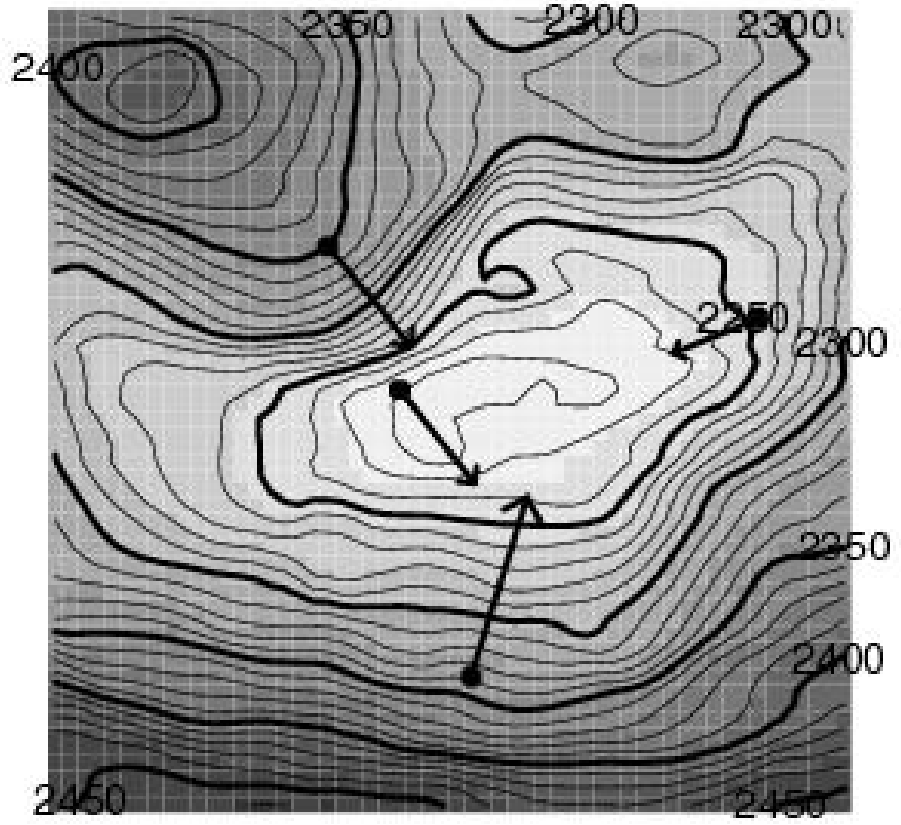
Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

Note

Kriging Derivatives



Note no SAND/11/2019
Author Petter Abrahamsen
Date November 15, 2019

Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

Norwegian Computing Center

Norsk Regnesentral (Norwegian Computing Center, NR) is a private, independent, non-profit foundation established in 1952. NR carries out contract research and development projects in information and communication technology and applied statistical-mathematical modelling. The clients include a broad range of industrial, commercial and public service organisations in the national as well as the international market. Our scientific and technical capabilities are further developed in co-operation with The Research Council of Norway and key customers. The results of our projects may take the form of reports, software, prototypes, and short courses. A proof of the confidence and appreciation our clients have in us is given by the fact that most of our new contracts are signed with previous customers.

Norsk Regnesentral
Norwegian Computing Center
Postboks 114, Blindern
NO-0314 Oslo, Norway

Besøksadresse
Office address
Gaustadalléen 23
NO-0373 Oslo, Norway

Telefon · telephone
(+47) 22 85 25 00
Telefaks · telefax
(+47) 22 69 76 60

Internett · internet
www.nr.no
E-post · e-mail
nr@nr.no

Title	**Kriging Derivatives**
Author	**Petter Abrahamsen**
Date	November 15, 2019
Publication number	SAND/11/2019

Abstract

Some properties of conditioning a Gaussian random field on point value data and derivative data is investigated. Linear predictors with or without a linear trend are considered. Derivative data are seen to improve predictions and the corresponding prediction errors are reduced. The predictors are extensions to the well known simple and universal kriging predictors. The covariance functions between different components of the derivative fields are the key component entering the predictors. They are given by partial derivatives of the covariance function of the Gaussian random field. Spatial symmetries such as stationarity and isotropy are used to restrict the number of covariance functions. In particular isotropy reduces complexity and a general framework for utilising this spatial symmetry is established. Properties of the predictors and the associated prediction error are studied in detail for some particular isotropic covariance functions. The importance of the smoothness of the random field on the predictor and the prediction error is illustrated. This text was first published as part of Abrahamsen (1997).

Keywords	derivatives, gradient data, Gaussian random field, isotropy, kriging, prediction, prediction error, spatial symmetries, stationarity
Target group	
Availability	Open
Number of pages	30

1 Introduction

The objective of this paper is to study the properties of kriging predictors when using derivative information such as slope and curvature.

The best linear unbiased predictors for random fields (random functions) are the so called kriging predictors. In the particular case of a Gaussian random field these are also optimal in the maximum likelihood sense. Kriging predictors were first described by the French mathematician George Matheron (Cressie, 1990) for use in mining applications. By now, kriging has become a standard technique described in several textbooks such as Journel and Huijbregts (1978), Ripley (1981), Ripley (1988), Cressie (1993), and Christakos (1992). The main applications of kriging are still within earth sciences such as mining, petroleum exploration, hydrology, and meteorology, see e.g. Soares (1993) or Baafi and Schofield (1997). Thus, the random fields considered are usually defined on $\mathbb{R}^d$ where $d = 2, 3$ or 4 in spatial-temporal settings. Recently, kriging techniques in a high dimensional parameter space has found its way into the exploration and utilisation of experimental designs (Sacks et al., 1989a,b).

Derivative data have hardly been considered in applications of kriging. An important reason is that genuine derivative data are rare, but also of importance is that few are aware of the possibility of using gradient and curvature data. Thus, its collection and utilisation are seldom promoted. There are a few exceptions found in the literature. Gradient data has been used within the petroleum exploration industry for mapping geological structures. Shiyi (1983) use gradient data for mapping a geological subsurface and Renard and Ruffo (1993) use gradient data obtained from a dip meter to improve the prediction of a seismic reflecting subsurface. Chauvet et al. (1976) use observations of atmospheric pressure and wind velocity—which to a certain approximation are proportional but perpendicular to the pressure gradient—to predict the pressure field. When performing computer experiments derivatives from sensitivity studies are occasionally present. Morris et al. (1993) exploit this by predicting a response surface in $\mathbb{R}^8$ using ordinary kriging. They report that using gradient information reduces empirical errors by a factor 4–10 depending on the experimental design. Mardia and Little (1994) and Mardia and Kent (1996) discuss kriging and splines with derivative and curvature information applied to deformations with landmark, tangent, and curvature constraints for medical imaging. None of these authors study the influence of derivative data on the prediction error.

Derivatives of any order are considered below, but special emphasis is given to first order derivatives (gradients). The conclusion is that derivative data carry valuable information and reduce prediction errors dramatically. The price of including the derivative data is that a larger linear equation system must be solved. The size of this system is proportional to the number of observations and should not be prohibitively large in applications where derivative data prove important, that is, in under-sampled spaces or in situations where data collection is expensive.

The number of covariance functions between derivatives becomes large. However, exploiting symmetries such as isotropy simplifies calculations of the necessary covariance functions considerably.

When considering gradient data the existence of the gradient field is tacitly assumed. Continuity and differentiability of a Gaussian random field is determined by the behaviour of the covariance functions for small separations (assuming a regular expectation). Two types of continuity (and differentiability) are usually considered; sample function continuity and mean square continuity (Adler, 1981). The first is a property related to the shape of the covariance functions whereas mean square continuity is ensured if the covariance function is continuous at the diagonal. For separable (Doob, 1953) Gaussian random fields the distinction becomes almost absent. Adler (1981, pp. 59ff) gives conditions ensuring sample path continuity for Gaussian random fields. These conditions are almost impossible to violate without adding a white noise component—which is mean square discontinuous. This carry over to differentiability; the crucial condition becomes mean square differentiability ensured by covariance functions having continuous deriva-

tives in each coordinate. For isotropic covariance functions this amounts to requiring that the covariance function is two times differentiable at the origin. This imposes restrictions on the choice of covariance functions. Some of the most widely used covariance functions—the spherical and the exponential—do not comply to these requirements. On the contrary, the widely used Gaussian covariance function possesses an infinite number of derivatives. This implies that sample functions are analytical with probability one. This could be a reasonable assumption when interpolating the response surface between points obtained from computer experiments. For natural phenomena however, something between non-differentiable and infinite differentiable sample functions should be more appropriate. Some alternatives will be given.

1.1 Organisation

The next section establishes the method of kriging for predicting Gaussian random fields and is proceeded by a short discussion of spatial symmetries and tensors. In Section 4 gradient fields and the properties of the associated covariance functions and tensors are treated. Especially the implications of isotropy is treated in detail. Some examples illustrating the use of gradient data in $\mathbb{R}^1$ and $\mathbb{R}^2$ are given. These include behaviour near an observation and the use of seismic and bore-hole data from a North Sea petroleum reservoir to predict the depth to a geologic subsurface. In Section 8 higher order derivatives are considered and an example in $\mathbb{R}^1$ including the 9th order derivative is given. This is followed by some discussion in Section 9.

Spatial symmetries such as stationarity and in particular isotropy simplify the various covariance functions significantly. The theoretical results are developed in Section A. Finally some alternative covariance functions for differentiable Gaussian random fields are given in Section B.

2 Kriging

2.1 Simple kriging

Consider a (real valued) Gaussian random field $X(t)$ where $t = [t_1, \ldots, t_d] \in \mathbb{R}^d$, and assume that

$$E\{X(t)\} = 0 \quad \text{and} \quad \operatorname{cov}\{X(t), X(s)\} = C(t, s),$$

where $C(t, s)$ is a known covariance function. Furthermore, assume that $X(t)$ has been observed at n distinct locations: $t^1, \ldots, t^n$. The objective of kriging is to predict the value of $X(t)$ at an arbitrary location t, given the observations $X^T = [X(t^1), \ldots, X(t^n)]$. The optimal predictor, in the minimum variance sense and the maximum likelihood sense, is the conditional expectation of $X(t)$ given the observations X:

$$X^*(t) = E\{X(t)|X\}.$$

The squared prediction error is the conditional variance:

$$\sigma^2_{X^*}(t) = E\{X^*(t) - X(t)\}^2 = \operatorname{var}\{X(t)|X\}.$$

Denote by K the covariance matrix of the observations and let the vector $k(t)$ be the covariances between $X(t)$ and the observation vector:

$$\begin{aligned} K_{ij} &= \operatorname{cov}\{X(t^i), X(t^j)\} = C(t^i, t^j); & i, j &= 1, \ldots, n \\ k_i(t) &= \operatorname{cov}\{X(t), X(t^i)\} = C(t, t^i); & i &= 1, \ldots, n. \end{aligned}$$

Using well known formulae for conditional expectation and variance of a multivariate Gaussian distribution give

$$X^*(t) = k^T(t)\, K^{-1} X \tag{1}$$

$$\sigma^2_{X^*}(t) = C(t, t) - k^T(t)\, K^{-1} k(t). \tag{2}$$

This predictor is commonly called the simple kriging predictor.

2.2 Cokriging

Within this framework, it is straightforward to amend the observation vector X with observations from some additional Gaussian random field, say $Y(t)$, provided the covariance function $\mathrm{cov}\{Y(t), Y(s)\}$ and the cross-covariance function $\mathrm{cov}\{X(t), Y(s)\}$ are known. The covariance matrix, K, and covariance vector, k, must be expanded to accommodate the new covariances. It is also clear that this extension can be pursued further to consider observations from a collection of correlated Gaussian random fields. Predicting a random field given additional observations from correlated random fields are usually referred to as cokriging.

In the following we will study a collection of random fields related in a particular manner; they are defined as the partial derivatives of a Gaussian random field X. Differential operators are linear so the derivatives are Gaussian random fields themselves. Moreover, all covariance functions and cross-covariance functions of the derivatives are obtained by differentiating $C(t, s)$. This means that given observations of $X(t)$ itself and observations of the derivatives (any order) of $X(t)$, it is possible to obtain predictions of $X(t)$ and prediction errors using Eq. 1 and Eq. 2 conditioned on all available data.

Replacing derivatives by closely spaced pairs of observations would for all practical purposes give identical predictions so using derivative data is essentially covered by the standard techniques. A practical problem however, is that the kriging matrix will be almost singular so numerical problems could occur.

2.3 Universal kriging

In many applications it is unrealistic to assume a vanishing expectation. It is straight forward to augment $X(t)$ with a linear trend and consider the model

$$Y(t) = f^T(t)\,\beta + X(t). \tag{3}$$

Here $X(t)$ is a Gaussian random field with expectation zero, β are p unknown parameters, and $f(t)$ is a vector of p known functions. These could be a collection of polynomials on $\mathbb{R}^d$ or some other functions reasonable for the particular problem at hand. The special case $p = 1$ and $f_1(t) = 1$ is called ordinary kriging. Assuming $Y(t)$ is observed at n distinct locations, $Y^T = \big[Y(t^1), \ldots, Y(t^n)\big]$, the optimal predictor is the universal kriging predictor (Ripley, 1981, pp. 47–50):

$$Y^*(t) = f^T(t)\,\widehat{\beta} + k^T(t)\,K^{-1}\big(Y - F\widehat{\beta}\big),$$

where $\widehat{\beta} = \big(F^T K^{-1} F\big)^{-1} F^T K^{-1} Y$ is the generalized least squares estimate, and the matrix F is formed by rows $f^T(t^i)$. The universal kriging predictor is seen to be the sum of the linear trend, $f^T(t)\,\widehat{\beta}$, and the simple kriging predictor for the fitted residuals: $\widehat{X} = Y - F\widehat{\beta}$. The corresponding squared prediction error is

$$\sigma^2_{Y^*}(t) = C(t,t) - k^T(t)\,K^{-1}k(t) + \big\|f(t) - k^T(t)K^{-1}F\big\|^2_{\widehat{S}},$$

where the norm is the Mahalanobis distance:

$$\widehat{S} = \mathrm{var}\{\widehat{\beta}\} = \big(F^T K^{-1} F\big)^{-1}.$$

Similar to the simple kriging predictor, it is possible to amend the observation vector Y with observations from correlated random fields such as derivatives of $Y(t)$. Then, the F matrix will contain additional rows formed from derivatives of $f^T(t)$. Thus, results obtained in the subsequent are applicable for both simple and universal kriging.

3 Scalars, vectors, and tensors

In the following space-vector is used rather than vector to stress that it is a vector in $\mathbb{R}^d$ not only a collection of d unrelated numbers. Any space-vector, v, is an entity in itself regardless of the

particular numerical representation in a specific Cartesian coordinate system. A space-vector is invariant to transformations of the coordinate system; the components will change when rotating or reflecting the coordinate system but the space-vector itself is unchanged. Thus, the components of a space-vector, v, obey specific transformations when transforming the coordinate system, i.e., $v' = Av$ where A is an orthogonal transform of the coordinates: $t' = At$. The corresponding transformation of the unit vectors defining the Cartesian coordinate system is $e' = A^{-1}e$. Note that for orthogonal transforms $A^T = A^{-1}$. Tensors are a generalisation of a space-vector.

Definition (Cartesian tensor)**.** A (Cartesian) tensor of rank r is a set of d^r values $V_{i_1\cdots i_r}$ which transform according to

$$V'_{i_1\cdots i_r} = \sum_{j_1\cdots j_r} V_{j_1\cdots j_r}\, a_{i_1 j_1} \cdots a_{i_r j_r}$$

where a_{ij} form an orthogonal transformation matrix A.

Just as for space-vectors, tensors are entities in itself regardless of its coordinate representation. In a particular Cartesian coordinate system a tensor of rank two in $\mathbb{R}^d$ is represented by a $d \times d$-dimensional matrix with components V_{ij}. The components of a tensor of rank two transform as $V' = AVA^T$. Note that space-vectors are tensors of rank one according to the definition .

A scalar is a tensor of rank zero. Scalars are represented by a single numerical value and is invariant to coordinate transformations. A scalar can be obtained from tensors by contracting space-vectors and tensors. An example is the Euclidian norm: $\|v\|^2 = v^T v = v^T A^T A v = v'^T v' = \|v'\|^2$.

In the following tensors defined on $\mathbb{R}^d$ will be considered.

Definition (Tensor field)**.** A tensor field of rank r is a tensor of rank r on $\mathbb{R}^d$: $V_{i_1\cdots i_r} = V_{i_1\cdots i_r}(t)$, where $t \in \mathbb{R}^d$ and t is a tensor of rank one.

Our main interest will be covariance functions of scalar, vector, and tensor fields. A few important results ensure that the covariance functions in the next sections behave like tensor fields.

Theorem 1 (Tensor properties)**.**

i) Tensors form a linear space so that linear combinations of tensors of rank r is a tensor of rank r.

ii) Differentiation of a tensor field of rank r with respect to Cartesian coordinates results in a tensor field of rank $r + 1$.

iii) The outer product of two tensors of rank r and s is a new tensor of rank $r + s$.

Proofs are found in Butkov (1968, Chapter 16).

3.1 Symmetries

A tensor of rank r has d^r components. This is a large number even for modest ranks. By imposing symmetries the number of independent components are reduced to a manageable number.

Definition (Symmetric tensor)**.** A symmetric tensor of rank r is invariant to permutations of the r indices $i_1, \ldots, i_r$.

A symmetric tensor has $\binom{d+r-1}{r}$ independent components (Feller, 1968, p. 38), which can be considerably less than the number of components. Table 2 in Section 8 contains some examples. We will see that covariance tensors of stationary gradient fields are symmetric.

Let $\overline{\tau} \in \mathbb{R}^d$ and let the Euclidian distance of $\overline{\tau}$ be $\tau = (\tau_1^2 + \cdots + \tau_d^2)^{1/2}$. An isotropic scalar function $C\colon \mathbb{R}^d \to \mathbb{R}$ is by definition dependent on $\overline{\tau}$ through the Euclidian norm τ alone, i.e. $C(\overline{\tau}) = C(\tau)$. Thus, isotropic scalar functions are invariant to orthogonal transformations of the coordinates.

Definition (Isotropic tensor field)**.** Isotropic tensor fields are isotropic functions of a tensor of rank one under orthogonal transformations, that is, $V'_{i_1\cdots i_r}(t) = V_{i_1\cdots i_r}(t')$, where t is a tensor of rank one (space-vector).

Isotropy impose even stronger restrictions on tensor fields. The following result is a corollary to Conjecture 1 in the Appendix which states the general form of an isotropic and symmetric tensor field. The conjecture is proved for all covariance tensors encountered in the following paragraphes.

Corollary. *An isotropic and symmetric tensor field of rank r has $\lceil \frac{r+1}{2}$ number of independent components.*

The ceiling function $\lceil r$ is the nearest integer above or equal r, i.e. $\lceil 5 = 5$ and $\lceil 5.5 = 6$. Thus, isotropy dramatically reduce the number of independent components. This is of great importance in practical applications. For instance, a covariance tensor between second order derivatives has rank four and a total of 81 components in $\mathbb{R}^3$. Imposing isotropy means that only three independent components exist.

4 Gradient fields

Consider a Gaussian random field $X(t)$ on $\mathbb{R}^d$ possessing differentiable sample functions. The associated gradient field, $\dot{X}(t)$, is defined by its components in a Cartesian coordinate system:

$$\dot{X}_i(t) = \frac{\partial X(t)}{\partial t_i}; \qquad i = 1, \ldots, d.$$

According to Theorem 1(ii) the gradients form a tensor of rank one which is a space-vector. Assuming $E\{X(t)\} = m(t)$, then

$$\dot{m}_i(t) = E\{\dot{X}_i(t)\} = \frac{\partial m(t)}{\partial t_i}; \qquad i = 1, \ldots, d,$$

which according to Theorem 1(i) is a space-vector since the expectation is a linear operator. Further, assume that the covariance function, $C(t,s) = \mathrm{cov}\{X(t), X(s)\}$, is simultaneously differentiable in t and s, that is, $X(t)$ is mean square differentiable. Then, the cross-covariance function between $X(t)$ and a component of $\dot{X}(t)$ is defined by the components

$$\dot{C}_i(t,s) = \mathrm{cov}\{X(t), \dot{X}_i(s)\} = \frac{\partial}{\partial s_i} C(t,s); \qquad i = 1, \ldots, d.$$

This is also a space-vector according Theorem 1(ii). The covariance functions between components of $\dot{X}(t)$ are

$$\ddot{C}_{ij}(t,s) = \mathrm{cov}\{\dot{X}_i(t), \dot{X}_j(s)\} = \frac{\partial^2}{\partial t_i \partial s_j} C(t,s); \quad i,j = 1, \ldots, d.$$

The matrix $\ddot{C}_{ij}(t,s)$ is a rank two tensor called the covariance tensor of the gradient field. This follows from Theorem 1(ii) but also from Theorem 1(iii) by noting that the covariance is the expectation of the outer product of the centred gradient fields.

4.1 Stationary random fields

Stationary scalar random fields have by definition a translation invariant distribution. This implies a constant expectation so that $\dot{m}_i(t) = 0$, and a stationary covariance function $C(t,s) = C(\overline{\tau})$, where $\overline{\tau} = t - s$. Since derivatives commute, the covariance tensor must be symmetric for stationary random fields: $\ddot{C}_{ij}(\overline{\tau}) = \ddot{C}_{ji}(\overline{\tau})$, thereby reducing the number of independent components to $\binom{d+1}{2} = d(d+1)/2$.

A subclass of stationary covariance functions are the separable ones defined such that $C(\overline{\tau}) = \prod_i C_i(\tau_i)$ (Vanmarcke, 1983). This class could be reasonable when considering computer experiments (Morris et al., 1993) where the coordinates, $t_1, \ldots, t_d$, have no canonical relationship.

4.2 Isotropic random vector fields

The expected value of any isotropic random space-vector field must be zero since the zero vector is the only rotation invariant space-vector. The covariances between the components of the gradient field form an isotropic tensor field. Conjecture 1 in the Appendix gives the general form of isotropic tensor fields. A convenient parameterization is (Yaglom, 1986, pp. 372ff); (Monin and Yaglom, 1975, pp. 29ff):

$$C_{ij}(\overline{\tau}) = \Big\{C_R(\tau) - C_T(\tau)\Big\}\frac{\tau_i\tau_j}{\tau^2} + C_T(\tau)\,\delta_{ij}, \tag{4}$$

where δ_{ij} is the Kronecker symbol. This means that the covariance tensor can be uniquely specified by two isotropic covariance functions $C_R(\tau)$ and $C_T(\tau)$. We call these the radial covariance function and the transverse covariance function respectively. The radial and transverse covariance functions have geometrical interpretations. The radial covariance function gives the covariance between components of $\dot{X}(t)$ and $\dot{X}(s)$ parallel to the direction $\overline{\tau} = t - s$, e.g.

$$C_R(\tau_i) = \operatorname{cov}\big\{\dot{X}_i(t), \dot{X}_i(s)\big\}; \qquad (\tau = \tau_i).$$

The transverse covariance function gives the covariance between components of $\dot{X}(t)$ and $\dot{X}(s)$ perpendicular to the direction $\overline{\tau} = t - s$, e. g.

$$C_T(\tau_j) = \operatorname{cov}\big\{\dot{X}_i(t), \dot{X}_i(s)\big\}; \qquad (\tau = \tau_j \text{ and } i \neq j).$$

These equations show that $C_T(0) = C_R(0)$ leading to $C_{ij}(0) = C_T(0)\,\delta_{ij}$.

4.3 Isotropic gradient fields

For a gradient field the curl vanishes and $\partial C_{ij}(\overline{\tau})/\partial\tau_l = \partial C_{lj}(\overline{\tau})/\partial\tau_i$. Then, Eq. 4 gives

$$C_R(\tau) = C_T(\tau) + \tau^2\left(\frac{1}{\tau}\frac{\mathrm{d}}{\mathrm{d}\tau}\right)C_T(\tau). \tag{5}$$

An immediate consequence of this is that $\int_0^\infty C_R(\tau)\,\mathrm{d}\tau = 0$ so that $C_R(\tau)$ must change sign. It is also possible to show that $C_T(\tau)$ must be a valid isotropic covariance function in $\mathbb{R}^{d+2}$ (Yaglom, 1986, p. 382).

It is convenient to express the partial derivatives by using derivatives of τ:

$$\dot{C}_i(\overline{\tau}) = -\tau_i\left(\frac{1}{\tau}\frac{\mathrm{d}}{\mathrm{d}\tau}\right)C(\tau), \tag{6}$$

$$\ddot{C}_{ij}(\overline{\tau}) = -\delta_{ij}\left(\frac{1}{\tau}\frac{\mathrm{d}}{\mathrm{d}\tau}\right)C(\tau) + \tau_i\tau_j\left(\frac{1}{\tau}\frac{\mathrm{d}}{\mathrm{d}\tau}\right)^2 C(\tau). \tag{7}$$

Comparing Eq. 7 to the general form Eq. 4 shows that

$$C_R(\tau) = -\left(\frac{1}{\tau}\frac{\mathrm{d}}{\mathrm{d}\tau}\right)C(\tau) - \tau^2\left(\frac{1}{\tau}\frac{\mathrm{d}}{\mathrm{d}\tau}\right)^2 C(\tau) = -\frac{\mathrm{d}^2}{\mathrm{d}\tau^2}C(\tau) \tag{8}$$

$$C_T(\tau) = -\left(\frac{1}{\tau}\frac{\mathrm{d}}{\mathrm{d}\tau}\right)C(\tau). \tag{9}$$

Since $C_R(0)$ is finite, Eq. 8 show that the second order derivative of $C(\tau)$ must be finite at $\tau = 0$. Moreover, since $C_T(0)$ is finite, $\frac{\mathrm{d}}{\mathrm{d}\tau}C(\tau) \sim \tau$ for small τ. In particular, this means that $\frac{\mathrm{d}}{\mathrm{d}\tau}C(\tau)$ vanish at $\tau = 0$.

Comparing Eq. 9 to the cross-covariance function Eq. 6 shows that

$$\dot{C}_i(\overline{\tau}) = \tau_i\,C_T(\tau).$$

Furthermore, combining Eq. 4 with Eq. 5 gives

$$\ddot{C}_{ij}(\overline{\tau}) = \tau_i\tau_j\left(\frac{1}{\tau}\frac{\mathrm{d}}{\mathrm{d}\tau}\right)C_T(\tau) + \delta_{ij}\,C_T(\tau).$$

The following statements are valid:

i) The variance of the gradient components are proportional to the curvature of $C(\tau)$ at 0, i.e., $\text{var}\{\dot{X}_i(t)\} = C_T(0) = C_R(0)$.

ii) The isotropic random field, $X(t)$, and its gradient field, $\dot{X}(t)$, are uncorrelated at $\tau = 0$. This is even true for stationary random fields (Vanmarcke, 1983, p. 112).

iii) The isotropic random field, $X(t)$, and a component, say $\dot{X}_i(t)$, of its gradient field are uncorrelated along the line $\tau_i = 0$.

iv) Two components of the gradient field, say $\dot{X}_i(t)$ and $\dot{X}_j(t)$, are uncorrelated along the lines $\tau_i = 0$ and $\tau_j = 0$.

v) The components of the gradient field are stationary but not isotropic.

Some examples of differentiable isotropic covariance functions are given in Section B.

5 Prediction—local properties

In this section prediction conditioned on observations from a single location are considered. The objective is to establish properties of the predictor near an observation.

5.1 Predictor for a single observation point

Consider an isotropic Gaussian random field $X(t)$ on $\mathbb{R}^d$ with expectation zero and a known covariance function $C(\tau)$. Assume that at a given location s the point value, $X(s)$, and the components of the gradient vector, $\dot{X}_i(s)$, are known. A prediction of $X(t)$ given the data vector $X^T = [X(s), \dot{X}_1(s), \ldots, \dot{X}_d(s)]$ is the expectation of the conditional distribution: $X^*(t) = E\{X(t)|X\}$. Thus the predictor and prediction error are given by Eq. 1 and Eq. 2 respectively. The necessary covariance vectors and covariance matrix are

$$k(\overline{\tau}) = \begin{bmatrix} C(\tau) \\ \overline{\tau} C_T(\tau) \end{bmatrix} \quad \text{and} \quad K = \begin{bmatrix} C(0) & 0_d^T \\ 0_d & I_d C_T(0) \end{bmatrix}, \tag{10}$$

where 0_d is a d-dimensional zero vector and I_d is an d-dimensional identity matrix. All off-diagonal elements of the covariance matrix vanish since cross-covariances are zero at $\tau = 0$.

5.2 Some properties of the predictor

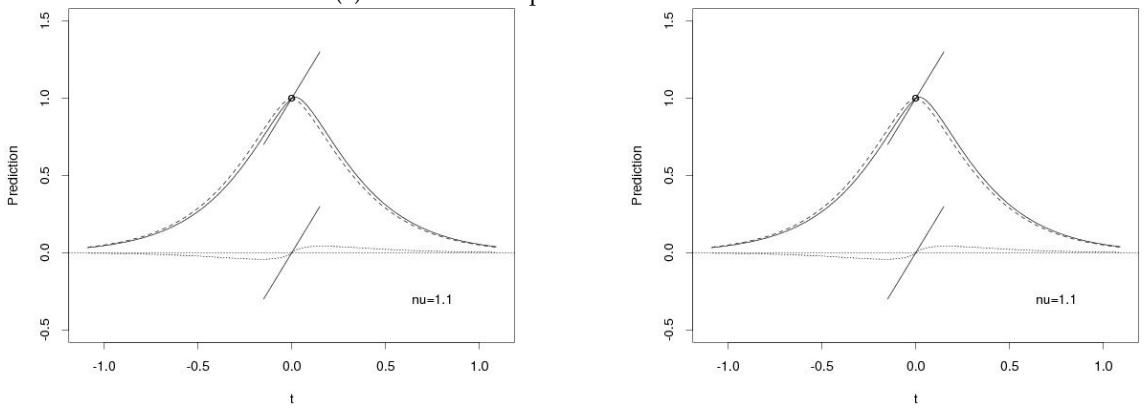
Inserting Eq. 10 in the predictor Eq. 1 gives

$$X^*(\overline{\tau}) = \frac{C(\tau)}{C(0)} X(s) - \frac{C_T(\tau)}{C_T(0)} \overline{\tau}^T \dot{X}(s).$$

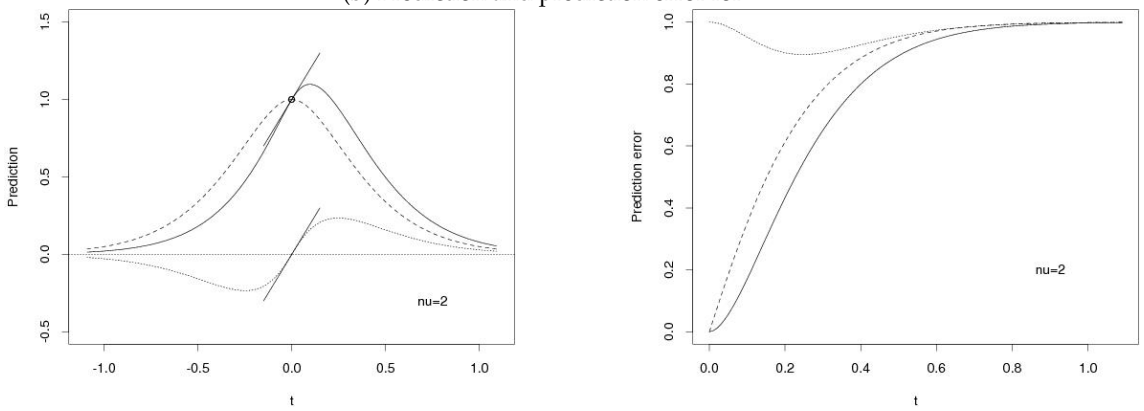
Note that the predictor splits into two independent parts since $\text{cov}\{X(0), \dot{X}_i(0)\} = 0$. The first part is from conditioning on the point value and the second part is from the conditioning on the gradient. When introducing several data locations, this partition survives within each point value and gradient set, but point values and gradient components from separate locations correlate. Fig. 1 illustrates single data predictions for different choices of covariance functions. Each figure contains three curves corresponding to conditioning on point value and gradient, point value alone, or gradient alone. A few comments are appropriate:

- The predictor interpolate the point value and the gradient of the predictor equals the gradient value.
- The predictor is independent of the variance.
- The contribution to the predictor from the gradient data is proportional to the cosine of the angle, θ, between the separation vector, $\overline{\tau}$, and the gradient vector $\dot{X}(s)$, i.e. $\overline{\tau}^T \dot{X}(s) = \cos\theta \, \tau \|\dot{X}(s)\|$.

(a) Prediction and prediction error for $\nu = 1.1$



(b) Prediction and prediction error for $\nu = 2$



(c) Prediction and prediction error for $\nu = \infty$ (Gaussian).

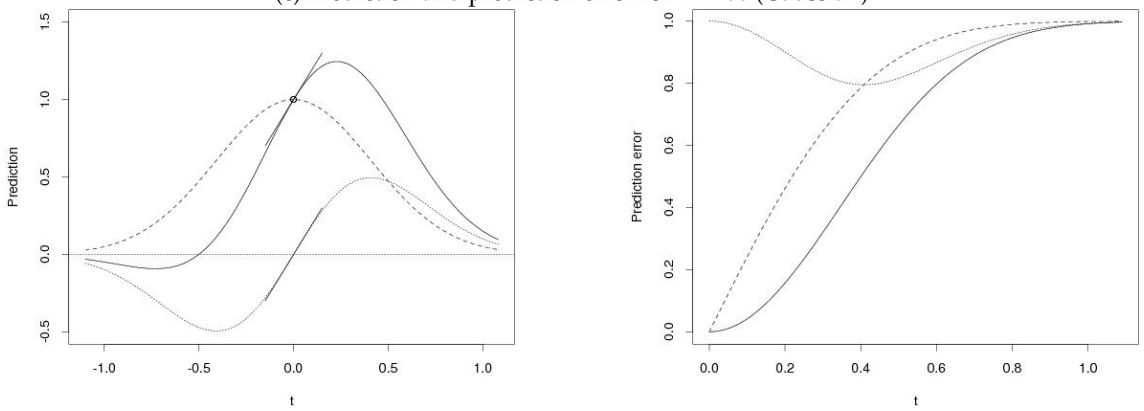


Figure 1. The left column of figures show predictions conditioned on point value and derivative (solid lines), point value alone (dashed line), and derivative alone (dotted line). Modified Bessel covariance functions with different smoothness parameter ν have been used. ($\nu = \infty$ corresponds to the Gaussian covariance function.) They have unit variance and they are scaled such that $C(1) = 0.05$ for simple comparison. Point value data are shown as a small circle whereas the derivative are drawn as a straight line. The corresponding prediction errors are illustrated in the right column.

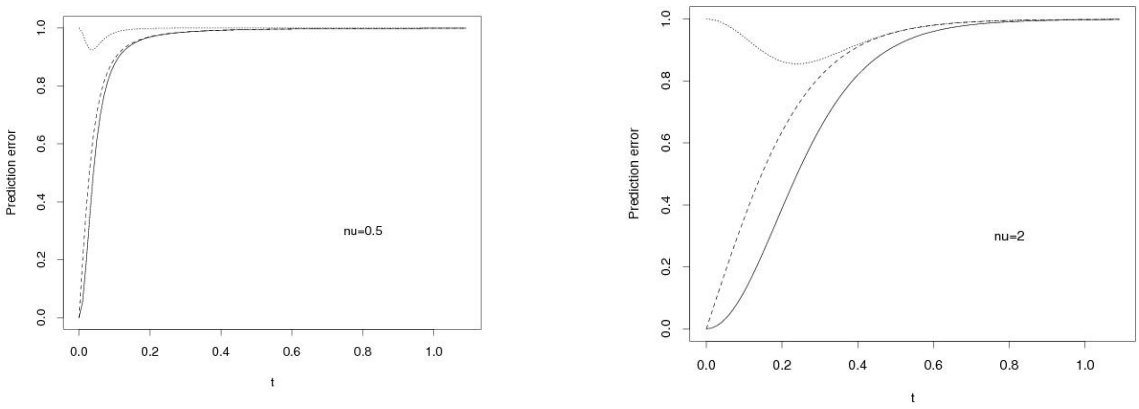


Figure 2. Prediction errors for rational quadratic covariance functions with $\nu = 0.25$ and 2. See Figure 1 for further explanation.

- The predictor approach the expectation (zero here) as τ increase.

- The dashed lines (conditioning on point value alone) shows the shape of the covariance function $C(\tau)$.

- The gradient data has minor influence on the predictor for irregular random fields; the gradient fields approach white noise fields as $\nu \to 1$.

5.3 Some properties of the prediction error

The squared prediction error for a single data becomes:

$$\sigma^2_{X^*}(\tau) = C(0) - \frac{C^2(\tau)}{C(0)} - \tau^2 \frac{C_T^2(\tau)}{C_T(0)}. \tag{11}$$

Thus, the squared prediction error is spherically symmetric. The squared prediction error also splits into two parts. Fig. 1 also illustrates the prediction error for different choices of covariance functions. Each plot contains three curves corresponding to conditioning on point value and gradient, point value alone, and gradient alone. Some remarks can be made:

- Using gradient information alone does not reduce the prediction error at $\tau = 0$.

- Gradient information looses its value as the random field becomes more irregular.

- Gradient data has good long-range prediction capabilities. For smooth random fields, gradient data even reduce the prediction error more than a point value data alone at large distances.

5.4 Short range behaviour of prediction error

The limiting form for small arguments of any continuous isotropic covariance function must be of the form

$$C(\tau) = C(0) - a\,\tau^\phi + \mathcal{O}\big(\tau^{2\phi}\big); \qquad 0 < \phi \leq 2.$$

Equation Eq. 9 implies that $\phi > 1$ for C_T to exist. Further, Eq. 8 determines that $\phi = 2$; $\phi > 2$ would give $C_R(0) = 0$ and $\phi < 2$ would give $C_R(0) = \infty$. Thus, covariance functions for random fields possessing gradient fields must have the limiting form

$$C(\tau) = C(0) - \frac{C_T(0)}{2}\,\tau^2 + \mathcal{O}\big(\tau^4\big).$$

Inserting this into Eq. 11 gives the limiting forms:

$$\begin{aligned}
&\sigma^2_{X^*}(\tau) = \mathcal{O}\big(\tau^4\big) \\
&\sigma^2_{X^*}(\tau) = C_T(0)\,\tau^2 + \mathcal{O}\big(\tau^4\big) && \text{(unknown gradient)} \\
&\sigma^2_{X^*}(\tau) = C(0) - C_T(0)\,\tau^2 + \mathcal{O}\big(\tau^4\big) && \text{(unknown point value).}
\end{aligned}$$

For non-differentiable covariance functions the limiting form is:

$$\sigma^2_{X^*}(\tau) = a\,\tau^\phi + \mathcal{O}\big(\tau^{2\phi}\big); \qquad 0 < \phi < 2.$$

The prediction error (when point data is used) has the following leading terms:

$$\begin{aligned}
&\sigma_{X^*}(\tau) = \mathcal{O}\big(\tau^2\big) \\
&\sigma_{X^*}(\tau) = C_T^{1/2}(0)\,\tau + \mathcal{O}\big(\tau^2\big) && \text{(unknown gradient)} \\
&\sigma_{X^*}(\tau) = a^{1/2}\,\tau^{\phi/2} + \mathcal{O}\big(\tau^\phi\big); && 0 < \phi < 2.
\end{aligned}$$

This shows that the slope of an error bound at a data position is zero when including gradient data, finite when using two times differentiable covariance functions, and infinite when using other covariance functions. We conclude that using gradient data reduces the prediction error dramatically near the observation.

6 A one dimensional example

In real applications a single data location is hardly interesting. So, consider n sets of point value and gradient data organised as a vector:

$$X^T = \big[X(t^1), \dot{X}_1(t^1), \ldots, \dot{X}_d(t^1), \ldots, X(t^n), \dot{X}_1(t^n), \ldots, \dot{X}_d(t^n)\big].$$

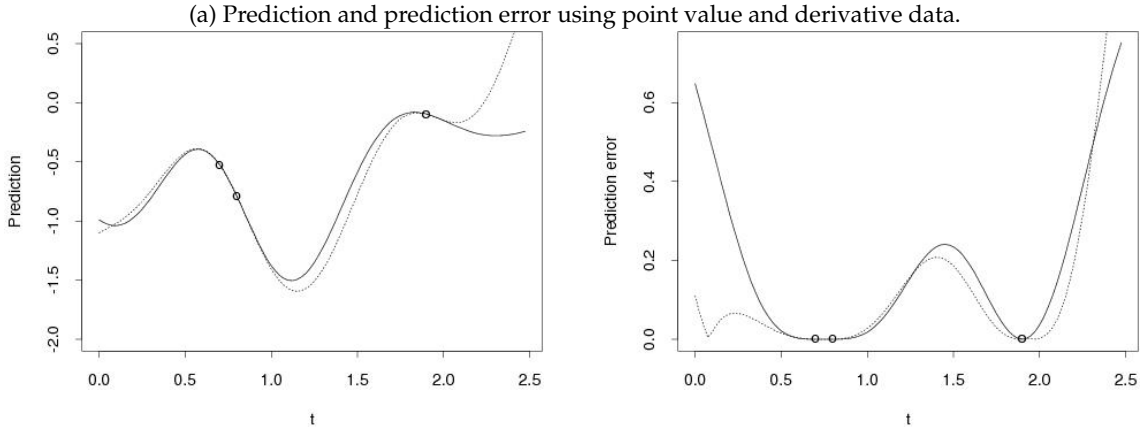
The conditional first and second order moments of $X(t)$ given the data, X, are once more given by the simple kriging predictor Eq. 1 and the associated squared prediction error Eq. 2.

Consider a one dimensional example. The objective is to predict a smooth function in the interval [0, 2.5], given $n = 3$ observations of point values and derivatives sampled at $t = 0.7$, 0.8 and 1.9. The function chosen is a simulated realisation obtained by using the Gaussian covariance function, Eq. B.3, with $a = 3$ giving $C(1) \approx 0.05$. This covariance function is used for prediction as well. Figure 3(a) shows the predictor and the associated prediction error. For comparison Figure 3(b) shows the corresponding results when ignoring the derivatives. The superiority of predictions using gradients are striking. Also worth noticing is the effect of the two close observations at $t = 0.7$ and 0.8. The difference between a narrow pair of observations and a single observation is clearly seen on the prediction error. When gradients are used the influence of this pair resembles a single observation of point value, derivative, and second order derivative.

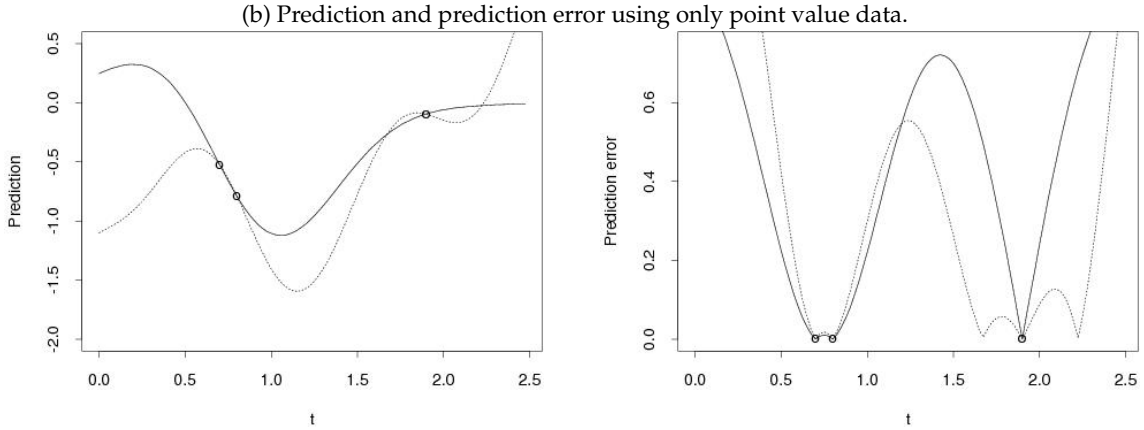
6.1 Observation errors in gradient measurements

It is reasonable to suspect that the precision of gradient data could be argued. Independent observation errors are included by adding the variance of the errors to the diagonal the covariance matrix, K. Figure 3(c) shows predictions and the associated prediction errors when assuming independent measurement errors in the gradient observations. The error (standard deviation) is chosen equal to 10 % of $\mathrm{var}\{\dot{X}\}^{1/2}$. The point value observations are assumed to be exact. As expected, the errors are something between Figure 3(a) (exact gradient data) and Figure 3(b) (no gradient data).

(a) Prediction and prediction error using point value and derivative data.



(b) Prediction and prediction error using only point value data.



(c) Prediction and prediction error using point value and noisy derivative data.

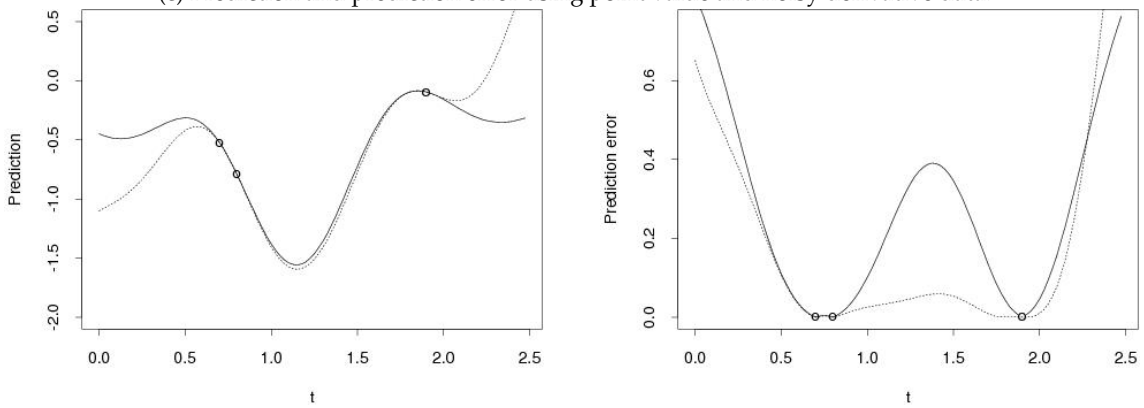


Figure 3. The left figures show predictions as solid lines and the dashed lines are the true function. Observation points are marked as small circles. The right column of figures show the corresponding prediction errors as solid lines whereas the dashed lines are the difference between the prediction and the true function (absolute value). In (a) both point values and derivatives are used, in (b) only the point values are used, and in (c) noise has been added to the derivative data.

7 Seismic depth conversion — a two dimensional case study

A contoured map of the seismic travel times to the top subsurface of a dome shaped North Sea petroleum reservoir is shown in Figure 4(b). The map is 4.2 by 4.2 kilometers and stored in 75 by 75 meter grid cells. The objective is to obtain a map of the depth to the subsurface given the travel times, a process commonly called 'depth conversion' (Abrahamsen, 1993; Hwang and McCorkindale, 1994; Jeffery et al., 1996). The relation between depth, z, and travel time, T is given by the simple kinematic relation $z(t) = v(t)\,T(t)$, where v is the average velocity. Velocities generally increase with depth (travel time) and a simple model including this effect is $v(t) = \beta_0 + \beta_1\{T(t) - 1.17\}$, where β_i are unknown coefficients. The second regressor is $T(t) - 1.17$ where 1.17 is subtracted to reduce collinearity. The residual between the depth model and the actual subsurface is assumed to be a Gaussian random field with vanishing expectation. A rational quadratic covariance model, Eq. B.4, has been chosen with parameters $\nu = 2$, $\sigma = 23\,\mathrm{m}$, and range parameter a such that $C(2000\,\mathrm{m})/C(0) = 0.05$. The particular choice of covariance function is based on experience from similar North Sea data and should be realistic.

Figure 4(a) shows a simulated realisation of the depth to the subsurface. It was obtained by using trend parameters $\beta_0 = 2000\,\mathrm{m/s}$ and $\beta_1 = 1000\,\mathrm{m/s^2}$ and simulating the residual according to the specified covariance function. Four chosen well locations within the area are marked by dots. Arrows indicate direction and magnitude of dip (gradients) at these locations. See Renard and Ruffo (1993) for details on how dip-meter measurements can be used to obtain gradient data. Comparing Figure 4(a) to the travel times in Figure 4(b) shows a rougher surface and a more pronounced dome shape due to the positive value of β_1.

Since the depth model is linear in the β's universal kriging discussed in § 2.3 apply. Figure 4(c) shows a map of depth predictions using depth and dip data while Figure 4(c) shows depth predictions using only depth data. There are significant differences between the two depth predictions. Figure 5 shows the residuals, that is, the the difference between the predictions and the 'true depth' in Figure 4(a). A possible quantitative measure of the difference is to consider the square root of the sum of squares of the two residual grids. Theses values are added above the two maps in Figure 5 and shows that in this case including dip data improves prediction significantly. Note however that this is based on one simulated realisation and a particular choice of well locations.

To check this further predictions were made using data from 100 simulated realizations of the 'true depth'. The square root of the sum of squares of the resulting residual grids and are presented as a scatter plot in Figure 6. It is seen that dip data improves depth predictions for almost all of the 100 realizations.

The (theoretical) prediction errors for the universal kriging predictor is also shown in Figure 4(c,d). It is seen that the regions with prediction errors less than 20m are significantly larger when including dip data.

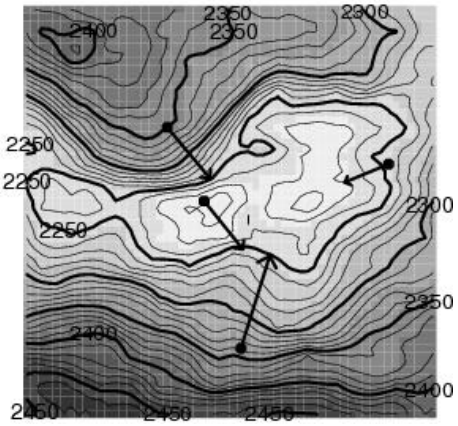
7.1 Estimation of trend parameters

The prediction of the depth to the subsurface is improved and moreover the estimation of the trend parameters, β, are also improved. Table 1 gives a comparison of the estimates using only depth data and using additional dip data.

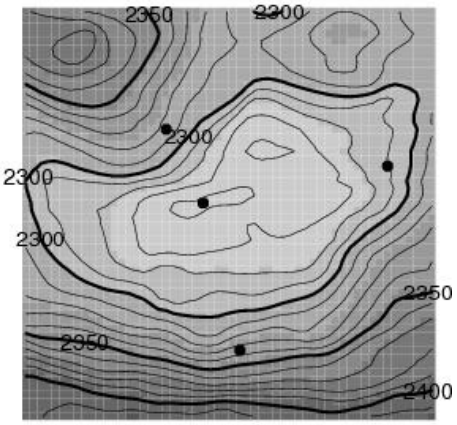
8 Higher order derivatives and tensor fields

Although higher order derivatives are rarely accessible, with the possible exception of second order derivatives, the theory easily extends despite some unpleasant bookkeeping. Consider the

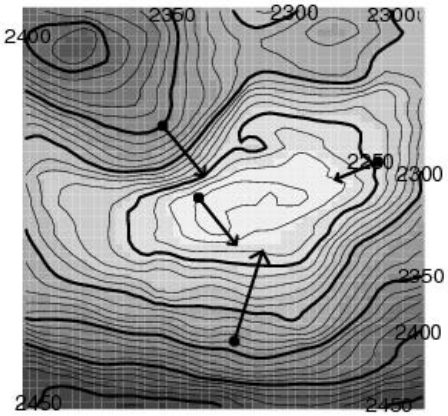
(a) Simulated 'True depth':

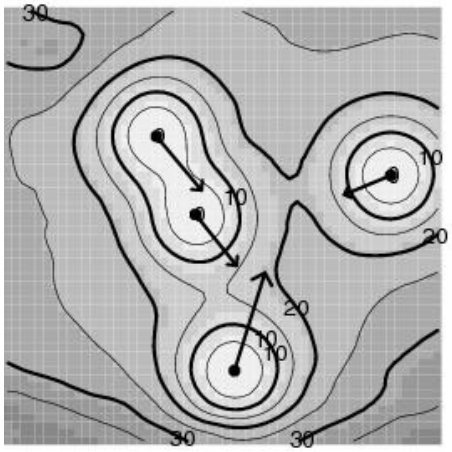


(b) Travel times (two-way msec.):

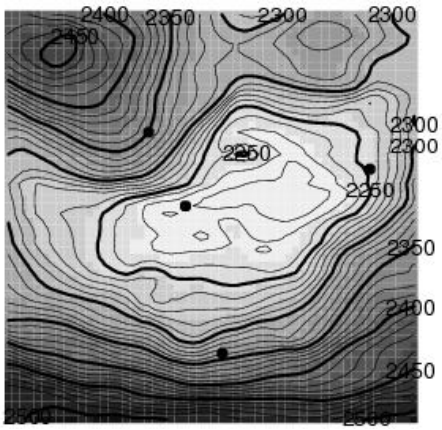


(c) Prediction and prediction error conditioned on depth and dip data:





(d) Prediction and prediction error conditioned on depth data:



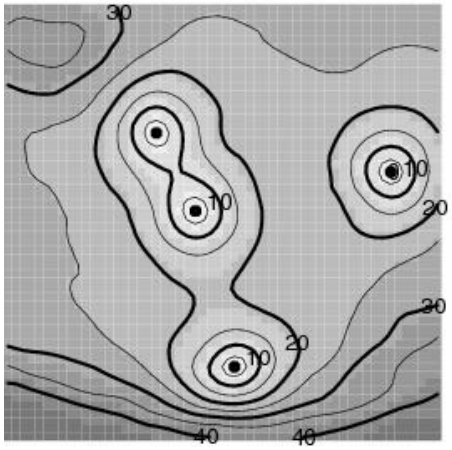
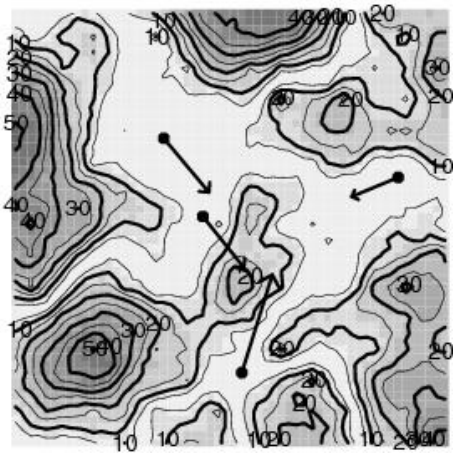


Figure 4. Contour maps of 'true depth' (a), travel time data (b), and depth predictions (c,d) with corresponding theoretical prediction error. The map area is 4.2 by 4.2 kilometers.

(a) Average error: 17.5 m



(b) Average error: 28.2 m

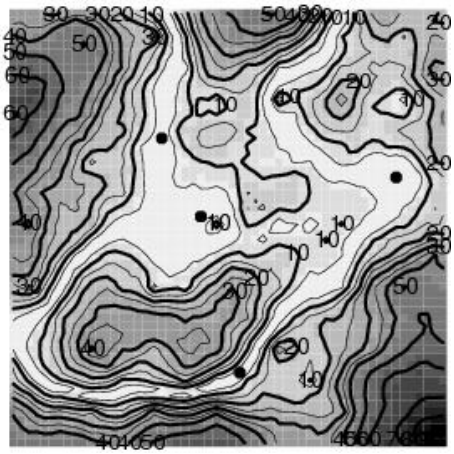


Figure 5. Contoured map of difference (absolute value) between predicted depth and 'true depth' for depth predictions conditioned on depth and dip data (a), and depth data (b). The map area is 4.2 by 4.2 kilometers.

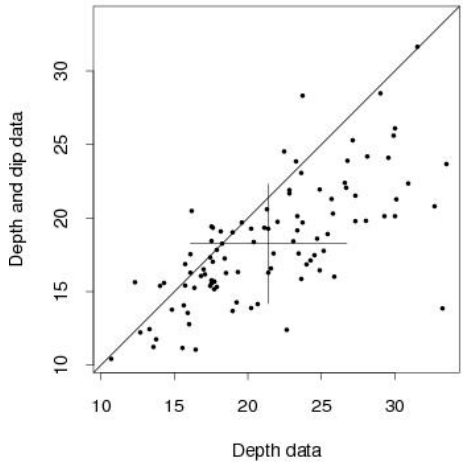


Figure 6. Scatter plot of the square root of the sum of squares of the residual grids for 100 different depth maps. The big cross marks the mean plus minus the empirical standard errors of the scatter data.

general derivative of $X(t)$ (assuming sufficient regularity conditions are satisfied)

$$X^{(\kappa)}(t) = \frac{\partial^{|\kappa|}}{\partial t_1^{\kappa_1} \dots \partial t_d^{\kappa_d}} X(t), \tag{12}$$

where $\kappa = (\kappa_1, \dots, \kappa_d)$ are d non-negative integers and $|\kappa| = \sum_i \kappa_i$. A collection of derivative fields $X^{(\kappa)}(t)$ where $|\kappa| = K$, form a tensor field of rank K. The gradient field is an example of a first rank tensor field.

8.1 General covariance tensors

The general cross-covariances are

$$\begin{aligned} C^{(\kappa,\lambda)}(t,s) &= \operatorname{cov}\{X^{(\kappa)}(t), X^{(\lambda)}(s)\} \\ &= \frac{\partial^{|\kappa|+|\lambda|}}{\partial t_1^{\kappa_1} \cdots \partial t_d^{\kappa_d}\, \partial s_1^{\lambda_1} \cdots \partial s_d^{\lambda_d}} C(t,s). \end{aligned} \tag{13}$$

Table 1. Estimated β values (mean) and corresponding estimation error (standard deviation) and correlation.

		Depth and dip data			Depth data		
	True	Mean	(SD)	Corr.	Mean	(SD)	Corr.
β_0 (m/s)	2000	2005	(13)	0.55	2021	(16)	0.68
β_1 (m/s^2)	1000	830	(450)		1556	(592)	

Table 2. Number of components and number of independent components for covariance tensors.

		Rank ($r = K + L$)							
		0	1	2	3	4	5	6	7
	Dimension	Number of components (d^r)							
	$d = 2$	1	2	4	8	16	32	64	128
	$d = 3$	1	3	9	27	81	243	729	2187
	$d = 4$	1	4	16	64	256	1024	4096	16384
		Number of independent components							
General:	$d = 2$	1	2	4	6	9	12	16	20
($\lvert K - L\rvert \leq 1$)	$d = 3$	1	3	9	18	36	60	100	150
	$d = 4$	1	4	16	40	100	200	400	700
Stationary:	$d = 2$	1	2	3	4	5	6	7	8
(symmetric)	$d = 3$	1	3	6	10	15	21	28	36
	$d = 4$	1	4	10	20	35	56	84	120
Isotropic:	any d	1	1	2	2	3	3	4	4

Now consider a collection of derivatives, $C^{(\kappa,\lambda)}$, defined such that $|\kappa| = K$ and $|\lambda| = L$. This collection form a $K + L$ rank tensor given by the components

$$C^{(K,L)}_{i\ldots k\, j\ldots l}(t,s) = \frac{\partial^{K+L}}{\partial t_i \cdots \partial t_k\, \partial s_j \cdots \partial s_l} C(t,s); \tag{14}$$

$$i, \ldots, k,\ j, \ldots, l \in \{1, \ldots, d\}.$$

Two examples were studied in detail in the previous section; the second rank covariance tensor, $\ddot{C}_{ij}(t,s) = C^{(1,1)}_{ij}(t,s)$, and the first rank (space-vector) cross-covariance function, $\dot{C}_i(t,s) = C^{(1,0)}_i(t,s)$.

The number of components in a general $K + L$ rank covariance tensor is d^{K+L}. The number of independent components however, are considerably reduced by the fact that for two given index sets, say $\{i, \ldots, k\}$ and $\{j, \ldots, l\}$, any permutation of each of the sets yields the same result since differential operators commute. See Table 2 for some examples. We see that the number of independent components is a large number even for modest d, K, and L.

8.2 Spatial symmetries

Imposing stationarity, i.e. $C(t,s) = C(\overline{\tau})$, simplifies Eq. 13 to

$$C^{(\kappa,\lambda)}(\overline{\tau}) = (-1)^{|\lambda|} C^{(\kappa+\lambda)}(\overline{\tau}) = (-1)^{|\lambda|} \frac{\partial^{|\kappa|+|\lambda|}}{\partial \tau_1^{\kappa_1+\lambda_1} \cdots \partial \tau_d^{\kappa_d+\lambda_d}} C(\overline{\tau}),$$

and the tensor Eq. 14 takes the less general form

$$C^{(K,L)}_{i\cdots k\, j\cdots l}(\overline{\tau}) = (-1)^L \frac{\partial^{K+L}}{\partial\tau_i\cdots\partial\tau_k\,\partial\tau_j\cdots\partial\tau_l} C(\overline{\tau});$$
$$i,\ldots,k,\ j,\ldots,l \in \{1,\ldots,d\}.$$

The reduction of number of independent components is significant since all permutations of the pooled index sets, say $\{i,\ldots,k,j,\ldots,l\}$, now yield identical results. Table 2 gives a few examples comparing the number of independent components for different ranks. It is seen that the number of independent components becomes large as the rank increases even for stationary covariance functions.

For separable stationary covariance functions, $C(\overline{\tau}) = \prod_i C_i(\tau_i)$, the higher order derivatives become more tractable:

$$C^{(\kappa,\lambda)}(\overline{\tau}) = (-1)^{|\lambda|}\prod_{i=1}^{d} C_i^{(\kappa_i+\lambda_i)}(\tau_i) = (-1)^{|\lambda|}\prod_{i=1}^{d}\frac{\partial^{\kappa_i+\lambda_i}}{\partial\tau_i^{\kappa_i+\lambda_i}} C_i(\tau_i),$$

and the tensor Eq. 14 takes the simpler form

$$C^{(K,L)}_{i\cdots k\, j\cdots l}(\overline{\tau}) = (-1)^L \prod_{i=1}^{d}\frac{\partial^{N_i}}{\partial\tau_i^{N_i}} C_i(\tau_i),$$

where $N_q = \delta_{qi}+\cdots+\delta_{qk}+\delta_{qj}+\cdots+\delta_{ql} \in [0, K+L]$. This means that $(K+L)$ derivatives of increasing order for each $C_i(\tau_i)$ must be evaluated. For the example above ($d=3$, $K=L=2$) a total of 12 derivatives must be evaluated. Assuming the $C_i(\tau_i)$'s have identical form reduces this number to 4. Note that this number does not necessarily coincide with the number of independent components.

8.3 Isotropy

Isotropic covariance tensors have the general form

$$C^{(K,L)}_{i\cdots k\, j\cdots l}(\overline{\tau}) = (-1)^L \frac{\partial^{K+L}}{\partial\tau_i\cdots\partial\tau_k\,\partial\tau_j\cdots\partial\tau_l} C(\tau); \tag{15}$$
$$i,\ldots,k,\ j,\ldots,l \in \{1,\ldots,d\}$$

The number of independent components are given by Conjecture 1. Table 2 gives a few examples revealing a simple pattern. It is seen that isotropy dramatically reduces the number of independent components, and moreover, the number of independent components are independent of d. Theorem 3 give a recipe for expressing the partial derivatives in Eq. 15 by the first $K+L$ derivatives of $C(\tau)$ with respect to τ.

As an example of the use of Theorem 3, consider a tensor field of rank two (second order derivatives) and the corresponding rank three and four covariance tensors (see Section A.1 for more details):

$$\begin{aligned}
C^{(1,2)}_{ijk}(\overline{\tau}) &= \left(\tau_i\delta_{jk}+\tau_j\delta_{ik}+\tau_k\delta_{ij}\right) C_{TT}(\tau) + \tau_i\tau_j\tau_k\frac{1}{\tau^2}\Big\{C_{RT}(\tau)-C_{TT}(\tau)\Big\}\\
C^{(2,2)}_{ijkl}(\overline{\tau}) &= \left(\delta_{ij}\delta_{kl}+\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk}\right) C_{TT}(\tau)\\
&\quad + \left(\delta_{ij}\tau_k\tau_l+\delta_{ik}\tau_j\tau_l+\delta_{il}\tau_j\tau_k+\delta_{jk}\tau_i\tau_l+\delta_{jl}\tau_i\tau_k+\delta_{kl}\tau_i\tau_j\right)\\
&\qquad \times\frac{1}{\tau^2}\Big\{C_{RT}(\tau)-C_{TT}(\tau)\Big\}\\
&\quad + \tau_i\tau_j\tau_k\tau_l\frac{1}{\tau^4}\Big\{C_{RR}(\tau)-6C_{RT}(\tau)+3C_{TT}(\tau)\Big\},
\end{aligned}$$

where the isotropic covariance functions similar to $C_R(\tau)$ and $C_T(\tau)$ are

$$C_{RR}(\tau) = \frac{\mathrm{d}^4}{\mathrm{d}\tau^4} C(\tau) \qquad C_{TT}(\tau) = \left(\frac{1}{\tau}\frac{\mathrm{d}}{\mathrm{d}\tau}\right)^2 C(\tau)$$
$$C_{RT}(\tau) = \frac{\mathrm{d}^2}{\mathrm{d}\tau^2}\left(\frac{1}{\tau}\frac{\mathrm{d}}{\mathrm{d}\tau}\right) C(\tau).$$

These have geometrical interpretations similar to $C_R(\tau)$ and $C_T(\tau)$:

$$\begin{aligned}
C_{RR}(\tau) &= \operatorname{cov}\{X_{ii}^{(2)}(t), X_{ii}^{(2)}(s)\}; && (\tau_i = \tau)\\
C_{TT}(\tau) &= \frac{1}{3}\operatorname{cov}\{X_{ii}^{(2)}(t), X_{ii}^{(2)}(s)\}; && (\tau_i = 0)\\
&= \operatorname{cov}\{X_{ii}^{(2)}(t), X_{kk}^{(2)}(s)\}; && (\tau_i = \tau_k = 0 \text{ and } i \neq k)\\
C_{RT}(\tau) &= \operatorname{cov}\{X_{ii}^{(2)}(t), X_{kk}^{(2)}(s)\}; && (\tau_i = \tau, \tau_k = 0, \text{and } i \neq k)
\end{aligned}$$

From these expressions it is seen that $3C_{RR}(0) = C_{TT}(0) = C_{RT}(0)$ so that

$$C_{ijkl}^{(2,2)}(0) = \left(\delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\right) C_{TT}(0).$$

Moreover, we see that $C(\tau)$ must be four times differentiable at 0 for $C_{RR}(0)$ to be well defined.

8.4 Prediction using higher order derivatives

To illustrate the use of higher order derivatives a simple one dimensional example is supplied. Although useless for practical purposes some peculiar effects are interesting. Only one data location is considered but the n first derivatives will be taken into account. Thus the data vector takes the form $X^T = \left[X(0), X^{(1)}(0), \ldots, X^{(n)}(0)\right]$. The necessary covariances are:

$$\begin{aligned}
K_{ij} &= \operatorname{cov}\{X^{(i)}(0), X^{(j)}(0)\} = (-1)^j \frac{\mathrm{d}^{i+j}}{\mathrm{d}t^{i+j}} C(t)|_{t=0}; && i,j = 0,\ldots,n\\
k_i(t) &= \operatorname{cov}\{X(t), X^{(i)}(0)\} = (-1)^i \frac{\mathrm{d}^i}{\mathrm{d}t^i} C(t); && i = 0,\ldots,n.
\end{aligned}$$

The Gaussian covariance function with unit variance, $C(t) = \exp(-t^2)$, is considered. It is infinitely differentiable so that sample functions are analytical, that is, any order of sample path derivatives exist. The covariances can be expressed as

$$K_{ij} = \begin{cases} (-1)^{(i-j)/2}(i+j)!/(i/2+j/2)!\,; & i+j \text{ even} \\ 0; & i+j \text{ odd}, \end{cases} \tag{16}$$

$$k_i(t) = H_i(t)\,\exp(-t^2). \tag{17}$$

where H_n is the nth Hermite polynomial (Abramowitz and Stegun, 1972, p. 934). Correlations and standard deviations in K_{ij} is given in Table 3.

The variances increase rapidly so that K becomes ill-conditioned for large n; numerical instability occurred for $n > 9$. Also it is seen that the negative correlations increase as the order of the derivatives increase. This suggests that the information added by adding higher order derivatives are successively reduced. Moreover it is seen that even and odd derivatives are uncorrelated and that correlations within even or odd derivatives alternates between negative and positive values.

The predictor Eq. 1 with K and k given by Eq. 16 and Eq. 17 yields the conditional expectation. Our goal of prediction will be to reproduce the function $\cos \pi t$ given the nine first derivatives and the point value at $t = 0$:

$$X^T = [1, 0, -\pi^2, 0, \pi^4, 0, -\pi^6, 0, \pi^8, 0].$$

Figure 7 shows predictions for an increasing number of included derivatives. The dashed line is the target; the cosine function. The initial prediction, using the point value alone, gives a Gaussian

Table 3. Standard deviations, $\sigma_i = K_{ii}^{1/2}$, along diagonal. Correlations, $K_{ij}/(\sigma_i\sigma_j)$, off diagonal.

	j						
i	0	1	2	3	4	5	6
0	1	0	–0.58	0	0.29	0	–0.15
1		1.4	0	–0.77	0	0.49	0
2			3.5	0	–0.85	0	0.59
3				11.0	0	–0.88	0
4					41.0	0	–0.99
5						173.9	0
6							815.6

shaped curve asymptotically approaching zero. As further derivatives are included the predictions approach the first part of the cosine. Note that even numbered derivatives of the cosine are zero at $t = 0$ so that conditioning on these has no effect on the predictor causing pair-wise curves to coincide. For a small number of derivatives, say n, it is possible to identify the n leading terms of the kriging predictor. These are, as one should expect, identical to the n leading terms of the Taylor expansion for the cosine function. Thus we conjecture that the kriging predictor approach the Taylor expansion as n is increased, the main difference being the damping given by the exponential function for large t. This is somewhat contradictory to the observations that the diagonal of K increase rapidly and that the correlations approach ± 1. The increase in variances are probably compensated by rapidly increasing values in k and X. The corresponding prediction error given by Eq. 2 are also illustrated in Figure 7. It is seen that the errors are reduced as more derivatives are considered. The reduction however, recede as more derivatives are added. This is probably a sign of the increasingly high correlations in the K matrix.

9 Discussion

9.1 Combinations of gradient data and point value data

In the preceding sections gradient data and point value data at the same location was considered. It is of course possible to have any combination of gradient data and point value data scattered independently in space. It is also possible to condition on a subset of the components of the gradient vector although this will remove the isotropic symmetry. In computer experiments—where isotropy is hardly reasonable to assume—some derivatives could be cheap to obtain while others are utterly expensive to calculate.

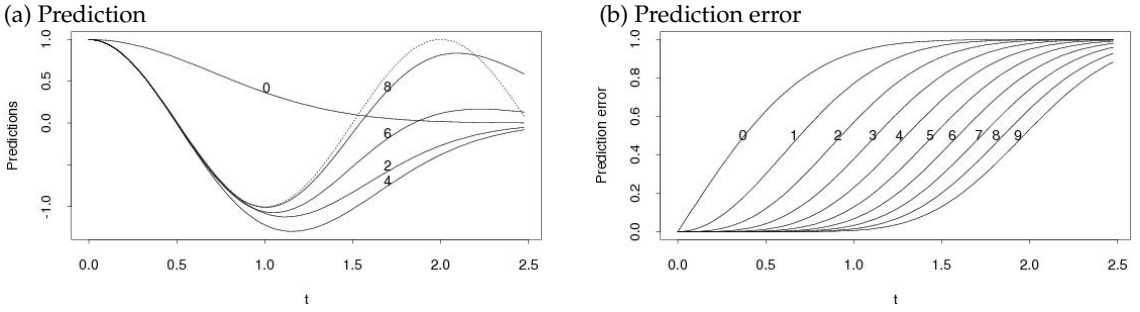


Figure 7. (a) shows an attempt to predict $\cos \pi t$ (dashed line) using an increasing number of derivatives. Predictions are labelled by the number of derivatives used. Even and odd numbered curves coincide since odd numbered derivatives are zero at $t = 0$. Prediction errors in (b).

9.2 Non stationary models

The universal kriging model Eq. 3 include a linear trend so that the random field $Y(t)$ is not isotropic nor stationary. Note however that the covariance functions are not affected and remains isotropic (or stationary). When considering Bayesian kriging (Abrahamsen, 1993; Kitanidis, 1986; Omre, 1987) even the stationarity of the covariances are actually abandoned; the linear parameters, β, are assumed to be random variables so that the covariance function of $Y(t)$ becomes:

$$C_Y(t,s) = \mathrm{cov}\{Y(t), Y(s)\} = f(t)^T S_0 \, f(s) + C(\tau),$$

where $S_0 = \mathrm{var}\{\beta\}$ is the prior covariances. So even though C is isotropic (or stationary), C_Y is in general not. Derivatives can be introduced as in the universal kriging case by considering the necessary derivatives of $f(t)$.

The isotropy of the covariance function C can also be somewhat relaxed. Consider the random field $\tilde{X}(t) = \sigma(t)\, X(t)$, where $\sigma(t) > 0$. The covariance function of $\tilde{X}(t)$ is

$$\tilde{C}(t,s) = \mathrm{cov}\{\tilde{X}(t), \tilde{X}(s)\} = \sigma(t)\, \sigma(s)\, C_X(\tau),$$

which is non-stationary. The correlation function however, will remain isotropic. This will remain true for the correlation tensor of the gradient field as well. Thus all results for isotropic covariance functions will be valid for the correlation functions; the correlation tensor must be rescaled to possess unit variance. The kriging predictors described above are still valid although the entries of the kriging matrices, K, and the covariance vectors, k, must be multiplied by appropriate $\sigma(t)\,\sigma(s)$-combinations.

9.3 Smoothness and precision of gradient data

The standard kriging predictors give perfect interpolation of data, and in the case of gradient data, perfect compliance to the gradient. When considering gradient data it is expedient to question whether this is reasonable. Also, it is occasionally unrealistic to assume that the random field under study is differentiable despite the existence of gradient information. A few approaches dealing with this problem are sketched below.

As Figure 1 shows, the influence of the gradient is very local for almost non-differentiable covariance functions such as the modified Bessel functions with the ν parameter close to 1. Thus using an erratic model reduces the impact of gradient data.

The measuring of gradients are more likely to be inaccurate than measurements of point values. It was shown in Section 6 that incorporating measuring errors is possible by adding appropriate variances to the corresponding diagonal variances of the kriging matrix.

An different approach considered by Renard and Ruffo (1993) is to consider gradient data as spatial averages thus requiring the support to be considered. This requires (numerical) spatial integration over some support volume of the covariance tensor as well as the cross-covariances. This will cause the predictors to tend to the specified gradient values rather than obey them perfectly. An additional nice feature is that the covariance function of $X(t)$ does not need to be differentiable even though the spatially averaged covariance functions are.

A final approach is to consider $X(t)$ as a sum of one differentiable component and one more erratic component and let the gradient data apply to the smooth component. This is closely related to the approach considered by Renard and Ruffo (1993). This approach is related to factorial kriging.

10 Closing remarks

A framework for using gradient data has been established. It consists of expanding the kriging matrix and cross-covariance vector to accommodate the new information. If gradient data are available at all point value observations this amounts to expanding the kriging system by a factor

$d+1$ where d is the dimension of the space under consideration. For surfaces this means that the size is three fold. Thus the size of computation could be prohibitively large.

A few examples show the large reduction in prediction error obtained by adding gradient information. This encourage the use of gradient data and hopefully justifies the additional need for computational resources.

Isotropic symmetry simplifies the treatment of the covariance tensors by reducing the number of independent components. The number of independent components are the same in all spatial dimensions.

Acknowledgments

This work was supported by a research fellowship from The Research Council of Norway. The author benefited from facilities at Norwegian Computing Centre. I am grateful to Petter Mostad who supplied the substantial parts of Section A.

A Isotropic covariance tensors

Definition (Isotropic tensor fields)**.** For integers $r \geq 0$, $n \geq 0$ and an index set $i_1, \ldots, i_r \in \{1, \ldots, d\}$, define a tensor field of rank r, $D^{(r)}_{i_1 \cdots i_r}(\overline{\tau}; n)$, by

$$D^{(r)}_{i_1 \cdots i_r}(\overline{\tau}; n) = \begin{cases} 1 & \text{if } r = n = 0 \\ 0 & \text{if } n > r \text{ or } n = 0 \text{ and } r > 0, \end{cases}$$

$$D^{(r)}_{i_1 \cdots i_r}(\overline{\tau}; n) = \tau_{i_r} D^{(r-1)}_{i_1 \cdots i_{r-1}}(\overline{\tau}; n-1) + \sum_{s=1}^{r-1} \delta_{i_r i_s} D^{(r-2)}_{i_1 \cdots \hat{i}_s \cdots i_{r-1}}(\overline{\tau}; n-1) \quad \text{if } 0 < n \leq r.$$

Here, $\hat{i}_s$ means that the index is missing.

The definition implies that $D^{(r)}_{i_1 \cdots i_r}(\overline{\tau}; n) = 0$ unless $0 \leq \lceil \frac{r}{2} \leq n \leq r$. The first non-trivial examples are

$$\begin{aligned}
D^{(1)}_{i}(\overline{\tau}; 1) &= \tau_i \\
D^{(2)}_{ij}(\overline{\tau}; 1) &= \delta_{ij} & D^{(2)}_{ij}(\overline{\tau}; 2) &= \tau_i \tau_j \\
D^{(3)}_{ijk}(\overline{\tau}; 2) &= \tau_i \delta_{jk} + \tau_j \delta_{ik} + \tau_k \delta_{ij} & D^{(3)}_{ijk}(\overline{\tau}; 3) &= \tau_i \tau_j \tau_k \\
D^{(4)}_{ijkl}(\overline{\tau}; 2) &= \delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk} \\
D^{(4)}_{ijkl}(\overline{\tau}; 3) &= \delta_{ij}\tau_k\tau_l + \delta_{ik}\tau_j\tau_l + \delta_{il}\tau_j\tau_k + \delta_{jk}\tau_i\tau_l + \delta_{jl}\tau_i\tau_k + \delta_{kl}\tau_i\tau_j \\
D^{(4)}_{ijkl}(\overline{\tau}; 4) &= \tau_i\tau_j\tau_k\tau_l.
\end{aligned}$$

Note that $D^{(r)}_{i_1 \cdots i_r}(\overline{\tau}; n)$ are symmetric and isotropic tensor fields. Moreover, it is possible to prove the following Lemma by induction.

Lemma 2. *The tensor fields $D^{(r)}_{i_1 \cdots i_r}(\overline{\tau}; n)$ can be written as*

$$D^{(r)}_{i_1 \cdots i_r}(\overline{\tau}; n) = \sum \cdots \tau_i \delta_{jk} \cdots,$$

where summation is over all combinations of τ_i and δ_{jk} such that

- *every index among $i_1, \ldots, i_r$ is represented once in every term,*
- *all terms have n factors.*

Conjecture 1. *Any symmetric and isotropic tensor field of rank r can be written as*

$$V_{i_1 \cdots i_r}(\overline{\tau}) = \sum_{n=\lceil \frac{r}{2}}^{r} D^{(r)}_{i_1 \cdots i_r}(\overline{\tau}; n)\, C_n(\tau),$$

where $C_n(\tau)$ are scalar fields.

The conjecture is a consequence of Theorem 3 below for isotropic and symmetric tensor fields of the form Eq. 15. To establish that $D^{(r)}_{i_1 \cdots i_r}(\overline{\tau}; n)$ forms a basis for the linear space of isotropic and symmetric tensors of rank r is beyond the scope of this paper. However, some arguments are supplied:

Sketch of proof. The terms in $D^{(r)}_{i_1 \cdots i_r}(\overline{\tau}; n);\ n = \lceil \frac{r}{2}, \ldots, r$ are the only tensors of rank r that can be formed from the rank one tensor $\overline{\tau}$ and the isotropic unit tensor δ_{ij}. This exhausts the possibilities for isotropic tensors. The tensors $D^{(r)}_{i_1 \cdots i_r}(\overline{\tau}; n);\ n = \lceil \frac{r}{2}, \ldots, r$ are the possible symmetric combinations of these isotropic terms. □

Table A.1. *The function* $F(k,n)$.

	k				
n	0	1	2	3	4
0	1	0	0	0	0
1	0	1	0	0	0
2	0	–1	1	0	0
3	0	3	–3	1	0
4	0	–15	15	–6	1

Additional constraints on the scalar fields, $C_n(\tau)$, must be imposed to ensure positive definiteness of the tensors in Conjecture 1 to obtain covariance tensors. The following theorem gives explicit formulae for isotropic covariance tensors for tensor fields of the form Eq. 15.

Definition. For integers $n \geq 0$, $k \geq 0$ define a function $F(k,n)$ by

- $F(0,0) = 1$,
- $F(0,n) = F(k,0) = 0 \quad$ for $k, n > 0$,
- $F(k,n) = (k-2n+2)F(k,n-1) + F(k-1,n-1)$.

The simplest examples are found in Table A.1.

Theorem 3. *For an r times differentiable isotropic function $C\colon \mathbb{R}^d \to \mathbb{R}$, we have*

$$\frac{\partial^r}{\partial\tau_{i_1}\cdots\partial\tau_{i_r}}C(\tau) = \sum_{n=\lceil\frac{r}{2}}^{r} D^{(r)}_{i_1\cdots i_r}(\overline{\tau};n)\sum_{k=0}^{n}\frac{F(k,n)}{\tau^{2n-k}}\frac{\partial^k}{\partial\tau^k}C(\tau).$$

Proof. We use induction on r. If $r=0$ the theorem is trivial. Assuming it is valid for $r-1 \geq 0$, it remains to prove it for r. Differentiating the formula for $r-1$:

$$\begin{aligned}\frac{\partial^r}{\partial\tau_{i_1}\cdots\partial\tau_{i_r}}C(\tau) &= \frac{\partial}{\partial\tau_{i_r}}\left(\sum_{n=\lceil\frac{r-1}{2}}^{r-1} D^{(r-1)}_{i_1\cdots i_{r-1}}(\overline{\tau};n)\sum_{k=0}^{n}\frac{F(k,n)}{\tau^{2n-k}}\frac{\partial^k}{\partial\tau^k}C(\tau)\right)\\ &= \sum_{n=\lceil\frac{r-1}{2}}^{r-1}\left(\frac{\partial}{\partial\tau_{i_r}}D^{(r-1)}_{i_1\cdots i_{r-1}}(\overline{\tau};n)\right)\sum_{k=0}^{n}\frac{F(k,n)}{\tau^{2n-k}}\frac{\partial^k}{\partial\tau^k}C(\tau)\\ &\quad+\sum_{n=\lceil\frac{r-1}{2}}^{r-1} D^{(r-1)}_{i_1\cdots i_{r-1}}(\overline{\tau};n)\frac{\tau_{i_r}}{\tau}\frac{\partial}{\partial\tau}\left(\sum_{k=0}^{n}\frac{F(k,n)}{\tau^{2n-k}}\frac{\partial^k}{\partial\tau^k}C(\tau)\right)\end{aligned} \tag{A.1}$$

The final term of Eq. A.1 can be rewritten using

$$\begin{aligned}\frac{\partial}{\partial\tau}\left(\sum_{k=0}^{n}\frac{F(k,n)}{\tau^{2n-k}}\frac{\partial^k}{\partial\tau^k}C(\tau)\right) &= \sum_{k=0}^{n}\frac{(k-2n)F(k,n)}{\tau^{2n-k+1}}\frac{\partial^k}{\partial\tau^k}C(\tau)+\sum_{k=0}^{n}\frac{F(k,n)}{\tau^{2n-k}}\frac{\partial^{k+1}}{\partial\tau^{k+1}}C(\tau)\\ &= \sum_{k=1}^{n+1}\frac{(k-2n)F(k,n)}{\tau^{2n-k+1}}\frac{\partial^k}{\partial\tau^k}C(\tau)+\sum_{k=1}^{n+1}\frac{F(k-1,n)}{\tau^{2n-k+1}}\frac{\partial^k}{\partial\tau^k}C(\tau)\\ &= \sum_{k=1}^{n+1}\frac{(k-2n)F(k,n)+F(k-1,n)}{\tau^{2n-k+1}}\frac{\partial^k}{\partial\tau^k}C(\tau)\\ &= \sum_{k=1}^{n+1}\frac{F(k,n+1)}{\tau^{2n-k+1}}\frac{\partial^k}{\partial\tau^k}C(\tau).\end{aligned}$$

So the second term of Eq. A.1 equals

$$\sum_{n=\lceil\frac{r-1}{2}\rceil}^{r-1} D^{(r-1)}_{i_1\cdots i_{r-1}}(\overline{\tau};n)\frac{\tau_{i_r}}{\tau}\sum_{k=0}^{n+1}\frac{F(k,n+1)}{\tau^{2n-k+1}}\frac{\partial^k}{\partial\tau^k}C(\tau)$$
$$= \sum_{n=\lceil\frac{r-1}{2}\rceil+1}^{r} \tau_{i_r} D^{(r-1)}_{i_1\cdots i_{r-1}}(\overline{\tau};n-1)\sum_{k=0}^{n}\frac{F(k,n)}{\tau^{2n-k}}\frac{\partial^k}{\partial\tau^k}C(\tau).$$

Looking at the whole expression in Eq. A.1, it is seen that both terms may be summed from $n=\lceil\frac{r-1}{2}$ to $n=r$, giving

$$\frac{\partial^r}{\partial\tau_{i_1}\cdots\partial\tau_{i_r}}C(\tau) = \sum_{n=\lceil\frac{r-1}{2}\rceil}^{r}\Bigg(\frac{\partial}{\partial\tau_{i_r}}D^{(r-1)}_{i_1\cdots i_{r-1}}(\overline{\tau};n) + \tau_{i_r}D^{(r-1)}_{i_1\cdots i_{r+1}}(\overline{\tau};n-1)\Bigg)\sum_{k=0}^{n}\frac{F(k,n)}{\tau^{2n-k}}\frac{\partial^k}{\partial\tau^k}C(\tau). \tag{A.2}$$

The following equation is easily verified from Lemma 2:

$$\frac{\partial}{\partial\tau_{i_r}}D^{(r-1)}_{i_1\cdots i_{r-1}}(\overline{\tau};n) = \sum_{s=1}^{r-1}\delta_{i_r i_s}D^{(r-2)}_{i_1\cdots\hat{i}_s\cdots i_{r-1}}(\overline{\tau};n).$$

Finally, using Definition A and observing that the summation in Eq. A.2 may start at $n=\lceil\frac{r}{2}$ completes the proof. □

A.1 Examples

A few examples illustrate the use of Theorem 3. First order derivative ($r=1$):

$$\frac{\partial}{\partial\tau_i}C(\tau) = D^{(1)}_i(1)\frac{F(1,1)}{\tau}\frac{\mathrm{d}}{\mathrm{d}\tau}C(\tau) = -\tau_i C_T(\tau).$$

Second order derivative ($r=2$):

$$\begin{aligned}\frac{\partial^2}{\partial\tau_i\partial\tau_j}C(\tau) &= D^{(2)}_{ij}(\overline{\tau};1)\frac{F(1,1)}{\tau}\frac{\mathrm{d}}{\mathrm{d}\tau}C(\tau)\\ &\quad + D^{(2)}_{ij}(\overline{\tau};2)\left(\frac{F(1,2)}{\tau^3}\frac{\mathrm{d}}{\mathrm{d}\tau}C(\tau)+\frac{F(2,2)}{\tau^2}\frac{\mathrm{d}^2}{\mathrm{d}\tau^2}C(\tau)\right)\\ &= -\delta_{ij}C_T(\tau) - \frac{\tau_i\tau_j}{\tau^2}\Big\{C_R(\tau)-C_T(\tau)\Big\}.\end{aligned}$$

We recognise the covariance tensor Eq. 7 except from the opposite sign.

Third order derivative ($r=3$):

$$\begin{aligned}\frac{\partial^3}{\partial\tau_i\partial\tau_j\partial\tau_k}C(\tau) &= D^{(3)}_{ijk}(\overline{\tau};2)\left(\frac{F(1,2)}{\tau^3}\frac{\mathrm{d}}{\mathrm{d}\tau}C(\tau)+\frac{F(2,2)}{\tau^2}\frac{\mathrm{d}^2}{\mathrm{d}\tau^2}C(\tau)\right)\\ &\quad + D^{(3)}_{ijk}(\overline{\tau};3)\left(\frac{F(1,3)}{\tau^5}\frac{\mathrm{d}}{\mathrm{d}\tau}C(\tau)+\frac{F(2,3)}{\tau^4}\frac{\mathrm{d}^2}{\mathrm{d}\tau^2}C(\tau)\right.\\ &\qquad\left.+\frac{F(3,3)}{\tau^3}\frac{\mathrm{d}^3}{\mathrm{d}\tau^3}C(\tau)\right)\\ &= D^{(3)}_{ijk}(\overline{\tau};2)C_{TT}(\tau) + D^{(3)}_{ijk}(\overline{\tau};3)\frac{1}{\tau^2}\Big\{C_{RT}(\tau)-C_{TT}(\tau)\Big\}.\end{aligned}$$

Fourth order derivative ($r = 4$):

$$\begin{aligned}
\frac{\partial^4}{\partial\tau_i\partial\tau_j\partial\tau_k\partial\tau_l}C(\tau) &= D^{(4)}_{ijkl}(\overline{\tau};2)\left(\frac{F(1,2)}{\tau^3}\frac{\mathrm{d}}{\mathrm{d}\tau}C(\tau) + \frac{F(2,2)}{\tau^2}\frac{\mathrm{d}^2}{\mathrm{d}\tau^2}C(\tau)\right)\\
&+ D^{(4)}_{ijkl}(\overline{\tau};3)\left(\frac{F(1,3)}{\tau^5}\frac{\mathrm{d}}{\mathrm{d}\tau}C(\tau) + \frac{F(2,3)}{\tau^4}\frac{\mathrm{d}^2}{\mathrm{d}\tau^2}C(\tau)\right.\\
&\left.+ \frac{F(3,3)}{\tau^3}\frac{\mathrm{d}^3}{\mathrm{d}\tau^3}C(\tau)\right)\\
&+ D^{(4)}_{ijkl}(\overline{\tau};4)\left(\frac{F(1,4)}{\tau^7}\frac{\mathrm{d}}{\mathrm{d}\tau}C(\tau) + \frac{F(2,4)}{\tau^6}\frac{\mathrm{d}^2}{\mathrm{d}\tau^2}C(\tau)\right.\\
&\left.+ \frac{F(3,4)}{\tau^5}\frac{\mathrm{d}^3}{\mathrm{d}\tau^3}C(\tau) + \frac{F(4,4)}{\tau^4}\frac{\mathrm{d}^4}{\mathrm{d}\tau^4}C(\tau)\right)\\
&= D^{(4)}_{ijkl}(\overline{\tau};2)\,C_{TT}(\tau) + D^{(4)}_{ijkl}(\overline{\tau};3)\frac{1}{\tau^2}\Big\{C_{RT}(\tau) - C_{TT}(\tau)\Big\}\\
&+ D^{(4)}_{ijkl}(\overline{\tau};4)\frac{1}{\tau^4}\Big\{C_{RR}(\tau) - 6C_{RT}(\tau) + 3C_{TT}(\tau)\Big\}.
\end{aligned}$$

B Differentiable covariance functions

The isotropic covariance functions mentioned below all apply in any dimension. Proofs of this is given in the supplied references.

B.1 Gaussian or second order exponential covariance function

The second order exponential covariance function is defined as (Yaglom, 1986, p. 364)

$$C(\tau) = \sigma^2\, e^{-a\tau^2}; \qquad a > 0. \tag{B.3}$$

(Choosing $a = 3/R^2$ where R is the correlation length is common.) The longitudinal and lateral covariance functions are:

$$\begin{aligned}
C_T(\tau) &= \sigma^2\, 2a\, e^{-a\tau^2}\\
C_R(\tau) &= C_T(\tau)\left(1 - \tau^2 2a\right).
\end{aligned}$$

so the cross-covariances and covariance tensor are:

$$\begin{aligned}
\dot{C}_i(\overline{\tau}) &= -\tau_i\, \sigma^2\, 2a\, e^{-a\tau^2}\\
\ddot{C}_{ij}(\overline{\tau}) &= \left(\delta_{ij} - 2a\,\tau_i\tau_j\right)\sigma^2\, 2a\, e^{-a\tau^2}.
\end{aligned}$$

B.2 Rational quadratic covariance functions

The covariance function is given by (Matérn 1986, p. 17 and Cressie 1993, p. 61)

$$C(\tau; a, \nu) = \sigma^2\Big(1 + a\,\tau^2\Big)^{-\nu}; \qquad a, \nu > 0. \tag{B.4}$$

Choosing $a = \mathcal{S}_\nu/R^2$ where $\mathcal{S}_\nu = 20^{1/\nu} - 1$ gives correlation 0.05 at $\tau = R$. The longitudinal and lateral covariance functions are

$$\begin{aligned}
C_T(\tau) &= \sigma^2\, 2a\nu\left(1 + a\tau^2\right)^{-(\nu+1)}\\
C_R(\tau) &= C_T(\tau)\left(1 - \frac{2a\tau^2(\nu+1)}{1 + a\tau^2}\right),
\end{aligned}$$

so the covariance tensor and cross-covariances are

$$\dot{C}_i(\overline{\tau}) = \tau_i\,\sigma^2\,2a\nu\left(1+a\tau^2\right)^{-(\nu+1)}$$
$$\ddot{C}_{ij}(\overline{\tau}) = \left(\delta_{ij} - \tau_i\tau_j\frac{2a(\nu+1)}{1+a\tau^2}\right)\sigma^2\,2a\nu\left(1+a\tau^2\right)^{-(\nu+1)}.$$

B.3 Modified Bessel covariance functions

The covariance function is defined as ()Matérn 1986, p. 17 and Yaglom 1986, pp 362–363)

$$C(\tau;a,\nu) = \sigma^2\frac{1}{\Gamma(\nu)2^{\nu-1}}(a\tau)^\nu K_\nu(a\tau), \qquad a,\nu>0,$$

where it is convenient to let $a = \mathcal{S}_\nu/R$ and choose an appropriate ν-dependent scaling $\mathcal{S}_\nu$. The longitudinal and lateral covariance functions are

$$C_T(\tau) = \frac{\sigma^2a^2}{\Gamma(\nu)2^{\nu-1}}(a\tau)^{\nu-1}K_{\nu-1}(a\tau)$$
$$C_R(\tau) = \frac{\sigma^2a^2}{\Gamma(\nu)2^{\nu-1}}\Big\{(a\tau)^{\nu-1}K_{\nu-1}(a\tau) - (a\tau)^\nu K_{\nu-2}(a\tau)\Big\}.$$

The covariance tensor and cross-covariances are

$$\dot{C}_i(\overline{\tau}) = \frac{\sigma^2a^2}{\Gamma(\nu)2^{\nu-1}}\,\tau_i\,(a\tau)^{\nu-1}K_{\nu-1}(a\tau)$$
$$\ddot{C}_{ij}(\overline{\tau}) = \frac{\sigma^2a^2}{\Gamma(\nu)2^{\nu-1}}\Big\{(a\tau)^{\nu-1}K_{\nu-1}(a\tau)\delta_{ij} - a^2\,\tau_i\,\tau_j\,(a\tau)^{\nu-2}K_{\nu-2}(a\tau)\Big\}.$$

The limiting behaviour for small z is

$$K_0(z) \sim -\ln z$$
$$K_\nu(z) \sim \Gamma(\nu)2^{\nu-1}z^{-\nu} \quad (\nu>0).$$

Also note that $K_\nu(z) = K_{-\nu}(z)$. The covariances C_T and C_R must be finite at $\tau = 0$ so $\nu > 1$ for this to be fulfilled.

References

Abrahamsen, P. (1993). Bayesian kriging for seismic depth conversion of a multi-layer reservoir. In Soares (1993), pages 385–398. Available from: http://dx.doi.org/10.1007/978-94-011-1739-5_31. 15, 22

Abrahamsen, P. (1997). *Prediction of Geological Subsurfaces Based on Gaussian Random Field Models*. Dr.scient. thesis, University of Oslo. 3

Abramowitz, M. and Stegun, I. A., editors (1972). *Handbook of Mathematical Functions*. 9th printing, Dover Publications Inc., New York. 20

Adler, R. J. (1981). *The Geometry of Random Fields*. John Wiley & Sons, New York. 4

Baafi, E. Y. and Schofield, N. A., editors (1997). *Geostatistics Wollongong '96*, Dordrecht. proc. '5th Inter. Geostat. Congr.', Wollongong Australia, 1996, Kluwer Academic Publ. 4

Butkov, E. (1968). *Mathematical Physics*. Addison-Wesley, Reading, Massachusetts. 7

Chauvet, P., Pailleux, J., and Chiles, J. P. (1976). Analyse objective des champs meteorologiques par cokrigeage. *Météorologie*, 6(4):37–54. 4

Christakos, G. (1992). *Random Field Models in Earth Sciences*. Academic Press Inc., San Diego, California. 4

Cressie, N. (1990). The origins of kriging. *Math. Geol.*, 22(3):239–252. 4

Cressie, N. (1993). *Statistics for Spatial Data*. John Wiley & Sons, New York, revised edition. 4, 27

Doob, J. L. (1953). *Stochastic Processes*. John Wiley & Sons, New York. 4

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, volume I. John Wiley & Sons, New York, 3rd edition. 7

Hwang, L. and McCorkindale, D. (1994). Troll field depth conversion using geostatistically derived average velocities. *The Leading Edge*, 13(4):561–569. 15

Jeffery, R. W., Stewart, I. C. F., and Alexander, D. W. (1996). Geostatistical estimation of depth conversion velocity using well control and gravity data. *First Break*, 14(8):313–320. 15

Journel, A. G. and Huijbregts, C. J. (1978). *Mining Geostatistics*. Academic Press Inc., London. 4

Kitanidis, P. K. (1986). Parameter uncertainty in estimation of spatial functions: Bayesian analysis. *Water Resources Res.*, 22(4):499–507. 22

Mardia, K. V. and Kent, J. T. (1996). Kriging and splines with derivative information. *Biometrika*, 83(1):207–221. 4

Mardia, K. V. and Little, J. A. (1994). Image warping using derivative information. In Bookstein, F. L., Duncan, J. S., Lange, N., and Wilson, D. C., editors, *Mathematical Methods in Medical Imaging*, volume 2299, pages 16–31, Bellingham, WA. Society of Photo-Optical Instrumentation Engineers. 4

Matérn, B. (1986). *Spatial Variation*. Springer-Verlag Inc., Berlin, 2nd edition. First edition published in 1960 by the Swedish Forest Research Institute. 27, 28

Monin, A. S. and Yaglom, A. M. (1975). *Statistical Fluid Mechanics: Mechanics of Turbulence*, volume 2. The MIT Press, Cambridge, Massachusetts. 9

Morris, M. D., Mitchell, T. J., and Ylvisaker, D. (1993). Bayesian design and analysis of computer experiments: Use of derivatives in surface prediction. *Technometrics*, 35(3):243–255. 4, 8

Omre, H. (1987). Bayesian kriging—merging observations and qualified guesses in kriging. *Math. Geol.*, 19(1):25–39. 22

Renard, D. and Ruffo, P. (1993). Depth, dip and gradient. In Soares (1993), pages 167–178. 4, 15, 22

Ripley, B. D. (1981). *Spatial Statistics*. John Wiley & Sons, New York. 4, 6

Ripley, B. D. (1988). *Statistical Inference for Spatial Processes*. Cambridge University Press, Cambridge. 4

Sacks, J., Schiller, S. B., and Welch, W. J. (1989a). Designs for computer experiments. *Technometrics*, 31(1):41–47. 4

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989b). Computer experiments. *Statistical Science*, 4(4):409–423. 4

Shiyi, Y. (1983). Essai d'utilisation du gradient. Technical Report S-141, École Nationale Supérieure des Mines de Paris, Centre de Géostatistique, 35 rue Saint-Honoré, 77305-Fontainebleau cedex, Paris. 4

Soares, A., editor (1993). *Geostatistics Tróia '92*, Dordrecht. proc. '4th Inter. Geostat. Congr.', Tróia Portugal, 1992, Kluwer Academic Publ. 4, 29, 30

Vanmarcke, E. (1983). *Random Fields: Analysis and Synthesis*. The MIT Press, Cambridge, Massachusetts. 8, 10

Yaglom, A. M. (1986). *Correlation Theory of Stationary and Related Random Functions I: Basic Results*. Springer Series in Statistics. Springer-Verlag Inc., New York. 9, 27, 28