Mathematical Geosciences
https://doi.org/10.1007/s11004-025-10252-y

SPECIAL ISSUE

The Grid-Free Spatial Kernel Predictor for Huge Observation Sets

Henning Omre[1] · **Mina Spremic**[1]

Received: 20 February 2025 / Accepted: 9 November 2025

Abstract
Assessment of a continuous spatial variable based on a set of m observations is usually performed in a Gaussian random field framework. The optimal predictor under this model can be presented either as a linear kriging predictor or as a dual kriging predictor. The spatial variable predictor is usually stored in a kriging grid representation of size n. Alternatively, one may define a kernel function representation based on the dual kriging formulation. The latter can be efficiently reduced to the former, but not vice versa. To provide a prediction at an arbitrary location, a piecewise planar interpolation in the actual grid unit is typically required. For the functional representation, the functional value in the actual location must be calculated. The computational challenge of both representations is primarily related to the inversion of the observation covariance $(m \times m)$-matrix. In large spatial studies with huge sets of observations, and thus huge m, this inversion may not be computationally feasible. Localized kriging predictors are then frequently used to generate the grid representation of the spatial variable. This approach has computational demands proportional to the grid size n. We present a localized kernel predictor to provide a functional representation of the spatial variable. The specification of this localized kernel predictor constitutes the major contribution of this paper. This predictor has computational demands proportional to the number of observations m. This is particularly beneficial in three-dimensional models and spatiotemporal studies where one typically has $n \gg m$. The characteristics of the kernel predictor are demonstrated in an example. A study on real observations indicates that the localized kernel function representation has substantial computational advantages over the localized kriging grid representation. Even generating the grid representation from a kernel function representation appears more computationally efficient than generating it directly using a localized kriging predictor.

Keywords Spatial statistics · Gaussian random fields · Kriging predictors · Big data analysis

✉ Henning Omre
henning.omre@ntnu.no

[1] Department of Mathematical Sciences, Norwegian University of Science and Technology, Alfred Getz' vei 1, Trondheim, Norway

Published online: 13 January 2026

1 Introduction

Many applications involve spatial prediction either as the ultimate goal or as an intermediate effort. Assessing a spatial variable $\{r(\mathbf{x}); \mathbf{x} \in \mathtt{D}\}$ based on a set of m observations $\mathbf{r}^d = (r(\mathbf{x}_1^d), \ldots, r(\mathbf{x}_m^d)); \mathbf{x}_i^d \in \mathtt{D}$ is often a challenge. The prediction of the spatial variable in an arbitrary location $\mathbf{x}_+ \in \mathtt{D}$ is usually performed by kriging (see Chiles and Delfiner (2012) and Omre et al. (2024)), which is optimal under a Gaussian random field (RF) model. If the entire spatial variable over $\mathtt{D}$ is of interest, a grid representation is often used, where kriging predictions and prediction variances are assigned to each grid node. The computational demand is dominated by the inversion, or Cholesky decomposition, of the observation covariance matrix of dimension $(m \times m)$. The grid is usually defined as sufficiently dense such that predictions at non-grid node locations are made by simple piecewise constant or piecewise planar interpolation. Particularly in studies with a huge number of observations, the grid needs to be dense enough to capture the information content in the observations. Thus, only predictions and prediction variances in each grid node are computed and stored, and one need not solve the entire kriging system to assess the spatial variable in an arbitrary location. This grid-based representation can also be the basis for contour or perspective presentations, and further processing of the predicted spatial variable.

The Gaussian Markov RF model (see Rue and Held (2005)) is defined on the actual grid system by specifying a high degree of conditional independence in the grid node values. Thus, the associated precision matrix for the spatially discretized Gaussian RF model appears sparse, and Cholesky decomposition can be efficiently applied. The predictor is, however, crucially dependent on the grid representation, and the computational demand is dominated by the inversion, or Cholesky decomposition, of the sparse grid precision matrix of dimension $(n \times n)$. To obtain a full spatial representation, simple piecewise constant or piecewise planar interpolations of the grid node values must be performed. The basis function predictor (see Cressie and Johannesson (2008)) provides a prediction on functional representation with computational demand related to Cholesky decomposition of the observation covariance matrix of dimension $(m \times m)$. Thus, the prediction in an arbitrary location in the domain can be assessed directly from the predictor, requiring no further interpolations. The challenge is that each function in the basis function set, which defines the predictor, requires a unique reference location. This set of reference locations defines an implicit grid system, and thus the predictor is grid-based.

We define a grid-free predictor for the spatial variable represented in functional form, with associated prediction variances. This predictor is related to the dual kriging predictor discussed in Matheron (1971) and Royer (1984), but we present an alternative development of the predictor. We refer to the predictor as the kernel predictor, and like the kriging predictor, it is optimal under Gaussian RF model assumptions. The grid-free kernel predictor actually appears as the grid infill asymptotic limit of a grid-based kriging predictor. The spatial predictor only involves operations on the observation covariance matrix of dimension $(m \times m)$, so no implicit grid representation is required, potentially providing large computational advantages.

To improve computational efficiency even further, we consider a class of finite-range spatial correlation function models (see Gneiting (2002)). Under these model

assumptions, the observation covariance matrix appears sparse, for which efficient Cholesky decomposition algorithms are available. In studies with huge numbers of observations, localized approximations of the predictor must be used, since inversion of the complete observation covariance matrix of dimension $(m \times m)$ is infeasible (see Chiles and Delfiner (2012)). We define an approximate localized kernel predictor which is far more computationally efficient than the traditional localized kriging predictor. The definition of this localized kernel predictor constitutes the major contribution of the paper. Model parameter inference is also defined for this case.

Aunon and Gomez-Hernandez (2000) and Vigsnes et al. (2017) present discussions of localized versions of the dual kriging predictor, but both of these predictors are based on spatial patching, which is different from our proposed approach. Spatial patching may introduce spatial artifacts in the spatial predictions. In Heaton et al. (2018), an overview and an evaluation of various alternative spatial predictors suitable for studies with large observation sets are presented.

In the notation, Latin letters represent random variables, whereas Greek letters represent model parameters. Vectors are denoted by bold lowercase letters, while matrices are denoted by bold uppercase letters, and upper-index T indicates the transpose. A random Gaussian n-vector $\mathbf{r}$ with expectation n-vector $\boldsymbol{E}[\mathbf{r}] = \boldsymbol{\mu}_r$ and covariance $(n \times n)$-matrix $\boldsymbol{Var}[\mathbf{r}] = \boldsymbol{\Sigma}_r$ has a probability density function (pdf) represented by $\phi_n(\mathbf{r}; \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r)$. The n-vector $\boldsymbol{i}_n$ is a vector with entries of 1s only. The set of real numbers is specified as $\mathbb{R}$, with a lower index sometimes indicating its different subsets, such as $\oplus$ non-negative numbers, $+$ positive numbers, and $[a, b]$ closed sets of numbers between a and b.

2 Model definition

Consider a continuous spatial variable $\{r(\mathbf{x}); \mathbf{x} \in \mathtt{D} \subset \mathbb{R}^q\}$; $r(\mathbf{x}) \in \mathbb{R}$ with spatial reference variable $\mathbf{x}$ in a bounded q-dimensional domain $\mathtt{D}$, with $q = 1, 2,$ or 3. Let the spatial variable be observed in m locations represented by the spatial reference set $\mathtt{M} : \{\mathbf{x}_1^d, \ldots, \mathbf{x}_m^d\}$; $\mathbf{x}_i^d \in \mathtt{D}$, also referred to as observation number $i = 1, \ldots, m$. The corresponding observed values are represented by the m-vector $\mathbf{r}^d = (r(\mathbf{x}_1^d), \ldots, r(\mathbf{x}_m^d))^T = \{r_{\mathbf{y}}^d; \mathbf{y} \in \mathtt{M}\}$. The observations are assumed to be location-wise and exact only to simplify the notation; the developments can be easily extended to observations with additive Gaussian error. The challenge is to assess the spatial variable given the observations; hence

$$\{[r(\mathbf{x})|\mathbf{r}^d]; \mathbf{x} \in \mathtt{D}\}. \tag{1}$$

Note in particular that a spatial representation of the conditional spatial variable is the objective.

Assign a prior probabilistic model to the spatial variable of interest $\{r(\mathbf{x}); \mathbf{x} \in \mathtt{D}\}$. Assume that it is a stationary Gaussian RF with expectation and variance levels $\mu_r \in \mathbb{R}$ and $\sigma_r^2 \in \mathbb{R}_{\oplus}$, respectively, and a spatial correlation function $\rho_r(\boldsymbol{\tau}) \in \mathbb{R}_{[-1,1]}$; $\boldsymbol{\tau} = \mathbf{x}'' - \mathbf{x}'$; $\mathbf{x}', \mathbf{x}'' \in \mathtt{D}$, being a non-negative definite function.

Consider an arbitrary location $\mathbf{x}_+ \in \mathtt{D}$ at which the value of the conditional spatial variable will be assessed. Then from the Gaussian RF assumptions above, we have

$$\begin{bmatrix} r(\mathbf{x}_+) \\ \mathbf{r}^d \end{bmatrix} \sim p(r_+, \mathbf{r}^d) = \phi_{1+m}\left(\begin{bmatrix} r_+ \\ \mathbf{r}^d \end{bmatrix}; \begin{bmatrix} \mu_r \\ \mu_r \boldsymbol{i}_m \end{bmatrix}, \begin{bmatrix} \sigma_r^2 & \boldsymbol{\sigma}_{+d} \\ \boldsymbol{\sigma}_{d+} & \boldsymbol{\Sigma}_d \end{bmatrix} \right), \tag{2}$$

where the observation covariance $(m \times m)$-matrix $\boldsymbol{\Sigma}_d = \sigma_r^2 \boldsymbol{\Sigma}_d^\rho$ and the observation prediction m-vector $\boldsymbol{\sigma}_{d+} = \sigma_r^2 \boldsymbol{\rho}_{d+}$ are defined by the correlation $(m \times m)$-matrix $\boldsymbol{\Sigma}_d^\rho$ containing the inter-correlations between the observations; hence $\rho_r(\boldsymbol{\tau}_{ij}^d)$; $\boldsymbol{\tau}_{ij}^d = \mathbf{x}_i^d - \mathbf{x}_j^d$; $i, j = 1, \ldots, m$. The correlation m-vector $\boldsymbol{\rho}_{d+}$ contains the corresponding correlations between the observations and the spatial variable value to be predicted; hence $\rho_r(\boldsymbol{\tau}_{i+}^d)$; $\boldsymbol{\tau}_{i+}^d = \mathbf{x}_i^d - \mathbf{x}_+$; $i = 1, \ldots, m$, and $\boldsymbol{\rho}_{+d} = \boldsymbol{\rho}_{d+}^T$.

It follows from familiar Gaussian theory that

$$[r(\mathbf{x}_+)|\mathbf{r}^d] \sim p(r_+|\mathbf{r}^d) = \phi_1(r_+; \mu_{+|d}, \sigma_{+|d}^2), \tag{3}$$

where

$$\begin{aligned} \boldsymbol{E}[r(\mathbf{x}_+)|\mathbf{r}^d] = \mu_{+|d} &= \mu_r + \boldsymbol{\sigma}_{+d}[\boldsymbol{\Sigma}_d]^{-1}(\mathbf{r}^d - \mu_r \boldsymbol{i}_m) \\ &= \mu_r + \boldsymbol{\rho}_{+d}[\boldsymbol{\Sigma}_d^\rho]^{-1}(\mathbf{r}^d - \mu_r \boldsymbol{i}_m) \\ \boldsymbol{Var}[r(\mathbf{x}_+)|\mathbf{r}^d] = \sigma_{+|d}^2 &= \sigma_r^2 - \boldsymbol{\sigma}_{+d}[\boldsymbol{\Sigma}_d]^{-1}\boldsymbol{\sigma}_{d+} \\ &= \sigma_r^2[1 - \boldsymbol{\rho}_{+d}[\boldsymbol{\Sigma}_d^\rho]^{-1}\boldsymbol{\rho}_{d+}]. \end{aligned}$$

Based on these relations, one obtains the prediction and prediction variance at the arbitrary location $\mathbf{x}_+ \in \mathtt{D}$

$$\begin{aligned} \hat{r}_+ = \mu_{+|d} &= \mu_r + \boldsymbol{\rho}_{+d}[\boldsymbol{\Sigma}_d^\rho]^{-1}(\mathbf{r}^d - \mu_r \boldsymbol{i}_m) \\ \sigma_p^2 = \sigma_{+|d}^2 &= \sigma_r^2[1 - \boldsymbol{\rho}_{+d}[\boldsymbol{\Sigma}_d^\rho]^{-1}\boldsymbol{\rho}_{d+}]. \end{aligned} \tag{4}$$

Under the current Gaussian RF model assumptions, these expressions are exact and optimal in both a maximum posterior and squared-error sense, given the model parameters $[\mu_r, \sigma_r^2, \rho_r(\boldsymbol{\tau})]$. Maximum marginal likelihood estimators for these model parameters are specified in Appendix A.

The objective is to provide a predictor of the spatial variable $\{r(\mathbf{x}); \mathbf{x} \in \mathtt{D}\}$ based on the observation set $\mathbf{r}^d$, with associated predictor variances. Consider the model parameters $[\mu_r, \sigma_r^2, \rho_r(\boldsymbol{\tau})]$ to be known. The predictor for the value of the spatial variable at an arbitrary location $\mathbf{x}_+ \in \mathtt{D}$, specified in Eq. (4), must be activated as $\mathbf{x}_+$ slides across the spatial domain $\mathtt{D}$. There are two alternative representations of the spatial predictor, either a spatially discretized grid representation or a spatially continuous functional representation.

The traditional kriging predictor (see Chiles and Delfiner (2012) and Omre et al. (2024)), being the best linear unbiased (BLU) predictor in the squared-error sense, coincides with the conditional expectation under Gaussian RF model assumptions. The spatial predictor is usually on a grid representation in the n nodes of a regular

grid $\mathtt{L}$ covering the reference domain $\mathtt{D}$. The grid $\mathtt{L} \subset \mathtt{D}$ must be dense relative to the observation density in order to capture the variability caused by the conditioning; hence $n \gg m$. From Eq. (4) one has the grid representation

$$\{\hat{r}(\mathbf{x}) = \mu_r + \boldsymbol{\alpha}^{xT}(\mathbf{r}^d - \mu_r \boldsymbol{i}_m) = \mu_r + \Sigma_{\mathbf{y}\in\mathtt{M}}\alpha^x_{\mathbf{y}}(r^d_{\mathbf{y}} - \mu_r); \mathbf{x} \in \mathtt{L} \subset \mathtt{D}\}, \quad (5)$$

with kriging weight m-vector $\boldsymbol{\alpha}^x = [\boldsymbol{\Sigma}_d]^{-1}\boldsymbol{\sigma}_{dx} = [\boldsymbol{\Sigma}^\rho_d]^{-1}\boldsymbol{\rho}_{dx} = (\alpha^x_1, \ldots, \alpha^x_m)^T$. Hence, the predictor appears as a linear combination of the observed values $\mathbf{r}^d$. The corresponding prediction variance expression in grid representation is

$$\{\sigma^2_p(\mathbf{x}) = \sigma^2_r[1 - \boldsymbol{\alpha}^{xT}\boldsymbol{\Sigma}^\rho_d\boldsymbol{\alpha}^x]; \mathbf{x} \in \mathtt{L} \subset \mathtt{D}\}. \quad (6)$$

The kriging weight m-vector $\boldsymbol{\alpha}^x$ is dependent on the prediction location $\mathbf{x} \in \mathtt{D}$ and the observation locations $\mathtt{M}$, but independent of the actual values of the observations $\mathbf{r}^d$. Hence, these weights are dependent only on the model parameters and are therefore deterministic. The $(\mathbf{r}^d - \mu_r \boldsymbol{i}_m)$-term is random, however, with $\boldsymbol{E}[\mathbf{r}^d - \mu_r \boldsymbol{i}_m] = 0\boldsymbol{i}_m$ and $\boldsymbol{Var}[\mathbf{r}^d - \mu_r \boldsymbol{i}_m] = \boldsymbol{\Sigma}_d$.

The weight m-vector $\boldsymbol{\alpha}^x = [\boldsymbol{\Sigma}^\rho_d]^{-1}\boldsymbol{\rho}_{dx}$ must be calculated for each grid node, hence n times. These calculations need not be too computationally demanding, however, since the computationally demanding term $[\boldsymbol{\Sigma}^\rho_d]^{-1}$ is global and common in all weight calculations, and therefore only needs to be calculated once. The spatial predictor is based on this grid representation, and it is usually defined as either piecewise constant or piecewise planar within each grid unit. The former predictor, at an arbitrary location, takes the value of the closest grid node. The latter predictor, at an arbitrary location, is based on a linear interpolator of values in the closest grid nodes. The associated prediction variance can be interpolated accordingly. These interpolations will of course only provide approximations of the optimal prediction and correct prediction variance.

The kernel predictor introduced in this study has a form similar to the dual kriging predictor (see Royer (1984) and Chiles and Delfiner (2012)). It appears as an alternative phrasing of the conditional expectation under the Gaussian RF model assumptions. From Eq. (4) one has the spatial predictor in functional representation

$$\{\hat{r}(\mathbf{x}) = \mu_r + \boldsymbol{\rho}_{xd}\boldsymbol{\alpha}^d = \mu_r + \Sigma_{\mathbf{y}\in\mathtt{M}}\alpha^d_{\mathbf{y}}\nu_{\mathbf{y}}(\mathbf{x}); \mathbf{x} \in \mathtt{D}\}, \quad (7)$$

with weight m-vector $\boldsymbol{\alpha}^d = \sigma^2_r[\boldsymbol{\Sigma}_d]^{-1}(\mathbf{r}^d - \mu_r \boldsymbol{i}_m) = [\boldsymbol{\Sigma}^\rho_d]^{-1}(\mathbf{r}^d - \mu_r \boldsymbol{i}_m) = (\alpha^d_1, \ldots, \alpha^d_m)^T$. The observation kernel functions are $\nu_{\mathbf{y}}(\mathbf{x}) = \rho_r(\mathbf{x}-\mathbf{y}); \mathbf{y} \in \mathtt{M}$. Hence, the predictor appears as a linear combination of the observation kernel functions, being identical to the correlations between the spatial variable value to be predicted and the observation values. Thus, $\nu_i(\mathbf{x}) = \rho_r(\mathbf{x} - \mathbf{x}^d_i); i = 1, \ldots, m$, with i and $\mathbf{y}$ being matching observation numbers and locations. The corresponding prediction variance expression in functional representation is

$$\begin{aligned}\{\sigma^2_p(\mathbf{x}) &= \sigma^2_r[1 - \boldsymbol{\rho}_{xd}[\boldsymbol{\Sigma}^\rho_d]^{-1}\boldsymbol{\rho}_{dx}] \\ &= \sigma^2_r[1 - \Sigma_{\mathbf{y}'\in\mathtt{M}}\Sigma_{\mathbf{y}''\in\mathtt{M}}\beta_{\mathbf{y}'\mathbf{y}''} \times \nu_{\mathbf{y}'}(\mathbf{x})\nu_{\mathbf{y}''}(\mathbf{x})]; \mathbf{x} \in \mathtt{D}\}, \end{aligned} \quad (8)$$

with weights $\beta_{\mathbf{y}'\mathbf{y}''} = [[\boldsymbol{\Sigma}_d^{\rho}]^{-1}]_{ij}; i, j = 1, \ldots, m$, with (i, j) and $(\mathbf{y}', \mathbf{y}'')$ being matching observation numbers and locations. The m-vector of weights in the kernel predictor $\boldsymbol{\alpha}^d$ is dependent on the observation locations $\mathtt{M}$ and the actual observed values $\mathbf{r}^d$, but independent of the prediction location $\mathbf{x} \in \mathtt{D}$. Thus these weights are dependent on the random observed values and are therefore random. Note further that $\boldsymbol{E}[\boldsymbol{\alpha}^d] = 0\boldsymbol{i}_m$ and $\boldsymbol{Var}[\boldsymbol{\alpha}^d] = [\sigma_r^2]^2[\boldsymbol{\Sigma}_d]^{-1} = \sigma_r^2[\boldsymbol{\Sigma}_d^{\rho}]^{-1}$. The observation kernel functions corresponding to the entries of $\boldsymbol{\rho}_{xd}$ only depend on the model parameters, and thus they are deterministic.

The weight m-vector $\boldsymbol{\alpha}^d = [\boldsymbol{\Sigma}_d^{\rho}]^{-1}(\mathbf{r}^d - \mu_r \boldsymbol{i}_m)$ is global and only has to be calculated once. The functional representation of the predictor appears as a weighted linear combination of the symmetric spatial correlation functions centered at the m observation locations. Thus, this set of observation kernel functions spans the prediction space for all possible observed values in the given observation design. Since the kernel predictor is in functional representation, no spatial grid discretization is required.

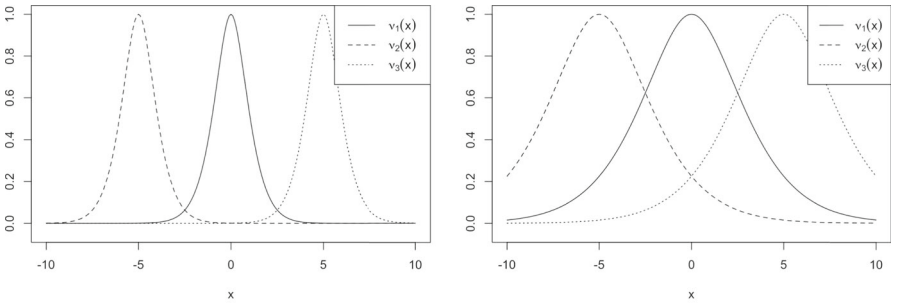
To summarize, both the kriging and the kernel predictor expressions are exact and optimal predictors for the values of the spatial variable at the grid nodes—and hence are identical. This is the case for all valid Gaussian RF models, and therefore for all non-negative definite spatial correlation functions. The functional representation of the kernel predictor is defined for all $\mathbf{x} \in \mathtt{D}$ and appears as the grid infill asymptotic limit (see Stein (1999)) for the grid representation of the traditional kriging predictor. This limit can be assessed by the kernel predictor, with computational demand dominated by the inversion of the observation correlation $(m \times m)$-matrix $\boldsymbol{\Sigma}_d^{\rho}$. This computational demand is similar to that for the kriging predictor. The storage demand for the kernel predictor is clearly preferable to that for the kriging one. To assess the prediction and prediction variance at an arbitrary location the functional value of the kernel functional representation must be calculated to obtain the optimal values. For the kriging grid representation, the piecewise planar interpolator in the actual grid unit must be activated to obtain approximate values. These two operations have comparable computational demands. Thus, the kernel predictor appears as favorable to the kriging one.

Since the functional representation is the grid infill asymptotic limit, it can be efficiently reduced to a grid representation in arbitrary design, but not vice versa. Moreover, the functional representation makes it possible to calculate analytically exact and optimal predictions, with associated prediction variances, for any linear operator on the Gaussian RF. Examples of such operators are integration over a specific area and differentiation in an arbitrary location, if they exist.

Example A

This example contains a demonstration of the characteristics of the kernel predictor in a case with a spatial variable defined in a one-dimensional reference domain $\{r(x); x \in \mathtt{D} : [-10, 10] \subset \mathbb{R}\}$. The stationary Gaussian RF model has an expectation level $\mu_r = 0.0$ and variance level $\sigma_r^2 = 1.0$. The spatial correlation function belongs to the Matérn class of functions with shape parameter $5/2$ and range parameter $\tau_M \in \mathbb{R}_+$. Thus, the spatial correlation function for $\tau = |x' - x''|/\tau_M; x', x'' \in \mathtt{D}$ is

$$\rho_r(\tau) = [1 + 5^{1/2}\tau + 5/3\tau^2] \times \exp(-5^{1/2}\tau).$$



(a) Range parameter $\tau_M = 1.0$ **(b)** Range parameter $\tau_M = 3.0$

Fig. 1 Observation kernel functions: Gaussian RF model with Matérn spatial correlation function—shape parameter $5/2$

The observation set is represented by the 3-vector $\mathbf{d}$ consisting of three exact location-wise observations at $x = 0$, $x = -5$, and $x = 5$

$$\begin{aligned} d_1 &= r(0) \\ d_2 &= r(-5) \\ d_3 &= r(5). \end{aligned}$$

The corresponding observation kernel functions are

$$\begin{aligned} \nu_1(x) &= \rho_r(x) \\ \nu_2(x) &= \rho_r(x-5) \\ \nu_3(x) &= \rho_r(x+5). \end{aligned}$$

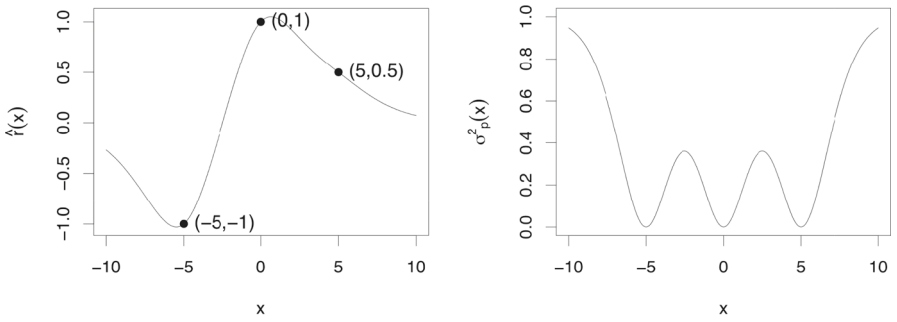
In Fig. 1, these observation kernel functions are displayed for $\tau_M = 1.0$ and $\tau_M = 3.0$, respectively. Note that the observation kernel function $\nu_1(x)$ is centered at $x = 0$, and thus the value $r(0)$ and the observation d_1 have a correlation of 1.0 at the center. The observation kernel function $\nu_2(x)$ is centered at $x = -5$, and the value $r(-5)$ and the observation d_2 have a correlation 1.0 at location $x = -5$, whereas the correlation between $r(5)$ and d_3 is 1.0 at location $x = 5$.

Consider the spatial correlation model with range parameter $\tau_M = 3.0$ displayed in Fig. 1b. Let the observed values be $\mathbf{d} = (d_1, d_2, d_3)^T = (1.0, -1, 0, 0.5)^T$. The weights in the predictor are calculated as $\boldsymbol{\alpha}^d = [\boldsymbol{\Sigma}_d^{\rho}]^{-1}\mathbf{d}$, and the actual values are computed as $\boldsymbol{\alpha}^d = (1.234, -1.282, 0.242)^T$. The kernel predictor in functional representation is then

$$\{\hat{r}(x) = [\boldsymbol{\nu}(x)]^T\boldsymbol{\alpha}^d = \Sigma_{i=1}^3 \alpha_i^d \nu_i(x);\, x \in \mathtt{D}\},$$

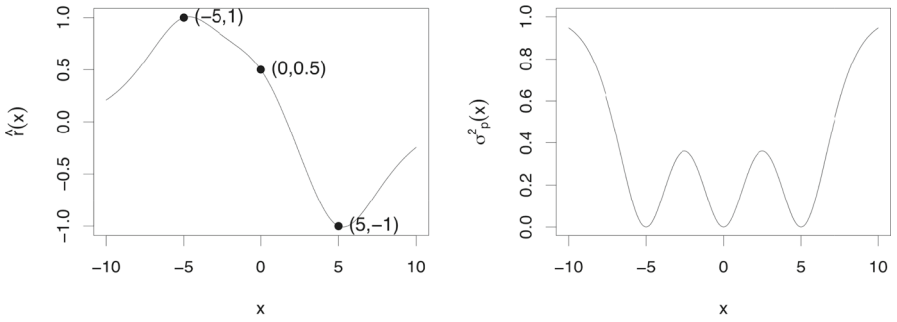
with associated prediction variance in functional representation

$$\{\sigma_p^2(x) = 1 - [\boldsymbol{\nu}(x)]^T[\boldsymbol{\Sigma}_d^{\rho}]^{-1}\boldsymbol{\nu}(x) = 1 - \Sigma_{i=1}^3\Sigma_{j=1}^3 \beta_{ij}\nu_i(x)\nu_j(x);\, x \in \mathtt{D}\},$$



(a) Kernel prediction **(b)** Kernel prediction variance

Fig. 2 Kernel predictor: Gaussian RF model with Matérn spatial correlation function—shape 5/2 and range 3.0—observations (1.0, −1.0, 0.5)



(a) Kernel prediction **(b)** Kernel prediction variance

Fig. 3 Kernel predictor: Gaussian RF model with Matérn spatial correlation function—shape 5/2 and range 3.0—observations (0.5, 1.0, −1.0)

with $\beta_{ij} = [[\boldsymbol{\Sigma}_d^{\rho}]^{-1}]_{ij}; i, j = 1, \ldots, 3$. In Fig. 2 the results from the kernel predictor are displayed. The left display contains the spatially continuous prediction with the exact observations correctly reproduced. The predictions regress towards $\mu_r = 0$ as the observation influence decreases at both sides of the spatial interval. The right display contains the spatially continuous prediction variances. The exact observations of the spatial variable at $x = 0$, $x = -5$, and $x = 5$ entail that the prediction variances are zero at these locations. The prediction variance approaches $\sigma_r^2 = 1.0$ as the observation influence decreases at both sides of the spatial interval.

Lastly, the same Gaussian RF model with identical observation design to that above is used, resulting in the same observation kernel functions as in Fig. 1b. The observation values are changed however, to $\mathbf{d} = (0.5, 1.0, -1.0)^T$. This change does influence the weights, which are calculated as $\boldsymbol{\alpha}^d = (0.555, 0.893, -1.139)^T$. In Fig. 3 the results from the kernel predictor are presented. The left display contains the prediction which reproduces the observed values. The prediction variances in the right display are identical to those in the previous case, however, since for Gaussian RF models, the prediction variances are independent of the actual observed values.

3 Spatial predictors for large observation sets

Both predictors discussed in Sect. 2 require the observation correlation $(m \times m)$-matrix $\mathbf{\Sigma}_d^\rho$ to be inverted, which is usually achieved by Cholesky decomposition. For large observation sets, and hence large m, this inversion may be computationally demanding, since general matrix inversion requires computer processing of order m^3. If the $(m \times m)$-matrix is sparse, however, efficient algorithms for Cholesky decomposition of orders $m^{3/2}$ for reference domain $\mathtt{D} \subset \mathbb{R}^2$ and m^2 for $\mathtt{D} \subset \mathbb{R}^3$ are available (see Rue and Held (2005)). The recommended mitigation strategy for obtaining a sparse observation correlation $(m \times m)$-matrix is to constrain the model space for the spatial variable $\{r(\mathbf{x}); \mathbf{x} \in \mathtt{D}\}$ (see Omre et al. (2024)). The Gaussian RF is assumed to have a finite-range non-negative definite spatial correlation function; hence

$$\rho_r(\boldsymbol{\tau}) = \begin{cases} 1.0 & |\boldsymbol{\tau}| = 0.0 \\ \in [-1.0, 1.0] & 0.0 < |\boldsymbol{\tau}| < \tau_0 \\ 0.0 & |\boldsymbol{\tau}| \geq \tau_0 \end{cases},$$

with finite-range value $\tau_0 < \infty$. This class of finite-range correlation functions is large (see Gneiting (2002)). A suitable choice of range values τ_0 will cause the observation correlation $(m \times m)$-matrix $\mathbf{\Sigma}_d^\rho$ to be sparse, and sometimes even blocked. Efficient algorithms can then be used for Cholesky decomposition, and thus the inverse of $\mathbf{\Sigma}_d^\rho$ can be calculated.

Moreover, the spatial kernel predictor (see Eq. (7)) will for the finite-range Gaussian RF model be a local predictor, since $\rho_r(\boldsymbol{\tau}) = 0$ for $|\boldsymbol{\tau}| > \tau_0$; hence

$$\{\hat{r}(\mathbf{x}) = \mu_r + \Sigma_{\mathbf{y} \in \mathtt{M}_x^{\tau_0}} \alpha_{\mathbf{y}}^d \rho_r(\mathbf{x} - \mathbf{y}); \mathbf{x} \in \mathtt{D}\},$$

with the observation location subset $\mathtt{M}_x^{\tau_0} = \{\mathbf{y} | \mathbf{y} \in \mathtt{M}, |\mathbf{x} - \mathbf{y}| < \tau_0\}$ containing observation locations which are closer than τ_0 to location $\mathbf{x}$. This spatial, local kernel predictor is exact and optimal for the specified finite-range Gaussian RF model. The corresponding spatial kriging predictor in Eq. (5) will not appear as a local predictor for this finite-range Gaussian RF model.

Consider studies with huge observation sets, hence huge m, for which inversion of the observation correlation $(m \times m)$-matrix $\mathbf{\Sigma}_d^\rho$ is computationally infeasible even for actual finite-range Gaussian RF models. We recommend a mitigation strategy that in addition to a finite-range spatial correlation function enforces some type of localization. Both predictors will then appear as approximate.

The localized kriging predictor at a given location $\mathbf{x} \in \mathtt{D}$ is based only on observations located within a distance $k\tau_0$ from $\mathbf{x}$ (see Chiles and Delfiner (2012)), with k being a user-specified constant. The size of this subset of observations is denoted as m_x. To develop the localized kriging predictor for location $\mathbf{x}$, the corresponding observation correlation $(m_x \times m_x)$-matrix must be inverted. This inversion is feasible if k is specified such that $m_x \ll m$. In a grid representation of the spatial variable, localized kriging predictors must be activated at each grid node location. The localization subsets of observations may vary from one grid node to the other. Generally, unique observation correlation matrices for these subsets must be inverted at each

grid node. Thus, the computational demand for generating the grid representation is proportional to the grid size n, with a fairly large proportionality constant.

The grid representation based on the localized kriging predictor is exact in the sense that if an exact observation at a grid node location is available, it is precisely reproduced. Exact observations located outside the grid nodes are of course only approximately reproduced. The grid representation will however contain prediction discontinuities caused by extreme observations being either included or excluded from the localization subsets of observations.

The localized kernel predictor, with user-specified localization factor k, replaces the troublesome inverse observation correlation $(m \times m)$-matrix $[\boldsymbol{\Sigma}_d^{\rho}]^{-1}$ by an approximate $[\boldsymbol{\Sigma}_d^{\rho}]^{-1*}$. First, each line in the $(m \times m)$-matrix, corresponding to one specific observation, is enforced as sparse. Nonzero values are eligible only for entries corresponding to observation locations within a $k\tau_0$ neighborhood of the actual observation. This assignment requires the inversion of a neighborhood observation correlation submatrix. After sparsifying all lines, a sparse and non-symmetric $(m \times m)$-matrix is obtained. This matrix is symmetrized, which retains sparseness but does not ensure non-negative definiteness. The resulting $(m \times m)$-matrix is defined as the approximate $[\boldsymbol{\Sigma}_d^{\rho}]^{-1*}$. The major computational saving is obtained by inverting m smaller neighborhood observation correlation sub-matrices instead of inverting one huge observation correlation $(m \times m)$-matrix. The algorithm for obtaining this approximate matrix is presented in Algorithm 1.

Algorithm 1 *(Inverse observation correlation matrix approximation)*

Initiate,

Localization range: $\Delta \in \mathbb{R}_+$

Support matrix: $\boldsymbol{\Psi} = 0\boldsymbol{I}_m$ *- dim* $(m \times m)$

For $\mathbf{x}_i^d; i = 1, \ldots, m$

Define neighborhood set,

$M_{\mathbf{x}_i^d}^{\Delta} = \{\mathbf{y} | \mathbf{y} \in M, |\mathbf{x}_i^d - \mathbf{y}| < \Delta\}$ *- dim* $m_{\mathbf{x}_i^d}^{\Delta}$

Construct matrix,

$\boldsymbol{\Sigma}_{d_{\mathbf{x}_i^d}^{\Delta}}^{\rho} =$ *Sub-matrix* $\{\boldsymbol{\Sigma}_d^{\rho}; M_{\mathbf{x}_i^d}^{\Delta}\}$ *- dim* $(m_{\mathbf{x}_i^d}^{\Delta} \times m_{\mathbf{x}_i^d}^{\Delta})$

Compute,

$[\boldsymbol{\Sigma}_{d_{\mathbf{x}_i^d}^{\Delta}}^{\rho}]^{-1}$ *- dim* $(m_{\mathbf{x}_i^d}^{\Delta} \times m_{\mathbf{x}_i^d}^{\Delta})$

Copy,

$[\boldsymbol{\Psi}]_{ij} = [[\boldsymbol{\Sigma}_{d_{\mathbf{x}_i^d}^{\Delta}}^{\rho}]^{-1}]_{\mathbf{x}_i^d \mathbf{y}}$ *- for corresponding* j *and* $\mathbf{y} \in M_{\mathbf{x}_i^d}^{\Delta}$ *entries*

End For

Define,

$[\boldsymbol{\Sigma}_d^{\rho}]^{-1*} = 1/2 \times [\boldsymbol{\Psi} + \boldsymbol{\Psi}^T]$.

The observation correlation $(m \times m)$-matrix $\boldsymbol{\Sigma}_d^\rho$ has lines and columns corresponding to the entries in the observation location set $\mathtt{M}$. The operator sub-matrix$\{\boldsymbol{\Sigma}_d^\rho; \mathtt{M}^\Delta_{\mathbf{x}_i^d}\}$ delivers the sub-matrix of $\boldsymbol{\Sigma}_d^\rho$ containing the line and columns corresponding to the observation location subset $\mathtt{M}^\Delta_{\mathbf{x}_i^d} \subset \mathtt{M}$, hence removing the line and columns corresponding to $\mathtt{M} \setminus \mathtt{M}^\Delta_{\mathbf{x}_i^d}$. The support $(m \times m)$-matrix $\boldsymbol{\Psi}$ will be sparse since the location set $\mathtt{M}^\Delta_{\mathbf{x}_i^d}$ is a subset of $\mathtt{M}$, and hence not all entries in line i will be assigned new values. In addition, the approximation $[\boldsymbol{\Sigma}_d^\rho]^{-1*}$ will be sparse and it will be symmetric, but usually not non-negative definite.

The corresponding localized kernel predictor is defined with localization range $\Delta = k\tau_0$, and the functional representation is

$$\begin{aligned}\{\hat{r}^*(\mathbf{x}) &= \mu_r^* + \boldsymbol{\rho}_{xd}\boldsymbol{\alpha}^{d*} = \mu_r^* + [\boldsymbol{\nu}(\mathbf{x})]^T\boldsymbol{\alpha}^{d*} \\ &= \mu_r^* + \Sigma_{\mathbf{y}\in\mathtt{M}}\alpha_{\mathbf{y}}^{d*}\nu_{\mathbf{y}}(\mathbf{x}) \\ &= \mu_r^* + \Sigma_{\mathbf{y}\in\mathtt{M}_x^{\tau_0}}\alpha_{\mathbf{y}}^{d*}\nu_{\mathbf{y}}(\mathbf{x});\ \mathbf{x} \in \mathtt{D}\},\end{aligned} \tag{9}$$

with the approximate weight m-vector $\boldsymbol{\alpha}^{d*} = [\boldsymbol{\Sigma}_d^\rho]^{-1*}(\mathbf{r}^d - \mu_r^*\boldsymbol{i}_m)$. The associated localized prediction variance in functional representation is

$$\begin{aligned}\{\sigma_p^{2*}(\mathbf{x}) &= \sigma_r^{2*}[1 - \boldsymbol{\rho}_{xd}[\boldsymbol{\Sigma}_d^\rho]^{-1*}\boldsymbol{\rho}_{dx}] = \sigma_r^{2*}[1 - [\boldsymbol{\nu}(\mathbf{x})]^T[\boldsymbol{\Sigma}_d^\rho]^{-1*}\boldsymbol{\nu}(\mathbf{x})] \\ &= \sigma_r^{2*}[1 - \Sigma_{\mathbf{y}'\in\mathtt{M}}\Sigma_{\mathbf{y}''\in\mathtt{M}}\beta^*_{\mathbf{y}'\mathbf{y}''}\nu_{\mathbf{y}'}(\mathbf{x})\nu_{\mathbf{y}''}(\mathbf{x})] \\ &= \sigma_r^{2*}[1 - \Sigma_{\mathbf{y}'\in\mathtt{M}_\mathbf{x}^{\tau_0}}\Sigma_{\mathbf{y}''\in\mathtt{M}_\mathbf{x}^{\tau_0}}\beta^*_{\mathbf{y}'\mathbf{y}''}\nu_{\mathbf{y}'}(\mathbf{x})\nu_{\mathbf{y}''}(\mathbf{x})];\ \mathbf{x} \in \mathtt{D}\},\end{aligned} \tag{10}$$

with $\beta^*_{\mathbf{y}'\mathbf{y}''} = [[\boldsymbol{\Sigma}_d^\rho]^{-1*}]_{ij}$, where $(i,\ j)$ and $(\mathbf{y}',\ \mathbf{y}'')$ are matching observation numbers and locations. The approximate prediction variances are not ensured to be in the eligible interval $[0.0,\ \sigma_r^{2*}]$; hence, post-adaption is necessary. Both predictions and prediction variances are on functional representations without any spatial grid discretization. The last equality in the expressions above follows from the spatial correlation function with finite range $\rho_r(\boldsymbol{\tau}) = 0.0;\ |\boldsymbol{\tau}| > \tau_0$, and it causes the expressions to be spatially local. The model parameters $[\mu_r,\ \sigma_r^2]$, being the expectation and variance levels, respectively, can be estimated by inserting the approximation $[\boldsymbol{\Sigma}_d^\rho]^{-1*}$ for $[\boldsymbol{\Sigma}_d^\rho]^{-1}$ in the estimator expressions in Appendix A.

The computational demands are extremely favorable. The processing time is proportional to the number of observations m, almost independent of the dimension of the reference space $\mathtt{D}$. Since no spatial discretization is involved, the processing time is of course independent of any grid geometry. Moreover, the approximation algorithm is very simple to implement and is suitable for parallel processing on a computer.

The localized kernel predictor appears with correct spatial characteristics according to the prior Gaussian RF model. The approximation will cause no prediction discontinuities. The localized predictor will however be non-exact in the sense that exactly observed values will not be precisely reproduced by the predictor. The estimated deviations variance σ_*^2 between exact observations and predictions in observation locations provides a quantification of the approximation error. The variance is zero

for the global kernel predictor and increases with decreasing k and rougher approximation. The prediction variances can be adjusted for this approximation uncertainty by defining $\{\hat{\sigma}_p^2(\mathbf{x}) = \sigma_p^{2*}(\mathbf{x}) + \sigma_*^2; \mathbf{x} \in \mathtt{D}\}$, which most likely takes values in the eligible domain for variances. The functional representation based on the localized kernel predictor can be efficiently reduced to a grid representation in arbitrary design. The respective grid values are assessed by computing the functional values at each grid node location.

To summarize, for moderate to large observation sets, and hence moderate to large m, the observation correlation $(m \times m)$-matrix $\mathbf{\Sigma}_d^\rho$ may be sparsified by assuming a finite-range Gaussian RF model. Thus, efficient algorithms for Cholesky decomposition can be used. For huge observation sets where not even sparsification of the observation correlation matrix $\mathbf{\Sigma}_d^\rho$ makes the inversion computationally feasible, localization approximations can be used. The localized kriging predictor requires a reduced observation correlation matrix to be inverted for each grid node $\mathbf{x} \in \mathtt{L}$, hence n times. The localized kernel predictor requires a reduced observation correlation matrix to be inverted for each observation location $\mathbf{x} \in \mathtt{M}$, hence m times. Note that usually $n \gg m$; thus, one will expect that the localized kernel predictor is more computationally efficient than the localized kriging predictor.

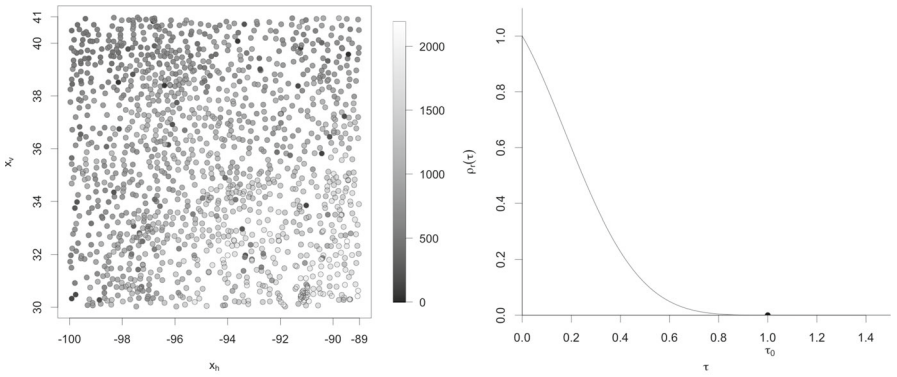
The functional representation based on the localized kernel predictor can be reduced to an arbitrary grid representation by assessing the functional value n times. Each assessment only involves calculation of a linear expression. The same grid representation, based on the localized kriging predictor, requires correlation matrices for subsets of the observation set to be inverted for each of the n grid nodes. Thus, for very large grid designs, taking the bypass through the localized Kernel function representation may be computationally preferable.

Example B

This example presents a demonstration of the characteristics of the localized kernel predictor on a real observation set with some challenging features. The computational demands for obtaining the localized kernel functional representation and its reduction into a grid representation are compared to the demand for assessing the corresponding localized kriging grid representation.

The spatial variable under study is accumulated precipitation in millimeters during January–December 1997 in a sub-region of the United States. The observation set consists of 1,330 locations with observed precipitation values, after removing missing and exact-zero observations. The set has its origin in the US National Climatic Data Center, and a thorough discussion can be found in Johns et al. (2003). The observation set is not huge, but it is easily tractable and sufficiently large to demonstrate the characteristics of the localized kernel predictor.

Figure 4a displays the $m = 1{,}330$ observation locations with exactly observed values of precipitation. The spatial reference $\mathbf{x} = (x_h, x_v) \in \mathtt{D} \subset \mathbb{R}^2$ is defined as (longitude,latitude), and the distance between two locations is defined as the Euclidean distance. The observations are relatively evenly located in the spatial reference domain. The values, assumed to be exact observations of the precipitation, seem to exhibit a slight trend in the northwest direction. Several observations with extreme values are in locations spread across the domain. These spike patterns are difficult to capture and reproduce in the predictions.



(a) Locations and values - 1330 observations

(b) Finite-range spatial correlation function

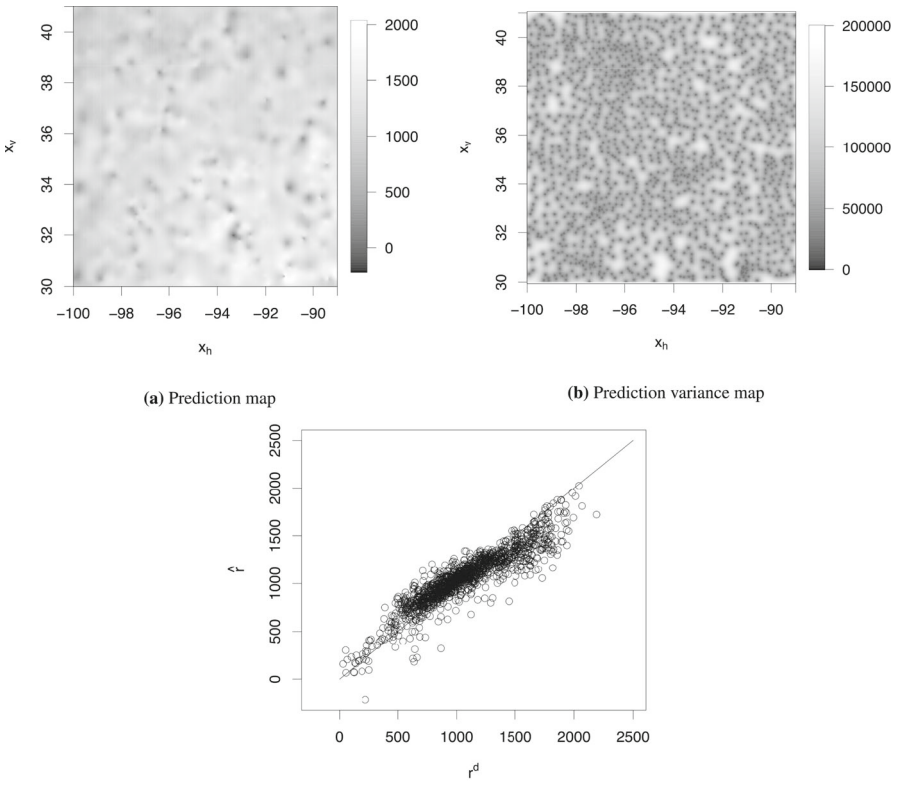
Fig. 4 Observation set and model parameters

Despite the slight trend, a stationary, isotropic Gaussian RF model is defined. The model parameters $[\mu_r, \sigma_r^2, \boldsymbol{\eta}_r]$ are the expectation and variance levels and the parameters of the spatial correlation function $\rho_r(\boldsymbol{\tau}; \boldsymbol{\eta}_r)$. After a short preliminary study, we assign a finite-range second-order exponential correlation function

$$\begin{aligned}\rho_r(\boldsymbol{\tau}) &= \rho_B(\boldsymbol{\tau}) \times \rho_R(\boldsymbol{\tau}) \\ &= \begin{cases} 1.0 & |\boldsymbol{\tau}| = 0.0 \\ \exp\{-(|\boldsymbol{\tau}|/0.5)^2\} \times (1 + |\boldsymbol{\tau}|/2)(1 - |\boldsymbol{\tau}|)^2 & 0.0 < |\boldsymbol{\tau}| < 1.0 \\ 0.0 & |\boldsymbol{\tau}| \geq 1.0. \end{cases}\end{aligned}$$

The correlation function is displayed in Fig. 4b, and the finite range is $\tau_0 = 1.0$. It is typically 30 observations from the set located within the range disc with radius τ_0. The two other model parameters $[\mu_r, \sigma_r^2 | \boldsymbol{\eta}_r]$ will later be estimated from the observations. The observation correlation $(m \times m)$-matrix $\boldsymbol{\Sigma}_d^\rho$ contains entries $[\boldsymbol{\Sigma}_d^\rho]_{\mathbf{y}'\mathbf{y}''} = \rho_r(\mathbf{y}' - \mathbf{y}'')$; $\mathbf{y}', \mathbf{y}'' \in \mathtt{M}$. The corresponding observation kernel function m-vector $\{\boldsymbol{\nu}(\mathbf{x}); \mathbf{x} \in \mathtt{D}\}$ contains entries $\nu_{\mathbf{y}}(\mathbf{x}) = \rho_r(\mathbf{x} - \mathbf{y})$; $\mathbf{y} \in \mathtt{M}$. Since the spatial correlation function is defined as having a finite range, both the matrix and the vector will be sparse.

The computations are performed on a regular MacBook Pro laptop computer with standard sequential R code without any refined implementation. The localized kernel predictor, as defined above, relies on the specification of a neighborhood parameter k. We present the prediction results for two alternative values of this parameter, $k = 0.5$ and $k = 1.0$. Thus, for the smaller neighborhood, the radius is $k\tau_0 = 0.5$, and it typically contains eight observations from the set, approximately one quarter of the number of observations within the range radius. The computational demand typically includes 1,330 inversions of (8×8)-matrices. The approximate inverse observation correlation $(1{,}330 \times 1{,}330)$-matrix $[\boldsymbol{\Sigma}_d^\rho]^{-1*}$ is obtained. The deliveries from the localized kernel predictor are prediction and prediction variance in functional representations over the study domain. The model parameters expectation and variance level are estimated as $\mu_r^* = 1{,}062$ and $\sigma_r^{2*} = 176{,}617$.



(a) Prediction map

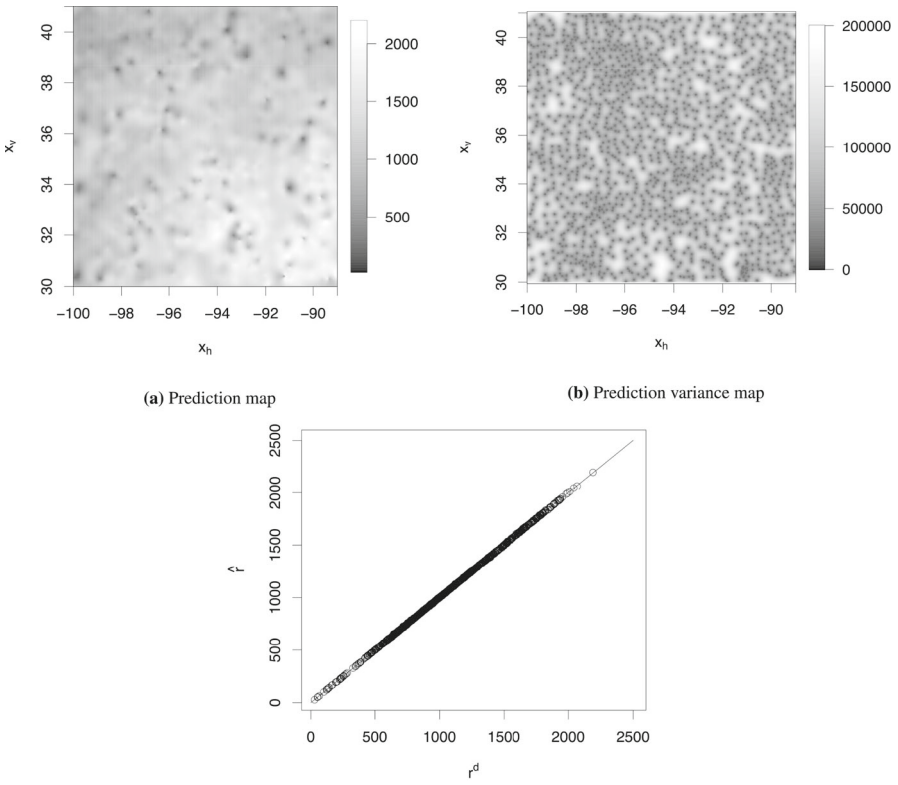
(b) Prediction variance map

(c) Value versus prediction cross-plot in observation locations

Fig. 5 Localized kernel predictor: neighborhood parameter $k = 0.5$

Figure 5 contains three displays: the prediction map, the prediction variance map, and the prediction error cross-plot. The prediction is based on Eq. (9), which is in functional representation. The functional representation is reduced to a grid representation of dimension $(1,101 \times 1,101)$ with grid size $n = 1,212,201$. This grid representation is used in the map construction. This conversion can be made very efficiently because the predictor is spatially local. The spatial smoothness of the localized kernel prediction in Fig. 5a is according to the Gaussian RF model without localization artifacts. Note that the spike patterns are reasonably well reproduced. In the global kernel predictor, the observed values will be exactly reproduced, but in the localized predictor they may not be. The prediction variance is based on Eq. (10), which is in functional form, whereas the map is based on a reduced grid representation as above. Note that the prediction variances in Fig. 5b approach zero at the observation locations, as they should. Some negative values appear, however, which of course are not eligible as prediction variances. The cross-plot in Fig. 5c displays the prediction errors in the observation locations, which appear significant, with deviation variance $\sigma_*^2 = 26,244$.

For the larger neighborhood, the radius is $k\tau_0 = 1.0$, and it typically contains 30 observations from the set, approximately the average number of observations within



(a) Prediction map

(b) Prediction variance map

(c) Value versus prediction cross-plot in observation locations

Fig. 6 Localized kernel predictor: neighborhood parameter $k = 1.0$

the range radius. The deliveries from the localized kernel predictor are prediction and prediction variance in functional representations over the study domain, and the computer runtime is 9 s. The expectation and variance levels of the model parameters are estimated as $\mu_r^* = 1{,}078$ and $\sigma_r^{2*} = 189{,}984$. The functional representations are reduced to the grid representation of size $n = 1{,}212{,}201$, with computer runtime of 994 s. The maps are based on this reduced grid representation.

Figure 6 contains three displays: the prediction map, the prediction variance map, and the prediction error cross-plot. The prediction is based on Eq. (9) which is in a functional representation, and the spatial smoothness is according to the Gaussian RF model without localization artifacts. Note that the spike patterns are reliably reproduced. The localized kernel predictor is not reproducing exact observations, however. The prediction variance is based on Eq. (10), which is in a functional representation. Note that the prediction variances approach zero at the observation locations, as they should. The prediction variances are in the range $[-4.09 \times 10^{-9}, 181{,}262]$; thus some ineligible, slightly negative values appear. The cross-plot in Fig. 6c displays the prediction errors at the observation locations, which appear almost on the diagonal with deviation variance $\sigma_*^2 = 189$. This deviation variance is very small compared to the

variance of the Gaussian RF model, so the exact observations are almost exactly reproduced. Lastly, a more reliable expression for the prediction variance, also capturing the approximation uncertainty, is $\{\hat{\sigma}_p^2(\mathbf{x}) = \sigma_p^{2*}(\mathbf{x}) + \sigma_*^2; \mathbf{x} \in \mathtt{D}\}$, which provides values in the eligible range for variances.

Lastly, the same grid representations of size $n = 1{,}212{,}201$ for both prediction and prediction variance are generated by the localized kriging predictor with the same $k = 1.0$. The computer code is notably comparable to the one for the localized kernel predictor. The computer runtime for the localized kriging predictor is 3,386 s. The corresponding maps appear very similar to the ones for the localized kernel predictor, and thus they are not displayed.

The relative computer runtime for generating the localized kernel functional representation versus the localized kriging grid representation with grid size $n = 1{,}212{,}201$ is $\frac{9}{994} = 9.0 \times 10^{-3}$. To provide a prediction at an arbitrary location, the functional value of the former must be calculated, whereas a piecewise planar interpolation in a grid unit of the latter is necessary. In this study, the former predictor appears significantly more computationally efficient than the latter. The relative computer runtime for generating the grid representation of size $n = 1{,}212{,}201$ by the localized kernel predictor versus the localized kriging predictor is $\frac{9+994}{3{,}386} = 3.0 \times 10^{-1}$. This relative computer runtime is expected to decrease as the ratio of the grid size to the observation set size increases. In this study, first assessing the localized kernel function representation and then reducing it to the grid representation appears to be more computationally efficient than assessing the grid representation directly using the localized kriging predictor.

4 Conclusions

The challenge is to assess the continuous spatial variable over a finite spatial domain from observations in a set of locations. Based on a Gaussian random field model, predictions and prediction variance over the domain can be presented as either a kriging grid representation or a grid-free kernel function representation. The computational demands for obtaining both these representations are dominated by the need to invert the observation correlation matrix, whereas the storage demand for the latter is significantly smaller than for the former. To assess the prediction and prediction variance in an arbitrary location, the former usually requires piecewise planar interpolations within the actual grid unit, whereas the latter requires functional values to be computed in the actual location. Note however that the former only provides approximations of the optimal values, whereas the latter provides the exact optimal values. The grid-free kernel function representation appears preferable to the kriging grid representation.

The kernel function representation constitutes the grid infill asymptotic limit of the kriging grid representation. Thus, the former can be reduced to the latter with arbitrary grid design, but not vice versa. For huge observation sets, the inversion of the observation correlation matrix may not be computationally feasible. The usual mitigation strategy is to define an approximate localized spatial predictor and use it to assess the spatial variable. The localized kriging predictor can be used to generate a grid representation with computational demand proportional to the grid size. A functional

representation can be generated by the localized kernel predictor, with computational requirements proportional to the number of observations. The latter is expected to be far more computationally efficient than the former.

The study in the example clearly indicates that the localized kernel predictor is computationally superior to the localized kriging predictor when generating the functional representation and comparable grid representation, respectively. The former representation is generated with a computer runtime 100 times that of the latter representation. The functional representation is reduced to the same grid representation with computer runtime three times as fast as that for the localized kriging predictor. The actual runtimes are of course crucially dependent on the computer configuration and the code implementation. Both approaches will gain efficiency by improving these factors. The relative computer runtimes are, however, expected to remain more stable.

Appendix A Model parameter inference

The parameters of the model are expectation level $\mu_r \in \mathbb{R}$, variance level $\sigma_r^2 \in \mathbb{R}_{\oplus}$, and the spatial correlation function $\rho_r(\boldsymbol{\tau}; \boldsymbol{\eta}_r) \in \mathbb{R}_{[-1,1]}$ parameterized by $\boldsymbol{\eta}_r \in \mathbb{R}^{\kappa}$. These model parameters may be assessed by a maximum marginal likelihood criterion

$$\begin{aligned}(\hat{\mu}_r, \hat{\sigma}_r^2, \hat{\boldsymbol{\eta}}_r) &= \arg \max_{\mu_r, \sigma_r^2, \boldsymbol{\eta}_r} \{p(\mathbf{r}^d; \mu_r, \sigma_r^2, \boldsymbol{\eta}_r)\} \\ &= \arg \max_{\mu_r, \sigma_r^2, \boldsymbol{\eta}_r} \{[2\pi]^{-m/2} [\sigma_r^2]^{-m/2} |\boldsymbol{\Sigma}_d^{\rho(\eta_r)}|^{-1/2} \\ &\quad \times \exp\{-[2\sigma_r^2]^{-1} (\mathbf{r}^d - \mu_r \boldsymbol{i}_m)^T [\boldsymbol{\Sigma}_d^{\rho(\eta_r)}]^{-1} (\mathbf{r}^d - \mu_r \boldsymbol{i}_m)\}. \end{aligned} \tag{11}$$

Note that the three maximum conditional marginal likelihood estimators are

$$\begin{aligned}(\hat{\mu}_r | \sigma_r^2, \boldsymbol{\eta}_r) &= [\boldsymbol{i}_m^T [\boldsymbol{\Sigma}_d^{\rho|\eta_r}]^{-1} \boldsymbol{i}_m]^{-1} \times \boldsymbol{i}_m^T [\boldsymbol{\Sigma}_d^{\rho|\eta_r}]^{-1} \mathbf{r}^d \\ (\hat{\sigma}_r^2 | \mu_r, \boldsymbol{\eta}_r) &= m^{-1} (\mathbf{r}^d - \mu_r \boldsymbol{i}_m)^T [\boldsymbol{\Sigma}_d^{\rho|\eta_r}]^{-1} (\mathbf{r}^d - \mu_r \boldsymbol{i}_m) \\ (\hat{\boldsymbol{\eta}}_r | \mu_r, \sigma_r^2) &= \arg \min_{\boldsymbol{\eta}_r} \{\log |\boldsymbol{\Sigma}_d^{\rho(\eta_r)}| + (\mathbf{r}^d - \mu_r \boldsymbol{i}_m)^T [\sigma_r^2 \boldsymbol{\Sigma}_d^{\rho(\eta_r)}]^{-1} (\mathbf{r}^d - \mu_r \boldsymbol{i}_m)\}. \end{aligned} \tag{12}$$

The first two optimizations are sequentially analytically tractable; hence, $[\hat{\mu}_r, \hat{\sigma}_r | \boldsymbol{\eta}_r]$ can be assessed analytically. The third optimization must be made numerically, but since the parameter κ-vector $\boldsymbol{\eta}_r$ is usually low-dimensional, this optimization is normally feasible. Hence, the maximum marginal likelihood estimator $[\hat{\mu}_r, \hat{\sigma}_r, \hat{\boldsymbol{\eta}}_r]$ can be assessed iteratively.

Acknowledgements This research is supported by the Department of Mathematical Sciences, NTNU, Trondheim, Norway. Mina Spremic has a PhD-grant from the Research Initiative GAMES—Geophysics and Applied Mathematics in Exploration and Safe production.

Funding Open access funding provided by NTNU Norwegian University of Science and Technology (incl St. Olavs Hospital - Trondheim University Hospital). Funding was supported by Norwegian Research Council (Grant No. 294404).

Data Availability The dataset is available as a GitHub repository at Nychka (2023).

Declarations

Conflict of interest No Conflict of interest to declare.

References

Aunon J, Gomez-Hernandez JJ (2000) Dual Kriging with local neighborhoods: application to the representation of surfaces. Math Geol 32(1):69–85

Chiles JP, Delfiner P (2012) Geostatistics–modeling spatial uncertainty. Wiley, Hoboken

Cressie N, Johannesson G (2008) Fixed rank Kriging for very large spatial data sets. J R Stat Soc - Ser B 70:209–226

Gneiting T (2002) Compactly supported correlation functions. J Multivar Anal 83:493–508

Heaton M, Datta A, Finley A, Furrer R, Guinness J, Guhaniyogi R, Gerber F, Gramacy R, Hammerling D, Katzfuss M, Lindgren F, Nychka D, Sun F, Zammit-Mangion A (2018) A case study competition among methods for analysing large spatial data. J Agric Biol Environ Stat 24(3):398–425

Johns C, Nychka D, Kittel T, Daly C (2003) Infilling sparse records of spatial fields. J Am Stat Assoc 98(464):796–806

Matheron G (1971) The theory of regionalized variables and its applications. École National Supérieure des Mines 5:212

Nychka D (2023) NCARData. https://github.com/dnychka/ncardata

Omre H, Fjeldstad TM, Forberg OB (2024) Bayesian spatial modelling with conjugate prior models. Springer Verlag, Berlin

Royer JJ, Vieira PC (1984) Dual formalism of Kriging. In: Verly G, David M, Journel AG, Marechal A (eds) Geostatistics for natural resources characterization, vol 2. Reidel, Dordrecht and Holland, pp 691–702

Rue H, Held L (2005) Gaussian Markov random fields: theory and applications. Monographs on statistics and applied probability, vol 104. Chapman & Hall, London

Stein ML (1999) Interpolation of spatial data: some theory for Kriging (Springer series in statistics). Springer, Berlin

Vigsnes M, Kolbjornsen O, Hauge V, Dahle P, Abrahamsen P (2017) Fast and accurate approximation to Kriging using common data neighborhoods. Math Geosci 49:619–634