

may vary. In the CVC predictor the scale of the normalized crossvalidation error is identical to unity but large deviations may be reduced by large prediction variances in this measure. We use PMSE as measure for precision in the CVC predictor \( \tilde{r}_+ \) and we prefer small values of PMSE, of course.

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

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

(18)

Recall that normalized crossvalidation squared error in this expression is centered exactly at unity. We use VMSE as measure for precision in the CVC prediction variance $\tilde{\sigma}_i^2$ and we prefer small values of VMSE, of course.

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

6 Empirical Studies

It is difficult to compare the various predictors analytically, hence we conduct an empirical comparision. We present two case: set of observations of yearly accumulated precipitation and a simulation study on a general Gaussian random field.

US Precipitation Study

We consider observations of 1997 accumulated precipitation in 1001 locations in an area of the US, see Figure 1. The observations is a subset of much larger data set, see Data (2014) and Johns et al (2003).

By inspecting the observations in Figure 1 we note relatively dense, uniform coverage of observations with a slight south-eastern trend in the values. The spatial correlation function is inferred under a stationary model, see Figure 2, and the following model is fitted:

$$\rho(\tau) = \exp \left\{ - \left[ \frac{\tau}{3.3} \right]^{1.4} \right\}; \quad \tau \geq 0$$

with $\tau = |x' - x''|$. This spatial correlation model is used throughout the study. The localized predictors require that the number of observations in the neighborhood $k$ is specified. We have, after a small preliminary study, chosen $k=10$. 

14
We compare the five alternative CVC predictors defined Section 4 by using the evaluation criteria defined in Section 5. The results from the evaluation are displayed in Figure 3 thru 9 and Table 1.

The results from the Glob/Stat/Trad CVC predictor are displayed in Figure 3 and 4 and in Table 1. Figure 3 display the actual crossvalidation predictions and crossvalidation prediction standard deviations in the observations locations. The predictions can be compared to the actual observations in Figure 1, and no dramatic deviations are seen since the observation coverage is dense. Figure 4 display the normalized crossvalidation errors both spatially and as a histogram. Note that the histogram is centred to 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 crossvalidation errors of the non-calibrated predictor. The predictor appears as well centred 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 base for comparison of the various CVC predictors.

The results from the Loc/Stat/Trad $k=10$ CVC predictor are displayed in Figure 5 and 6 and in Table 1. The formats are identical to the ones discussed in the previous paragraph. The crossvalidation predictions and prediction standard deviations in Figure 5 are very similar to the results for the global predictor in Figure 3. The normalized crossvalidation errors in Figure 6 do deviate from the results for the global predictor in Figure 4. Note that the large errors tend to be more uniformly located in the area and that the histogram have somewhat lighter tails. From Table 1 we see that the non-calibrated predictor is well centred 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. Not much favorable for prediction but clearly favorable for prediction variance.

The results from the Loc/Stat/Shr $k=10$ CVC predictor are displayed in
Figure 7 thru 9, and in Table 1, on 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 Figure 9. The crossvalidation predictions and crossvalidation standard deviations in Figure 7 appear as very similar to the results for the other CVC predictors in Figure 3 and 5. The normalized crossvalidation errors in Figure 8 appear as similar to the ones for the traditional localized predictor in Figure 6. Note, however, that the histograms are different in the sense that the histogram of the shrinkage predictor have lighter tails than for 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 centre of the model. From Table 1, we observe that the non-calibrated predictor is well centred but with downward biased prediction variances. The evaluation criteria PMSE and VMSE for the Loc/Stat/Shr $k=10$ CVC predictor are both favorable to both the traditional global and localized predictors. Minor improvement on 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 are only 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 prediction and prediction variance.

To summarize the current study, we conclude that localized predictors can be clearly favorable to the global one. Favorable both in prediction precision and particularly in precision of prediction variance. The localized models are robust with respect to deviations from assumptions of global stationary, and this robustness causes improvements of the localized predictors. Localized stationary predictors are favorable 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 we must estimate expectation and variance in each observation location, which introduces additional uncertainty in the
model. This uncertainty is dominating 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 we need to make bias/variance trade-offs when selecting the localization. Using a regularizer representing the global variability in the parameter 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 we have used a localization with $k=10$. In the extended study, see Asfaw and Omre (2014), we have evaluated the sensitivity to choice of $k$-value. If we reduce $k$ considerably, to for example $k=4$, the localized predictor are poorer than the global one. This is probably caused by overfitting to the observations. Results for $k$ in the range 8 to 16 are consistent with the results in the current study with localized shrinkage predictors clearly favorable. By increasing $k$ we will of course in the limit have the localized and global predictors to coincide.

**Simulation Study**

We conduct a simulation study on a general Gaussian random field model \( \{r(x); x \in D \subset \mathbb{R}^1\} \) with \( D = [1, \ldots, 200] \) discretized to grid \( L_D = \{1, \ldots, 200\} \). The expectation and variance fields are displayed in Figure 10.a, and note the non-stationarity with inter-dependence. The known correlation function is \( \rho(\tau) = \exp\{-0.2\tau^{1.5}\} \) with \( \tau = |x'' - x'| \). We generate one realization from this random field model and use realizations at locations \( L_o = \{1, 10, \ldots, 190, 200\} \) as observations, hence \( n = 21 \).

With known model parameters, the correct predictions and prediction variances in locations \( L_D \) are analytically assessable, see Figure 10.b. The predictor reproduces the observations and the prediction variances reflects the non-stationarity. We use the various CVC predictors Glob/Stat/Trad, Loc/Stat/Trad and Loc/Stat/Shr with localization $k=\pm 4$ when relevant, and obtain results as in Figure 10.c-e. Note that the global predictor have 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 version appears as somewhat dampened relative to the traditional one. This
dampening is caused by the prior models on local expectation and variance which are assessed in an empirical Bayesian spirit. The prior model for this realization are displayed in Figure 11. We also perform prediction based on the non-stationary models but the results are not displayed in Figure 10.

To summarize, the global predictor, which corresponds to classical ordinary Kriging, appears as highly unreliable due to deviations from the stationarity assumptions, and will not be further evaluated. The localized predictors are difficult to distinguish by visual inspection.

The evaluation criteria discussed in Section 5 can be revised to characterize the deviations between the CVC predictions and prediction variances and the correct ones in Figure 10.a. We repeat 1000 realizations and average the criteria to obtain APMSC and AVMSC as evaluation criteria for prediction and prediction variance respectively, see Table 2 with $k=\pm4$. We repeat this procedure for varying localizations $k$. Note that for localized, stationary predictors the shrinkage-versions are favorable to the traditional one for both criteria and all $k$, except one extreme case with large $k$. For the localized, non-stationary predictors the shrinkage-versions are uniformly favorable to the traditional one.

One typical feature is observed for localized/stationary predictors for criterion APMSC, characterizing precision in prediction, for varying localizations $k$. The traditional predictor makes bias/variance trade-offs, with poor performances for large $k$ due to bias and for small $k$ due to instability in parameter estimates. Localization $k=\pm4$ appears as favorable. The shrinkage predictor stabilizes the variance estimates by shrinkage and tolerates smaller localizations with $k=\pm2$. We consider the Loc/Stat/Shr $k=\pm4$ CVC predictor to be the favorable one, since the APMSC criterion is seen as more important than the AVMSC one.

The evaluation criterion discussed above require the correct predictors to be available, which is not the case in real studies. We have also computed the crossvalidation criteria discussed in Section 5 which always are available. We average over 1000 realizations to obtain APMSE and AVMSE, and the values are listed in Table 2. For these criteria the shrinkage-versions are uniformly favorable to the traditional ones for all cases. We consider the Loc/Stat/Shr $k=\pm2$ CVC predictor to be best based on these criteria, of
To summarize, the crossvalidation based criteria are representative for the exact criteria based on the correct predictions. Note that the former can be computed in real studies with only one set of observations available. The Loc/Stat/Shr CVC predictors are favorable to other predictors, although the localization $k$ tends to be somewhat underestimated by crossvalidation.

In the current study we have only used one expectation and variance function. In the extended study, see Asfaw and Omre (2014), we have considered many other cases. 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 of course. Lastly, recognize that the synthetic simulation case is Gaussian, no outliers and no heavy-tailed distributions are involved. In spite of this, we find that the shrinkage predictors are favorable to the traditional ones. In presence of outliers and heavy-tailed models, we expect the shrinkage predictors to perform even more favorably.

Figure 1: The 1997 accumulated precipitation observations in the US with sub-area to be studied (top). The 1001 observations used in the study (bottom).
Figure 2: Spatial correlation function used with estimated correlation values.

Figure 3: Glob/Stat/Trad CVC Predictor-ordinary kriging.
Figure 4: Glob/Stat/Trad crossvalidation errors.

Figure 5: Loc/Stat/Trad k=10 CVC Predictor.
Figure 6: Loc/Stat/Trad k=10 crossvalidation errors.

Figure 7: Loc/Stat/Shr k=10 CVC Predictor.
Figure 8: Loc/Stat/Shr k=10 crossvalidation errors.

Figure 9: US Precipitation study. Priors model for expectation and variance.
## Table 1: US Precipitation study. Evaluation criteria: Mean normalized error (MNE), Mean square normalized error (MSNE), Prediction mean squared error (PMSE) and Variance mean squared error (VMSE).

<table>
<thead>
<tr>
<th>Model</th>
<th>Localized/Stationary</th>
<th>Localized/Non-stationary</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Traditional</td>
<td>Shrinkage</td>
</tr>
<tr>
<td>Test D.</td>
<td>$k = 1000$</td>
<td>$k = 10$</td>
</tr>
<tr>
<td>MNE</td>
<td>$1.0148e-17$</td>
<td>$-5.3792e-18$</td>
</tr>
<tr>
<td>MSNE</td>
<td>$1.5399$</td>
<td>$2.9740$</td>
</tr>
<tr>
<td>PMSE</td>
<td>$6.8758e + 03$</td>
<td>$6.8745e + 03$</td>
</tr>
<tr>
<td>VMSE</td>
<td>$9.8749$</td>
<td>$5.9746$</td>
</tr>
</tbody>
</table>

Figure 10: General Gaussian random field. Expectation and variance field (a), Predictions and prediction variances for one realization (b-e).
Figure 11: General Gaussian random field. Prior model for expectation and variance for one realization.

<table>
<thead>
<tr>
<th>Model</th>
<th>Localized/Stationary</th>
<th>Shrinkage</th>
<th>Localized/Non-stationary</th>
<th>Shrinkage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test D.</td>
<td>±6 ±4 ±2</td>
<td>±6 ±4 ±2</td>
<td>±6 ±4 ±2</td>
<td>±6 ±4 ±2</td>
</tr>
<tr>
<td>APMSC</td>
<td>0.3665</td>
<td>0.3557</td>
<td>1.0100</td>
<td>0.3968</td>
</tr>
<tr>
<td>AVMSC</td>
<td>0.5287</td>
<td>1.9339</td>
<td>159.3760</td>
<td>0.3927</td>
</tr>
<tr>
<td>APMSE</td>
<td>18.236</td>
<td>17.8311</td>
<td>19.8851</td>
<td>16.9502</td>
</tr>
<tr>
<td>AVMSE</td>
<td>1.8936</td>
<td>2.9277</td>
<td>2.9888</td>
<td>1.7271</td>
</tr>
</tbody>
</table>

Table 2: General Gaussian random field. Evaluation criteria: 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).

7 Conclusion

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

The localized/shrinkage kriging predictors are compared to the traditional kriging predictors, both global and localized, in a study on real precipitation data and in a synthetic simulation study. We use two crossvalidation based criteria 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 make local predictors suitable. The localized/shrinkage kriging predictors appear as clearly favorable to traditional localized kriging predictors. The shrinkage predictors based on local stationarity seems to be the superior ones.

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 focus is on computational demands.

References


Appendix A
Hierarchical, Gaussian model

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

The pdf are:

\[
[r_o | m, s^2] \sim f[r_o | m, s^2] = Gauss_n[mi_n, s^2\Omega_{oo}]
\]

\[
= [2\pi]^{-\frac{n}{2}} |s^2\Omega_{oo}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} [r_o - mi_n]^T [s^2\Omega_{oo}]^{-1} [r_o - mi_n] \right\}
\]

\[
[m | s^2] \sim f[m | s^2] = Gauss_1[\mu_m, \tau_m s^2]
\]

\[
= [2\pi]^{-\frac{1}{2}} [\tau_m s^2]^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} [\tau_m s^2]^{-1} [m - \mu_m]^2 \right\}
\]

\[
[s^2] \sim f[s^2] = InvGam_1[\xi_s, \gamma_s]
\]

\[
= [\Gamma(\xi_s)]^{-1} \gamma_s^{\xi_s} [s^2]^{-[\xi_s+1]} \exp \left\{ -\gamma_s [s^2]^{-1} \right\}
\]

Note that:

\[
E[s^2] = [\xi - 1]^{-1} \gamma_s \quad ; \quad \xi_s > 1, \gamma_s > 0
\]

\[
Var[s^2] = [(\xi - 1)^2(\xi - 2)]^{-1} \gamma_s^2 \quad ; \quad \xi_s > 2, \gamma_s > 0
\]

From the pdfs above

\[
[m | s^2, r_o] \sim f[m | s^2, r_o] = \frac{f[r_o | m, s^2]f[m | s^2]}{\int f[r_o | m, s^2]f[m | s^2]dm}
\]

\[
= Gauss_1(\mu_{m|o}, \sigma_{m|o}^2)
\]
\[
\begin{align*}
\mu_{m|o} &= \mu_m + \tau_m \mathbf{i}_m^T \left[ \tau_m \mathbf{i}_m \mathbf{i}_m^T + \Omega_{oo} \right]^{-1} \left[ \mathbf{r}_o - \mu_m \mathbf{i}_n \right] \\
\sigma_{m|o}^2 &= \tau_m s^2 - \tau_m s^2 \mathbf{i}_n^T \left[ \tau_m s^2 \mathbf{i}_n \mathbf{i}_n^T + s^2 \Omega_{oo} \right]^{-1} \left[ \tau_m s^2 \mathbf{i}_n \right] \\
[s^2 | \mathbf{r}_o] &\sim f[s^2 | \mathbf{r}_o] = \frac{\int \int f[\mathbf{r}_o | m, s^2]f[m | s^2]f[s^2]dmds^2}{\int \int \int f[\mathbf{r}_o | m, s^2]f[m | s^2]f[s^2]dmds^2} \\
&= InvGam_1(\xi_{s|o}, \gamma_{s|o}) \\
\xi_{s|o} &= \xi_s + \frac{n}{2} \\
\gamma_{s|o} &= \gamma_s + \frac{1}{2} \left[ \mathbf{r}_o - \mu_m \mathbf{i}_n \right]^T \left[ \tau_m \mathbf{i}_m \mathbf{i}_m^T + \Omega_{oo} \right]^{-1} \left[ \mathbf{r}_o - \mu_m \mathbf{i}_n \right]
\end{align*}
\]