Localized/Shrinkage kriging-Prediction in a Bayesian, non-stationary Gaussian random field

Zeytu Gashaw Asfaw*
School of Mathematical and Statistical Sciences, Hawassa University, Hawassa, Ethiopia - zeytugashaw@yahoo.com

Abstract

Assume that a set of exact observations from a regionalized variable is available. Focus is on spatial prediction in an unobserved location with associated prediction variance,i.e classical kriging prediction. Consider a classical, stationary traditional kriging model with spatially constant expectation and variance. Let the spatial correlation function be shift invariant and let it be known. This constitutes global ordinary kriging model, and global predictors are optimal. A more flexible and robust spatial predictor can be defined by applying the traditionalkriging predictor locally. This entails using only observations in a finite neighborhood around the location of the predictor. This robustfies the predictor with respect to devations from the assumptions of globally constant expectation and variance. The challenge in localized prediction is to make a bias/ variance trade-off in selection of the size of the neighborhood, i.e the number of observations involved in the predictor. Unwanted abrupt changes in the predicted regionalized variable and the associated prediction variance can appear.

The objective of the study is to improve on the flexibility and robustness of the spatial kriging predictors with respect to deviations from spatial stationarity assumptions. A predictor based on a non-stationary Gausian 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 global ordinary kriging predictors to traditional localized kriging predictors in a case study with observations of annual accumulated precipitation. A crossvalidation criterion is used in the comparision. 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 the predictor parameters like neighborhood. Lastly, the computational demands of localized predictors are very modest, hence the localized/ shrinkage predictors are suitable for large scale spatial prediction problems.

Keywords: Spatial statistics; Gaussian random fields; Bayesian inference; Conjugate models.

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.

2 Model Parameter Inference

The stationary Gaussian RF model and the hierarchical, stationary Gaussian RF model depends on a set of model parameters. These model parameters must be assessed from the avaliable 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.

2.1 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:

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

Hence the maximum likelihood estimates are:

$$\hat{\mu}_r = [i_{n_o}^T \Omega_{oo}^{-1} \mathbf{r}_o][i_{n_o}^T \Omega_{oo}^{-1} i_{n_o}]^{-1}$$

$$\hat{\sigma_r^2} = \tfrac{1}{n_o}(\mathbf{r}_o - \hat{\mu}_r i_{n_o})^T \Omega_{oo}^{-1}(\mathbf{r}_o - \hat{\mu}_r i_{n_o})$$

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:

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

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

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

$$\hat{\Gamma}_o^k = \begin{bmatrix} \hat{\sigma}_1^k & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \hat{\sigma}_{n_0}^k \end{bmatrix}$$

2.2 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:

$$\hat{\mu}_m^k = \tfrac{1}{n_o} i_{n_o}^T \hat{\boldsymbol{\mu}}_o^k$$

$$\hat{\sigma}_m^{k2} = \tfrac{1}{n_o}[\hat{\boldsymbol{\mu}}_o^k - \hat{\mu}_m^k i_{n_o}]^T [\hat{\boldsymbol{\mu}}_o^k - \hat{\mu}_m^k i_{n_o}]$$

$$\hat{\sigma}_{r|.}^{k2} = \tfrac{1}{n_o} Tr[\hat{\Gamma}_o^k]$$

$$\hat{\tau}_m^k = \tfrac{\hat{\sigma}_m^{k2}}{\hat{\sigma}_{r|.}^{k2}}$$

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:

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

which is independent of s^2.

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

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

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 :

$$\mu_s = E[s^2] = \tfrac{\gamma_s}{\xi_s - 1}$$

$$\sigma_s^2 = Var[s^2] = \tfrac{\gamma_s^2}{[\xi_s-1]^2[\xi_s-2]}$$

Consquently,

$$\xi_s = \tfrac{\mu_s^2}{\sigma_s^2} + 2$$

$$\gamma_s = \mu_s \left[\frac{\mu_s^2}{\sigma_s^2} + 1\right]$$

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

$$\hat{S}_o^k = \begin{bmatrix} \hat{s}_1^k & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \hat{s}_{n_0}^k \end{bmatrix}$$

3 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 $\mathbf{r}_o = [r(x_{o1}), \cdots, r(x_{on_o})] = [r_{o1}, \cdots, r_{on_o}]$. 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.

3.1 Localized/Stationary Model

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

3.1.1 Traditional Predictor

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

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

$$\hat{\sigma}_{STP+}^{k2} = \hat{\sigma}_+^{k2} \left[1 - [G_+^k \omega_{o+}]^T [G_+^k \Omega_{oo}[G_+^k]^T]^{-1} G_+^k \omega_{o+}\right]$$

with the parameter estimators defined in Section 2.1.

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.

3.1.2 Shrinkage Predictor

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

$$\hat{r}^k_{SSP+} = \hat{m}^k_+ + [G^k_+\omega_{o+}]^T[G^k_+\Omega_{oo}[G^k_+]^T]^{-1}\left[G^k_+\mathbf{r}_o - \hat{m}^k_+ i_k\right]$$

$$\hat{\sigma}^{k2}_{SSP+} = \hat{s}^{k2}_+\left[1 - [G^k_+\omega_{o+}]^T[G^k_+\Omega_{oo}[G^k_+]^T]^{-1}G^k_+\omega_{o+}\right]$$

with the parameters estimators defined in Section 2.2.

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 regulizer when estimating the local expectation and variance. The prior models for m_+ and s_+^2 are assessed from the avaliable observations in an empirical Bayesian setting. This makes it possible to use smaller neighborhoods which hopefully provides predictors with less bias.

3.2 Localized/Non-stationary Model

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

3.2.1 Traditional Predictor

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

$$\hat{r}^k_{NTP+} = \hat{\mu}^k_+ + \hat{\sigma}^k_+[G^k_+\hat{\Gamma}^k_o\omega_{o+}]^T\left[G^k_+\hat{\Gamma}^k_o\Omega_{oo}\hat{\Gamma}^k_o[G^k_+]^T\right]^{-1}G^k_+[\mathbf{r}_o - \hat{\boldsymbol{\mu}}^k_o]$$

$$\hat{\sigma}^{k2}_{NTP+} = \hat{\sigma}^{k2}_+\left[1 - [G^k_+\hat{\Gamma}^k_o\omega_{o+}]^T\left[G^k_+\hat{\Gamma}^k_o\Omega_{oo}\hat{\Gamma}^k_o[G^k_+]^T\right]^{-1}G^k_+\hat{\Gamma}^k_o\omega_{+o}\right]$$

with the parameter estimators defined in Section 2.1.

The expectation and variance is locally and uniquely estimated for each observation according to the General Gaussian RF model. 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.

3.2.2 Shrinkage Predictor

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

$$\hat{r}^k_{NSP+} = \hat{m}^k_+ + \hat{s}^k_+[G^k_+\hat{S}^k_o\omega_{o+}]^T\left[G^k_+\hat{S}^k_o\Omega_{oo}\hat{S}^k_o[G^k_+]^T\right]^{-1}G^k_+[\mathbf{r}_o - \hat{\boldsymbol{m}}^k_o]$$

$$\hat{\sigma}^{k2}_{NSP+} = \hat{s}^{k2}_+\left[1 - [G^k_+\hat{S}^k_o\omega_{o+}]^T\left[G^k_+\hat{S}^k_o\Omega_{oo}\hat{S}^k_o[G^k_+]^T\right]^{-1}G^k_+\hat{S}^k_o\omega_{o+}\right]$$

with the parameter estimators defined in Section 2.2.

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 regulizers 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.

4 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 also in this simulation study. 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.

5 References

Asfaw, Z. G. and Omre, H.; 2014: "Localized/Shrinkage Kriging Predictors". Preprint Series in Statistics no.1, 2014, Department of Mathematical Sciences, NTNU, Trondheim, Norway.

Chiles, J.P and Delfiner,J.P; 1999: *Geostatistics: Modeling spatial uncertainty(wiley series in probability and statistics)*, Wiley, New York, 695p.

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.

Fuentes M. A.; 2005: "formal test for nonstationarity of spatial stochastic processes", Journal of Multivariate Analysis, 96, pp. 30-55.

Gribov, A. and Krivoruchko, K. ; 2004: "Geostatistical Mapping with Continuous Moving Neighborhood", Published in Mathematical Geology, Volume 36, Number 2.

Journel, A.G and Huijbregts, C.J; 1978: *Mining Geostatistics*, Academic Press Inc, London.

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.

Stein, M. L. ; 2002: "The Screening Effect in Kriging", The Annals of Statistics, Vol. 30, No. 1, pp. 298-323