Bayesian Geostatistical Design: Task-Driven Optimal Site Investigation When the Geostatistical Model is Uncertain

W. Nowak

Department of Civil and Environmental Engineering, University of California, Berkeley, USA.

Institute of Hydraulic Engineering / LH^2, University of Stuttgart, Germany

F. P. J. de Barros and Y. Rubin

Department of Civil and Environmental Engineering, University of California, Berkeley, USA

W. Nowak, Department of Civil and Environmental Engineering, University of California, Berkeley, 626 Davis Hall, Berkeley, CA, 94720-1710, USA.

Now at the Institute of Hydraulic Engineering / LH^2 (SimTech), Pfaffenwaldring 61, University of Stuttgart, 70569 Stuttgart, Germany (wolfgang.nowak@iws.uni-stuttgart.de)

F. P. J. de Barros, Department of Civil and Environmental Engineering, University of California, Berkeley, 626 Davis Hall, Berkeley, CA, 94720-1710, USA. (barros@berkeley.edu)

Yoram Rubin, Department of Civil and Environmental Engineering, University of California, Berkeley, 627 Davis Hall, Berkeley, CA, 94720-1710, USA. (rubin@ce.berkeley.edu)

Abstract. Geostatistical optimal design optimizes subsurface exploration for maximum information towards task-specific prediction goals. Until recently, most geostatistical design studies have assumed that the geostatistical description (i.e., the mean, trends, covariance models and their parameters) is given a priori. This contradicts, as emphasized in *Rubin and Dagan* [1987b], the fact that only few or even no data at all offer support for such assumptions prior to the bulk of exploration effort. We believe that geostatistical design should (1) avoid unjustified a priori assumptions on the geostatistical description, (2) instead reduce geostatistical model uncertainty as secondary design objective, (3) rate this secondary objective optimal for the overall prediction goal and (4) be robust even under inaccurate geostatistical assumptions. Bayesian Geostatistical Design [*Diggle and Lophaven*, 2006] follows these guidelines by considering uncertain covariance model parameters. We transfer this concept from kriging-like applications to geostatistical inverse problems. We also deem it inappropriate to consider parametric uncertainty only within a single covariance model. The Matérn family of covariance functions has an additional shape parameter. Controlling model shape by a parameter converts covariance model selection to parameter identification and resembles Bayesian model averaging over a continuous spectrum of covariance models. This is appealing since it generalizes Bayesian model averaging from a finite number to an infinite number of models. We illustrate how our approach fulfills the above four guidelines in a series of synthetic test cases. The underlying scenarios are to minimize the prediction vari-

ance of (1) contaminant concentration or (2) arrival time at an ecologically sensitive location by optimal placement of hydraulic head and log-conductivity measurements. Results highlight how both the impact of geostatistical model uncertainty and the sampling network design vary according to the choice of objective function.

1. Introduction

Scarcity of data and subsurface variability lead to the understanding of hydraulic conductivity as a random space function [e.g., *Journel and Huijbregts*, 1978; *de Marsily*, 1986; *Kitanidis*, 1997; *Rubin*, 2003]. This acknowledges the uncertainty in flow and transport models stemming from unresolved heterogeneity of aquifer parameters. Adopting the model-based geostatistical approach [e.g., *Diggle and Ribeiro*, 2007], the random space function is defined by the global mean value, trend coefficients, and parameters in for covariance models often called structural parameters.

Incorporating hydrogeological measurements (e.g., conductivity, flow and tracer data) helps to reduce the involved uncertainties. Two types of information are required: (1) hydrogeological measurements and (2) the underlying geostatistical model to interpolate between unsampled positions. Given limited financial resources, this information need has to be satisfied in an efficient manner via geostatistical optimal design (see *Massmann and Freeze* [1987]; *Freeze et al.* [1990]; *James and Gorelick* [1994]; *Herrera and Pinder* [2005] for applications in groundwater hydrology). Optimal design finds sampling schemes that maximize the expected gain of information, measured in various ways. The importance of setting task-oriented objectives, for example risk-driven approaches, is outlined by *Maxwell et al.* [1999]; *de Barros and Rubin* [2008]; *de Barros et al.* [2009].

Most of these studies presume prefect prior knowledge on the structural parameters. In realistic scenarios, however, such strong a priori assumptions are hard to justify. Structural parameters tend to be poorly identifiable, especially from data sets limited in size and accuracy. *Pardo-Iguzquiza* [1999] illustrated the inadequacy of point estimates for

structural parameters in synthetic case studies. The principle of Bayesian geostatistics [*Kitanidis*, 1986; *Feinerman et al.*, 1986] acknowledges this fact and, instead, treats structural parameters as yet another set of random variables. Uncertain structural parameters ("structural uncertainty") increase the uncertainty of model predictions (such as contaminant levels or fluxes) because they have a substantial influence on macroscopic flow, plume dilution and dispersion [e.g., *Rubin*, 2003], and covariance shape has a high impact on prediction uncertainty [*Riva and Willmann*, 2009]. Conditional simulation and geostatistical inverse modeling with structural uncertainty are provided by *Rubin and Dagan* [1987b, 1992a]; *Woodbury and Ulrych* [2000]; *Pardo-Iguzquiza and Chica-Olmo* [2008]. We believe that geostatistical optimal design should fulfill the following four guidelines:

1. The objective of optimal design is to minimize uncertainty predictions. Structural uncertainty contributes to the overall prediction uncertainty, and hence must be assessed and accounted for.

2. The potential of the planned data collection to reduce structural uncertainty must be considered in finding the optimal design.

3. Estimating structural parameters should be "*treated as a means to the primary end of spatial* [or hydrogeological] *prediction, rather than as an end in itself*" [*Diggle and Lophaven*, 2006]. This asks for an optimal resource allocation between collecting spatial and structural information.

4. Optimal design patterns are sensitive to structural parameter values and should be made robust towards structural parameters [e.g., *Christakos*, 1992, p. 438].

Our main concern is that structural uncertainty has to be accounted for in geostatistical optimal design. Although structural uncertainty is all the more relevant when setting

out to plan site investigation prior to data collection, it has hardly been recognized in geostatistical optimal design studies. Only few optimal design studies in hydrogeology [e.g., *Criminisi et al.*, 1997] analyzed the issue of robustness. *Müller* [2007, chapter 3] included only uncertain trend parameters (no uncertain covariance) into geostatistical design.

Most geostatistical design studies serve either the collection of spatial information or the identification of structural parameters. The former requires coverage of certain areas of the domain with samples, while the latter requires sampling certain lag distances. These seemingly contradictory objectives have sometimes been combined in multi-objective optimization [e.g., *Müller*, 2007, pp. 173].

Diggle and Lophaven [2006] introduced the concept of Bayesian Geostatistical Design, which accommodates for uncertainty in covariance parameters within the design procedure in a most natural manner. They featured the averaged kriging variance as objective function and limited their study to direct measurements of the estimated quantity. The more recent work by *Marchant and Lark* [2007b] introduced a first-order approximation for the influence of structural parameters on the kriging variance. Similar approximations [e.g., *Zimmerman*, 2006], are summarized by *Müller* [2007, pp. 178].

Our study may claim, to the best knowledge of the authors, to be the transfer and first-time application of Bayesian Geostatistical Design to geostatistical inverse problems. In Section 2, we extend the Bayesian Geostatistical Design framework to measurements of dependent state variables (such as hydraulic heads) and the prediction uncertainty of yet other state variables (such as future solute concentrations).

Moreover, we wish to become more independent of arbitrarily chosen model shapes of covariance functions in geostatistical optimal design. *Neuman* [2003] stressed that the choice of geostatistical models will always be uncertain, and could be accounted for by Bayesian Model Averaging [*Hoeting et al.*, 1999]. In our work, we opt for the Matérn family of covariance functions [*Matérn*, 1986] because it has an additional shape parameter. *Feyen et al.* [2003] mentioned briefly that this shape parameter could be used to represent uncertainty in the shape of covariances. Following their rationale, we utilize the parametric control on the Matérn covariance shape to transform the model selection problem to a stochastic parameter inference problem. This approach resembles Bayesian Model Averaging over a continuous spectrum of models. Details and the relation to recent literature are provided in Section 3.

We illustrate the resulting optimal design framework in a synthetic case study. Technical details of an exemplary implementation are provided in Section 4. In Sections 5 and 6, we optimize sampling strategies (conductivity and head data) for predicting (1) future contaminant levels and (2) arrival times at an ecologically sensitive location. On that basis, we demonstrate and discuss the fulfillment of our four suggested guidelines. It is important to stress that neither is Bayesian Optimal Design limited to our implementational choice nor is it restricted to our exemplary choice of unknown parameters, data types and the assumption of multi-Gaussianity taken in our test case.

2. Bayesian Geostatistical Design

2.1. Model-Based Bayesian Geostatistics

Consider $\mathbf{s}$ a $n_s \times 1$ random space vector $\mathbf{s} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_\mathbf{s}$ (e.g., log-conductivity discretized on a numerical grid), with a trend model $E[\mathbf{s}] = \mathbf{X}\boldsymbol{\beta}$ plus zero-mean fluctuations $\boldsymbol{\varepsilon}_\mathbf{s}$.

$\mathbf{X}$ is a $n_s \times p$ matrix containing p deterministic trend functions with p corresponding trend coefficients $\boldsymbol{\beta}$. $\boldsymbol{\theta}$ are structural parameters in the distribution of $\boldsymbol{\varepsilon}_\mathbf{s}$ (e.g., such as scale and variance parameters of a covariance function, so that $\boldsymbol{\varepsilon}_\mathbf{s}$ has a covariance matrix $\mathbf{C} = \mathbf{C}(\boldsymbol{\theta})$).

Conventional model-based geostatistics [*Diggle and Ribeiro*, 2002] consider known structural parameters, and the distribution of $\mathbf{s}$ is $p(\mathbf{s}|\boldsymbol{\beta}, \boldsymbol{\theta})$. Bayesian geostatistics reflect the uncertainty of structural parameters by their joint distribution $p(\boldsymbol{\beta}, \boldsymbol{\theta})$. This is in contrast to classical variogram analysis [e.g., *Matheron*, 1971] and maximum likelihood estimation methods [e.g., *Schweppe*, 1973; *Kitanidis*, 1995]. The Bayesian distribution (marked by a tilde) is obtained by marginalization [e.g., *Kitanidis*, 1986]:

$$\tilde{p}(\mathbf{s}) = \int_{\boldsymbol{\beta}} \int_{\boldsymbol{\theta}} p(\mathbf{s}|\boldsymbol{\beta}, \boldsymbol{\theta})\, p(\boldsymbol{\beta}, \boldsymbol{\theta})\, d\boldsymbol{\theta}\, d\boldsymbol{\beta}\,. \tag{1}$$

Now consider the $n_y \times 1$ vector $\mathbf{y}$ of measurements at locations $\mathbf{x}_m$ according to $\mathbf{y} = \mathbf{f}_y(\mathbf{s}) + \boldsymbol{\varepsilon}_\mathbf{r}$. Here, $\mathbf{f}_y(\mathbf{s})$ is a process model (e.g., the groundwater flow equation) that relates observable variables (e.g., hydraulic heads) to $\mathbf{s}$. $\boldsymbol{\varepsilon}_\mathbf{r}$ is a vector of random measurement errors with known distribution $p(\varepsilon_\mathbf{r})$. According to Bayes theorem, the distribution of $\mathbf{s}$ conditional on a given measurement vector $\mathbf{y}_o$ is:

$$p(\mathbf{s}|\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{y}_o) \propto p(\mathbf{y}_o|\mathbf{s})\, p(\mathbf{s}|\boldsymbol{\beta}, \boldsymbol{\theta})\,. \tag{2}$$

Again, the Bayesian distribution is obtained by marginalization:

$$\tilde{p}(\mathbf{s}|\mathbf{y}_o) = \int_{\boldsymbol{\beta}} \int_{\boldsymbol{\theta}} p(\mathbf{s}|\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{y}_o)\, p(\boldsymbol{\beta}, \boldsymbol{\theta}|\mathbf{y}_o)\, d\boldsymbol{\theta}\, d\boldsymbol{\beta}\,. \tag{3}$$

Note that the entire distribution $p(\mathbf{s}, \boldsymbol{\beta}, \boldsymbol{\theta})$ has been jointly conditioned on $\mathbf{y}_o$ (see *Kitanidis* [1986]; *Pardo-Iguzquiza* [1999]; *Woodbury and Ulrych* [2000]; *Diggle and Ribeiro* [2002]).

The final purpose is the prediction of yet a different variable c (e.g., concentration), related to $\mathbf{s}$ via $c = f_c(\mathbf{s})$ (e.g., the transport equation):

$$\tilde{p}(c|\mathbf{y}_o) \propto \int_{\boldsymbol{\beta}} \int_{\boldsymbol{\theta}} \int_{\mathbf{s}} p(c|\mathbf{s})\, p(\mathbf{s}|\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{y}_o)\, p(\boldsymbol{\beta}, \boldsymbol{\theta}|\mathbf{y}_o)\, d\mathbf{s}\, d\boldsymbol{\theta}\, d\boldsymbol{\beta} \tag{4}$$

with Bayesian mean $\tilde{c}$ and increased variance $\tilde{\sigma}^2_{c|\mathbf{y}}$

$$\tilde{c}(\mathbf{y}_o) = E_{\boldsymbol{\beta},\boldsymbol{\theta}|\mathbf{y}_o}\left[E_{\mathbf{s}}\left[f_c(\mathbf{s})|\mathbf{y}_o, \boldsymbol{\beta}, \boldsymbol{\theta}\right]\right] \tag{5}$$

$$\begin{aligned}\tilde{\sigma}^2_{c|\mathbf{y}}(\mathbf{y}_o) &= E_{\boldsymbol{\beta},\boldsymbol{\theta}|\mathbf{y}_o}\left[V_{\mathbf{s}}\left[f_c(\mathbf{s})|\mathbf{y}_o, \boldsymbol{\beta}, \boldsymbol{\theta}\right]\right] \\ &\quad + V_{\boldsymbol{\beta},\boldsymbol{\theta}|\mathbf{y}_o}\left[E_{\mathbf{s}}\left[f_c(\mathbf{s})|\mathbf{y}_o, \boldsymbol{\beta}, \boldsymbol{\theta}\right]\right], \end{aligned} \tag{6}$$

where $E_a[\cdot]$ is the expected value operator over the distribution of a random variable a, and $V_a[\cdot]$ is the respective variance.

2.2. Optimal Design

Optimal design theory originated from the context of linear and non-linear regression *Silvey* [e.g., 1980]; *Box* [e.g., 1982]; *Federov and Hackl* [e.g., 1997]; *Pukelsheim* [e.g., 2006]. Application to geostatistics is explained by *Uciński* [2005]; *Müller* [2007]; *Nowak* [2009a]. A design is a set of decision variables $\mathbf{d}$ that specify the number, location and types of measurements to be collected in the data vector $\mathbf{y}$. The objective is to minimize the uncertainty inherent in the predictive distributions $p(\mathbf{s}|\mathbf{y}_o)$ or $p(c|\mathbf{y}_o)$, before even knowing the actual data values $\mathbf{y}_o$. To this end, a task-specific measure of prediction uncertainty $\phi(\mathbf{d}, p)$ is defined [e.g., *Müller*, 2007; *Nowak*, 2009a] and minimized. For Bayesian Geostatistical Design [*Diggle and Lophaven*, 2006], these distributions are simply replaced by their Bayesian counterparts $\tilde{p}(\mathbf{s}|\mathbf{y}_o)$ or $\tilde{p}(c|\mathbf{y}_o)$ (Eqs. 3 and 4):

$$\phi(\mathbf{d}, \tilde{p}) = E_{\mathbf{y}}\left[\phi(\mathbf{y}(\mathbf{d}), \tilde{p})\right] = \int \phi(\mathbf{y}(\mathbf{d}), \tilde{p})\, \tilde{p}(\mathbf{y})\, d\mathbf{y}\,. \tag{7}$$

Eq. (7) implicitly includes averaging over all possible values of the structural parameters, because:

$$\tilde{p}\left(\mathbf{y}\right) = \int_{\boldsymbol{\beta}} \int_{\boldsymbol{\theta}} p\left(\mathbf{y}|\boldsymbol{\theta},\boldsymbol{\beta}\right) p\left(\boldsymbol{\theta},\boldsymbol{\beta}\right)\, d\boldsymbol{\theta}\, d\boldsymbol{\beta}\,. \tag{8}$$

In our illustrative test case (but not as a limitation of the general framework), we will choose to minimize the expected Bayesian prediction variance of c:

$$\begin{aligned} E_{\mathbf{y}}\left[\tilde{\sigma}^2_{c|\mathbf{y}}\right] \;=\; & E_{\mathbf{y}}\left\{E_{\boldsymbol{\beta},\boldsymbol{\theta}|\mathbf{y}}\left[V_{\mathbf{s}|\mathbf{y}(\mathbf{d}),\boldsymbol{\beta},\boldsymbol{\theta}}\left[f_c\left(\mathbf{s}\right)\right]\right]\right\} \\ & + E_{\mathbf{y}}\left\{V_{\boldsymbol{\beta},\boldsymbol{\theta}|\mathbf{y}}\left[E_{\mathbf{s}|\mathbf{y}(\mathbf{d}),\boldsymbol{\beta},\boldsymbol{\theta}}\left[f_c\left(\mathbf{s}\right)\right]\right]\right\}\,. \end{aligned} \tag{9}$$

It is important to highlight the two individual contributions to overall prediction uncertainty in the right-hand-side of Eq. (9): The first term resembles the prediction variance of concentration, averaged over all possible values of potential data and structural parameters. The second term reflects how the estimate of concentration varies due to the uncertainty of structural parameters. These two terms result directly from Bayesian principles and weight the objectives of interpolation and structural identification in a natural manner. Note that this form of the minimum variance design criterion complies with all four of our guidelines stated in the introduction. More details on the fulfillment of our four guidelines are provided in Sections 6 and 7.

The second term vanishes only at the theoretical (and mostly unrealistic) limit of known structural parameters. For that case, Eq. (9) degenerates to the C-criterion for geostatistical optimal design [e.g., *Müller*, 2007; *Nowak*, 2009a] with exemplary applications by *Cirpka et al.* [2004] and *Herrera and Pinder* [2005].

3. Continuous Bayesian Model Averaging and the Matérn Family of Covariance Functions

Bayesian model averaging [*Hoeting et al.*, 1999] considers several model alternatives and assigns prior probabilities to each of them, reflecting their respective credibility level. The modeling task is performed with all model alternatives, and posterior credibilities are assigned after comparison with available data. The final result is the ensemble of model outcomes, each one weighted by its posterior credibility. The overwhelming advantage is the increased robustness towards errors in individual conceptual models or in model selection.

This principle can be applied to geostatistical model selection [e.g., *Neuman*, 2003]. One could pick an arbitrary choice from the entire list of traditional parametric covariance models, and then proceed with Bayesian Model Averaging. We believe, however, that the choice of model alternatives should not be restricted by traditional adherence to a small set of mathematically preferred covariance models.

Instead, we recommend a more elegant approach based on the Matérn family of covariance functions [*Matérn*, 1986]. The works of *Handcock and Stein* [1993]; *Diggle and Ribeiro* [2002]; *Marchant and Lark* [2007a] suggest to use the flexibility of the Matérn family in order to include uncertainty in covariance shape and smoothness into geostatistical inversion The Matérn function is given by:

$$C(\ell) = \frac{\sigma_Y^2}{2^{\kappa-1}\Gamma(\kappa)} \left(2\sqrt{\kappa}\ell\right)^{\kappa} B_{\kappa}\left(2\sqrt{\kappa}\ell\right)$$
$$\ell = \sqrt{\left(\frac{\Delta x_1}{\lambda_1}\right)^2 + \left(\frac{\Delta x_2}{\lambda_2}\right)^2 \ldots}, \tag{10}$$

where σ_Y^2 is the variance of log-conductivity, ℓ is the anisotropic effective separation distance, and $\kappa \geq 0$ is an additional shape parameter. $\Gamma(\cdot)$ is the Gamma function, and

$B_k(\cdot)$ is the modified Bessel function of the third kind (Bessel's k) of order κ [*Abramowitz and Stegun*, 1972, section 10.2]. ℓ has λ_i as scale parameters for each spatial dimension. In the form provided here, ℓ is scaled by a factor $2\sqrt{\kappa}$ to make the integral scale roughly independent of κ [e.g., *Handcock and Stein*, 1993]. For the specific values of $\kappa = 0.5$, 1, ∞, the Matérn family simplifies to the exponential, Whittle and Gaussian covariance models, respectively (Figure 1). The combination of $\kappa = 1$ and $\lambda \to \infty$ approximates a power-law covariance [*Minasny and McBratney*, 2005], and arbitrary constructions with other models are allowed. More details on properties and specific additional advantages of the Matérn family are provided by *Stein* [1999].

The novelty of this approach is the following: If one treats κ as a discrete random variable to resemble model selection, one arrives back at the principle of Bayesian Model Averaging. However, we suggest to keep κ a continuous parameter on the positive real line, introducing a continuous spectrum of model alternatives. We then simply include κ in the vector $\boldsymbol{\theta}$ and treat it no different than the other uncertain structural parameters. This way, we convert the problem of model selection to a problem of stochastic parameter inference, embedded in the Bayesian approach, with a long list of available methods to draw from. We refer to this approach as Continuous Bayesian Model Averaging.

4. Implementational Choices for the Illustrative Test Case

In the present section, we provide a computationally efficient first-order approximation of structural uncertainty applied to a multi-Gaussian geological description. First-order approaches are computationally very efficient and useful (within their range of validity). Examples are adjoint-state sensitivities [*Cirpka et al.*, 2004] in conjunction with FFT-based error propagation [*Nowak et al.*, 2003] or the static Ensemble Kalman Filter

in *Herrera and Pinder* [2005]. We are aware of the limitations present in the multi-Gaussianity assumption. However, avoiding it would prohibit any linear approximation and would cause a substantial increase in computational costs. Note that Bayesian Geostatistical Design is not restricted to any of the choices and approximations taken in the following.

4.1. Multi-Gaussian First-Order Second-Moment Approximation

We model log-conductivity as a multi-Gaussian vector $\mathbf{s}$ of discrete cell-wise values with $\mathbf{s}|\boldsymbol{\beta},\boldsymbol{\theta} \sim \mathbf{N}\left(\mathbf{X}\boldsymbol{\beta}, \mathbf{C}_{\mathbf{ss}}\left(\boldsymbol{\theta}\right)\right)$, i.e., with mean vector $\mathbf{X}\boldsymbol{\beta}$ and covariance matrix $\mathbf{C}_{\mathbf{ss}}\left(\boldsymbol{\theta}\right)$. In the generalized intrinsic case, uncertain $\boldsymbol{\beta}$ is absorbed in the distribution of $\mathbf{s}$. We assume a Gaussian prior distribution $\boldsymbol{\beta} \sim \mathbf{N}\left(\boldsymbol{\beta}^*, \mathbf{C}_{\boldsymbol{\beta}\boldsymbol{\beta}}\right)$ with expected value $\boldsymbol{\beta}^*$ and covariance $\mathbf{C}_{\boldsymbol{\beta}\boldsymbol{\beta}}$. By assuming $\boldsymbol{\beta}$ multi-Gaussian and independent of $\boldsymbol{\theta}$, we can integrate over $p\left(\boldsymbol{\beta}\right)$ in Eqs. (3) to (6) analytically [*Kitanidis*, 1986]: $\mathbf{s}|\boldsymbol{\theta} \sim \mathbf{N}\left(\mathbf{X}\boldsymbol{\beta}^*, \mathbf{G}_{\mathbf{ss}}\left(\boldsymbol{\theta}\right)\right)$, where $\mathbf{G}_{\mathbf{ss}} = \mathbf{C}_{\mathbf{ss}}\left(\boldsymbol{\theta}\right) + \mathbf{X}\mathbf{C}_{\boldsymbol{\beta}\boldsymbol{\beta}}\mathbf{X}^T$ is a generalized covariance matrix [*Kitanidis*, 1993]. This approach has already proven useful to generalize geostatistical inversion [*Nowak and Cirpka*, 2004; *Fritz et al.*, 2009]. The individual steps of linearizing $\mathbf{f}_y\left(\mathbf{s}\right)$ and $f_c\left(\mathbf{s}\right)$ in $\mathbf{s}$ are summarized in Appendix A, leading to:

$$E_{\mathbf{y}}\left[\tilde{\sigma}^2_{c|\mathbf{y}}\right] \;=\; E_{\boldsymbol{\theta}}\left[\sigma^2_{c|\mathbf{y}}\left(\boldsymbol{\theta}\right)\right] + E_{\mathbf{y}}\left\{V_{\boldsymbol{\theta}|\mathbf{y}}\left[\hat{c}\left(\mathbf{y}\left(\mathbf{d}\right),\boldsymbol{\theta}\right)\right]\right\}\,, \tag{11}$$

where $\sigma^2_{c|\mathbf{y}}\left(\boldsymbol{\theta}\right)$ is the conditional variance of concentration for given $\boldsymbol{\theta}$.

To further simplify Eq. (11), we expand in $\boldsymbol{\theta}$ about its prior mean value $\bar{\boldsymbol{\theta}}$, truncate after first order, and assume a prior covariance $\mathbf{C}_{\boldsymbol{\theta}\boldsymbol{\theta}}$ to specify the structural uncertainty, similar to *Rubin and Dagan* [1987a]. After executing $E_{\mathbf{y}}\left\{\cdot\right\}$, we obtain:

$$E_{\mathbf{y}}\left[\tilde{\sigma}^2_{c|\mathbf{y}}\left(\mathbf{d}\right)\right] \;=\; \sigma^2_{c|\mathbf{y}}\left(\bar{\boldsymbol{\theta}}\right) + \sum_i \sum_j \left\langle \mathbf{C}_{\boldsymbol{\theta}\boldsymbol{\theta}|\mathbf{y}} \right\rangle_{ij} \Bigg\{ \cdots$$

$$\cdots \left.\frac{\partial \bar{c}}{\partial \theta_i}\right|_{\bar{\theta}_i} \left.\frac{\partial \bar{c}}{\partial \theta_j}\right|_{\bar{\theta}_j} + \left(\left.\frac{\partial \boldsymbol{\kappa}}{\partial \theta_i}\right|_{\bar{\theta}_i}\right) \mathbf{G}_{\mathbf{yy}}\left(\bar{\boldsymbol{\theta}}\right) \left(\left.\frac{\partial \boldsymbol{\kappa}}{\partial \theta_j}\right|_{\bar{\theta}_j}\right)^T \Bigg\} , \tag{12}$$

where $\boldsymbol{\kappa} = \mathbf{H}_c \mathbf{G}_{\mathbf{ss}}\left(\boldsymbol{\theta}\right) \mathbf{H}^T \mathbf{G}_{\mathbf{yy}}^{-1}\left(\boldsymbol{\theta}\right)$ is the Kalman gain of concentration in Eq. (B3), $\bar{c} = E_{\mathbf{s}}\left[c\right]$, and $\bar{\boldsymbol{\theta}} = E_{\boldsymbol{\theta}}\left[\boldsymbol{\theta}\right]$. $\left\langle \mathbf{C}_{\boldsymbol{\theta}\boldsymbol{\theta}|\mathbf{y}} \right\rangle_{ij}$ is the i,j-the element in the conditional covariance of $\boldsymbol{\theta}$, here approximated by the inverse of the Fisher information $\mathbf{F}$. Eq. (12) is a linearized version of Eq. (9), similar to what *Marchant and Lark* [2007b] found in the simpler kriging-like design context. Details of the derivation are provided in Appendix B.

Once actual data values become available after the optimal design task, we can update the structural parameters with the technique by *Kitanidis and Lane* [1985] and *Kitanidis* [1995], later upgraded to the generalized intrinsic case by *Nowak and Cirpka* [2006]. The conditional covariance of $\boldsymbol{\theta}$ is again approximated by the inverse of $\mathbf{F}$, and the conditional mean $\widehat{\boldsymbol{\theta}}$ is approximated by

$$\widehat{\boldsymbol{\theta}} \approx \bar{\boldsymbol{\theta}} - \mathbf{F}^{-1}\mathbf{g} \tag{13}$$

where $\mathbf{g}$ is the gradient, $\mathbf{F}$ is the Fisher information matrix as specified in Appendix B and $\mathbf{C}_{\boldsymbol{\theta}\boldsymbol{\theta}|\mathbf{y}} \approx \mathbf{F}^{-1}$.

4.2. Implementation

Eq. (12) and the equations in the appendices merely require auto- and cross-covariances between data and predicted variables, and their derivatives with respect to the structural parameters. We entrust this task to the static Ensemble Kalman Filter (sEnKF) by *Herrera* [1998], and obtain the derivatives with respect to $\boldsymbol{\theta}$ from additional parallel sEnKF's with slightly different parameter values. Ensemble Kalman Filters are based on a certain type of optimal linearization [*Nowak*, 2009b] that outmatches traditional first-order expansions in accuracy, adequately represent dispersion and dilution of solute

transport, and hence avoid the non-trivial choice of dispersion coefficients when using estimated conductivity fields [*Rubin et al.*, 1999; *Nowak and Cirpka*, 2006]. Once the design is decided upon and the data become available, we condition the log-conductivity field by the Kalman Ensemble Generator (KEG) [*Nowak*, 2009b], which is an adaptation of the EnKF to geostatistical inversion. The main steps of the analysis are:

1. Find a near-optimal design using the techniques described in Section 2 and 4;

2. Generate a synthetic data set for the suggested design by unconditional random simulation of a synthetic "true" aquifer (see section 5.5).

3. With the synthetic data, compute the conditional ensemble statistics (e.g. mean and variance for concentration and arrival times) using KEG;

4. Analyze the results for compliance with our four guidelines (see Section 7).

The above framework is implemented in MATLAB. For groundwater flow, solute transport and random field generation, we use the same codes as in *Nowak et al.* [2008]. Each EnKF ensemble had a size of 4000 realizations, which is more than sufficient for Ensemble Kalman Filters in hydrogeological applications [*Chen and Zhang*, 2006]. We optimize our sampling patterns using the sequential exchange algorithm [e.g., *Christakos*, 1992, p. 411]. Both the first-order approximation and the chosen optimization algorithm allow only to obtain so-called "near-optimal" designs (compare with *Janssen et al.* [2008]).

5. Synthetic Case Study

5.1. Scenario Definition and Relevance

Consider a potential future groundwater contamination at an ecologically sensitive location due to a hypothetical upstream groundwater contamination as part of a risk scenario.

This type of scenario is relevant, for example, in the probabilistic assessment of human health risk, where hydrogeological data acquisition helps to reduce the uncertainty in risk estimates [*Rubin et al.*, 1994; *Maxwell et al.*, 1999; *de Barros and Rubin*, 2008; *de Barros et al.*, 2009]. We will follow two different prediction objectives: to minimize the prediction variance of (1) steady-state contaminant concentration after continuous release (same scenario as *McKinney and Loucks* [1992]) and (2) contaminant arrival time at the sensitive location.

We chose two objectives, because different objectives can yield fundamentally different design patterns. The actual choice of design objectives (and also multi-objective design, see *Müller* [2007]) of course depends on the specific modeling and management goals at the site under consideration. Our main point of using these two different objectives is to illustrate the role of structural uncertainty varies under different objectives.

We will place 24 boreholes to obtain both core-scale measurements of transmissivity (e.g., from slug tests or disturbed-core grain-size analysis) and additional co-located measurements of hydraulic head (e.g., from minimum-cost groundwater level monitoring wells at the cored locations), with their respective measurement errors provided in Table 1. To demonstrate the effect of structural uncertainty, we compare the results between (a) known and (b) uncertain structural parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ in the geostatistical model. Combined with our two prediction objectives, this yields four different cases (1a, 1b, 2a and 2b; see Table 2).

5.2. Flow and Transport Configuration

For simplicity we limit our solute transport problem to the steady-state concentration and the arrival time down-gradient of a continuous line source in a depth-integrated

2D setting, and consider a point-like sensitive location. Depth-integrated steady-state groundwater flow is described by:

$$\nabla \cdot [T(\mathbf{x}) \nabla h] = 0\,, \tag{14}$$

where $T\,[L^2/t]$ is locally isotropic transmissivity and $h\,[L]$ is hydraulic head. The space coordinates are represented by $\mathbf{x} = (x_1, x_2)$. Boundary conditions are specified later. For the steady-state transport we use:

$$\mathbf{v} \cdot \nabla c - \nabla \cdot (\mathbf{D}_d \nabla c) = 0\,, \tag{15}$$

where $c\,[M/L^3]$ is concentration, $\mathbf{v} = \mathbf{q}/n_e$ is velocity, $\mathbf{q}$ is the Darcy specific flux, n_e is porosity, and $\mathbf{D}_d\,[L^2/t]$ is the pore-scale-dispersion tensor. We simulate the arrival time t_{50} using moment-generating equations [*Harvey and Gorelick*, 1995]:

$$\mathbf{v} \cdot \nabla m_k - \nabla \cdot (\mathbf{D}_d \nabla m_k) = k m_{k-1}\,, \tag{16}$$

with $t_{50} = m_1/m_0$, where m_0 and m_1 are the zeroth and first temporal moments of breakthrough for the related instantaneous release problem [e.g., *Cirpka and Kitanidis*, 2000; *Cirpka and Nowak*, 2004; *Nowak and Cirpka*, 2006].

The domain geometry, contaminant source and the sensitive location are illustrated in Figure (2). Relevant parameter values are provided in Table 1. Boundary conditions are $\hat{h} = 1m$ and $\hat{h} = 0m$ at $x_1 = 0m$ and $x_1 = 600m$, respectively. Uncontaminated groundwater with $\hat{c} = 0mg/\ell$ enters at $x_1 = 0m$, and the outflow boundary at $x_1 = 600m$ is unrestricted. The remaining two boundaries at $x_2 = 0m$ and $x_2 = 200m$ are no-flux boundaries for both flow and transport. We consider a fixed-concentration source with unit concentration $c_0 = 1$ along a $50m$ ($\sim$ 3 1/3 integral scales) wide line centered at $x_1 = 150m$. A sensitive location is located at a longitudinal travel distance of $300m$ ($\sim$

20 integral scales) down-gradient from the source, located at x_2 =118.75m (12.5m or $\sim$ 5/6 integral scales offset from the center line).

5.3. Bayesian Geostatistical Setup and Test Cases

Predicting contaminant transport over some distance in heterogeneous formations requires assumptions on the structure of variability. For illustration, we assume a stationary model in cases 1a and 2a, and an intrinsic model (due to a trend model with uncertain coefficients) in cases 1b and 2b, see Table 2. Cases 1b and 2b are less arbitrary and less subjective in their prior model assumptions: Following the Bayesian rationale, they do not claim to deterministically know the global mean, trend or the covariance function in absence of information, i.e., prior to design and data collection. The remaining assumptions chosen for our illustration are that a single, domain-wide, intrinsic and multi-Gaussian geostatistical model applies. Less parsimonic descriptions with varying covariances, or even more complex multi-variate dependence, could be adopted if deemed necessary. This would increase the number of structural parameters and result in different sampling patterns. We generate random log-transmissivity fields using the Matérn family of covariances plus a global mean and a linear trend. The two linear trend functions have a spatial mean of zero and cause a total variation of ± 0.5 over the respective length of the domain.

Only for cases 1a and 2a, the structural parameters are considered known. For cases 1b and 2b, their values are uncertain, with squared coefficients of variation $CV^2 = 0.5$ for covariance parameters and unity variance for mean and trend parameters. Uncertain parameters are the global mean value β_1, the trend coefficients in x_1 and x_2 directions (β_2 and β_3, respectively), the variance σ_Y^2, the scale parameters in x_1 and x_2 directions (λ_1 and λ_2, respectively), and the Matérn shape parameter κ. Their prior mean values and

variances are specified in Table 1. For simplicity, we assume prior stochastic independence among the structural parameters.

5.4. Effect of Structural Uncertainty on Prediction Mean and Variance

Figures 2 and 3 compare prior mean values and standard deviations of Y, h, c and t_{50} for the case of known and uncertain structural parameters, respectively. They are obtained from Monte-Carlo analysis with 16000 realizations each, using the geostatistical settings for cases 1 and 2 described in Section 5.3.

The impact of uncertain mean and trend manifests in the form of a prior standard deviation of log-conductivity with values larger than $\sigma_Y = 1$ in the center of the domain, with increasing values towards the domain boundaries (see Figure 3). The standard deviation of h for uncertain structure is dominated by the uncertain trend in x_1 direction and by the given head boundaries [*Rubin and Dagan*, 1988].

Structural uncertainty also affects the standard deviation of concentration, compare Figures 2 and 3 and analytical expressions [e.g., *Fiorotto and Caroni*, 2002; *Caroni and Fiorotto*, 2005; *Schwede et al.*, 2008]. Our explanation is that macrodispersion and the approach rate to ergodicity become uncertain when the variance, integral scales and anisotropy are uncertain. Results from different Monte-Carlo analyses (not shown here) indicate that the global trend functions have almost no impact on concentration variance.

With uncertain structure, the standard deviation for arrival time explodes by a factor of roughly ten; this can mostly be traced back to the uncertain global mean of Y, which dictates the average velocity. Variance, integral scales and anisotropy have an impact on large-scale effective conductivity [*Zhang*, 2002; *Rubin*, 2003], so that uncertain covariance parameters further increase the uncertainty of arrival time [*Rubin and Dagan*, 1992b].

5.5. Synthetic True Aquifer

Figure 4 depicts spatial maps of log-conductivity and the corresponding heads, concentrations and arrival times for the "true" aquifer generated with random structural parameters (see Table 3). We will read values of Y and h at the near-optimal sampling locations and add random measurement error to obtain synthetic data. This way, we can compare the conditional results to fully known reference fields of Y, hydraulic heads, concentrations and arrival times, and to the random values of structural parameters used for generation.

6. Results: Near-Optimal Sampling Patterns with Uncertain Structural Parameters

In this section, we present the sampling patterns resulting from Bayesian Geostatistical Design, considering the structural parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ as uncertain (cases 1b and 2b), and then compare with the non-Bayesian cases 1a and 2a. The methodological steps are given in Section 4.2.

6.1. Sampling Pattern Optimized for Predicting Concentration (Case 1b)

In case 1b (see Table 2), all structural parameters considered in our geostatistical model are uncertain, and we optimize the sampling pattern for optimal prediction of late-time concentration at the sensitive location. The resulting sampling pattern is shown in Figure 5. The figure also shows the conditional mean (left column) and standard deviations (right column) after applying the design and using the synthetic measurement values obtained from the synthetic "true" aquifer shown in Figure 4.

In principle, the original non-Bayesian prediction purpose leads to information needs in those regions of the domain where the statistical dependence between measurable quanti-

ties and the prediction goal is highest [e.g., *Cirpka et al.*, 2004; *Herrera and Pinder*, 2005; *Zhang et al.*, 2005; *Nowak*, 2009a]. At the same time, the Bayesian approach requires a diversification of sampled lag distances in order to reduce structural uncertainty. Hence, the sampling pattern found for case 1b is in essence similar to the one found by *McKinney and Loucks* [1992], with only small modifications due to structural uncertainty that will be discussed in Section 6.3. All our patterns are asymmetrical due to the transverse offset of the sensitive location relative to the center of the source.

For case 1b, all sampling locations fall into two groups, each providing a specific set of information that is explained below:

1. Measurements flanking the average migration pattern of the hypothetical plume.

2. Samples in and around the source.

The objective is to minimize the concentration fluctuations, σ_c^2, at a sensitive location. As explained in *Rubin* [1991b], the prediction of low concentrations at the periphery of the plume is subject to the largest uncertainty. If we could obtain concentration measurements (which we cannot do because we predict a future contamination), we would sample along the flanks of the plume [see further discussion in *Rubin*, 2003], as depicted in Figure 6. Our results sample the flanks of the expected plume trajectory for different but highly related reasons. The plume's flanks have the largest concentration gradients and, combined with uncertain positions of the streamlines, these gradients are converted to a high concentration variance. The head and conductivity measurements flanking the plume infer the position of streamlines in the plane of the sensitive location, which then translates to a reduced concentration variance [*Rubin*, 1991a].

The importance of small-scale fluctuations increases towards the sensitive target, while large-scale plume meandering is important at larger distances from the sensitive target. Therefore, the two rows of samples flanking the plume converge to the plume's center in the vicinity of sensitive location. The above mechanisms are an outcome of solving the flow and transport equations [e.g., *Rubin*, 1991a; *Kapoor and Kitanidis*, 1997]. The same holds for the reduction of variance and coefficient of variation for concentration induced by conditioning (Figure 6).

Figure 5 also shows a tendency to place the samples near the contaminant source. Besides resolving the directionality at short travel distances, measurements at these locations capture the volumetric flow rate through the source area. The latter is a key to predicting the evolving plume: A source area with below-average volumetric flow rate produces a narrow plume (in the ensemble sense) because streamlines are likely to converge downstream of the source. Vice-versa, when a large portion of the domain's total discharge is focused though the source area, a wider plume evolves. These facts emphasize the importance of source zone characterization even for far-field predictions. For the current objective function, the area within the expected plume trajectory turns out to be least significant.

The availability of different data types is yet another factor that influences sampling patterns, not less important than the choice of prediction objective. For example, when monitoring contaminations that have already occurred, concentration data become available. In studies on optimal plume monitoring, the resulting sampling patterns typically try to determine the current outline of the plume, i.e., find its fringes and the current

front [e.g., *Criminisi et al.*, 1997; *Herrera and Pinder*, 2005; *Wu et al.*, 2005; *Zhang et al.*, 2005].

When comparing to the synthetic "true" aquifer (see Figure 4) with the resulting conditional statistics (Figure 5), the global mean and trend of conductivity have been captured well. The contaminant source happens to be in an area of slow flow, such that a narrow plume leaves the source area, with peak concentrations prevailing only over a short travel distance (see the relatively short $c = 0.5$ isoline in Figures 4 and 5). This effect has been captured by the measurements in and around the source. The large-scale features of the flow field have also been captured: Like in our reference field, the conditional ensemble mean plume is accurately hitting the sensitive location, with its center line passing only slightly south of the sensitive location. Also, the uncertainty of structural parameters is reduced by conditioning, moving closer to the case of known structural parameters. Therefore, the map of concentration variance starts to exhibit the two distinct lines along the fringes of the plume (compare with Figure 2).

6.2. Sampling Pattern Optimized for Predicting Arrival Time (Case 2b)

Now we repeat the above analysis, this time minimizing the prediction variance of arrival time t_{50} at the sensitive location (case 2b). Figure 7 shows the resulting near-optimal sampling pattern, the conditional mean (left column) and standard deviations (right column). The synthetic measurement values for conditioning are again taken from the random simulation in Figure 4. The main sampling effort goes to the area between the source and the sensitive location, because arrival time is an integral outcome of the transport velocity along the entire distance. The seemingly random scattering of measurements within and outside the area between source and sensitive location mainly addresses structural un-

certainty. Some samples are scattered throughout the domain for better identification of the global mean and trend coefficients. Comparison of the conditional standard deviation between case 1b and 2b (Figures 5 and 7, respectively) shows that the pattern for case 1b is indeed better in reducing the uncertainty of concentration, while pattern 2b performs better in reducing the uncertainty of arrival time.

6.3. Comparison to Non-Bayesian Cases (Cases 1a and 2a)

To illustrate the impact of structural uncertainty on sampling patterns, we repeated the same analysis with known structural parameters (cases 1a and 2a) and compare the resulting patterns (Figure 8, left column) and sampled lag distances (Figure 8, right column).

The Bayesian approach to structural uncertainty honors the need for model identification, leading to a diversification of sampled lag distances [e.g., *Uciński*, 2005; *Diggle and Lophaven*, 2006; *Müller*, 2007]. The structural parameters $\boldsymbol{\theta} = [\sigma_Y^2,\, \lambda_1,\, \lambda_2,\, \kappa]$ require lag distances where they have the strongest impact on the covariance function (Figure 1). This information need appears in Eq. (12) as the derivatives of covariance functions with respect to structural parameters. For the global mean and variance, uncorrelated samples at great spacing are best, while covariance shape and scale additionally require a variety of low- to intermediate-range lags [e.g., *Bogaert and Russo*, 1999]. For the trend parameters, the most sensitive locations are close to the domain boundaries and corners, where the trend functions $\mathbf{X}$ have the largest impact on the expected value of $Y = \ln K$.

The pattern for case 1a (with known structural parameters) does already offer a variety of lag distances, so that the patterns in case 1a and 1b do not differ much. Minor changes include a better coverage of long lag distances, which help to better identify the trend

components. This is drastically different between cases 2a and 2b. The pattern for case 2a is extremely narrow in the x_2-direction, and therefore does not support inference of the transverse trend or the transverse integral scale. Also, the samples are highly correlated due to their proximity along a single line, so that identification of the mean and variance are compromised. For these reasons, the pattern for case 2b is substantially different, offering a much wider range of lag distances for better identification of covariance parameters, and samples closer to the corners of the domain for better identification of the trend coefficients.

7. Analysis and Discussion

This section discusses the impact of added samples on the prediction variance (Section 7.1), the reduction of structural uncertainty through sampling (model identification, Section 6.3) as well as the property of robustness (Section 7.3).

7.1. Effect of Sampling on Prediction Variance

How well did the near-optimal sampling patterns reduce the prediction variance of concentration? For the Bayesian cases 1b and 2b, the design criterion (Eq. 12) promised (in the expected sense) a reduction of prediction variances from $\sigma_c^2 = 0.0329$ to $\sigma_c^2 = 0.0215$ for concentration, and from $\sigma_{t50}^2 = 2466.4\, a^2$ to $\sigma_{t50}^2 = 1192.7\, a^2$ for arrival time. σ_c^2 is dimensionless because we used $c_0 = 1[-]$ for generality. An important caveat about expected prediction variances (such as Eqs. 9 and (12)) lies in their nature as expected value over yet unobserved data values (compare with *Feyen and Gorelick* [2005]). In addition, the Bayesian geostatistical framework averages over uncertain structural parameters that will be updated with yet unobserved data only later. By this property and by using Bayesian

prediction variances, Bayesian Geostatistical Design fulfills guideline 1 mentioned in the introduction. Using the synthetic data set, the near-optimal designs reduced the variances from $\sigma_c^2 = 0.0585$ to $\sigma_c^2 = 0.0204$ and from $\sigma_{t50}^2 = 2466.4\,a^2$ to $\sigma_{t50}^2 = 41.121\,a^2$ according to the conditional ensemble statistics.

Figure 9a (*total*) shows the expected prediction variance of concentration according to Eq. (12) evaluated for different numbers of near-optimal sampling locations. Across all cases, the planned sampling at 24 borehole locations reduce prediction uncertainties to between 50 and 70 percent of the initial uncertainty. The most effective samples are, of course, the first few ones that occupy the most informative locations. Samples placed later are displaced to less informative locations or suffer from redundancy of information if placed close by. Due to the presence of measurement error, even exhaustive sampling of the entire domain prevents a deterministic description of the system.

7.2. Effect of Sampling on Structural Uncertainty

The randomly generated structural parameters used to generate the synthetic "true" aquifer (Figure 4) are provided in Table 3 together with prior and posterior mean values after conditioning on the synthetic data from case 1b. Given the relatively small number of measurements and their level of measurement error (see Table 1), most structural parameters have been estimated very well.

The right half of Figure 9b shows how structural uncertainty (measured by distribution entropy) decreases with increasing number of samples placed. We approximate the entropy difference by [*Nowak*, 2009a]:

$$\Delta E\left(\boldsymbol{\beta},\boldsymbol{\theta}\right) \;=\; \det\left[\mathbf{C}_{\boldsymbol{\beta\theta}|\mathbf{y}}\mathbf{C}_{\boldsymbol{\beta\theta}}^{-1}\right]^{\frac{1}{d}}\,, \tag{17}$$

where $\mathbf{C}_{\boldsymbol{\beta\theta}|\mathbf{y}}$ is the joint conditional covariance matrix of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, $\mathbf{C}_{\boldsymbol{\beta\theta}}$ is its prior version, and d is the total number of structural parameters. Apart from a sign flip, the same curves are called Information Yield Curves by *de Barros et al.* [2009]. They illustrate how the data narrow down structural uncertainty and identify the geostatistical model (see guideline 2 in the introduction). Model identification is merely an implicit sub-goal to gain prediction confidence. Eq. (9) and (12) contains this sub-goal in a natural manner, and hence do not require a user-defined (and hence subjective) ranking between prediction and model identification (see guideline 3 in the introduction).

7.3. Cross-Case Validation, Robustness and Regular Sampling Grid

In order to discuss design robustness, we applied each near-optimal design pattern to the conditions of all other test cases. We then scaled all performances (reduction of prediction variance) by the performance of the pattern that was designed for each specific case. This yields the performance indices summarized in Table 4. Of course, each sampling pattern performs best when applying it to the respective case it was designed for, surpassing all other patterns. In the cross-comparison, pattern 1b outperforms pattern 1a when applied to the respective other test case (see Table 4, first two rows): The under-achievements when designing for structural uncertainty are smaller than the under-achievements when falsely pretending a known structure. Quite contrarily, pattern 2a outperforms pattern 2b: Due to the high impact of structural uncertainty onto the prediction objective, pattern 2b is dominated by model identification. When applied to the rather unrealistic case 2a, most of its sampling effort is spent uselessly.

The comparison between prediction variances with and without structural uncertainty in Section 5.4 indicated that some structural parameters do not contribute to one or

both of the prediction variances discussed here. One may now ask why to consider a seemingly irrelevant geostatistical parameter as uncertain. We will discuss the role of trend parameters in the prediction of concentration as an example. The trends add to the variability of both log-conductivity and hydraulic head. Even if the physical presence of a trend is known, we doubt that its actual magnitude could be specified deterministically. Without properly de-trended data, the data values would falsely be interpreted towards a larger overall variance σ_Y^2, resulting in false interpretation of the data. In similar fashions, any unjustified assumption or mis-specification of geostatistical structure may introduce spurious error into data interpretation, and hence into either spatial interpolation or into the estimation of other structural parameters. In conclusion, even seemingly irrelevant structural parameters should be accounted for, thus providing robustness against mis-specified geostatistical models (guideline 4 listed in the introduction).

For additional illustration, reference and comparison to simplistic designs, we also tested a regular sampling grid (8×3 grid with $40m \times 36m$ spacing) shown in (Figure 8, bottom row). The regular grid is clearly defeated in all cases. Neither can it provide detailed information on the release conditions, nor does it cover the variety of lag distances to identify the structural parameters, nor does it focus on the process-specific most sensitive regions of the domain.

8. Summary and Conclusions

This study transferred the concept of Bayesian Geostatistical Design [*Diggle and Lophaven*, 2006] to geostatistical inverse problems. Like other geostatistical design techniques, it optimizes site investigation or monitoring plans (called designs) for contaminated sites, while accounting for heterogeneous subsurface parameters as geostatistical random

space functions. The optimal design is defined to achieve a minimal expected prediction uncertainty with respect to a given prediction objective.

In contrast to conventional techniques, Bayesian Geostatistical Design allows for uncertainties in the geostatistical model itself. Uncertainties in the geostatistical model may include uncertain mean values, uncertain trend coefficients, uncertain choices of covariance models, and uncertain parameters within the covariance model, all summarized under the term of structural uncertainty. Even non-Multi-Gaussian descriptions can be tackled, given adequate computational resources.

In realistic situations of site investigation, initial information on geostatistical model parameters such as the variance or integral scale of log-conductivity is extremely scarce. This makes it illegitimate to assume fixed values a priori, and forces to treat them as uncertain. Otherwise, overly optimistic small levels of uncertainty would be specified, and the design would be optimized under unjustified (and possible false) assumptions. We argued that, under these premises, an adequate optimal design technique should fulfill four guidelines:

1. Structural uncertainty has a significant impact on prediction uncertainty, which must be accounted for.

2. Sampling helps to reduce structural uncertainty. This potential should be utilized in finding a sampling design.

3. Reduction of structural uncertainty should be ranked versus the primary design objective in an optimal and natural manner.

4. Designs should be robust towards mis-specified structural assumptions.

We showed that Bayesian Geostatistical Design indeed reduces the number or a priori assumptions on geostatistical structure, and also fulfills the above four guidelines. The only remaining assumptions are that the variability of the site can be described by a reasonable parametric geostatistical model (regardless of its parameter values). However, several different parametric models may cover different parts of the domain, and there is little restriction to the complexity of the parametric models.

A key point is minimum arbitrariness when choosing a covariance model prior to sampling. To this end, we suggest the Matérn family of geostatistical covariance models. It offers an additional shape parameter, and includes the exponential, Whittle and Gaussian covariance function as special cases. This way, as indicated earlier by [*Feyen et al.*, 2003], the problem of model selection becomes a problem of parameter estimation, with a wide range of methods available. We treat the shape parameter as yet another uncertain structural parameter, providing seamless integration of model uncertainty into the optimal design framework. We call this approach Continuous Bayesian Model Averaging because it is the limiting case of Bayesian Model Averaging over a continuous parametrized spectrum of models.

In a series of test cases, we demonstrated how structural uncertainty influences the optimal design. The test scenario featured the placement of 24 co-located hydraulic head and log-conductivity measurements, optimized for minimal prediction variance of (1) steady-state contaminant concentration and (2) contaminant arrival time at an ecologically sensitive location. Structural uncertainty was represented by an uncertain global mean, uncertain coefficients of a linear trend model, and the Matérn covariance function

582 with uncertain shape, variance and anisotropic integral scales. A variation of the test
583 cases considered the structural parameters to be known for comparison.

584 Only a few samples placed optimally were sufficient to largely eliminate the additional
585 uncertainty stemming from structural uncertainty. The list of uncertain structural pa-
586 rameters was shown to leave a distinct diversification in the fingerprint of the spatial
587 pattern of the resulting optimal sampling layouts. The required diversification showed
588 most clearly in the lag distances covered by the individual sampling patterns.

589 Within the risk assessment application context, Bayesian Geostatistical Design aligns
590 well with the TRIAD principle of site investigation suggested by the US EPA [*Crumbling*,
591 2001]. The TRIAD principle argues that information from ongoing site investigation
592 should provide immediate feedback to adjust the sampling campaign in real-time, by
593 continuously updating the site's conceptual model during the ongoing investigation effort.
594 Bayesian Geostatistical Design extends the TRIAD principle by a continuous updating of
595 the site's geostatistical model.

596 It is important to emphasize that the Bayesian Geostatistical Design framework is
597 not in any way limited to the implementational choices taken in our illustrative test
598 cases. Our implementation used a first-order approximation for structural uncertainty,
599 Ensemble Kalman Filters, and a sequential exchange optimization algorithm. We are
600 aware of the limited range of validity given by first-order approximations and we do
601 not claim generality within the results obtained in our illustrative test case. Willing to
602 accept substantially increased computational costs, our approximations can be removed in
603 exchange for brute-force Monte-Carlo or particle filter techniques, combined with genetic
604 or simulated annealing optimization algorithms.

Appendix A

We define a linearized representation for $\mathbf{f}_y(\mathbf{s})$ in the following form:

$$\mathbf{y} = \mathbf{f}_y(\mathbf{s}) \approx E[\mathbf{f}(\mathbf{s})] + \mathbf{H}(\mathbf{s} - \bar{\mathbf{s}}) , \tag{A1}$$

where $\bar{\mathbf{y}} = E[\mathbf{f}_y(\mathbf{s})]$ and $\bar{\mathbf{s}}$ are the mean values of $p(\mathbf{y})$ and $p(\mathbf{s})$, respectively. $\mathbf{H}$ takes the role of a sensitivity matrix. Within the linearized framework, the relevant mean values and covariances become:

$$\mathbf{G}_{\mathbf{yy}}(\boldsymbol{\theta}) = \mathbf{H}\mathbf{G}_{\mathbf{ss}}(\boldsymbol{\theta})\mathbf{H}^T + \mathbf{R} \tag{A2}$$

$$\hat{\mathbf{s}}(\mathbf{y}_o, \boldsymbol{\theta}) = \mathbf{X}\boldsymbol{\beta}^* + \mathbf{G}_{\mathbf{ss}}(\boldsymbol{\theta})\mathbf{H}^T\mathbf{G}_{\mathbf{yy}}^{-1}(\boldsymbol{\theta})(\mathbf{y}_o - \bar{\mathbf{y}}) \tag{A3}$$

$$\mathbf{G}_{\mathbf{ss|y}}(\boldsymbol{\theta}) = \mathbf{G}_{\mathbf{ss}}(\boldsymbol{\theta}) - \mathbf{G}_{\mathbf{ss}}(\boldsymbol{\theta})\mathbf{H}^T\mathbf{G}_{\mathbf{yy}}^{-1}(\boldsymbol{\theta})\mathbf{H}\mathbf{G}_{\mathbf{ss}}(\boldsymbol{\theta}) \tag{A4}$$

$$\mathbf{C}_{\boldsymbol{\beta}\boldsymbol{\beta}|\mathbf{y}} = \left(\mathbf{X}^T\mathbf{H}_{\mathbf{y}}^T\mathbf{G}_{\mathbf{yy}}^{-1}\mathbf{H}_{\mathbf{y}}\mathbf{X} + \mathbf{C}_{\boldsymbol{\beta}\boldsymbol{\beta}}\right)^{-1} , \tag{A5}$$

where $\mathbf{G}_{\mathbf{yy}}$ is the generalized covariance of $\mathbf{y}$, $\hat{\mathbf{s}}$ and $\mathbf{G}_{\mathbf{ss|y}}$ are the conditional mean and generalized covariance of $\mathbf{s}$, and $\mathbf{C}_{\boldsymbol{\beta}\boldsymbol{\beta}|\mathbf{y}}$ is the conditional covariance of $\boldsymbol{\beta}$. Employing a likewise linearized representation of $c = f_c(\mathbf{s})$ with coefficient matrix $\mathbf{H}_c$, the conditional predictive distribution for c becomes:

$$\hat{c}(\mathbf{y}_o, \boldsymbol{\theta}) = \bar{c} + \mathbf{H}_c(\hat{\mathbf{s}}(\mathbf{y}_o, \boldsymbol{\theta}) - \mathbf{X}\boldsymbol{\beta}^*)$$

$$\sigma_{c|\mathbf{y}}^2(\boldsymbol{\theta}) = \mathbf{H}_c\mathbf{G}_{\mathbf{ss|y}}(\boldsymbol{\theta})\mathbf{H}_c^T . \tag{A6}$$

Due to linearization, the prediction variances for known $\boldsymbol{\theta}$ are independent of data values, and Eq. (9) simplifies to:

$$E_{\mathbf{y}}\left[\tilde{\sigma}_{c|\mathbf{y}}^2\right] = E_{\boldsymbol{\theta}}\left[\sigma_{c|\mathbf{y}}^2(\boldsymbol{\theta})\right] + E_{\mathbf{y}}\left\{V_{\boldsymbol{\theta}|\mathbf{y}}\left[\hat{c}(\mathbf{y}(\mathbf{d}), \boldsymbol{\theta})\right]\right\} . \tag{A7}$$

Linearization of $\mathbf{f}_y(\mathbf{s})$ is exact for direct measurements of $\log K$, and overwrites the responsible rows of $\mathbf{H}$ by a sampling matrix [e.g., *Fritz et al.*, 2009]. *Dagan* [1985] showed

analytically that linearized $\mathbf{f}_y(\mathbf{s})$ for hydraulic heads is highly accurate for variances of $\log K$ up to unity, and *Nowak et al.* [2008] demonstrated its reliability for up to $\sigma_Y^2 = 5$ by Monte-Carlo analysis.

Appendix B

We now derive Eq. (12) from Eq. (11). For simplicity of notation, let $\omega(\boldsymbol{\theta}) \equiv \sigma^2_{c|\mathbf{y}}(\boldsymbol{\theta})$. Expanding $\omega(\boldsymbol{\theta})$ up to first-order in $\boldsymbol{\theta}$ yields:

$$\omega(\boldsymbol{\theta}) \approx \omega\left(\bar{\boldsymbol{\theta}}\right) + \nabla_{\boldsymbol{\theta}}\omega\boldsymbol{\theta}' \tag{B1}$$

where $\bar{\boldsymbol{\theta}} = E_{\boldsymbol{\theta}}[\boldsymbol{\theta}]$, $\boldsymbol{\theta}' = \boldsymbol{\theta} - \bar{\boldsymbol{\theta}}$ and $\nabla_{\boldsymbol{\theta}}\omega$ is the row-vector Jacobian of ω evaluated at $\boldsymbol{\theta} = \bar{\boldsymbol{\theta}}$. Due to $E[\boldsymbol{\theta}'] = 0$, the first term in Eq. (11) becomes:

$$E_{\boldsymbol{\theta}}\left[\sigma^2_{c|\mathbf{y}}(\boldsymbol{\theta})\right] \approx \sigma^2_{c|\mathbf{y}}\left(\bar{\boldsymbol{\theta}}\right) = \mathbf{H}_c\mathbf{G}_{\mathbf{ss}|\mathbf{y}}\left(\bar{\boldsymbol{\theta}}\right)\mathbf{H}_c^T . \tag{B2}$$

The second term in Eq. (11) is obtained in a similar fashion by setting

$$\begin{aligned}\bar{c}(\boldsymbol{\theta}) + \boldsymbol{\kappa}(\boldsymbol{\theta})\mathbf{y}' &\equiv \bar{c}(\boldsymbol{\theta}) + \mathbf{H}_c\mathbf{G}_{\mathbf{ss}}(\boldsymbol{\theta})\mathbf{H}^T\mathbf{G}^{-1}_{\mathbf{yy}}(\boldsymbol{\theta})(\mathbf{y} - \bar{\mathbf{y}}) \\ &= \hat{c}(\mathbf{y}(\mathbf{d}), \boldsymbol{\theta}) ,\end{aligned} \tag{B3}$$

with $\mathbf{y}' = (\mathbf{y} - \bar{\mathbf{y}})$. $\boldsymbol{\kappa}$ can be interpreted as the Kalman gain of predicted concentration, and $\bar{c}(\boldsymbol{\theta})$ is the ensemble mean concentration given $\boldsymbol{\theta}$, prior to sampling. Now, we expand $\bar{c}(\boldsymbol{\theta})$ and $\boldsymbol{\kappa}(\boldsymbol{\theta})$ up to first order in $\boldsymbol{\theta}$:

$$\bar{c}(\boldsymbol{\theta}) \approx \bar{c}\left(\bar{\boldsymbol{\theta}}\right) + \nabla_{\boldsymbol{\theta}}\bar{c}\boldsymbol{\theta}' \tag{B4}$$

$$\boldsymbol{\kappa}(\boldsymbol{\theta}) \approx \boldsymbol{\kappa}\left(\bar{\boldsymbol{\theta}}\right) + \nabla_{\boldsymbol{\theta}}\boldsymbol{\kappa}\boldsymbol{\theta}' . \tag{B5}$$

The first-order perturbation of $\hat{c}(\mathbf{y}(\mathbf{d}), \boldsymbol{\theta})$ is:

$$\hat{c}' = \nabla_{\boldsymbol{\theta}}\bar{c}\boldsymbol{\theta}' + \nabla_{\boldsymbol{\theta}}\boldsymbol{\kappa}\boldsymbol{\theta}'\mathbf{y}' , \tag{B6}$$

and its variance over the distribution $p(\boldsymbol{\theta}|\mathbf{y})$ is (accurate to first order in $\boldsymbol{\theta}$):

$$\begin{aligned} & V_{\boldsymbol{\theta}|\mathbf{y}}\left[\hat{c}\left(\mathbf{y}\left(\mathbf{d}\right),\boldsymbol{\theta}\right)\right] \\ & \approx E_{\boldsymbol{\theta}|\mathbf{y}}\left[\sum_i\sum_j \theta_i'\theta_j'\left\{\frac{\partial\boldsymbol{\gamma}}{\partial\theta_i}\left(\frac{\partial\boldsymbol{\gamma}}{\partial\theta_j}\right)^T+\frac{\partial\boldsymbol{\kappa}}{\partial\theta_i}\mathbf{y}'\mathbf{y}'^T\left(\frac{\partial\boldsymbol{\kappa}}{\partial\theta_j}\right)^T\right\}\right] \\ & = \sum_i\sum_j \left\langle\mathbf{C}_{\boldsymbol{\theta}\boldsymbol{\theta}|\mathbf{y}}\right\rangle_{ij}\left\{\frac{\partial\boldsymbol{\gamma}}{\partial\theta_i}\left(\frac{\partial\boldsymbol{\gamma}}{\partial\theta_j}\right)^T+\frac{\partial\boldsymbol{\kappa}}{\partial\theta_i}\mathbf{y}'\mathbf{y}'^T\left(\frac{\partial\boldsymbol{\kappa}}{\partial\theta_j}\right)^T\right\} \end{aligned} \tag{B7}$$

where $\mathbf{C}_{\boldsymbol{\theta}\boldsymbol{\theta}|\mathbf{y}}$ is the conditional covariance of $\boldsymbol{\theta}$ and $\langle\cdot\rangle$ denotes the i,j-the element. $\mathbf{C}_{\boldsymbol{\theta}\boldsymbol{\theta}|\mathbf{y}}$ is independent of actual data values when expressed via the inverse of the Fisher information $\mathbf{F}$ [e.g., *Kitanidis and Lane*, 1985]:

$$\mathbf{F} \;=\; E_{\mathbf{y}}\left[\left(\frac{\partial}{\partial\boldsymbol{\theta}}\ln p\left(\mathbf{y}|\boldsymbol{\theta}\right)\right)^T\left(\frac{\partial}{\partial\boldsymbol{\theta}}\ln p\left(\mathbf{y}|\boldsymbol{\theta}\right)\right)\right]. \tag{B8}$$

In the current context, we assume that the $\boldsymbol{\theta}$ has a prior covariance matrix $\mathbf{C}_{\boldsymbol{\theta}\boldsymbol{\theta}}$, so that the elements F_{ij} of $\mathbf{F}$ are given by [*Nowak and Cirpka*, 2006]:

$$F_{ij}=\frac{1}{2}Tr\left[\frac{\partial\mathbf{G}_{\mathbf{yy}}}{\partial\theta_i}\mathbf{G}_{\mathbf{yy}}^{-1}\frac{\partial\mathbf{G}_{\mathbf{yy}}}{\partial\theta_j}\mathbf{G}_{\mathbf{yy}}^{-1}\right]+\mathbf{e}_i^T\mathbf{C}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}\mathbf{e}_j\,, \tag{B9}$$

where $\mathbf{e}_i$ is the i-th unit vector. $\mathbf{G}_{\mathbf{yy}}$ and its derivatives are evaluated at $\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}$. Now, we take the expected value over $p(\mathbf{y})$ to obtain the second term in Eq. (11):

$$\begin{aligned} & E_{\mathbf{y}}\left\{V_{\boldsymbol{\theta}|\mathbf{y}}\left[\hat{c}\left(\mathbf{y}\left(\mathbf{d}\right),\boldsymbol{\theta}\right)\right]\right\} \\ & \approx \sum_i\sum_j \left\langle\mathbf{C}_{\boldsymbol{\theta}\boldsymbol{\theta}|\mathbf{y}}\right\rangle_{ij}\left\{\frac{\partial\boldsymbol{\gamma}}{\partial\theta_i}\left(\frac{\partial\boldsymbol{\gamma}}{\partial\theta_j}\right)^T+\frac{\partial\boldsymbol{\kappa}}{\partial\theta_i}\mathbf{G}_{\mathbf{yy}}\left(\bar{\boldsymbol{\theta}}\right)\left(\frac{\partial\boldsymbol{\kappa}}{\partial\theta_j}\right)^T\right\} \end{aligned} \tag{B10}$$

Updating the structural parameters once data become available requires the gradient $\mathbf{g}$ [*Kitanidis and Lane*, 1985]. For the case of prior covariance $\mathbf{C}_{\boldsymbol{\theta}\boldsymbol{\theta}}$, its entries are [*Nowak and Cirpka*, 2006]:

$$g_i \;=\; \frac{1}{2}Tr\left[\frac{\partial\mathbf{G}_{\mathbf{yy}}}{\partial\theta_i}\mathbf{G}_{\mathbf{yy}}^{-1}\right]-\frac{1}{2}\left(\mathbf{y}_o-\bar{\mathbf{y}}\right)^T\mathbf{G}_{\mathbf{yy}}^{-1}\frac{\partial\mathbf{G}_{\mathbf{yy}}}{\partial\theta_j}\mathbf{G}_{\mathbf{yy}}^{-1}\left(\mathbf{y}_o-\bar{\mathbf{y}}\right)+\mathbf{e}_i^T\left(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}}\right) \tag{B11}$$

Acknowledgments.

This study has been funded in parts by the Deutsche Forschungsgemeinschaft (DFG) under grant No 805/1-1, under the Cluster of Excellence in Simulation Technology (EXC 310/1) at the University of Stuttgart, by the Coordenação de Aperfeiçoamento de Pessoal e Nível Superior (CAPES) from Brazil and by the U.S. DOE Office of Biological and Environmental Research, Environmental Remediation Science Program (ERSP), through DOE-ERSP grant DE-FG02-06ER06-16 as part of Hanford 300 Area Integrated Research Challenge Project. We would like to acknowledge Harrie-Jan Hendricks-Franssen and the other 2 reviewers for their constructive comments.

References

Abramowitz, M., and I. A. Stegun (1972), *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing*, Dover, New York.

Bogaert, P., and D. Russo (1999), Optimal spatial sampling design for the estimation of the variogram based on a least squares approach, *Water Resour. Res.*, *35*(4), 1275–1289.

Box, G. E. P. (1982), Choice of response surface design and alphabetic optimality, *Utilitas Math.*, *21*, 11–55.

Caroni, E., and V. Fiorotto (2005), Analysis of concentration as sampled in natural aquifers, *Trans. Porous Med.*, *59*, 19–45, doi:10.107/s11,242–004–1119–x.

Chen, Y., and D. Zhang (2006), Data assimilation for transient flow in geologic formations via ensemble kalman filter, *Adv. Water Resour.*, *29*, 1107–1122.

Christakos, G. (1992), *Random Field Models in Earth Sciences*, 4th ed., Dover, New York.

Cirpka, O. A., and P. K. Kitanidis (2000), Characterization of mixing and dilution in heterogeneous aquifers by means of local temporal moments, *Water Resour. Res.*, *36*(5),

1221–1136.

Cirpka, O. A., and W. Nowak (2004), First-order variance of travel time in non-stationary formations, *Water Resour. Res.*, *40*, doi:10.1029/2003WR002,851.

Cirpka, O. A., C. M. Bürger, W. Nowak, and M. Finkel (2004), Uncertainty and data worth analysis for the hydraulic design of funnel-and-gate systems in heterogeneous aquifers, *Water Resour. Res.*, *40*(W11502), doi:10.1029/2004WR003,352.

Criminisi, A., T. Tucciarelli, and G. P. Karatzas (1997), A methodology to determine optimal transmissivity measurement locations in groundwater quality management models with scarce field information, *Water Resour. Res.*, *33*(6), 1265–1274.

Crumbling, D. (2001), Using the triad approach to improve the cost-effectiveness of hazardous waste site cleanups, *Tech. Rep. EPA 542-R-01-016*.

Dagan, G. (1985), A note on higher-order corrections of the head covariances in steady aquifer flow, *Water Resour. Res.*, *21*(4), 573–578.

de Barros, F. P. J., and Y. Rubin (2008), A risk-driven approach for subsurface site characterization, *Water Resour. Res.*, *44*(W01414), doi:10.1029/2007WR006,081.

de Barros, F. P. J., Y. Rubin, and R. Maxwell (2009), The concept of comparative information yield curves and its application to risk-based site characterization, *Water Resour. Res.*, *45*(W06401), doi:10.1029/2008WR007,324.

de Marsily, G. (1986), *Quantitative Hydrology*, Academic Press, San Diego, CA.

Diggle, P., and S. Lophaven (2006), Bayesian geostatistical design, *Scandinavian J. Statist.*, *33*, 53–64, doi:10.1111/j.1467–9469.2005.00,469.x.

Diggle, P., and P. J. Ribeiro (2002), Bayesian inference in Gaussian model-based geostatistics, *Geogr. and Environ. Mod.*, *6*(2), 129–146.

Diggle, P. J., and P. J. Ribeiro (2007), *Model-based geostatistics*, Springer series in statistics, Springer, New York.

Federov, V., and P. Hackl (1997), *Model-Oriented Design of Experiments*, Springer-Verlag, New York.

Feinerman, E., G. Dagan, and E. Bresler (1986), Statistical inference of spatial random functions, *Water Resour. Res.*, *22*(6), 953–942.

Feyen, L., and S. M. Gorelick (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:10.1029/2003WR002,901.

Feyen, L., J. J. Gómez-Hernández, P. J. R. Jr., K. J. Beven, and F. D. Smedt (2003), A Bayesian approach to stochastic capture zone delineation incorporating tracer arrival times, conductivity measurements, and hydraulic head observations, *Water Resour. Res.*, *39*(5), doi:10.1029/2002WR001,544.

Fiorotto, V., and E. Caroni (2002), Solute concentration statistics in heterogeneous aquifers for finite Péclet values, *Transport in Porous Media*, *48*, 331–351.

Freeze, R., J. Massmann, L. Smith, T. Sperling, and B. James (1990), Hydrogeological decision analysis: 1. A framework, *Ground Water*, *28*(5), 738–766.

Fritz, J., W. Nowak, and I. Neuweiler (2009), Application of FFT-based algorithms for large-scale universal Kriging problems, *Math. Geosci.*, pp. 10.1007/s11,004–009–9220–x.

Handcock, M. S., and M. L. Stein (1993), A Bayesian analysis of Kriging, *Technometrics*, *35*(4), 403–410.

Harvey, C. F., and S. M. Gorelick (1995), Temporal moment-generating equations: Modeling transport and mass-transfer in heterogeneous aquifers, *Water Resour. Res.*, *31*(8),

703 1895–1911.

704 Herrera, G. S. (1998), Cost effective groundwater quality sampling network design, Ph.D.
705 thesis, University of Vermont, Burlington.

706 Herrera, G. S., and G. F. Pinder (2005), Space-time optimiziation of groundwater quality
707 sampling networks, *Water Resour. Res.*, *41*(W12407), doi:10.1029/2004WR003,626.

708 Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Colinsky (1999), Bayesian model
709 averaging: A tutorial, *Statisc. Sc.*, *14*(4), 382–417.

710 James, B., and S. Gorelick (1994), When enough is enough: the worth of monitoring data
711 in aquifer remediation design, *Water Resour. Res.*, *30*(12), 3499–3513.

712 Janssen, G. M. C. M., J. R. Valstar, and S. E. A. T. M. van der Zee (2008), Measure-
713 ment network design including traveltime determinations to minimize model prediction
714 uncertainty, *Water Resour. Res.*, *44*(W02405), doi:10.1029/2006WR005,462.

715 Journel, A., and C. Huijbregts (1978), Mining geostatistics, *New York*.

716 Kapoor, V., and P. Kitanidis (1997), Advection-diffusion in spatially random flows: For-
717 mulation of concentration covariance, *Stochastic Environmental Research and Risk As-*
718 *sessment (SERRA)*, *11*(5), 397–422.

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

721 Kitanidis, P. K. (1993), Generalized covariance functions in estimation, *Math. Geol.*,
722 *25*(5), 525–540.

723 Kitanidis, P. K. (1995), Quasi-linear geostatistical theory for inversing, *Water Resour.*
724 *Res.*, *31*(10), 2411–2419.

Kitanidis, P. K. (1997), *Introduction to Geostatistics*, Cambridge University Press, Cambridge.

Kitanidis, P. K., and R. W. Lane (1985), Maximum likelihood parameter estimation of hydrologic spatial processes by the Gauss-Newton method, *J. Hydrol.*, *79*, 53–71.

Marchant, B., and R. Lark (2007a), The Matérn variogram model: Implications for uncertainty propagation and sampling in geostatistical surveys, *Geoderma*, *140*(4), 337–345.

Marchant, B. P., and R. M. Lark (2007b), Optimized sample schemes for geostatistical surveys, *Math. Geol.*, *39*(1), doi: 10.1007/s11,004–006–9069–1.

Massmann, J., and R. Freeze (1987), Groundwater contamination from waste management sites: The interaction between risk-based engineering design and regulatory policy. 1. Methodology, *Water resources research (USA).*

Matérn, B. (1986), *Spatial variation*, Springer, Berlin, Germany.

Matheron, G. (1971), *The Theory of Regionalized Variables and Its Applications*, Ecole de Mines, Fontainebleau, France.

Maxwell, R., W. E. Kastenberg, and Y. Rubin (1999), A methodology to integrate site characterization information into groundwater-driven health risk assessment, *Water Resour. Res.*, *35*(9), 2841–2885.

McKinney, D. C., and D. P. Loucks (1992), Network design for predicting groundwater contamination, *Water Resour. Res.*, *28*(1), 133–147.

Minasny, B., and A. McBratney (2005), The Matern function as a general model for soil variograms, *Geoderma*, *128*(3-4), 192–207.

Müller, W. G. (2007), *Collecting spatial data. Optimum design of experiments for random fields*, 3 ed., Springer, Berlin, Germany.

748 Neuman, S. P. (2003), Maximum likelihood Bayesian averaging of uncertain model pre-
749 dictions, *Stoch. Environ. Res. Risk Assess.*, *17*, doi10.1007/s00,477–003–0151–7.
750 Nowak, W. (2009a), Measures of parameter uncertainty in geostatistical estimation and
751 design, *Math. Geosci.*, (doi: 10.1007/s11004-009-9245-1), In Press.
752 Nowak, W. (2009b), Best unbiased ensemble linearization and the quasi-linear Kalman
753 ensemble generator, *Water Resour. Res.*, *45*(W04431), doi:10.1029/2008WR007,328.
754 Nowak, W., and O. A. Cirpka (2004), A modified Levenberg-Marquardt algorithm for
755 quasi-linear geostatistical inversing, *Adv. Water Resour.*, *27*(7), 737–750.
756 Nowak, W., and O. A. Cirpka (2006), Geostatistical inference of conductivity and disper-
757 sion coefficients from hydraulic heads and tracer data, *Water Resour. Res.*, *42*(W08416),
758 doi:10.1029/2005WR004,832.
759 Nowak, W., S. Tenkleve, and O. A. Cirpka (2003), Efficient computation of linearized
760 cross-covariance and auto-covariance matrices of interdependent quantities, *Math. Geol.*,
761 *35*(1), 53–66.
762 Nowak, W., R. L. Schwede, O. A. Cirpka, and I. Neuweiler (2008), Probability density
763 functions of hydraulic head and velocity in three-dimensional heterogeneous porous
764 media, *Water Resour. Res.*, *44*(W08452), doi:10.1029/2007WR006,383.
765 Pardo-Iguzquiza, E. (1999), Bayesian inference of spatial covariance parameters, *Math.*
766 *Geol.*, *31*(1), 47–65.
767 Pardo-Iguzquiza, E., and M. Chica-Olmo (2008), Geostatistical simulation when the num-
768 ber of experimental data is small: an alternative paradigm, *Stoch. Environ. Res. Risk*
769 *Assess*, *22*, 325–337.

Pukelsheim, F. (2006), *Optimal Design of Experiments*, Classics in Applied Mathematics, classic edition ed., SIAM, Philadelphia.

Riva, M., and M. Willmann (2009), Impact of log-transmissivity variogram structure on groundwater flow and transport predictions, *Advances in Water Resources*, *32*(8), 1311–1322.

Rubin, Y. (1991a), Prediction of tracer plume migration in heterogeneous porous media by the method of conditional probabilities, *Water Resour. Res.*, *27*(6), 1291–1308.

Rubin, Y. (1991b), Transport in heterogeneous porous media: Prediction and uncertainty, *Water Resour. Res*, *27*(7), 1723–1738.

Rubin, Y. (2003), *Applied Stochastic Hydrogeology*, Oxford University Press, Oxford.

Rubin, Y., and G. Dagan (1987a), Stochastic identification of transmissivity and effective recharge in steady state groundwater flow. 1. Theory, *Water Resour. Res.*, *23*(7), 1185–1192.

Rubin, Y., and G. Dagan (1987b), Stochastic identification of transmissivity and effective recharge in steady groundwater flow: 2. Case Study, *Water Resources Research*, *23*(7).

Rubin, Y., and G. Dagan (1988), Stochastic Analysis of Boundaries Effects on Head Spatial Variability in Heterogeneous Aquifers: I. Constant Head Boundary, *Water Resources Research*, *24*(10).

Rubin, Y., and G. Dagan (1992a), A note on head an velocity covariances in three-dimensional flow through heterogeneous anisotropic porous media, *Water Resour. Res.*, *28*(5), 1463–1470.

Rubin, Y., and G. Dagan (1992b), Conditional estimates of solute travel time in heterogenous formations: impact of transmissivity measurements, *Water Resour. Res.*, *28*(4),

1033–1040.

Rubin, Y., M. A. Cushey, and A. Bellin (1994), Modeling of transport in groundwater for environmental risk assessment, *Stochastic Hydrol. Hydraul.*, *8*(1), 57–77.

Rubin, Y., A. Sun, R. Maxwell, and A. Bellin (1999), The concept of block-effective macrodispersivity and a unified approach for grid-scale- and plume-scale-dependent transport, *J. Fluid Mech.*, *395*, 161–180.

Schwede, R. L., O. A. Cirpka, W. Nowak, and I. Neuweiler (2008), Impact of sampling volume on the probability density function of steady state concentration, *Water Resour. Res.*, *44*(W12433), doi:10.1029/2007WR006,668.

Schweppe, F. C. (1973), *Uncertain Dynamic Systems*, Prentice-Hall, Englewood Cliffs, NJ.

Silvey, S. (1980), *Optimal Designs*, Chapman & Hall, London.

Stein, M. L. (1999), *Interpolation of Spatial Data: Some Theory for Kriging*, Springer, Berlin, Germany.

Uciński, D. (2005), *Optimal Measurement Methods for Distributed Parameters System Identification*, CRC Press, Boca Raton, FL, USA.

Woodbury, A. D., and T. J. Ulrych (2000), A full-Bayesian approach to the groundwater inverse problem for steady state flow, *Water Resour. Res.*, *36*(8), 2081–2093.

Wu, J., C. Zheng, and C. C. Chien (2005), Cost-effective sampling network design for contaminant plume monitoring under general hydrogeological conditions, *J. Contam. Hydrol.*, *77*(1), 41–65.

Zhang, D. (2002), *Stochastic Methods for Flow in Porous Media*, Academic Press, San Diego.

816 Zhang, Y., G. F. Pinder, and G. S. Herrera (2005), Least cost design of
817 groundwater quality monitoring networks, *Water Resour. Res.*, *41*(W08412),
818 doi:10.1029/2005WR003,936.
819 Zimmerman, D. L. (2006), Optimal network design for spatial prediction, covariance pa-
820 rameter estimation, and empirical prediction, *Environmetrics*, *17*, 635–652.

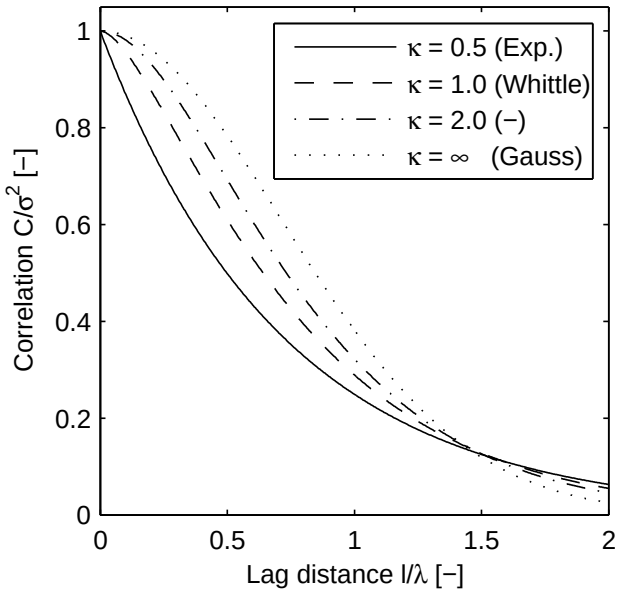

Figure 1. Examples from the Matérn family of covariance functions for different values of the shape parameter κ, including some special cases.

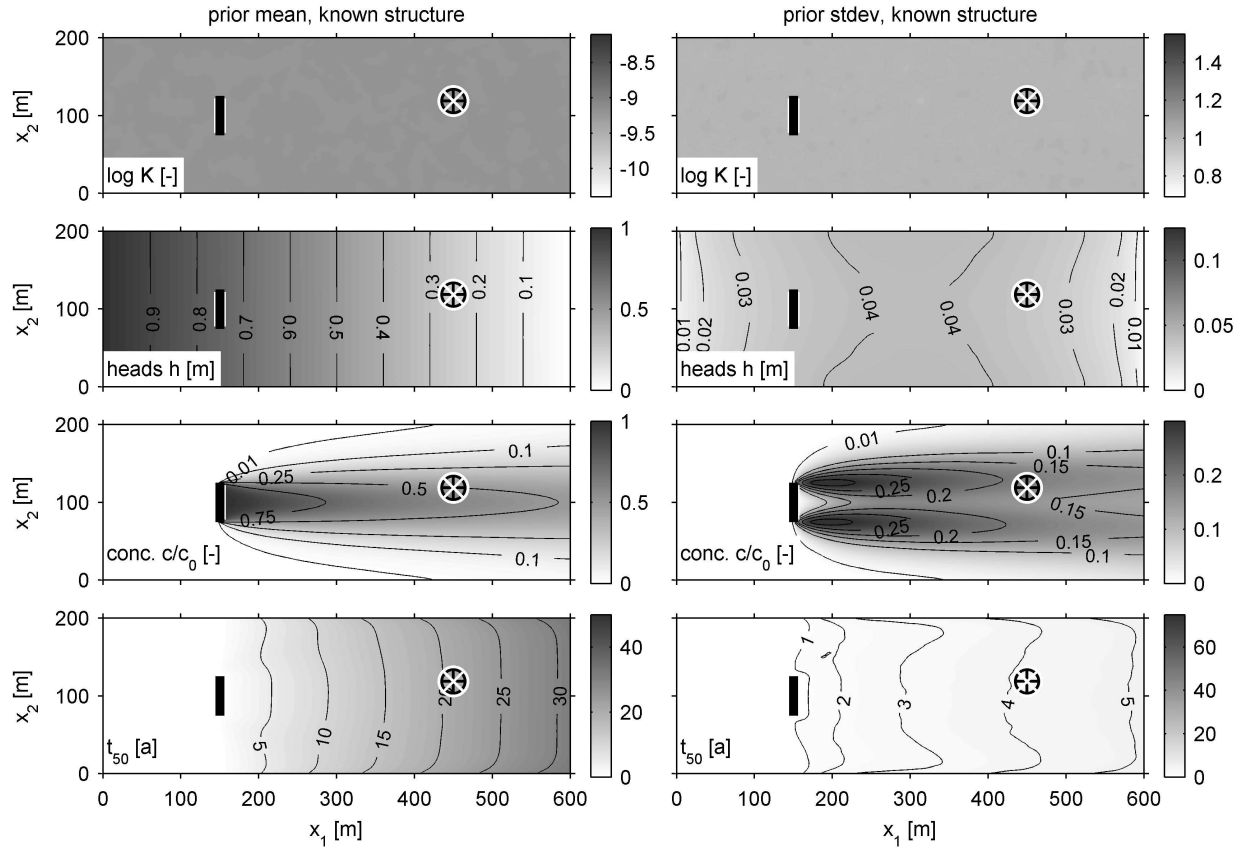

Figure 2. Illustration of the scenario for known structural parameters. Left: prior mean values of $Y = \ln K$, corresponding hydraulic heads h and the hypothetical plume (steady-state concentration c and arrival time t_{50}). Right: prior standard deviation. Crossed circle: sensitive location. Thick black line: hypothetical contaminant source. For parameter values, see Table 1 and Table 2 (cases 1a and 2a). Grey-scale is identical to Figure 3 for direct comparison.

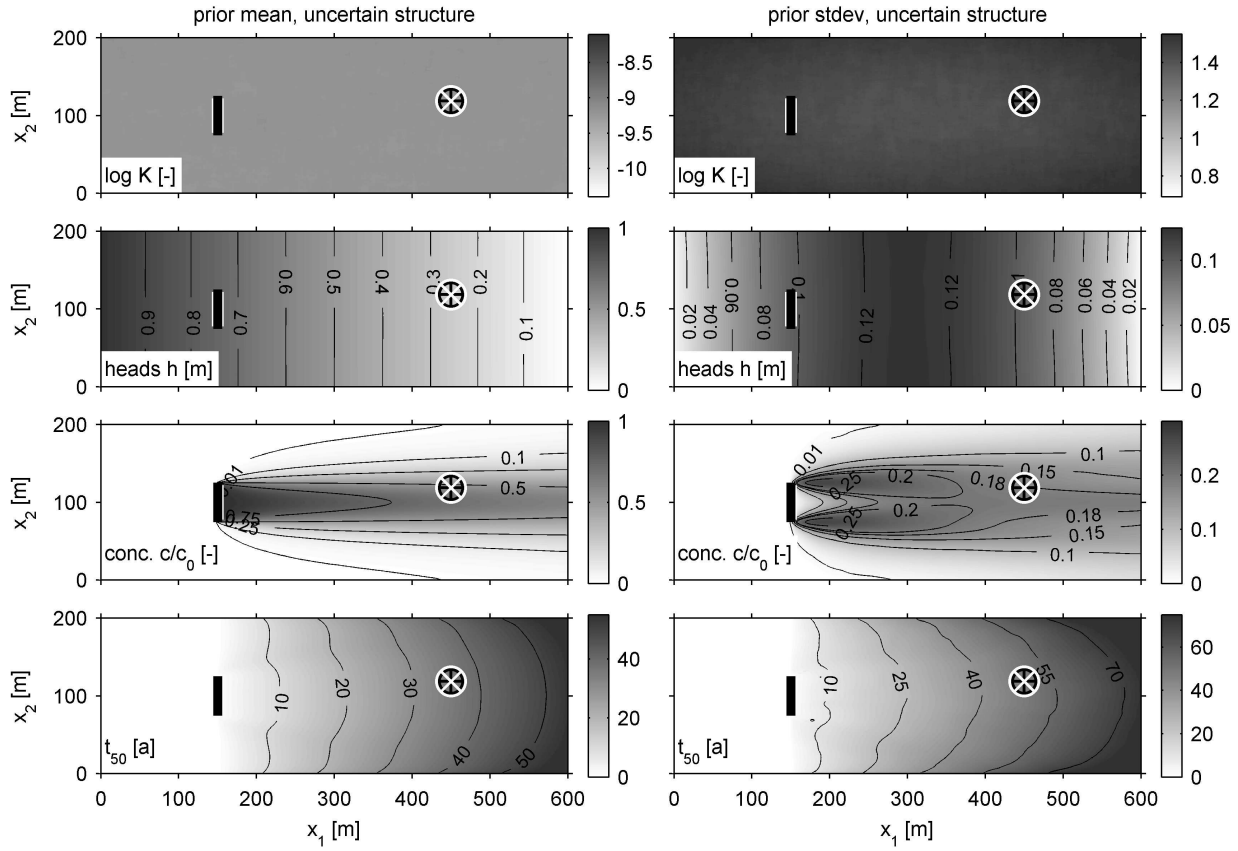

Figure 3. Same as in Figure 2 but for uncertain structural parameters (cases 1b and 2b in Table 2).

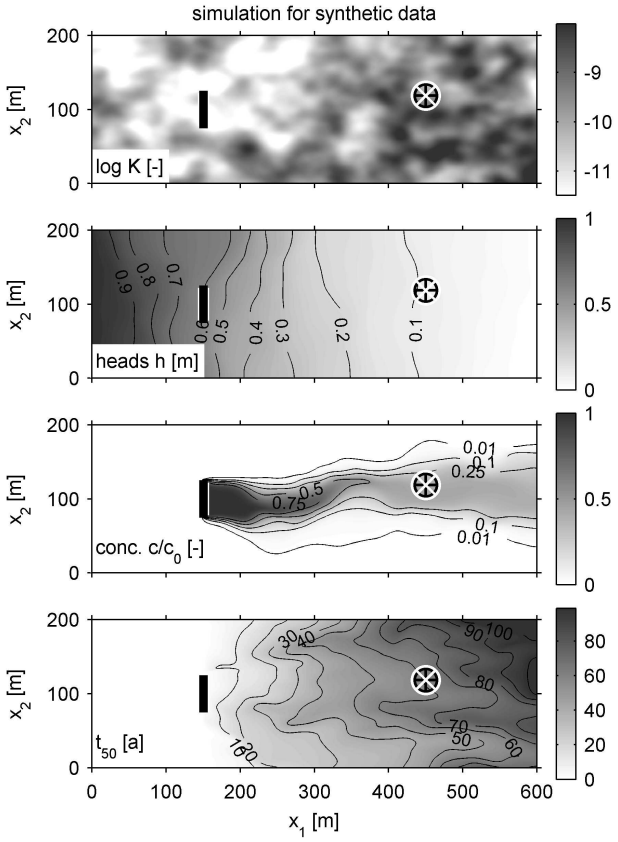

Figure 4. Synthetic "true" aquifer used to obtain measurement values: a realization of $Y = \ln K$ and corresponding simulated hydraulic heads, steady-state concentration and arrival time. Crossed circle: sensitive location. Thick black line: hypothetical contaminant source. For parameter values, see Tables 1, 2 and 3.

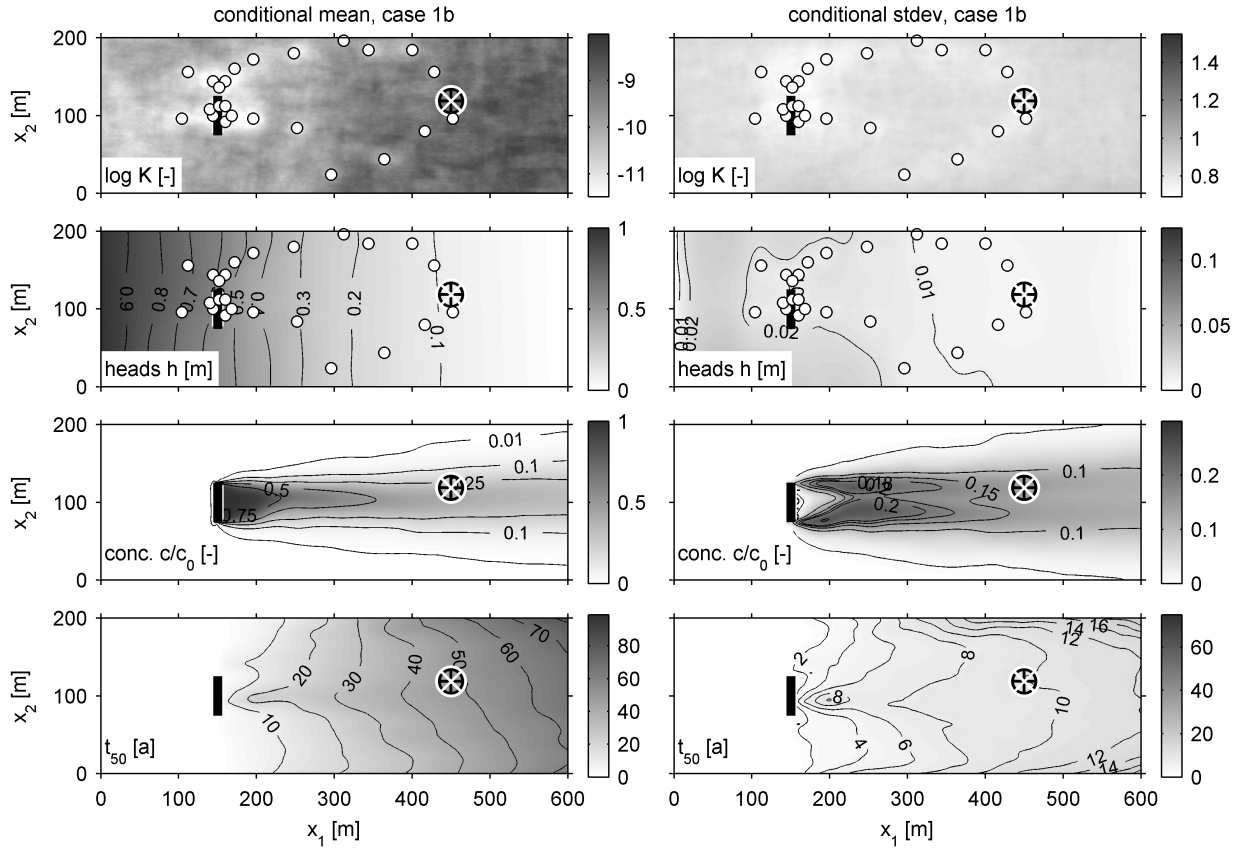

Figure 5. Results for case 1b. Left: conditional mean for $\ln K$, hydraulic heads h, steady-state concentration c and arrival time t_{50} of hypothetical plume. Right: corresponding conditional standard deviations. Crossed circle: sensitive location. Solid white circles: near-optimal sampling locations ($Y = \ln K$ and head measurements). Thick black line: hypothetical contaminant source. For parameter values, see Tables 1 and 2.

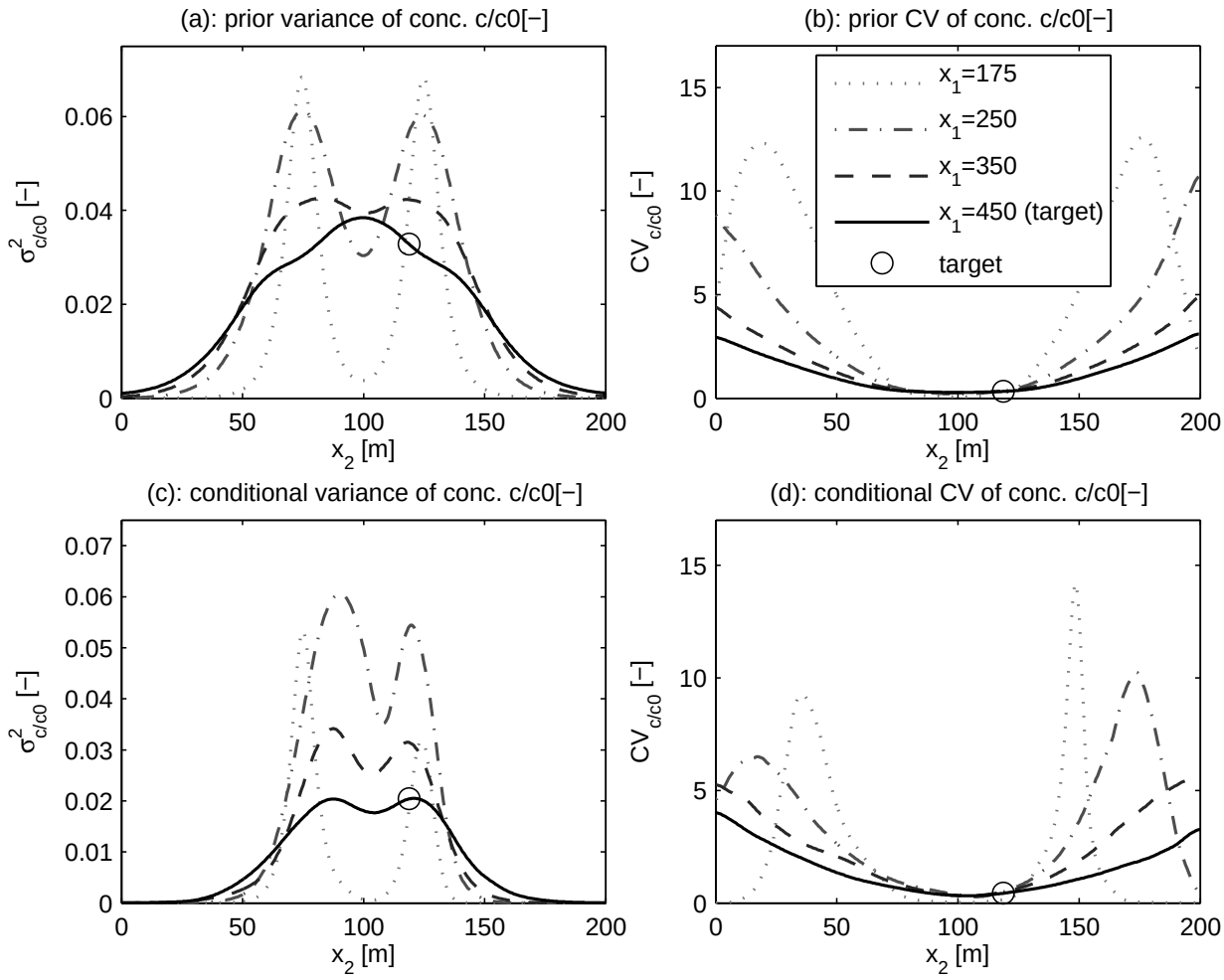

Figure 6. Concentration variance and coefficient of variation (CV_c) as a function of the transverse direction. Curves are presented for 4 transects at different distances from the source: $x_1 = 175$, 250, 350 and 450 (in dimensionless numbers: $\xi \approx 12$, 17, 23 and 30). (a) Prior concentration variance; (b) prior coefficient of variation; (c) conditional concentration variance; and (d) conditional coefficient of variation.

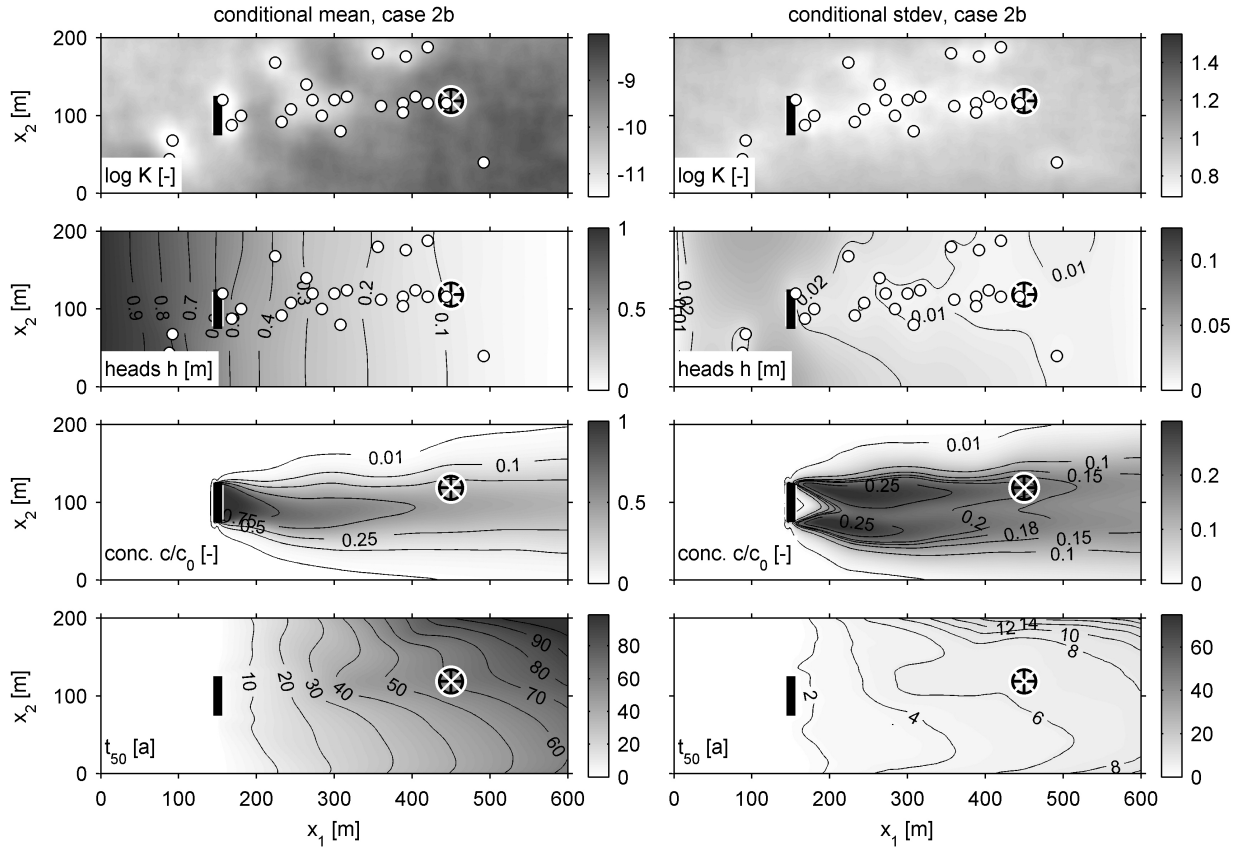

Figure 7. Results for case 2b. See Figure 5 for description.

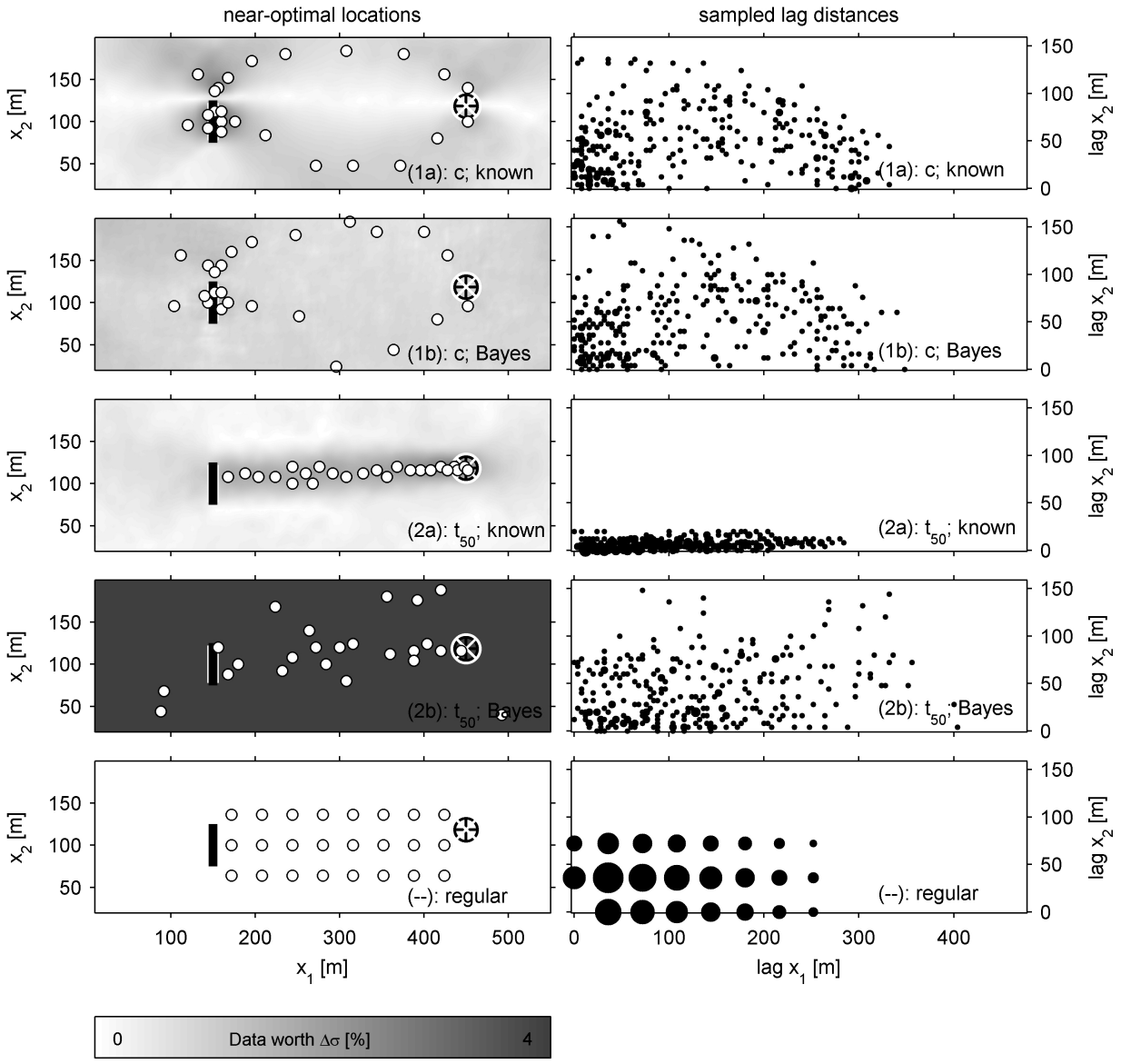

Figure 8. Left: Near-optimal design patterns for cases 1a-2b and a regular sampling grid. Right: respective sampled lag distances. Crossed circles (left): sensitive location. Solid white circles: 24 sampling locations; log-conductivity and hydraulic head measured jointly. Thick black line: hypothetical contaminant source. Grey-scale background: Maps of expected data worth (here: percent reduction of Bayesian predictive variance), evaluated before the first sample. Black dots (right): sampled lag distances. Dot area increases with multiple sampling of the same lag. Zero lag is not shown.

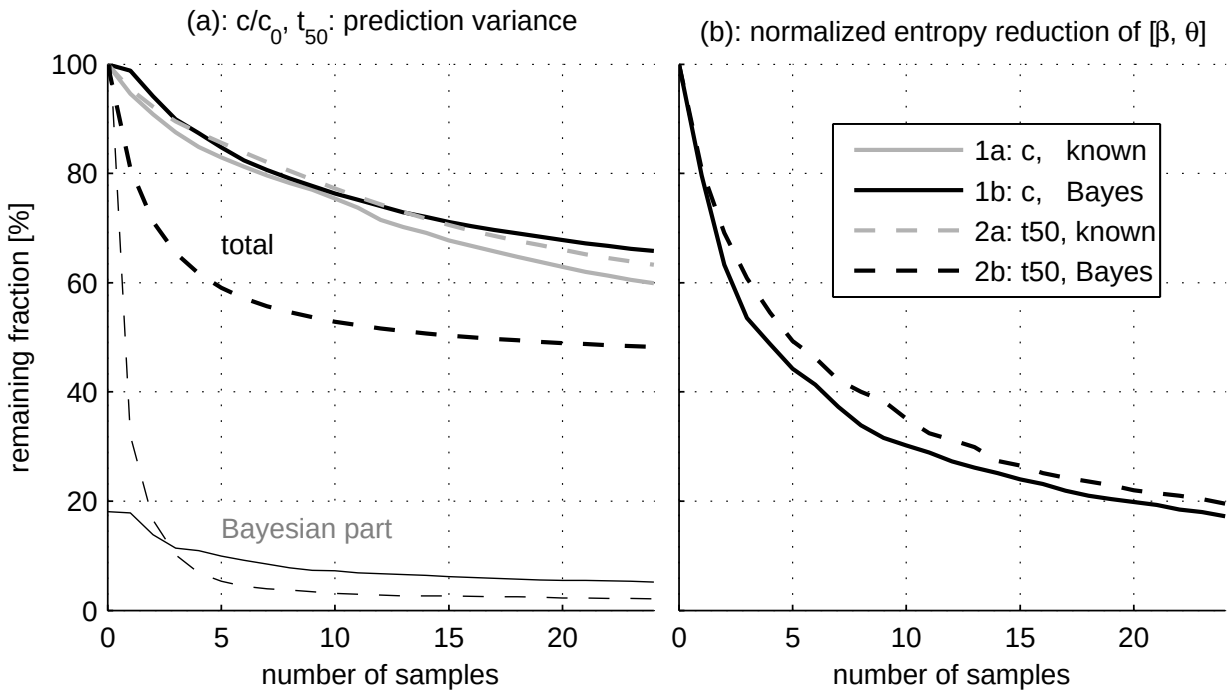

Figure 9. (a): Reduction of prediction variance with increasing number of samples, normalized to the initial prediction variance. Upper curve set ("total", thick lines): expected prediction variance of c (solid) and t_{50} (dashed) according to Eq. (12). Lower set of curves ("Bayesian part", thin lines): only second term of Eq. (12). (b): Relative entropy of structural parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ with increasing number of samples.

Table 1. Parameter values used for the synthetic test cases.

Numerical domain			
Domain size	$[L_1, L_2]$	$[m]$	[600, 200]
Grid spacing	$[\Delta_1, \Delta_2]$	$[m]$	[2, 0.5]
Transport parameters			
Head difference	Δh	$[m]$	1
Effective porosity	n_e	$[-]$	0.35
Pore-scale dispersivities	$[\alpha_\ell, \alpha_t]$	$[m]$	[2, 0.25]
Diffusion coefficient	D_m	$[m^2/s]$	10^{-9}
Transversal plume dimension	ℓ_S	$[m]$	50m
Geostatistical model parameters (prior mean values)			
Global mean	$\beta_1 = \ln K_g$	$[-]$	$\ln(10^{-5})$
Trend x_1	β_2	$[-]$	0
Trend x_2	β_3	$[-]$	0
Variance	σ_Y^2	$[-]$	1.00
Integral scales (see Eq. 10)	$[\lambda_1, \lambda_2]$	$[m]$	[15, 15]
Matérn's kappa (see Eq. 10)	κ	$[-]$	2.50
Measurement error standard deviations			
$Y \equiv \ln K$	$\sigma_{r,Y}$	$[-]$	1.00
Head h	$\sigma_{r,h}$	$[m]$	0.01
Dimensionless numbers			
Longitudinal travel distance source-target	$\xi = x_1/\lambda_1$	$[-]$	20.00
Transverse offset from center line	$\eta = x_2/\lambda_2$	$[-]$	1.2
Contaminant source width	$\zeta = \ell_s/\lambda_2$	$[-]$	3.33
Longitudinal Péclet	$Pe_\ell = \lambda_1/\alpha_\ell$	$[-]$	7.50
Transverse Péclet	$Pe_t = \lambda_2/\alpha_t$	$[-]$	60.00

Table 2. Definition of test cases in our scenario. Objective: the quantity to be minimized by sampling (prediction variance of contaminant concentration or of arrival time at the sensitive location, respectively). Symbols: β_1 [$-$]: global mean of $\ln K$; β_2 and β_3 [$-$]: linear trend parameters; λ_1 and λ_2 [m]: scale parameters (spatial correlation); κ [$-$]: shape parameter of the Matérn function.

Case Number	Objective	Assumptions	structural uncertainty
1a	σ_c^2	$\boldsymbol{\beta}, \boldsymbol{\theta}$ known	none
1b	σ_c^2	$\boldsymbol{\beta}, \boldsymbol{\theta}$ uncertain	$var\left[\beta_1, \beta_2, \beta_3, \sigma_Y^2, \lambda_1, \lambda_2, \kappa\right]$ $= [1, 1, 1, 0.5, 112.5, 112.5, 1.75]$
2a	σ_{t50}^2	$\boldsymbol{\beta}, \boldsymbol{\theta}$ known	none
2b	σ_{t50}^2	$\boldsymbol{\beta}, \boldsymbol{\theta}$ uncertain	$var\left[\beta_1, \beta_2, \beta_3, \sigma_Y^2, \lambda_1, \lambda_2, \kappa\right]$ $= [1, 1, 1, 0.5, 112.5, 112.5, 1.75]$

Table 3. Comparison of structural parameters: prior mean, synthetic reality and posterior mean values identified with synthetic data from case 1b. 95% confidence intervals are estimated from two times the posterior standard deviation, assuming a Gaussian distribution.

Parameter			prior mean (and 95% CI)	synthetic values	posterior mean (and 95% CI)
global mean	β_1	[−]	-9.32 (± 2)	-9.98	-9.50 (± 0.14)
trend x_1	β_2	[−]	0 (± 2)	+2.16	+2.24 (± 0.23)
trend x_2	β_3	[−]	0 (± 2)	-1.11	-0.39 (± 0.93)
variance	σ_Y^2	[−]	1.00 (± 1.41)	0.62	0.71 (± 0.53)
integral scale λ_1	λ_1	$[m]$	15.00 (±21.21)	21.53	21.49 (±15.00)
integral scale λ_2	λ_2	$[m]$	15.00 (±21.21)	29.63	24.92 (±15.06)
shape parameter	κ	[−]	2.50 (± 3.53)	3.89	2.12 (± 3.36)

Table 4. Performance index of different patterns in different cases

	case 1a	case 1b	case 2a	case 2b
pattern 1a	100%	60%		
pattern 1b	95%	100%		
pattern 2a			100%	98%
pattern 2b			68%	100%
regular grid	59%	75%	48%	96%