Math Geosci
DOI 10.1007/s11004-009-9245-1

# Measures of Parameter Uncertainty in Geostatistical Estimation and Geostatistical Optimal Design

**Wolfgang Nowak**

Received: 7 December 2007 / Accepted: 12 August 2009


**Abstract** Studies of site exploration, data assimilation, or geostatistical inversion measure parameter uncertainty in order to assess the optimality of a suggested scheme. This study reviews and discusses measures for parameter uncertainty in spatial estimation. Most measures originate from alphabetic criteria in optimal design and were transferred to geostatistical estimation. Further rather intuitive measures can be found in the geostatistical literature, and some new measures will be suggested in this study. It is shown how these measures relate to the optimality alphabet and to relative entropy. Issues of physical and statistical significance are addressed whenever they arise. Computational feasibility and efficient ways to evaluate the above measures are discussed in this paper, and an illustrative synthetic case study is provided. A major conclusion is that the mean estimation variance and the averaged conditional integral scale are a powerful duo for characterizing conditional parameter uncertainty, with direct correspondence to the well-understood optimality alphabet. This study is based on cokriging generalized to uncertain mean and trends because it is the most general representative of linear spatial estimation within the Bayesian framework. Generalization to kriging and quasi-linear schemes is straightforward. Options for application to non-Gaussian and non-linear problems are discussed.

**Keywords** Optimal design · Kriging · Cokriging · Conditioning

W. Nowak (✉)
Institute of Hydraulic Engineering, LH2, Universität Stuttgart, Pfaffenwaldring 61, 70659 Stuttgart, Germany
e-mail: wolfgang.nowak@iws.uni-stuttgart.de

W. Nowak
Dept. of Civil and Environmental Engineering, University of California at Berkeley, 760 Davis Hall, 94720-1710 Berkeley, CA, USA

Published online: 10 October 2009

## 1 Introduction

The importance of rigorous uncertainty assessment in general modeling and decision making is widely recognized (e.g., Pappenberger and Beven 2006). To this end, the conditional uncertainty of model parameters or model predictions (i.e., the uncertainty after exploiting available data) must be quantified. This may include the application of scalar measures for parameter or prediction uncertainty. The very same measures for parameter uncertainty are inseparably connected to the field of optimal experimental design. An experimental design (or sampling strategy) is a specific choice of measurement types, numbers, locations, and experimental conditions. Intuitively, it is straightforward to require some sort of maximum benefit from laborious or expensive experiments. To this end, optimal design techniques define scalar-valued measures of conditional uncertainty, and then minimizes these measures as criterion for optimality. The most common criteria include $A$-, $C$-, $D$-, $E$-, and $T$-optimality, often subsumed under the term of alphabetic optimality (Box 1982). The ubiquitous need for such optimal designs in most science and engineering disciplines has triggered a vast series of studies in the statistical literature, mostly focusing on regression-like problems (e.g., Silvey 1980; Federov and Hackl 1997; Pukelsheim 2006). Chaloner and Verdinelli (1995) review the topic within the Bayesian framework. In the typical context of regression-like data analysis, these measures characterize the conditional covariance matrix of a few regression parameters. In the geostatistical context, however, the targets of estimation are random space functions discretized on fine grids (e.g., $10^4$ to $10^7$ cell-wise values, rising with newly available computer power) rather than a few regression parameters. Available methods for spatial estimation are kriging, cokriging (e.g., Matheron 1971; de Marsily 1986) and various techniques of geostatistical inverse modeling (e.g., Keidser and Rosbjerg 1991; Zimmerman et al. 1998). If translated to geostatistical estimation, criteria for optimal design and their respective measures of uncertainty are applied to the conditional covariance matrix of the discretized parameter fields (e.g., Müller 2007).

For well-resolved two- or three-dimensional parameter fields, the sheer size of the conditional covariance matrix of the parameters constitutes a problem of its own right. It may strictly inhibit explicit storage of the entire matrix, let alone any further processing. No matrix operation will lend itself to such large and dense matrices (compare Zimmerman et al. 1998) unless specific attention is paid to exploitable properties of conditional covariance matrices (Zimmerman 1989). Other than computational aspects, there are more fundamental problems with the physical and statistical significance of the optimality alphabet when used in spatial estimation. Regression typically features overdetermined problems with more observations than unknown parameters. Quite contrarily, spatial estimation usually features underdetermined problems with only few observations but with many unknown discrete values of random space functions. As a matter of fact, it is questionable that measures derived for overdetermined problems maintain their original significance when applied to a different class of underdetermined problems. More details on these problems are provided in a later section.

Within the geostatistical context, there are two primary objectives in measuring conditional uncertainty or looking at optimal designs. First, in theoretical studies of model reduction, data compression (e.g., reducing transient equations and time series to temporal moments) or data assimilation, parameter uncertainty itself is measured or minimized. The most commonly used criterion is the estimation variance (e.g., Journel and Huijbregts 1978; Kitanidis 1997), and the most simplistic and intuitive choice for a scalar measure is its spatial average (Diggle and Ribeiro 2007; Fritz et al. 2009). Other frequently used measures in applied hydro-geostatistical studies compare synthetic random fields and their estimates through the $L_2$-norm of their difference (e.g., Li and Yeh 1999; Zhu and Yeh 2005; Woodbury and Ulrych 2000). The second objective aims at a somewhat optimal predictive power of stochastic model predictions (e.g., Herrera and Pinder 2005). In such cases, the definition of optimality depends on the specific prediction purpose (e.g., Sun 1994). Possible scenarios may also entail minimization of more complex criteria, such as exploration costs (e.g., Bakr et al. 2003), the overall sum of exploration and remediation costs (e.g., Cirpka et al. 2004; Feyen and Gorelick 2005), or the risk of management decisions (e.g., McPhee and Yeh 2006). The cross-reference to parameter identifiability is given by Sun and Yeh (1990b). For this second objective, parameter uncertainty is often assessed in some intermediate step, either implicitly or explicitly, and then propagated onto model predictions. Specified by the individual context, this is a typically low number of predicted quantities. Seen from the perspective of optimal design, requiring minimal uncertainty for model predictions is merely a specific form to define minimum parameter uncertainty. Hence, these two groups are not distinguished any further in the remainder of this study.

This study reviews scalar measures of parameter uncertainty that have been transferred from the optimality alphabet to geostatistical estimation, and compares them to the rather intuitive measures found in hydro-geostatistical applications. Additionally, several new measures will be introduced. Five main questions will be addressed:

1. What is the physical and statistical significance of the measures transferred from classical optimal design to spatial estimation? This discussion will disclose which of the transferred measures lose their intended character.
2. How would the rather intuitive measures from hydro-geostatistical practice be judged from the well-founded theoretical perspective of optimal design? This discussion seeks to reconcile hydro-geostatistical practice with optimal design theory.
3. Which measures are computationally feasible for numerical evaluation when applied to large problem sizes commonly found in spatial estimation? Whenever possible, efficient ways to evaluate the considered measures even for prohibitively large problems are presented.
4. Based on the above three questions, new measures will be suggested. The forth question is then, how do the newly suggested measures perform in theoretical soundness, physical significance, and computational feasibility? To answer this question, the current study will clarify their relation to the optimality alphabet and related fundamental principles such as relative entropy (e.g., Woodbury and Ulrych 1993).

5. How easily applicable are these measures when applied to non-Gaussian and non-linear problems? Issues of physical significance and alternative methods for numerical evaluation will be presented.

For simplicity and generality, this study draws on a simple notation for cokriging with uncertain mean and trends as introduced in Nowak and Cirpka (2004). It is a special case of Bayesian geostatistics, where elements of the geostatistical model itself (here the spatial function providing the mean value) are uncertain (Kitanidis 1986; Handcock and Stein 1993). When setting spatial variability to zero within this generalized framework, the remaining uncertain trend contributions resemble standard linear regression with prior knowledge. The special cases of kriging or cokriging with known and unknown mean (e.g., Matheron 1971; Journel and Huijbregts 1978; de Marsily 1986) or universal kriging (Olea 1974) are obtained easily from this general case, just as quasi-linear extensions to weakly non-linear problems (e.g., Kitanidis 1996b; Hughson and Yeh 2000). The following two sections summarize cokriging and alphabetic criteria in the traditional context. Only a minimum of basic definitions is provided in order to install the notation used in the subsequent discussions. More details can be found in the literature listed above.

## 2 Summary of Cokriging with Uncertain Mean and Trends

Consider an $n \times 1$ random space vector $\mathbf{s}$ representing an unknown discretized parameter field. Its distribution is $\mathbf{s} \sim \mathbf{N}(\mathbf{X}\boldsymbol{\beta}, \mathbf{C_{ss}})$, i.e., multi-Gaussian with mean vector $\mathbf{X}\boldsymbol{\beta}$ and covariance matrix $\mathbf{C_{ss}}$. $\mathbf{X}$ is an $n \times p$ matrix containing $p$ deterministic trend functions and $\boldsymbol{\beta}$ is the corresponding $p \times 1$ vector of trend coefficients. In the uncertain mean case, these coefficients are again random variables, distributed $\boldsymbol{\beta} \sim \mathbf{N}(\boldsymbol{\beta}^*, \mathbf{C}_{\boldsymbol{\beta}\boldsymbol{\beta}})$ (e.g., Kitanidis 1986; Handcock and Stein 1993). In the latter distribution, $\boldsymbol{\beta}^*$ defines the uncertain prior value, and $\mathbf{C}_{\boldsymbol{\beta}\boldsymbol{\beta}}$ defines the prior uncertainty about $\boldsymbol{\beta}^*$. While $\mathbf{s} \sim \mathbf{N}(\mathbf{X}\boldsymbol{\beta}, \mathbf{C_{ss}})$ holds for known values of $\boldsymbol{\beta}$, $\mathbf{s}$ for uncertain $\boldsymbol{\beta}$ follows the distribution $\mathbf{s} \sim \mathbf{N}(\mathbf{X}\boldsymbol{\beta}^*, \mathbf{G_{ss}})$ with

$$\mathbf{G_{ss}} = \mathbf{Q_{ss}} + \mathbf{X}\mathbf{C}_{\boldsymbol{\beta}\boldsymbol{\beta}}\mathbf{X}^T. \tag{1}$$

$\mathbf{G_{ss}}$ is the $n \times n$ generalized covariance matrix of $\mathbf{s}$ (Matheron 1971; Cressie 1991; Kitanidis 1993).

Now, consider the $m \times 1$ vector of error-prone observations $\mathbf{y}$ related to $\mathbf{s}$ via a linearized transfer function $\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{r}$, where $\mathbf{r} \sim \mathbf{N}(\mathbf{0}, \mathbf{R})$ is a vector of measurement errors. Measurement error is typically formalized as white noise with zero mean and the $m \times m$ covariance matrix $\mathbf{R}$. Then, for known $\mathbf{s}$, the measurements have the distribution $\mathbf{y}|\mathbf{s} \sim \mathbf{N}(\mathbf{H}\mathbf{s}, \mathbf{R})$. Linear error propagation yields the distribution of $\mathbf{y}$ for unknown $\mathbf{s}$ to be $\mathbf{y} \sim \mathbf{N}(\mathbf{H}\mathbf{X}\boldsymbol{\beta}^*, \mathbf{G_{yy}})$, where

$$\mathbf{G_{yy}} = \mathbf{H}\mathbf{G_{ss}}\mathbf{H}^T + \mathbf{R} \tag{2}$$

is the generalized covariance matrix of $\mathbf{y}$. Using Bayes theorem,

$$p(\mathbf{s}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{s})\,p(\mathbf{s})}{p(\mathbf{y})} \tag{3}$$

gives the distribution of the parameters $\mathbf{s}$ conditional on the measurements $\mathbf{y}$, $\mathbf{s}|\mathbf{y} \sim \mathbf{N}(\hat{\mathbf{s}}, \mathbf{G}_{\mathbf{ss}|\mathbf{y}})$. This distribution has a conditional mean $\hat{\mathbf{s}}$ and a conditional covariance $\mathbf{G}_{\mathbf{ss}|\mathbf{y}}$ given by

$$\begin{aligned} \hat{\mathbf{s}} &= \mathbf{X}\boldsymbol{\beta}^* + \mathbf{G}_{\mathbf{ss}}\mathbf{H}^T\boldsymbol{\xi}, \\ \boldsymbol{\xi} &= \mathbf{G}_{\mathbf{yy}}^{-1}(\mathbf{y} - \mathbf{H}\mathbf{X}\boldsymbol{\beta}^*), \\ \mathbf{G}_{\mathbf{ss}|\mathbf{y}} &= \left(\mathbf{G}_{\mathbf{ss}}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)^{-1} \\ &= \mathbf{G}_{\mathbf{ss}} - \mathbf{G}_{\mathbf{ss}}\mathbf{H}^T\mathbf{G}_{\mathbf{yy}}^{-1}\mathbf{H}\mathbf{G}_{\mathbf{ss}}. \end{aligned} \tag{4}$$

The latter equality is based on the often helpful identities for inverting partitioned matrices (e.g., Schweppe 1973). $\mathbf{G}_{\mathbf{ss}|\mathbf{y}}$ is again a generalized covariance because it accounts for the remaining uncertainty in both the fluctuations about the mean and the mean itself. This study will focus on measures that characterize the conditional covariance matrix $\mathbf{G}_{\mathbf{ss}|\mathbf{y}}$. Mostly, the second form of $\mathbf{G}_{\mathbf{ss}|\mathbf{y}}$ in (4) should be used because it requires no inversion of large $n \times n$ matrices, and because it is also defined in the special case of $\mathbf{R} = \mathbf{0}$. For highly efficient evaluation of all products involving $\mathbf{G}_{\mathbf{ss}}$ and $\mathbf{G}_{\mathbf{ss}|\mathbf{y}}$ in the stationary and several non-stationary cases, we refer to Nowak et al. (2003) and to Cirpka and Nowak (2004).

When tracing the normalizing factors of the involved probability density functions through Bayes theorem, one obtains for the determinants of the covariance matrices

$$\det(\mathbf{G}_{\mathbf{ss}|\mathbf{y}}) = \det\left(\mathbf{R}\mathbf{G}_{\mathbf{yy}}^{-1}\right)\det(\mathbf{G}_{\mathbf{ss}}). \tag{5}$$

Sometimes, the same estimator is also denoted in the form of $\hat{\mathbf{s}} = \mathbf{X}\widehat{\boldsymbol{\beta}} + \mathbf{C}_{\mathbf{ss}}\mathbf{H}^T\boldsymbol{\xi}$ where $\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\xi}$ are obtained by solving an $(m + p) \times (m + p)$ block matrix system (e.g., Kitanidis 1996a; Nowak and Cirpka 2004). The advantage of the generalized covariance notation introduced here is the formal identity to the known-mean case. The absence of symbols for estimating the mean $\mathbf{X}\widehat{\boldsymbol{\beta}}$ simplifies all subsequent derivations. In the limiting case of $\mathbf{C}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{-1} = 0$ (i.e., for entirely unknown mean and trends), $\mathbf{G}_{\mathbf{ss}}$ is not defined. Then, one has to revert to block matrix notation, but all statements based on the generalized notation issued in this study are still valid.

### 2.1 Transfer to Non-Gaussian and Non-linear Problems

The above equations are equivalent to the dual formulation (Galli et al. 1984) of universal kriging (Olea 1974), if $\mathbf{C}_{\boldsymbol{\beta}\boldsymbol{\beta}} = 0$ and if $\mathbf{H}$ is a sampling matrix that simply extracts the measurement-parameter cross-covariance from the parameter auto-covariance. The difference is that kriging does not require the assumption of multi-Gaussianity. It provides a best estimate (the conditional mean $\hat{\mathbf{s}}$) and an estimation variance (the main diagonal of $\mathbf{G}_{\mathbf{ss}|\mathbf{y}}$) regardless of the actual statistical distribution. Other than that, the results are identical. If the estimation problem is indeed linear, all results discussed later in this study will apply to non-Gaussian distributions without further modification, unless stated otherwise. For non-linear problems, (4) is only approximate, and various restrictions will apply. In the

weakly non-linear range (e.g., Mosegaard and Tarantola 2002), quasi-linear methods may be applied to obtain the conditional covariance (e.g., Kitanidis 1996b; Hughson and Yeh 2000) without further modifications. If the problem exceeds the weakly non-linear range, Monte Carlo techniques will mostly be required. Specific hardships will be highlighted in the discussion. For example, a respective measure may lose its original physical meaning.

## 3 Alphabetic Optimality in the Traditional Context

Because optimal design originates from regression-like applications, it seems appropriate to temporarily revert from spatial estimation to conventional regression-like problems. This is done by stripping off the geostatistical part from (1) to (5), i.e., by setting $\mathbf{C_{ss}} = \mathbf{0}$ and regarding $\boldsymbol{\beta}$ rather than $\mathbf{s}$ as primary unknowns. In the common notion of experimental design, information is defined by the Fisher information matrix

$$\mathbf{F}\big(\mathbf{u}|\mathbf{H}(\mathbf{u}), \mathbf{X}, \mathbf{R}\big) = E\left[\left(\frac{\partial}{\partial \boldsymbol{\beta}} \log p(\mathbf{y}|\boldsymbol{\beta})\right)\left(\frac{\partial}{\partial \boldsymbol{\beta}} \log p(\mathbf{y}|\boldsymbol{\beta})\right)^T\right], \tag{6}$$

where $\mathbf{u}$ is a chosen design formalized as a set of control or decision variables, $\mathbf{H}(\mathbf{u})$ indicates the dependence of $\mathbf{H}$ on the design $\mathbf{u}$, and all of $\mathbf{H}(\mathbf{u})$, $\mathbf{X}$ and $\mathbf{R}$ have to be known. $\mathbf{F}$ quantifies the information on the parameters $\boldsymbol{\beta}$ obtainable from the design $\mathbf{u}$. Under the assumptions of normality and linearity made in (1) through (4), $\mathbf{F}$ becomes

$$\mathbf{F}(\mathbf{u}, \cdot) = \mathbf{C}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{-1} + \mathbf{X}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{X}, \tag{7}$$

where non-Bayesian regression can be recovered by setting $\mathbf{C}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{-1}$ to the zero matrix and hence disregarding the prior. In the classical regression context, $\mathbf{HX}$ is replaced by an $m \times 1$ version of $\mathbf{X}$ because only direct measurements are considered, and it is $\mathbf{X}(\mathbf{u})$ rather than $\mathbf{H}(\mathbf{u})$ that depends on the design $\mathbf{u}$. Further, many publications assume homoscedasticity, substituting $\mathbf{R} = \sigma_r^2\mathbf{I}$.

The Cramer–Rao inequality identifies the inverse of $\mathbf{F}$ as a lower bound for the conditional covariance of the parameters, here denoted as $\mathbf{C}_{\boldsymbol{\beta}\boldsymbol{\beta}|\mathbf{y}}$. For linear regression and normal distributions, the relation is exact. In that case, $\mathbf{F}$ is also the precision matrix (i.e., the inverse of the covariance matrix) for the estimate of $\boldsymbol{\beta}$ and, at the same time, the moment matrix of the design specified by $\mathbf{HX}$. Thus, or alternatively derived from (1) through (4) with $\mathbf{C_{ss}} = \mathbf{0}$, $\mathbf{C}_{\boldsymbol{\beta}\boldsymbol{\beta}|\mathbf{y}}$ is given by

$$\begin{aligned}\mathbf{C}_{\boldsymbol{\beta}\boldsymbol{\beta}|\mathbf{y}}(\mathbf{u}, \cdot) &= \big(\mathbf{C}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{-1} + \mathbf{X}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{X}\big)^{-1} \\ &= \mathbf{C}_{\boldsymbol{\beta}\boldsymbol{\beta}} - \mathbf{C}_{\boldsymbol{\beta}\boldsymbol{\beta}}\mathbf{X}^T\mathbf{H}^T\mathbf{G}_{\mathbf{yy}}^{-1}\mathbf{H}\mathbf{X}\mathbf{C}_{\boldsymbol{\beta}\boldsymbol{\beta}},\end{aligned} \tag{8}$$

where $\mathbf{G_{yy}} = \mathbf{R} + \mathbf{HX}\mathbf{C}_{\boldsymbol{\beta}\boldsymbol{\beta}}\mathbf{X}^T\mathbf{H}^T$ under the current conditions. After agreeing to maximize information in the Fisherian sense, the only remaining question is in what sense $\mathbf{F}(\mathbf{u}, \cdot)$ should be maximized or, alternatively, in what sense $\mathbf{C}_{\boldsymbol{\beta}\boldsymbol{\beta}|\mathbf{y}}(\mathbf{u}, \cdot)$ should be

minimized (since matrix inversion preserves matrix ordering). In most derivations, alphabetic optimality criteria minimize corresponding measures of conditional uncertainty that are defined by context-specific functionals $\phi(\mathbf{C}_{\boldsymbol{\beta}\boldsymbol{\beta}|\mathbf{y}}(\mathbf{u}))$

*A-Optimality* requires the quadratic penalty function $(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_{\text{true}})^T \mathbf{A} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_{\text{true}})$ between the estimate and the unknown true value to be minimized. This leads to minimization of the measure

$$\phi_A = \text{trace}[\mathbf{A}\mathbf{C}_{\boldsymbol{\beta}\boldsymbol{\beta}|\mathbf{y}}], \tag{9}$$

where $\mathbf{A}$ is a problem-specific non-negative definite matrix. A special case is $\mathbf{A} = \mathbf{I}$, to which we assign the notation $A_I$ and $\phi_{AI}$ for later reference. The meaning of the $A$-criterion is not changed when normalizing it by $p$, which denotes the number of identified parameters in $\boldsymbol{\beta}$.

*C-Optimality* is based on a (linearly) dependent model prediction $z$ and requires its prediction variance to be minimal. Given the $p \times 1$ vector $\mathbf{c} = \partial z / \partial \boldsymbol{\beta}$, it minimizes the measure

$$\phi_C = \mathbf{c}^T \mathbf{C}_{\boldsymbol{\beta}\boldsymbol{\beta}|\mathbf{y}} \mathbf{c}. \tag{10}$$

This is a special case of $A$-optimality with $\text{rank}(\mathbf{A}) = 1$, $\mathbf{A} = \mathbf{c}\mathbf{c}^T$. $C$-optimality offers an incentive to rewrite the $A$-criterion for $\text{rank}(\mathbf{A}) > 1$ using $\mathbf{A} = (\partial \mathbf{z} / \partial \boldsymbol{\beta})(\partial \mathbf{z} / \partial \boldsymbol{\beta})^T$ for a vector $\mathbf{z}$ of quantities to be predicted. To this formulation, we assign the notation $A_C$ and $\phi_{AC}$ for later reference.

*D-Optimality* minimizes the measure

$$\phi_D = \det[\mathbf{C}_{\boldsymbol{\beta}\boldsymbol{\beta}|\mathbf{y}}] \tag{11}$$

and is derived by maximizing the expected gain in information entropy on $\boldsymbol{\beta}$. Its most intuitive significance is that it minimizes the hypervolume within any confidence hyper-ellipsoid of the parameters. $D_A$- and $D_S$-optimality do the same for linear combinations ($D_A$) and subsets ($D_S$) of $\boldsymbol{\beta}$, hence being the determinant-based equivalent of the $A_C$ criterion. The meaning of the $D$-criterion is not changed when taking the $p$th root. This is the most commonly used criterion in the statistical literature. If the underlying distributions are not Gaussian, the connection to entropy and confidence intervals is lost.

*E-Optimality* minimizes the measure

$$\phi_E = \max\left(\text{eig}[\mathbf{C}_{\boldsymbol{\beta}\boldsymbol{\beta}|\mathbf{y}}]\right) \tag{12}$$

and is derived from $C$-optimality with unknown $\mathbf{c}$. The motivation is to allow for a yet unknown predictive purpose of the experiment, then choosing the worst-case $\mathbf{c}$ subject to the normalization $\mathbf{c}^T \mathbf{c} = 1$. The worst-case $\mathbf{c}$ is the eigenvector of $\mathbf{C}_{\boldsymbol{\beta}\boldsymbol{\beta}|\mathbf{y}}$ that corresponds to the maximum eigenvalue.

*T-Optimality* minimizes the measure

$$\phi_T = \frac{1}{\text{trace}[\mathbf{M}]} \tag{13}$$

in which $\mathbf{M} = \mathbf{C}_{\boldsymbol{\beta}\boldsymbol{\beta}|\mathbf{y}}^{-1}$ is the precision matrix of the estimator, and the moment matrix of the design with $\mathbf{M} = \mathbf{X}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{X}$ in the non-Bayesian case. A basic drawback is that redundancy or mutual correlation among the measurements is not accounted for. Within this study, it is seen as a measure for the signal-to-noise ratio of measurements, where $\mathbf{X}^T\mathbf{H}^T[\ldots]\mathbf{H}\mathbf{X}$ is the squared $L_2$-norm of the signal and $\mathbf{R}$ is the squared $L_2$-norm of noise.

Further criteria (e.g., the $B$- and $G$-criteria) are used much less frequently and will not be considered in this study.

## 4 Relevant Differences Between Regression and Spatial Estimation

Quite obviously, regression-like problems are usually overdetermined ($m > p$), while spatial estimation problems are usually underdetermined ($n > m$) with very large $n$. This simple difference has two consequences which receive a major focus in the next section. First, the underdetermined character of spatial estimation may affect the significance of several measures when they are transferred to spatial estimation. The $n \times 1$ estimate $\hat{\mathbf{s}}$ has only $m < n$ degrees of freedom within the subspace spanned by $\mathbf{G}_{\mathbf{ss}}\mathbf{H}^T$, parametrized by the $m \times 1$ vector $\boldsymbol{\xi}$ (compare Nowak and Cirpka 2004, or see (4)). Equivalently, the conditional covariance matrix in (4) differs from the prior only within that very same subspace. The null space of $\mathbf{G}_{\mathbf{ss}}\mathbf{H}^T$ remains unaffected by conditioning. Measures of uncertainty that display extreme sensitivity either to this subspace or to the unaffected null space will have to be handled with care. Three limiting cases will provide insight into the behavior of questionable measures. First, completely unresolved variability for no or only uninformative measurements, i.e., $m \to 0$ or $\mathbf{R} \gg \mathbf{H}\mathbf{G}_{\mathbf{ss}}\mathbf{H}^T$. Second, a few but perfect measurements with $\mathbf{R} = \mathbf{0}$. And third, complete resolution of the parameter field (compare Tarantola 1987, p. 63).

Second, fine resolution in spatial estimation renders some alphabetic criteria computationally intractable. A general problem often encountered in geostatistical estimation is that the sheer size of covariance matrices may overstrain even powerful computers. For a grid size of 1000 cells along two dimensions, i.e., $n = 10^6$, a covariance matrix has $n^2 = 10^{12}$ elements and requires 8 TB (terabyte) of memory at double precision format. Such storage restrictions may become less relevant with future technological improvements, but the past has shown that problem sizes grow with the newly available computing resources. Some advantageous properties can be exploited though, allowing to handle stationary $\mathbf{C}_{\mathbf{ss}}$ (i.e., intrinsic $\mathbf{G}_{\mathbf{ss}}$) on regular grids via highly efficient methods (Zimmerman 1989). The reduced memory requirements forced to store only one row of the matrix, i.e., 8 MB instead of 8 TB in the above example. Extensions to specific non-stationary and conditional covariance matrices are provided in Cirpka and Nowak (2004). For further extension to irregular grids, see Pegram (2004) or Fritz et al. (2009). When showing computationally efficient ways of evaluation, links to respective existing methods will be provided.

## 5 Alphabetic Optimality Applied to Spatial Estimation

For geostatistical estimation, the alphabetic measures named above are applied to the $n \times n$ posterior covariance matrix $\mathbf{C}_{\mathbf{ss|y}}$ (e.g., Müller 2007) or, here, to the generalized version $\mathbf{G}_{\mathbf{ss|y}}$. In this section, the relevant differences between regression and spatial estimation are discussed for each individual alphabetic measure.

### 5.1 $A$-Optimality

The $A$-criterion translates to

$$\phi_A = \frac{1}{n}\text{trace}(\mathbf{A}\mathbf{G}_{\mathbf{ss|y}}). \tag{14}$$

Evaluating $\phi_A$ has only moderate computational costs since only the diagonal of $\mathbf{G}_{\mathbf{ss|y}}$ needs to be evaluated and summed up. Extremely efficient ways for evaluation are provided later in the context of the averaged estimation variance. The matrix trace is invariant under the similarity transform (e.g., Lang 2002, p. 511), and hence is the sum of all eigenvalues (e.g., Golub and van Loan 1996). This reveals $\phi_A$ to be the arithmetic mean of the eigenvalues of $\mathbf{A}\mathbf{G}_{\mathbf{ss|y}}$. For the $\phi_{AI}$ criterion ($\mathbf{A} = \mathbf{I}$), this yields

$$\phi_{AI} = \text{mean}\big(\text{eig}(\mathbf{G}_{\mathbf{ss|y}})\big). \tag{15}$$

It is bounded between zero and $\sigma_s^2$ (the prior variance), and assumes no special value for $\mathbf{R} = \mathbf{0}$.

The $A$-measure does not change its meaning for non-Gaussian or non-linear problems. For non-linear problems, the $A$-criterion must be evaluated via conditional simulation in a Monte Carlo framework.

### 5.2 $C$-Optimality

The $C$-criterion directly translates to

$$\phi_C = \mathbf{c}^T \mathbf{G}_{\mathbf{ss|y}} \mathbf{c}, \tag{16}$$

with $\mathbf{c} = \partial z / \partial \mathbf{s}$ being the sensitivity of a model prediction $z$ with respect to the estimated parameters, yielding the prediction variance for $z$. If more than a single quantity $z$ is to be predicted with $\mathbf{H}_\mathbf{z} = \partial \mathbf{z} / \partial \mathbf{s}$, the $A_C$-measure applies with

$$\phi_{AC} = \text{trace}\big[\mathbf{H}_\mathbf{z}^T \mathbf{G}_{\mathbf{ss|y}} \mathbf{H}_\mathbf{z}\big] = \text{trace}\big[\mathbf{H}_\mathbf{z} \mathbf{H}_\mathbf{z}^T \mathbf{G}_{\mathbf{ss|y}}\big] = \text{trace}[\mathbf{A}\mathbf{G}_{\mathbf{ss|y}}]. \tag{17}$$

An efficient method to evaluate $\phi_C$ uses the adjoint-state to obtain $\mathbf{c} = \partial z / \partial \mathbf{s}$ (e.g., Sykes et al. 1985; LaVenue and Pickens 1992; Sun and Yeh 1990a) and then evaluates $\phi_C$ in the following order:

$$\phi_C = \mathbf{c}^T \mathbf{d} - \mathbf{p}^T \mathbf{q}, \tag{18}$$

with $\mathbf{d} = \mathbf{G_{ss}c}$, $\mathbf{p} = \mathbf{Hd}$ and $\mathbf{q} = \mathbf{G_{yy}^{-1}p}$ (compare Cirpka et al. 2004; Cirpka and Nowak 2004). The convolution $\mathbf{d} = \mathbf{G_{ss}c}$ can be swiftly evaluated using the same methods based on the Fast Fourier Transform (FFT) as used for $\mathbf{G_{ss}H}^T$ (Cirpka et al. 2004; Nowak et al. 2003). As a special case, setting $\mathbf{c} = \mathbf{v}$ (an $n \times 1$ vector of unit entries) yields the prediction variance for the global mean value of $\mathbf{s}$. For non-Gaussian and non-linear problems, the same mild restrictions apply as for the $A$-measure.

### 5.3 $D$-Optimality

$D$-optimality applied to spatial estimation yields

$$\phi_D = \det(\mathbf{G_{ss|y}})^{\frac{1}{n}} = \operatorname{prod}\big(\operatorname{eig}(\mathbf{G_{ss|y}})\big)^{\frac{1}{n}}. \tag{19}$$

Normalization by the exponent of $1/n$ manifests the criterion as the geometric mean of the eigenvalues, while leaving the original properties of the criterion untouched. For numerical reasons, its logarithm (i.e., the sum of log-eigenvalues) might be considered for large matrices. The computational costs for $\phi_D$ in highly resolved fields are unbearable in the general case. Only for the stationary case discretized on a regular grid, Dietrich and Osborne (1996) offer an efficient method to evaluate the determinant of covariance matrices. When the measurements are uninformative, $\phi_D$ has a value of $\det(\mathbf{G_{ss}})^{1/n}$, i.e., $\phi_D = \sigma_s^2$ for the stationary regular-grid case. It is zero whenever one or more measurements are exact with $\det(\mathbf{R}) = 0$ (see Appendix A). This extreme sensitivity to the presence of a single highly informative measurement gives reason to doubt the practical use of this measure. For non-Gaussian or non-linear problems, the $D$-measure loses its direct connection to information entropy and to confidence intervals. For non-linear problems, Monte Carlo evaluation of covariance determinants or high-dimensional multivariate entropy has not been reported in the literature and will be subject of future studies.

### 5.4 $E$-Optimality

The measure based on $E$-optimality becomes

$$\phi_E = \max\big(\operatorname{eig}(\mathbf{G_{ss|y}})\big). \tag{20}$$

The eigenvalues of stationary covariance matrices are most graphically understood by their relation to the power spectrum in the Fourier space (e.g., Dietrich and Newsam 1993). Most covariance functions are dominated by the eigenvalues that correspond to low frequencies. From this perspective, the $E$-measure will mostly assess whether large-scale variability has been eliminated. Its practical utility needs to be considered in the respective context of individual applications. The computational costs of evaluating some largest eigenvalues is significantly smaller than those for all eigenvalues. Still, they will be rather unbearable for very large matrices, if the conditional covariance matrix fits into the computer memory at all. For non-Gaussian and non-linear cases, see $A$-criterion.

### 5.5 $T$-Optimality

The measure based on $T$-optimality is

$$\phi_T = \frac{1}{\text{trace}(\mathbf{G}_{\mathbf{ss}}^{-1}) + \text{trace}(\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})}. \tag{21}$$

Since the first term is a constant, it may as well be defined as

$$\phi_T^* = \frac{1}{\text{trace}(\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})} = \frac{1}{\text{trace}(\mathbf{H}\mathbf{H}^T \mathbf{R}^{-1})}. \tag{22}$$

This small modification shrinks the size of the matrices to be handled, yielding a computationally efficient measure. Also, the modification clarifies the character of an inverse signal-to-noise ratio of the measurements. Just like in the regression-like case, this measure is insensitive to redundancies or correlation among the measurements. The resulting inability to optimize measurement spacing is a severe drawback. For exact measurements with $\mathbf{R} \to \mathbf{0}$, this measure is useless due to $\phi_T \to 0$. For non-Gaussian and non-linear cases, see $A$-criterion.

### 5.6 $P$-Optimality

A generalization of $A$-, $D$-, $E$-, and $T$-optimality is

$$\phi_P = \left[\text{sum}\left(\text{eig}\left(\mathbf{G}_{\mathbf{ss}|\mathbf{y}}^P\right)\right)\right]^{\frac{1}{P}}, \tag{23}$$

where the special cases are recovered for $P = 1, 0, -\infty, -1$ in the alphabetic order listed above (compare Pukelsheim 2006, Chap. 6). This general measure seems to be practically useless for large problems. Explicit computation of all $n$ eigenvalues of $\mathbf{G}_{\mathbf{ss}|\mathbf{y}}$ is unfeasible, even for the stationary regular-grid case. For application to non-Gaussian and non-linear problems, see the respective special cases.

### 5.7 Relative $P$-Optimality

In order to obtain a measure independent of the magnitude of the prior covariance and with extremely reduced computational costs, measures relative to the prior covariance may be desirable. The author suggests to normalize the conditional covariance by the prior

$$\phi_{P,\text{rel}} = \left[\text{sum}\left(\text{eig}\left(\mathbf{G}_{\mathbf{ss}|\mathbf{y}}^P \mathbf{G}_{\mathbf{ss}}^{-P}\right)\right)\right]^{\frac{1}{P}}. \tag{24}$$

The special cases of $P = 1, 0, -\infty, -1$ result in the relative versions of the $A$-, $D$-, $E$-, and $T$-measures.

As shown in Appendix A, $n - m$ eigenvalues of the matrix $\mathbf{B} = \mathbf{G}_{\mathbf{ss}|\mathbf{y}} \mathbf{G}_{\mathbf{ss}}^{-1}$ are unity, and the $m$ other eigenvalues are those of the small $m \times m$ matrix $\mathbf{M} = {\mathbf{G}_{\mathbf{yy}}}^{-1} \mathbf{R}$. For the product of all eigenvalues, this can immediately be derived from (5). The simple

behavior of the eigenvalues allows to formulate the relative measure for any $P \in \mathbb{R}$ at small computational costs

$$\phi_{P,\text{rel}} = \left[ (n-m) + \sum_{i=1}^{m} \lambda_i^P(\mathbf{M}) \right]^{\frac{1}{P}}. \tag{25}$$

For all $P \in \mathbb{R}$, it approaches a value of unity when the measurements are uninformative, i.e., $\mathbf{R} \gg \mathbf{H}\mathbf{G}_{\mathbf{ss}}\mathbf{H}^T$ and hence $\mathbf{M} \to \mathbf{I}$. Its lower bound for the limit $\mathbf{R} \to \mathbf{0}$ is $(n-m)/n$ and 0 for $P = 1$ and $P = 0$, respectively. The case $P = -\infty$ (the relative $E$-measure) is mostly useless because $\phi_{P=1,\text{rel}} = 1$, unless there are $m \geq n$ measurements. The most interesting case will be $P = 0$, i.e., the relative $D$-measure

$$\phi_{D,\text{rel}} = \prod_{i=1}^{m} \lambda_i(\mathbf{M}). \tag{26}$$

As a side-product, an interesting related measure solely based on the measurement-related matrix $\mathbf{M} = {\mathbf{G}_{\mathbf{yy}}}^{-1}\mathbf{R}$ is

$$\phi^*_{AI,\text{rel}} = \frac{1}{m}\text{sum}\big(\text{eig}\big(\mathbf{G}_{\mathbf{yy}}^{-1}\mathbf{R}\big)\big). \tag{27}$$

The case $\phi^*_{AI,\text{rel}} = 1$ implies that the measurements are uninformative. $\mathbf{R} = 0$ is sufficient for $\phi^*_{AI,\text{rel}} = 0$, regardless of $m$ and $n$. It is a measure for the average noise-to-signal ratio of the measurements. However, in contrast to the $T$-measure, it does account for mutual correlation and redundancy of data. For non-Gaussian and non-linear problems, see the respective special cases for the original $P$-measure.

## 6 Other Measures and Their Relation to the Alphabet

### 6.1 Average Estimation Variance

The estimation variance is the conditional covariance for a separation distance of zero, defined by $\text{diag}(\mathbf{G}_{\mathbf{ss}|\mathbf{y}})$. A seeming ad-hoc measure for uncertainty is its spatial average (e.g., Diggle and Ribeiro 2007). Assuming a regular grid of the estimated parameters, it is

$$\bar{\sigma}^2_{\text{est}} = \text{mean}\big(\text{diag}(\mathbf{G}_{\mathbf{ss}|\mathbf{y}})\big) = \frac{1}{n}\text{trace}(\mathbf{G}_{\mathbf{ss}|\mathbf{y}}) = \phi_{AI}. \tag{28}$$

This unmasks $\bar{\sigma}^2_{\text{est}}$ as the $A_I$-criterion and hence provides a sound theoretical basis. Fritz, Nowak and Neuweiler (2009) showed an efficient order of evaluation that leads to

$$\frac{1}{n}\text{trace}\big(\mathbf{G}_{\mathbf{ss}|\mathbf{y}}\big) = \sigma_s^2 - \frac{1}{n}\text{trace}\big(\mathbf{G}^*_{\mathbf{yy}}\mathbf{G}^{-1}_{\mathbf{yy}}\big), \tag{29}$$

where $\mathbf{G}^*_{\mathbf{yy}} = \mathbf{H}\mathbf{G}_{\mathbf{ss}}\mathbf{G}_{\mathbf{ss}}\mathbf{H}^T$ is evaluated just like $\mathbf{G}_{\mathbf{yy}}$, only with a modified covariance function $\mathbf{G}^* = \mathbf{G}_{\mathbf{ss}}\mathbf{G}_{\mathbf{ss}}$, and $\sigma_s^2$ is the prior variance. Intuition may suggest that $\bar{\sigma}^2_{\text{est}}$ has limited information about the entire matrix because it merely represents the zero separation distance elements on the diagonal. The relation to $\phi_{AI}$, however, shows that $\bar{\sigma}^2_{\text{est}}$ is the average of all eigenvalues, and so quantifies the entire conditional uncertainty. For properties in non-Gaussian and non-linear cases, see the $A$-measure.

### 6.2 Empirical Covariance

The empirical covariance $C_{\text{es}}$ between a random synthetic field $\mathbf{s}_r$ and the corresponding estimated field $\hat{\mathbf{s}}$ is

$$C_{\text{es}} = \text{Cov}[\mathbf{s}_r, \hat{\mathbf{s}}] = \frac{1}{n}(\mathbf{s}_r - \mathbf{X}\boldsymbol{\beta}*)^T(\hat{\mathbf{s}} - \mathbf{X}\boldsymbol{\beta}*). \tag{30}$$

It is mostly applied graphically in scatter plots rather than by direct numerical evaluation. Various authors used it to assess the asymptotic exactness of data assimilation techniques in theoretical studies, or to assess the power of design proposals in synthetic test cases (e.g., Li and Yeh 1999; Zhu and Yeh 2005; Woodbury and Ulrych 2000). As shown in Appendix B, it is related to the mean estimation variance through its expected value

$$E[C_{\text{es}}] = \sigma_s^2 - \bar{\sigma}^2_{\text{est}} = \sigma_s^2 - \phi_{AI}. \tag{31}$$

$C_{\text{es}}$ is affected by the actual values used in the realization $\mathbf{s}_r$, while the mean estimation variance is not. This renders the mean estimation variance the statistically more significant criterion, unless the domain under consideration is large enough to be ergodic or the estimate $\hat{\mathbf{s}}$ and $C_{\text{es}}$ are evaluated multiple times for different $\mathbf{s}_r$ in a Monte Carlo framework. This measure is entirely unaffected by non-Gaussianity or non-linearity.

### 6.3 Empirical $L_2$-Measure

The $L_2$-norm of the difference between a synthetic field and its estimate, used by the same authors as the empirical covariance, is

$$L_2 = \text{MSE}[\mathbf{s}_r, \hat{\mathbf{s}}] = \frac{1}{n}(\hat{\mathbf{s}} - \mathbf{s}_r)^T(\hat{\mathbf{s}} - \mathbf{s}_r). \tag{32}$$

Its expected value is identical to $\bar{\sigma}^2_{\text{est}} = \phi_{AI}$ because (co-)kriging is a minimum variance estimator

$$E[L_2] = \bar{\sigma}^2_{\text{est}} = \phi_{AI}. \tag{33}$$

Therefore, similar to $C_{\text{es}}$, the same advantages and disadvantages apply.

### 6.4 Averaged Conditional Integral Scale

The above measures have quantified the magnitude of the conditional covariance. Besides the magnitude, the character in terms of a characteristic length scale of correlation may also be of interest. For stationary covariance functions $C_{ss}(h)$, the integral scale $\ell_s$ provides a measure for the characteristic length

$$\ell_s = \frac{1}{\sigma_s^2} \int_0^\infty C_{ss}(h)\, dh, \tag{34}$$

where $\sigma_s^2$ is the field variance. Its significance lies in the effect of variability on flow- and transport processes through heterogeneous porous media, where the integral scale often appears in analytical solutions (e.g., Rubin 2003). The current study suggests applying the same concept to the (non-stationary) conditional covariance, leading to an averaged conditional integral scale $\bar{\ell}_{\mathbf{ss|y}}$. The spatial averaging is required to overcome the non-stationarity induced by conditioning. The idea is similar to using the spatially averaged conditional variance $\bar{\sigma}^2_{\text{est}}$ as a measure. Cirpka and Nowak (2003) devised the stationary counterpart of the conditional covariance. They achieved stationarity by averaging the non-stationary conditional covariance matrix along its diagonals, i.e., among identical separation vectors. Since the order of summation (replacing the integral in (34)) and taking the averages along the diagonals (removing the non-stationarity) is irrelevant, (34) leads to

$$\bar{\ell}_{\mathbf{ss|y}} \approx \frac{1}{\Delta L n \bar{\sigma}^2_{\text{est}}} \mathbf{v}^T \mathbf{G}_{\mathbf{ss|y}} \mathbf{v} \propto \phi_C, \tag{35}$$

where $\mathbf{v}$ is an $n \times 1$ vector of ones and $\Delta L$ is the length of individual discrete section of $\mathbf{s}$.

$\bar{\ell}_{\mathbf{ss|y}}$ can be seen as the ratio between off-diagonal and diagonal entries of the conditional covariance matrix and therefore complements the $A$-measure which considers only the diagonal entries.

The above expression does not distinguish between fluctuations about the mean and uncertainty of the mean itself. The most meaningful results are obtained when not using the generalized covariance in this expression. The resulting value is only approximate due to geometrical boundary effects. These boundary effects cancel out when both the initial and the conditional integral scale are evaluated according to this equation for direct comparison.

In higher dimensional cases, the equation becomes (here: 3D)

$$\bar{\ell}^3_{\mathbf{ss|y}} \approx \frac{1}{\Delta V n \bar{\sigma}^2_{\text{est}}} \mathbf{v}^T \mathbf{G}_{\mathbf{ss|y}} \mathbf{v}, \tag{36}$$

and offers a characteristic volume instead of a characteristic length. Apparently, $\bar{\ell}_{\mathbf{ss|y}}$ is proportional to $\phi_C$ when setting $\mathbf{c} = \mathbf{v}$. The same efficient method for evaluation applies. The identity to $\phi_C$ underpins the new quantity $\bar{\ell}_{\mathbf{ss|y}}$ with a well-known theoretical basis to be found in the statistical literature. It also reveals $\bar{\ell}_{\mathbf{ss|y}}$ to be proportional to the variance of predicting the global spatial mean (not to be confused

with the true population mean). For non-Gaussian and non-linear cases, $\bar{\ell}_{\mathbf{ss|y}}$ is easily evaluated via the variance of the global spatial mean in a Monte Carlo framework. The meaning is not affected.

### 6.5 Entropy and Relative Entropy

Minimizing entropy or maximizing relative entropy (often called the Kullback–Leiber divergence in the context of optimal design or information theory) has been used to derive Bayesian $D$-optimality by some authors (e.g., Chaloner and Verdinelli 1995). Within the subsurface community, the concepts of entropy and relative entropy have received high attention, e.g., to select parametric distributions for hyperparameters (Woodbury and Ulrych 1993; Woodbury and Rubin 2000), to characterize parametric uncertainty de Barros and Rubin 2008, and to quantify the dilution of dissolved compounds (Kitanidis 1994). Therefore, the connection between entropy and the $D$-measure shall be highlighted. The $D$-measure represents the entropy of the conditional distribution $p(\mathbf{s}|\mathbf{y})$, which is a well-known result. The logarithm of the relative $D$-measure is the entropy difference between $p(\mathbf{s})$ and $p(\mathbf{s}|\mathbf{y})$ and the expected relative entropy of $p(\mathbf{s}|\mathbf{y})$ over $p(\mathbf{s})$. The relative and absolute $D$-measure yield equivalent design criteria, since they differ only by a constant. The required proofs are provided in Appendix C. The difference between these two criteria is that the relative measure is evaluated at speed because it is entirely based on the small $m \times m$ matrix $\mathbf{M} = \mathbf{G_{yy}}^{-1}\mathbf{R}$. The relation between the relative $D$-measure and relative entropy supplies a solid theoretical basis and a profound physical meaning. Due to its equivalence as design criterion, it may replace the computationally much more costly original $D$-measure. Still, it shares the same restrictions for non-Gaussian and non-linear problems as the original $D$-measure.

## 7 Illustration

For illustration, four design test cases are provided. The intention is to demonstrate how the different measures react to elementary changes in geostatistical design. A domain sized $1{,}000 \times 1{,}000$ meters is discretized on a regular grid with $500 \times 500$ cells ($n = 250{,}000$). Log-conductivity $Y = \log K$ is assumed to be multi-Gaussian, second-order stationary with isotropic exponential covariance function (variance $\sigma_Y^2 = 1$, correlation length $\lambda = 100$ m) and uncertain mean ($\beta^* = \log 10^{-5}$, $\sigma_\beta^2 = 1$). Synthetic conductivity fields and data sets are generated by random simulation. Available measurement types are direct measurements of $Y$ with a standard deviation of measurement error $\sigma_r = 0.5$ ($\sigma_r = 0$ in Case 2), and measurements of hydraulic heads $h$ with $\sigma_r = 0.01$ m. Hydraulic heads are defined by the confined depth-integrated groundwater flow equation with no-flow conditions on the north and south boundaries and fixed head boundary conditions of $h = 1$ and $h = 0$ m on the east and west boundaries. Conditioning on hydraulic head data is achieved by the quasi-linear geostatistical approach (Kitanidis 1995) with known covariance function, adjoint-state sensitivities (e.g., Sykes et al. 1985), FFT-based computation of cross- and auto-covariances (Nowak et al. 2003) and stabilized by a modified Levenberg–Marquardt

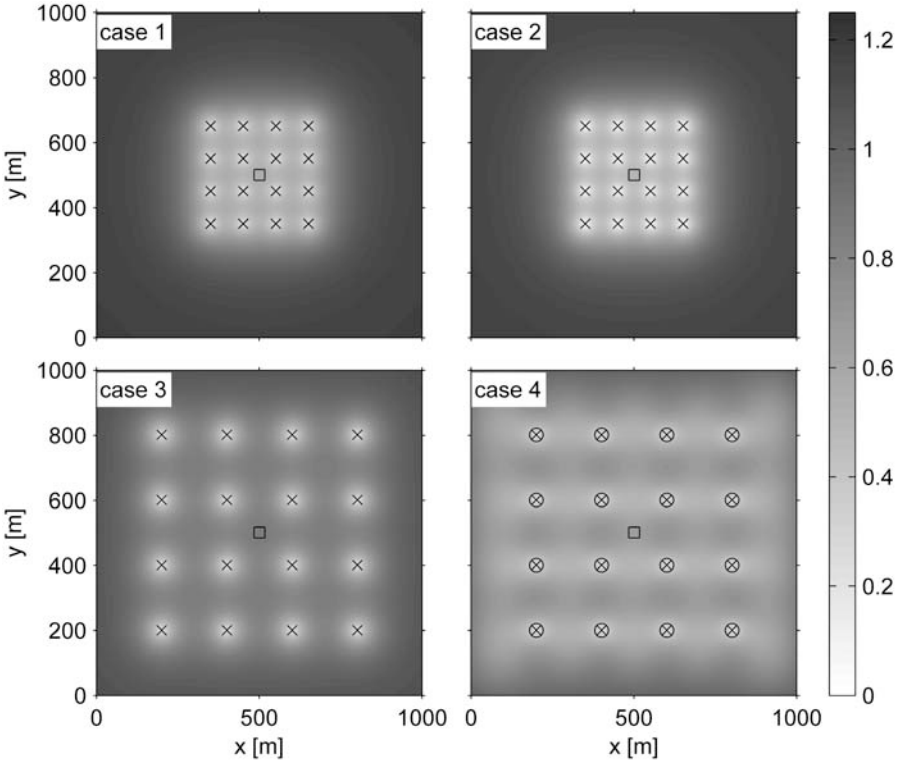


**Fig. 1** Estimation variances and measurement locations for four test cases. X-*marks*, measurements of log-conductivity $Y$; *circles*, measurements of hydraulic heads; *squares*, position of unknown hydraulic head used to define the $C$-measure. Case 2 features exact measurements of $Y$, while Case 1 features uncertain measurements

algorithm (Nowak and Cirpka 2004). Figure 1 shows the four different designs. Each design is evaluated using several of the measures featured in the current study. Table 1 provides the resulting values. The CPU times using the suggested methods for efficient computation are also provided. The $D$-, $E$- and $P$-measures are not considered because the involved $n \times n = 250{,}000 \times 250{,}000$ covariance matrices would require 500 GB of memory, preventing the computation of eigenvalues within a feasible time frame (estimated CPU time on the reference computer would be one month to compute the determinant). The empirical covariance and $L_2$ norm are not included because they react to changes in the design just like the $A$-measure. Depending on actual data values and the respective synthetic parameter field, a significant discussion of the covariance and the $L_2$ norm would have to consider their average values over many realizations and synthetic data sets.

The $A$-measure reacts to the four different design setups as desired, more precise measurements (Case 2) yield a smaller value than less precise ones (Case 1). An areal better coverage of the domain with less correlated measurements lead to a smaller parameter uncertainty (Case 3), and adding hydraulic head measurements as additional information leads to a further reduction of parameter uncertainty. The

**Table 1** Values of diverse measures for the test cases show in Fig. 1. Values for $\phi_A$, $\phi_C$ and $\bar{\ell}_{\mathbf{ss}|\mathbf{y}}$ are normalized by their values at the prior stage (in absence of all measurements). CPU time (milliseconds) evaluated in MATLAB 2006b on a Dual Core Machine at 2 GHz

| Case # | $\phi_A$ | $\phi_C$ | $\phi_T$ | $\phi_{D,\mathrm{rel}}$ | $\bar{\ell}_{\mathbf{ss}|\mathbf{y}}$ |
|---|---|---|---|---|---|
| 1 | 0.4987 | 0.6133 | 0.1963 | 0.9999 | 0.9095 |
| 2 | 0.4826 | 0.5591 | 0 | 0 | 0.9075 |
| 3 | 0.4283 | 0.4874 | 0.1963 | 0.9998 | 0.8051 |
| 4 | 0.3345 | 0.0526 | 0.1847 | 0.9997 | 1.0297 |
| CPU [ms] | 200 | 0.1 | 100 | 10 | 500 |

direct connection to the estimation variances plotted in Fig. 1 allows an intuitive understanding of this measure. The $C$-measure is more application-specific than the $A$-measure. Here, it is defined as the prediction variance of the hydraulic head value in the center of the domain. Hydraulic heads have a domain-wide dipole-like sensitivity function and cross-covariance to conductivity. Due to the global (sometimes called non-local) character, it reacts similarly to the $A$-measure in Cases 1 through 3. The strongest reduction of its value is achieved when measuring similar quantities (here neighboring head values) because similar measurements restrict the variance of similar components of the parameter field. The inadequacy of the $T$-measure is clearly demonstrated in the four cases. Different measurement spacing (Case 1 and Case 3) do not lead to a different value, although less correlated measurements convey more information. Perfect measurements with $\mathbf{R} = \mathbf{0}$ (Case 2) lead to a value of zero, although the system is not fully determined.

The relative $D$-measure does reflect the correlation between measurements and the amount of information they contain (Cases 1, 3 and 4). Its value hardly changes because out of $n = 250{,}000$ dimensions of the parameter space only $m = 16$ (Case 1 and 3) or $m = 32$ (Case 4) are restricted by the measurements. For precise measurements (Case 2), the measure assumes a value of zero. This is because the $D$-measure is the product of variances in the individual subspaces, and a single precise measurement overrules all remaining uncertainty. This harsh reaction to precise measurements is out of proportion to the remaining uncertainty. In presence of precise measurements, all alternative designs are $D$-equivalent and cannot be distinguished. The averaged conditional integral scale $\bar{\ell}_{\mathbf{ss}|\mathbf{y}}$ is a complementary measure to characterize the scale of remaining uncertainty. The most intuitive cases are Cases 1 and 3, in which sampling on a coarser mesh clearly suppresses uncertainties in longer-range components, leading to a smaller integral scale. The increased integral scale in Case 4 is non-trivial and cannot be understood intuitively. This behavior results from an interference between the dipole-character of hydraulic heads and the spectrum of the exponential covariance function. In general, measurements with a large support volume (i.e., with wide-range sensitivities) will lead to less uncertainty in large-scale components of the parameter field, and hence to smaller integral scales. Examples are hydraulic tomography (e.g., Yeh and Liu 2000; Li et al. 2005) or tracer tests (e.g., Nowak and Cirpka 2006).

## 8 Summary and Conclusions

This study reviewed measures for parameter uncertainty in spatial estimation which are taken from optimal design criteria. Their properties were illustrated in a synthetic case study. As a representative method for spatial estimation, cokriging in the generalized case of uncertain mean and trends was chosen. The design criteria were taken from the alphabetic optimality (the $A$-, $C$-, $D$-, $E$-, $T$-, and $P$-criteria). A specific peculiarity of conditional covariance matrices in geostatistical estimation is that they differ from the prior covariance matrix only in a subspace with rank limited to the number of measurements. This property is the key to swift and efficient evaluation of eigenvalues and helps drastically speed up some computations. Unfortunately, it also gives rise to some adverse properties of the reviewed measures that required further attention. When applying the optimality alphabet to geostatistical estimation, the computational effort of evaluating the associated measures of uncertainty plays a crucial role. Ways of swift evaluation for the $A$-, $C$- and $T$-measures were presented. The $D$- and generalized $P$-measure (and the $E$-measure to a lesser extent) were found to be not applicable in finely resolved cases. Discussion of their properties in the geostatistical context revealed that the $D$-measure is overly sensitive to single exact measurements. The $E$-measure is sensitive only to the largest contribution in the power spectrum of the prior covariance. The $T$-measure is insensitive to redundancy or correlation among the data. Similar measures relative to the prior covariance matrix have been defined in this study. It was shown that the involved eigenvalues are related to much smaller matrices and hence can be evaluated at speed. These relative measures are computationally much more efficient than their original versions, but mostly lack theoretical support from existing statistical literature.

While the original $D$-measure receives most attention outside the geostatistical area, computational costs rule it out for application to geostatistical problems. The current study suggests using the new relative $D$-measure instead, which can be computed in virtually no time. It was also shown to differ from the original $D$-criterion by only a constant. Its coincidence with relative entropy provides a sound theoretical basis and profound physical meaning. As a downside, the $D$-measure loses its attractive connection to information entropy in non-Gaussian or non-linear cases, as inherited from the original $D$-measure. Several rather intuitive measures of spatial uncertainty have also been reviewed. They find more frequent application in the subsurface, but less in the statistical community. These intuitive measures include the averaged estimation variance, and the empirical $L_2$-norm or covariance between synthetic fields and their corresponding estimates. Their evaluation is bare of significant computational effort. For non-linear problems, they are easily evaluated from Monte Carlo simulations. They maintain their original significance for non-Gaussian and non-linear problems. Most importantly, the averaged estimation variance is identical to the well-founded $A$-measure. $L_2$-norm and covariance are directly linked to the $A$-measure. The averaged conditional integral scale is a new measure complimentary to these intuitive measures. It quantifies how the character of uncertainty changes under conditioning. It is theoretically supported by its identity to a specific case of the $C$-measure, and holds for both non-linear and non-Gaussian cases.

The results of this study recommend the use of spatial average of the estimation variance as an all-purpose measure (i.e., unless the design task at hand suggests using

a different context-specific criterion). This is due to its intuitive understanding, its identity to the $A$-criterion and the associated theoretical support, due to its computational efficiency, and its validity for non-Gaussian and non-linear problems. Being the arithmetic mean of all eigenvalues of the conditional covariance matrix, it subsumes more information than an intuitive understanding of the zero-separation estimation variance distance may suggest. If desired, it can be escorted by the averaged conditional integral scale in order to characterize the changing character of uncertainty under conditioning. The averaged conditional integral scale shares the latter list of advantages. The above recommendation contradicts the fact that statistical studies commonly use the $D$-criterion. However, the new relative $D$-measure offers an attractive alternative since it represents relative entropy, differs from the original $D$-criterion only by a constant, and is evaluated at speed.

**Acknowledgements** This study is funded by the Deutsche Forschungsgemeinschaft (DFG) under grant No. 805-1/1. The author is indebted to Wolfgang Hein for his contributions to Appendix A, to Felipe de Barros and Olaf Cirpka for their productive comments on the manuscript, and to the anonymous reviewers.

## Appendix A

**Proposition 1** The matrix $\mathbf{B} = \mathbf{G_{ss|y}}\mathbf{G_{ss}}^{-1}$ has $n-m$ eigenvalues of unity and $m$ eigenvalues given by those of the $m \times m$ matrix $\mathbf{M} = \mathbf{G_{yy}}^{-1}\mathbf{R}$ for any non-negative finite $\mathbf{R}$.

*Proof* Set $\mathbf{U} = \mathbf{G_{ss}}\mathbf{H}^T$, $\mathbf{V} = \mathbf{G_{yy}}^{-1}\mathbf{H}$, $\mathbf{Q} = \mathbf{UV}$, and $\mathbf{S} = \mathbf{VU}$, where $\mathbf{U}$ is $n \times m$ and $\mathbf{V}$ is $m \times n$. It is easily checked that

$$\begin{aligned} \mathbf{B} &= \mathbf{I} - \mathbf{Q}, \\ \mathbf{M} &= \mathbf{I} - \mathbf{S}, \end{aligned} \tag{37}$$

and that $\mathbf{QU} = \mathbf{US}$ and $\mathbf{SV} = \mathbf{VQ}$. For given ranks $\mathrm{rank}(\mathbf{H}) = m_h \leq m$, $\mathrm{rank}(\mathbf{G_{ss}}) = n$ and $\mathrm{rank}(\mathbf{G_{yy}}^{-1}) = m$, it follows that $\mathrm{rank}(\mathbf{U}) = \mathrm{rank}(\mathbf{V}) \leq m$ and $\mathrm{rank}(\mathbf{Q}) \leq m$, so that the nullity of $\mathbf{Q}$ is $\mu(\mathbf{Q}) = n - \mathrm{rank}(\mathbf{Q}) \geq n - m$, and thus the multiplicity of the unit eigenvalue for $\mathbf{B} = \mathbf{I} - \mathbf{Q}$ is $\dim(\mathrm{eig}_1(\mathbf{B})) \geq n - m$ with equality for $\mathrm{rank}(\mathbf{H}) = m$.

Now let $\mathbf{x}$ be an eigenvector of $\mathbf{Q}$ with corresponding eigenvalue $\lambda \neq 0$. Due to $\mathbf{SV} = \mathbf{VQ}$, the same $\lambda$ is an eigenvalue of $\mathbf{S}$ with eigenvector $\mathbf{Vx}$:

$$\mathbf{Qx} = \lambda\mathbf{x} \quad \Longrightarrow \quad \mathbf{VQx} = \mathbf{SVx} = \lambda\mathbf{Vx}, \tag{38}$$

where, at the same time, $\mathbf{Vx} \neq \mathbf{0}$ because $\mathbf{Vx} = \mathbf{0} \Rightarrow \mathbf{UVx} = \mathbf{0} \Rightarrow \mathbf{Qx} = \mathbf{0}$ was excluded by choosing $\lambda \neq 0$. Since (38) holds for all $\lambda \neq 0$ from the spectrum $\lambda(\mathbf{Q})$, each non-zero eigenvalue of $\mathbf{S}$ is also an eigenvalue of $\mathbf{Q}$. Due to (37), each non-unit eigenvalue of $\mathbf{B} = \mathbf{I} - \mathbf{Q}$ is also an eigenvalue of $\mathbf{M} = \mathbf{I} - \mathbf{S}$, which completes the proof. □

*Remark 1* For the special case $\mathbf{R} = \mathbf{0}$, $\mathbf{Q}$ is idempotent with $\mathbf{Q}^k = \mathbf{Q}\ \forall k$, i.e., a projection matrix, and $\mathbf{V}$ is a reflexive generalized inverse of $\mathbf{U}$ as expected for a minimum-norm solution of (4). For the general case of $\mathbf{R} \neq \mathbf{0}$, $\mathbf{Q}$ is what might be called an incomplete projection matrix with general equation $\mathbf{Q}^k = \mathbf{US}^{(k-1)}\mathbf{V}$.

*Remark 2* For $\mathbf{R} = \mathbf{0}$ it follows that $\mathbf{M} = \mathbf{G}_{\mathbf{yy}}^{-1}\mathbf{R} = \mathbf{0}$, so that matrix $\mathbf{B}$ has $n - m$ eigenvalues of unity and $m$ eigenvalues of zero. For $\mathbf{R} \gg \mathbf{H}\mathbf{G}_{\mathbf{ss}}\mathbf{H}^T$ it follows that $\mathbf{S} \rightarrow \mathbf{0}$, and hence from (37) that $\mathbf{M} \rightarrow \mathbf{I}$, so that all eigenvalues of $\mathbf{B}$ are unity.

## Appendix B

**Proposition 2** $C_{\text{es}}$, as given by (30), has an expected value according to (31).

*Proof* Decompose the random vector $\mathbf{s}$ into

$$\begin{aligned} \mathbf{s} &= \mathbf{z} + \mathbf{X}\boldsymbol{\beta}^*, \\ E[\mathbf{z}] &= \mathbf{0}, \qquad \text{Cov}[\mathbf{z}] = \mathbf{G}_{\mathbf{ss}} \end{aligned} \tag{39}$$

and formulate the measurements in terms of this decomposition:

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{r} = \mathbf{H}\mathbf{z} + \mathbf{H}\mathbf{X}\boldsymbol{\beta}^* + \mathbf{r}. \tag{40}$$

Since $C_{\text{es}}$ is a scalar, it is identical to its trace. The trace is invariant for cyclic permutations, so that

$$C_{\text{es}} = \frac{1}{n}\text{Trace}\big[(\mathbf{s}_r - \mathbf{X}\boldsymbol{\beta}^*)(\hat{\mathbf{s}} - \mathbf{X}\boldsymbol{\beta}^*)^T\big]. \tag{41}$$

Next, we apply the expected value operator inside the trace, use (4), (39) and (40) and simplify to obtain

$$\begin{aligned} E[C_{\text{es}}] &= \frac{1}{n}\text{Trace}\big[E\big[(\mathbf{s}_r - \mathbf{X}\boldsymbol{\beta}^*)(\hat{\mathbf{s}} - \mathbf{X}\boldsymbol{\beta}^*)^T\big]\big] \\ &= \frac{1}{n}\text{Trace}\big[E\big[\mathbf{z}\mathbf{z}^T\big]\mathbf{H}^T\mathbf{G}_{\mathbf{yy}}{}^{-1}\mathbf{H}\mathbf{G}_{\mathbf{ss}}\big] \\ &= \frac{1}{n}\text{Trace}\big[\mathbf{G}_{\mathbf{ss}}\mathbf{H}^T\mathbf{G}_{\mathbf{yy}}{}^{-1}\mathbf{H}\mathbf{G}_{\mathbf{ss}}\big]. \end{aligned} \tag{42}$$

Comparison to (28) and $\mathbf{G}_{\mathbf{ss}|\mathbf{y}}$ in (4) completes the proof. □

## Appendix C

1. The entropy of a random variable $\mathbf{s}$ with distribution $p(\mathbf{s})$ is defined as

$$h(\mathbf{s}) = -\int_{-\infty}^{+\infty} p(\mathbf{s}) \log\big(p(\mathbf{s})\big)\, d\mathbf{s}. \tag{43}$$

For the Gaussian distributions at hand, this yields

$$\begin{aligned} h(\mathbf{s}) &= \frac{n}{2} + \frac{n}{2}\log(2\pi) + \log\big(\det(\mathbf{G}_{\mathbf{ss}})\big), \\ h(\mathbf{s}|\mathbf{y}) &= \frac{n}{2} + \frac{n}{2}\log(2\pi) + \log\big(\det(\mathbf{G}_{\mathbf{ss}|\mathbf{y}})\big). \end{aligned} \tag{44}$$

Requiring minimum entropy $h(\mathbf{s}|\mathbf{y})$ (or maximum information) leads to minimizing $\phi_D$.

2. (a) The difference of entropy $\Delta h = h(\mathbf{s}|\mathbf{y}) - h(\mathbf{s})$ simplifies to the log of the relative $D$-measure (26):

$$\Delta h(\mathbf{s}|\mathbf{y}, \mathbf{s}) = \log\bigl(\det\bigl(\mathbf{G}_{\mathbf{ss}|\mathbf{y}}\mathbf{G}_{\mathbf{ss}}^{-1}\bigr)\bigr) = \log\phi_{D,\mathrm{rel}}. \tag{45}$$

(b) The relative entropy of a distribution $q(\mathbf{s})$ over $p(\mathbf{s})$ is defined as

$$h_{\mathrm{rel}}(q, p) = -\int_{-\infty}^{+\infty} q(\mathbf{s}) \log\left(\frac{q(\mathbf{s})}{p(\mathbf{s})}\right) d\mathbf{s}. \tag{46}$$

Applying this to posterior $p(\mathbf{s}|\mathbf{y})$ over $p(\mathbf{s})$ requires some rearrangement before reaching the expression

$$h_{\mathrm{rel}}(\mathbf{s}|\mathbf{y}, \mathbf{s}) = \Delta h(\mathbf{s}|\mathbf{y}, \mathbf{s}) - \frac{1}{2}\mathrm{trace}\bigl(\mathbf{G}_{\mathbf{ss}}\mathbf{H}^T\bigl(\mathbf{G}_{\mathbf{yy}}^{-1} - \boldsymbol{\xi}\boldsymbol{\xi}^T\bigr)\mathbf{H}\bigr), \tag{47}$$

where $\boldsymbol{\xi}$ contains the cokriging weights. This is still a function of the actual data values in $\mathbf{y}$ via $\boldsymbol{\xi}$. Taking its expected value over the distribution $p(\mathbf{y})$ with $E[\boldsymbol{\xi}\boldsymbol{\xi}^T] = \mathbf{G}_{\mathbf{yy}}^{-1}$ again leads to the relative $D$-measure:

$$E\bigl[h_{\mathrm{rel}}(\mathbf{s}|\mathbf{y}, \mathbf{s})\bigr] = \Delta h(\mathbf{s}|\mathbf{y}, \mathbf{s}) = \log\phi_{D,\mathrm{rel}}. \tag{48}$$

3. Since $h(\mathbf{s})$ is a constant for a given prior, the relative and the absolute $D$-measure yield equivalent design criteria. This follows directly from comparison of (45) and (47) and the fact that $h(\mathbf{s})$ is a constant for a given problem.

## References


Bakr MI, te Stroet CBM, Meijerink A (2003) Stochastic groundwater quality management: role of spatial variability and conditioning. Water Resour Res 39(4). doi:10.1029/2002WR001745

Box GEP (1982) Choice of response surface design and alphabetic optimality. Util Math 21:11–55

Chaloner K, Verdinelli I (1995) Bayesian experimental design: a review. Stat Sci 10(3):273–304

Cirpka OA, Nowak W (2003) Dispersion on kriged hydraulic conductivity fields. Water Resour Res 39(2). doi:10.1029/2001WR000598

Cirpka OA, Nowak W (2004) First-order variance of travel time in non-stationary formations. Water Resour Res 40. doi:10.1029/2003WR002851

Cirpka OA, Bürger CM, Nowak W, Finkel M (2004) Uncertainty and data worth analysis for the hydraulic design of funnel-and-gate systems in heterogeneous aquifers. Water Resour Res 40(W11502). doi:10.1029/2004WR003352

Cressie NAC (1991) Statistics for spatial data. Wiley, New York

de Barros FPJ, Rubin Y (2008) A risk-driven approach for subsurface site characterization. Water Resour Res 44(W01414). doi:10.1029/2007WR006081

de Marsily G (1986) Quantitative hydrology. Academic Press, San Diego

Dietrich CR, Newsam GN (1993) A fast and exact method for multidimensional Gaussian stochastic simulations. Water Resour Res 29(8):2861–2869

Dietrich CR, Osborne MR (1996) $o(n \log^2 n)$ determinant computation of a Toeplitz matrix and fast variance estimation. Appl Math Lett 9(2):29–31

Diggle PJ, Ribeiro PJ (2007) Model-based geostatistics. Springer series in statistics. Springer, New York

Federov V, Hackl P (1997) Model-oriented design of experiments. Springer, New York

Feyen L, Gorelick SM (2005) Framework to evaluate the worth of hydraulic conductivity data for optimal groundwater resources management in ecologically sensitive areas. Water Resour Res 41(W03019). doi:10.1029/2003WR002901
Fritz J, Nowak W, Neuweiler I (2009) Application of FFT-based algorithms for large-scale universal kriging problems. Math Geosci 41(5):509–533. doi:10.1007/s11004-009-9220-x
Galli A, Murillo E, Thomann J (1984) Dual kriging, its properties and its uses in direct contouring. In: Verly G et al (eds) Geostatistics for natural resources characterization, part 2. Reidel, Dordrecht, pp 621–634
Golub GH, van Loan CF (1996) Matrix computations, 3rd edn. John Hopkins Press, Baltimore
Handcock MS, Stein ML (1993) A Bayesian analysis of kriging. Technometrics 35(4):403–410
Herrera GS, Pinder GF (2005) Space–time optimization of groundwater quality sampling networks. Water Resour Res 41(W12407). doi:10.1029/2004WR003626
Hughson D, Yeh T-CJ (2000) An inverse model for three-dimensional flow in variably saturated porous media. Water Resour Res 36(4):829–840
Journel AG, Huijbregts CJ (1978) Mining geostatistics. Academic Press, San Diego
Keidser A, Rosbjerg D (1991) A comparison of four inverse approaches to groundwater flow and transport parameter identification. Water Resour Res 27(9):2219–2232
Kitanidis PK (1986) Parameter uncertainty in estimation of spatial functions: Bayesian analysis. Water Resour Res 22(4):499–507
Kitanidis PK (1993) Generalized covariance functions in estimation. Math Geol 25(5):525–540
Kitanidis PK (1994) The concept of the dilution index. Water Resour Res 30(7):2011–2026
Kitanidis PK (1995) Quasi-linear geostatistical theory for inversing. Water Resour Res 31(10):2411–2419
Kitanidis PK (1996a) Analytical expressions of conditional mean, covariance, and sample functions in geostatistics. Stoch Hydrol Hydraul 12:279–294
Kitanidis PK (1996b) On the geostatistical approach to the inverse problem. Adv Wat Resour 19(6):333–342
Kitanidis PK (1997) Introduction to geostatistics. Cambridge University Press, Cambridge
Lang S (2002) Algebra. graduate texts in mathematics, 3rd edn. Springer, New York,
LaVenue AM, Pickens JF (1992) Application of a coupled adjoint sensitivity and Kriging approach to calibrate a groundwater flow model. Water Resour Res 28(6):1543–1559
Li B, Yeh T-CJ (1999) Cokriging estimation of the conductivity field under variably saturated flow conditions. Water Resour Res 35(12):3663–3674
Li W, Nowak W, Cirpka OA (2005) Geostatistical inversing of transient pumping tests using temporal moments of drawdown. Water Resour Res 41(W08403). doi:10.1029/2004WR003874
Matheron G (1971) The theory of regionalized variables and its applications. Ecole de Mines, Fontainebleau, France
McPhee J, Yeh WW-G (2006) Experimental design for groundwater modeling and management. Water Resour Res 42(W02408). doi:10.1029/2005WR003997
Mosegaard K, Tarantola A (2002) Probabilistic approaches to inverse problems. International handbook of earthquake and engineering seismology, vol 81A, pp 237–265
Müller WG (2007) Collecting spatial data. Optimum design of experiments for random fields, 3rd edn. Springer, Berlin
Nowak W, Cirpka OA (2004) A modified Levenberg–Marquardt algorithm for quasi-linear geostatistical inversing. Adv Water Resour 27(7):737–750
Nowak W, Cirpka OA (2006) Geostatistical inference of conductivity and dispersion coefficients from hydraulic heads and tracer data. Water Resour Res 42(W08416). doi:10.1029/2005WR004832
Nowak W, Tenkleve S, Cirpka OA (2003) Efficient computation of linearized cross-covariance and auto-covariance matrices of interdependent quantities. Math Geol 35(1):53–66
Olea RA (1974) Optimal contour mapping using universal kriging. J Geophys Res 79(5):695–702
Pappenberger F, Beven KJ (2006) Ignorance is bliss: Or seven reasons not to use uncertainty analysis. Water Resour Res 42(W05302). doi:10.1029/2005WR004820
Pegram GGS (2004) Spatial interpolation and mapping of rainfall (SIMAR) vol 3: Data merging for rainfall map production. Water Research Commission Report (1153/1/04)
Pukelsheim F (2006) Optimal design of experiments. Classics in applied mathematics. SIAM, Philadelphia. Classic edition
Rubin Y (2003) Applied stochastic hydrogeology. Oxford University Press, London
Schweppe FC (1973) Uncertain dynamic systems. Prentice-Hall, New York
Silvey S (1980) Optimal designs. Chapman & Hall, London

Sun N-Z (1994) Inverse problems in groundwater modeling. Theory and applications of transport in porous media. Kluwer Academic, Dordrecht

Sun N-Z, Yeh WW-G (1990a) Coupled inverse problems in groundwater modeling. 1. Sensitivity analysis and parameter identification. Water Resour Res 26(10):2507–2525

Sun N-Z, Yeh WW-G (1990b) Coupled inverse problems in groundwater modeling. 2. Identifiability and experimental design. Water Resour Res 26(10):2527–2540

Sykes JF, Wilson JL, Andrews RW (1985) Sensitivity analysis for steady-state groundwater-flow using adjoint operators. Water Resour Res 21(3):359–371

Tarantola A (1987) Inverse problem theory. Elsevier, Amsterdam

Woodbury AD, Rubin Y (2000) A full-Bayesian approach to parameter inference from tracer travel time moments and investigation of scale effects at the Cape Cod experimental site. Water Resour Res 36(1):159–171

Woodbury AD, Ulrych TJ (1993) Minimum relative entropy: forward probabilistic modeling. Water Resour Res 29(8):2847–2860

Woodbury AD, Ulrych TJ (2000) A full-Bayesian approach to the groundwater inverse problem for steady state flow. Water Resour Res 36(8):2081–2093

Yeh T-CJ, Liu S (2000) Hydraulic tomography: development of a new aquifer test method. Water Resour Res 36(8):2095–2105

Zhu J, Yeh T-CJ (2005) Characterization of aquifer heterogeneity using transient hydraulic tomographs. Water Resour Res 41(W07028). doi:10.1029/2004WR003790

Zimmerman DL (1989) Computationally exploitable structure of covariance matrices and generalized covariance matrices in spatial models. J Stat Comput Simul 32(1/2):1–15

Zimmerman DA, de Marsily G, Gotway CA, Marietta MG, Axness CL, Beauheim RL, Bras RL, Carrera J, Dagan G, Davies PB, Gallegos DP, Galli A, Gomez-Hernandez J, Grindrod P, Gutjahr AL, Kitanidis PK, Lavenue AM, McLaughlin D, Neuman SP, RamaRao BS, Ravenne C, Rubin Y (1998) A comparison of seven geostatistically based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow. Water Resour Res 34(6):1373–1413