

A. M. Michalak · P. K. Kitanidis

A method for the interpolation of nonnegative functions with an application to contaminant load estimation

Abstract The objective of this work is to extend kriging, a geostatistical interpolation method, to honor parameter nonnegativity. The new method uses a prior probability distribution based on reflected Brownian motion that enforces this constraint. The work presented in this paper focuses on interpolation problems where the unknown is a function of a single variable (e.g. time), and is developed both for the case with and without measurement error in the available data. The algorithms presented for conditional simulations are computationally efficient, particularly in the case with no measurement error. We present an application to the interpolation of dissolved arsenic concentration data from the North Fork of the Humboldt River, Nevada.

Keywords Inference under constraints · Nonnegativity · Reflected Brownian motion · Geostatistics · Kriging · Gibbs sampler

1 Introduction

Interpolation and inverse modeling techniques are gaining increased exposure as researchers and practitioners strive to make optimal use of limited data (see, for example, Silverman 1985; Yakowitz and Szidarovsky 1985; Eubank 1988; Hardle 1990; Handcock and Stein 1993; Schimek 2000). In stochastic approaches to interpolation, the unknown parameters are described through statistical distributions, and meaningful confi-

dence intervals can often be identified. Kriging (see, e.g., Cressie 1991; Kitanidis 1997) is a geostatistical minimum-mean-squared error method of spatial or temporal prediction that usually depends on the second-order properties of the modeled process and that has been used extensively for interpolation in environmental applications.

One of the challenges of stochastic approaches is the need to select a statistical model for each application that is consistent with our physical understanding of the problem. In most environmental applications, data are limited and any information available about an unknown parameter or function should be used to improve the analysis. One useful piece of information is that, in many cases, the unknown parameter is nonnegative. Examples of such parameters from environmental applications include chemical concentration, hydraulic conductivity, porosity and storage coefficients. However, traditional kriging methods (e.g. simple and ordinary kriging) cannot enforce such a constraint, allowing confidence intervals about the best estimate as well as conditional realizations of the unknown function to have negative values. Nonnegativity-enforcing methods used in the past present certain limitations, as will be discussed in the next section. This paper presents a mathematically rigorous and computationally efficient methodology for enforcing parameter nonnegativity that extends ordinary kriging, thereby leading to physically reasonable estimates of parameter values.

2 Background

Many of the methods traditionally available for enforcing parameter nonnegativity in interpolation and inverse problems were derived for deterministic problems, and are not easily transferable to a stochastic setting. Although little literature on this topic is available for environmental applications, this matter has been explored in other fields such as image deblurring, investment portfolio optimization, and emission

A. M. Michalak (✉) · P. K. Kitanidis
Department of Civil and Environmental Engineering,
Stanford University,
Stanford, CA 94305-4020, USA
E-mail: Anna.Michalak@NOAA.gov

Present address: A.M. Michalak
Climate Monitoring and Diagnostics Laboratory (CMDL),
National Oceanic and Atmospheric Administration (NOAA),
Mailcode R/CMDL1, 325 Broadway,
Boulder, Colorado 80305-3328, USA

tomography, to name a few (see Vardi and Lee 1993 for a partial review of these methods).

The popularization of Markov chain Monte Carlo (MCMC) methods has allowed for stochastic methods for modeling positive functions. Applications in areas such as image restoration (Besag et al. 1995; Weir 1997), spatial modeling (Diggle et al. 1998) and disease mapping (Besag et al. 1991; Waller et al. 1997), for example, are becoming common. The developed methods involve computationally expensive numerical algorithms, but allow for the use of constrained probability density functions. Poisson or binomial distributions are often used to enforce nonnegativity constraints, for example. MCMC-based methods have also been applied in a kriging context in a few instances, such as, for example, in Diggle et al. (1998). In this work, spatially indexed Poisson counts and binomial counts were used instead of Gaussian distributions, thereby enforcing nonnegativity. Although the MCMC-based methods allow for modeling positive functions in many instances, they do not benefit from the computational efficiency associated with working in a multi-Gaussian setting. Partly as a result of this limitation, their application in practical settings has been relatively limited.

In environmental and geostatistical applications, methods for dealing with nonnegativity have been almost exclusively limited to (a) the transformation of the original variable, yielding a variable that is defined over an infinite domain but that corresponds to the original variable defined in the nonnegative domain only; and (b) the use of Lagrange multipliers to obtain nonnegative best estimates from probability density functions that would otherwise allow the function estimates to become negative.

Data transformations are the classical method for dealing with nonnegativity in statistics, and involve using a link function to transform the original variable. Such methods have been applied in geostatistical interpolation and inverse problems in, for example, Kitanidis and Shen (1996), Kitanidis (1997, p. 70) and Saito and Goovaerts (2000). The most common of these is the power transformation, which is defined as

$$\tilde{\mathbf{s}} = \begin{cases} (\mathbf{s}^\kappa - 1)/\kappa & \kappa > 0 \\ \ln(\mathbf{s}) & \kappa = 0 \end{cases} \quad (1)$$

where \mathbf{s} is the vector of values in the original domain, $\tilde{\mathbf{s}}$ is the transformed data vector, and κ is a constant selected based on the application. The commonly used logarithmic transformation is included as a special case of the power transformation, obtained at the limit of κ tending to zero. For example, it is common in hydrogeologic analysis to use the logarithm of the hydraulic conductivity field (see, for example, Hoeksema and Kitanidis 1985). A few alternate data transformation methods have also been proposed for environmental optimization problems (see, e.g., Kauffmann and Kinzelbach 1989).

In general, however, data transformations lead to highly non-symmetric probability density functions for the unknown function values in the untransformed

space, which is unrealistic for certain physical systems. Such difficulties have been documented, among others, by Snodgrass and Kitanidis (1997). For example, for systems where the uncertainty is not expected to be a function of the magnitude of the modeled parameter, the confidence intervals obtained using a power transformation tend to be unreasonably narrow for low values of the parameter and wide for high values. In addition, some data transformation methods, such as the logarithmic transformation, have problems dealing with measurements of zero.

The other popular method is the use of Lagrange multipliers for obtaining nonnegative best estimates of unknown functions. Lagrange multipliers have been applied to bounded geostatistical interpolation in, for example, Barnes and You (1992). This method amounts to restricting a common Gaussian process by adding, *a posteriori*, a constraint that the process is greater than or equal to zero everywhere. The method consists of replacing the original objective function $f(\mathbf{s})$ by the Lagrange function

$$h(\mathbf{s}, \boldsymbol{\lambda}) = f(\mathbf{s}) - \sum_{i=1}^k \lambda_i [g_i(\mathbf{s}) - b_i] \quad (2)$$

where \mathbf{s} must satisfy the constraints $g_i(\mathbf{s}) = b_i$ or $g_i(\mathbf{s}) \geq b_i$, k is the total number of active constraints, and $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_k)$ are Lagrange multipliers. The solution method involves setting the derivative of the Lagrange function with respect to s_j and λ_i equal to zero. For inequality constraints, the Lagrange multipliers of the points corresponding to active constraints must be positive (see, e.g., Gill et al. 1986). If all that is required is a single estimate of the unknown function, then the use of Lagrange multipliers can be justified as a means to make this estimate nonnegative. When confidence intervals or conditional realizations are needed, however, the applicability of this method is not always clear. Specifically, if confidence intervals are specified using only a variance, this tends to imply a Gaussian distribution and the confidence intervals will therefore not honor the constraints. Furthermore, if Lagrange multipliers are used to constrain conditional realizations, the form of the posterior distribution from which these realizations are sampled, and whether they are equiprobable, is not clear. Finally, the conditional realizations will tend to have a relatively large number of “zero” values (at locations where the constraint is active), which is unrealistic for some applications. For example, in the case of geostatistical kriging where a multigaussian model is assumed *a priori*, the probability of occurrence of several zero values is very low, and the conditional realizations will therefore not have the spatial covariance structure specified by the underlying model.

In short, data transformations and Lagrange multipliers have the advantage of preserving some of the computational efficiency associated with kriging, but they have inevitable disadvantages relative to the use of a non-transformed Gaussian model.

Contrary to data transformations and the use of Lagrange multipliers, the method presented in this paper is based on a prior probability density function that has a non-zero value only in the nonnegative parameter range. This pdf is derived based on the method of images applied to reflected Brownian motion. For negligible measurement error, an exact solution is presented that has computational costs comparable to those associated with generating conditional realizations using ordinary kriging. In addition, even if there is some measurement error, this method is empirically shown to be a good approximation to the exact solution. The exact solution in the case where measurement error is taken into account does rely on an MCMC approach to obtain an exact estimate, but the algorithm is only applied at measurement locations, thereby maximizing computational efficiency.

3 Objective

The objective of this work is to develop a methodology based on kriging for enforcing parameter nonnegativity in interpolation problems. The proposed method is to behave similarly to a Gaussian process with a linear variogram (or Brownian motion) for parameter values significantly greater than zero. The properties of the linear variogram make it amenable to the development of the method presented here. In addition, the linear variogram imposes the least number of assumptions on the function being estimated, making it one of the most commonly used models in kriging and inverse modeling. Furthermore, as a Markov process, Brownian motion has convenient properties that are useful, for example, in the development of efficient methods for the generation of conditional realizations. Therefore, the method preserves much of the computational efficiency associated with ordinary kriging, while at the same time enforcing the nonnegativity constraint.

The work presented in this paper focuses on interpolation problems where the unknown is a function of a single variable (e.g. time). The main practical contribution of this work is the ability to efficiently generate realizations of unknown nonnegative functions, conditional on available measurements. Ensemble properties of these conditional realizations can then be used to determine statistical properties of the unknown function, such as a best estimate and confidence intervals.

The remainder of the paper is organized as follows. Section 4 presents the development of the statistical model. Section 5 presents the interpolation methodology for the case where measurement error is not considered. Specifically, Sect. 5.2 presents an algorithm for optimizing the structural parameter D which represents the variance of the process being modeled, and Sect. 5.3 presents an algorithm for generating conditional realizations of the unknown function. Section 6 presents the interpolation methodology for the case where measurement error is considered. Specifically, Sect. 6.2 presents an algorithm for optimizing the structural parameter D , and Sect. 6.3

presents the algorithm for generating conditional realizations. In Sect. 7, the method is applied to the interpolation of dissolved arsenic concentration data from the North Fork of the Humboldt River, Nevada. Finally, Sect. 8 presents conclusions and future directions. In a second paper, we have developed the tools necessary for the application of the method to geostatistical inverse modeling (Michalak and Kitanidis 2003).

4 Model Development

This section covers the development of the probability density function that is used for the stochastic interpolation application.

4.1 Governing partial differential equation

Brownian motion is a continuous time, continuous state space, simple Markov process. The probability law governing the transitions is stationary in time and the probability density function describing a parameter s at time t does not depend on the initial time (Karlin and Taylor 1975, p. 340). The partial differential equation satisfied by the probability density function of the location of a particle undergoing Brownian motion with no drift can be derived by starting from a random walk perspective (see Zauderer 1989 for a derivation). The limiting partial differential equation is:

$$\frac{\partial p}{\partial t} - D \frac{\partial^2 p}{\partial s^2} = 0 \quad (3)$$

where p is the probability density function of a particle location, and D is the variance of the process. For the application that we are interested in, the probability density function will describe the distribution of concentration values s , as a function of time t . Note that the derivation and method presented in this work would be identical for a process varying in space rather than time.

4.2 Selection of probability density function

The PDE defined in Eq. 3, subject to the boundary and initial conditions

$$p(s, T_i) = \delta(s - s_i) \quad p(s \rightarrow \pm\infty, t_o) = 0 \quad (4)$$

where T_i is initial time, t_o is a later time and $\delta(s - s_i)$ is the Dirac function, is satisfied by the probability density function:

$$p(s_o, t_o) = p(s_o | s_i) = \frac{1}{\sqrt{4\pi D(t_o - T_i)}} \exp\left(-\frac{(s_o - s_i)^2}{4D(t_o - T_i)}\right) \quad (5)$$

where s_o is located at a point t_o , the parameter value s_i at time T_i is known, and the vertical bar means ‘‘given’’; for example $p(s_o | s_i)$ represents the probability of s_o given s_i .

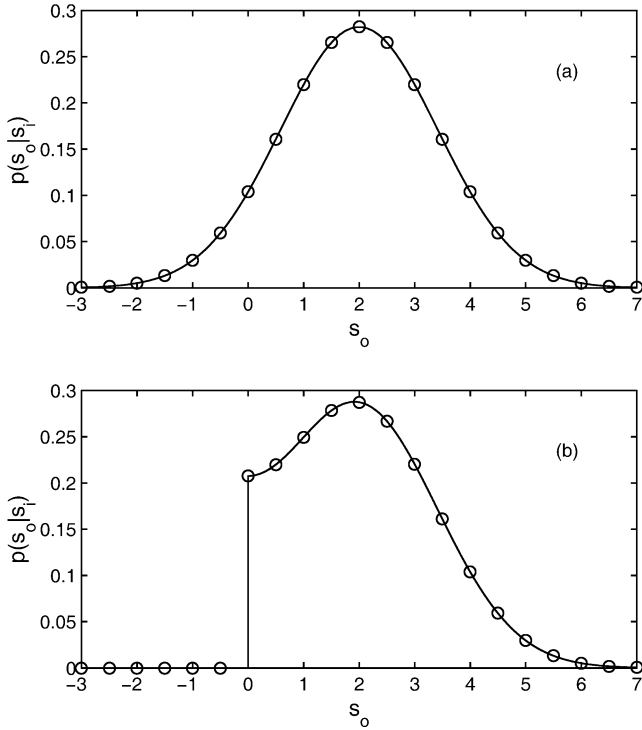


Fig. 1 Probability density functions for unrestricted and reflected Brownian motion with $s_i = 2$, variance $D = 1$, and separation distance $(t_o - T_i) = 1$. **a** Gaussian probability density function (unrestricted Brownian motion); **b** Method of images derived probability density function (reflected Brownian motion)

This pdf is plotted in Fig. 1a for the case of $D = 1$, $(t_o - T_i) = 1$, and $s_i = 2$. This pdf represents unrestricted Brownian motion, and is the one assumed when using kriging with a linear variogram model. A sample unconditional realization obtained using a Markov chain following this probability density function is presented in Fig. 2a. Note that, throughout this paper, T will be used to denote points in time at which observations are available, whereas t will denote other points in the discretized unknown function.

We now change the boundary conditions to:

$$\left. \frac{\partial p}{\partial s} \right|_{s=0, t_o} = 0 \quad p(s \rightarrow \infty, t_o) = 0 \quad p(s < 0, t_o) = 0 \quad (6)$$

The solution is obtained by adding an image pdf, with a mean located at $-s_i$, and setting the probability of $s < 0$ equal to zero (Karlin and Taylor 1975, p. 352–354):

$$p(s_o | s_i) = \begin{cases} \frac{1}{\sqrt{4\pi D(t_o - T_i)}} \left[\exp\left(-\frac{(s_o - s_i)^2}{4D(t_o - T_i)}\right) + \exp\left(-\frac{(s_o + s_i)^2}{4D(t_o - T_i)}\right) \right] & \text{for } s_o \geq 0 \\ 0 & \text{for } s_o < 0 \end{cases} \quad (7)$$

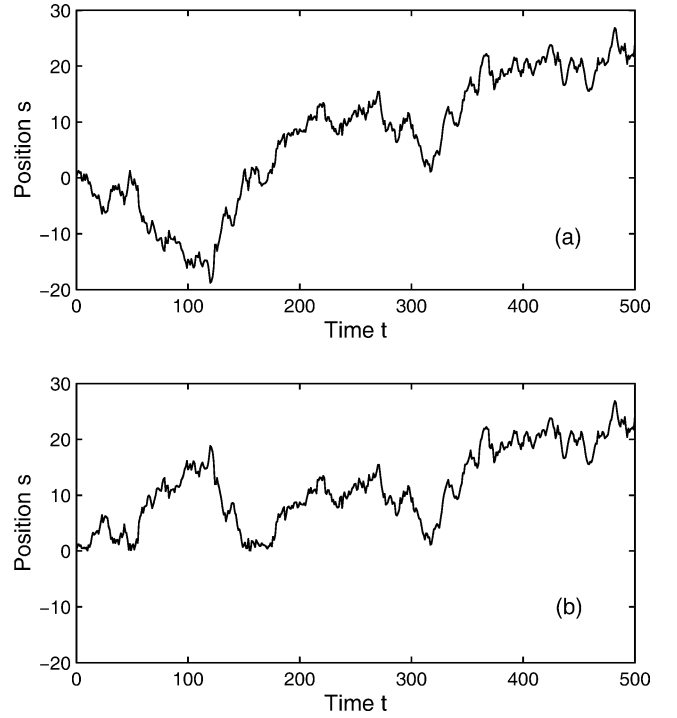


Fig. 2 Sample unconditional realizations with variance $D = 1$. **a** Realization generated using linear variogram model; **b** Realization generated using nonnegativity enforcing probability density function

This pdf is plotted in Fig. 1b for the case of $D = 1$, $(t_o - T_i) = 1$, and $s_i = 2$. This pdf represents reflected Brownian motion about the boundary $s = 0$, and is the one that is used in the current work.

A sample unconditional realization obtained using this distribution is presented in Fig. 2b. Equation 7 suggests that, in order to generate a realization of s_o given s_i , one can simply generate a Gaussian variate with mean s_i and variance $2D(t_o - T_i)$, and take its absolute value. Therefore, an unconditional realization can be obtained by generating an unconditional realization of a simple Brownian motion according to Eq. 3 and reflecting any negative values about the $s = 0$ axis. Note that the realization in Fig. 2b has the same random number seed as the realization presented in Fig. 2a.

5 Interpolation Equations – Error-free Observations

We are interested in the interpolation of a function to a certain set of points, given sparse but exact measurements at other points. Because we are working within a Brownian motion framework, only measurements at points immediately adjacent to the point at which we wish to estimate the function are relevant, if we assume that available measurements are error-free. For example, given measurements s_i and s_{i+1} at times

T_i and T_{i+1} , respectively, we are interested in the best estimate of, and confidence intervals about, the parameter value s_o at point t_o , or, more generally, in the probability density function of s_o at time t_o , where $T_i \leq t_o \leq T_{i+1}$.

5.1 Derivation of probability density function

The probability density function of s_o given s_i and s_{i+1} can be defined in terms of the joint probability of s_o , s_i , and s_{i+1} , and the joint probability of s_i and s_{i+1} :

$$p(s_o | s_i, s_{i+1}) = \frac{p(s_o, s_i, s_{i+1})}{p(s_i, s_{i+1})} \quad (8)$$

Using standard manipulations and the assumption of conditional independence (i.e. $p(s_{i+1} | s_o, s_i) = p(s_{i+1} | s_o)$), this expression becomes:

$$p(s_o | s_i, s_{i+1}) = \frac{p(s_{i+1} | s_o) p(s_o | s_i)}{p(s_{i+1} | s_i)} \quad (9)$$

Expressed in terms of the measurements and their locations as described in Eq. 7, this pdf becomes:

$$p(s_o | s_i, s_{i+1}) = \begin{cases} C \left(\exp\left(-\frac{s_{i+1}-s_o}{4D(T_{i+1}-t_o)} - \frac{(s_o-s_i)^2}{4D(t_o-T_i)}\right) \right. \\ \quad + \exp\left(-\frac{(s_{i+1}-s_o)^2}{4D(T_{i+1}-t_o)} - \frac{(s_o+s_i)^2}{4D(t_o-T_i)}\right) \\ \quad + \exp\left(-\frac{(s_{i+1}+s_o)^2}{4D(T_{i+1}-t_o)} - \frac{(s_o-s_i)^2}{4D(t_o-T_i)}\right) \\ \quad \left. + \exp\left(-\frac{(s_{i+1}+s_o)^2}{4D(T_{i+1}-t_o)} - \frac{(s_o+s_i)^2}{4D(t_o-T_i)}\right) \right) & \text{for } s_o \geq 0 \\ 0 & \text{for } s_o < 0 \end{cases} \quad (10)$$

where the constant C is:

$$C = \frac{\sqrt{\frac{(T_{i+1}-T_i)}{4\pi D(T_{i+1}-t_o)(t_o-T_i)}}}{\exp\left(-\frac{(s_{i+1}-s_i)^2}{4D(T_{i+1}-T_i)}\right) + \exp\left(-\frac{(s_{i+1}+s_i)^2}{4D(T_{i+1}-T_i)}\right)} \quad (11)$$

5.2 Determination of parameter D

In some applications, D is known. Otherwise, it can be estimated from the available data. We have shown that the probability density function of parameter value s_{i+1} at observation time T_{i+1} given the value of D and the value of the parameter at an adjacent observation point s_i is:

$$p(s_{i+1} | s_i, D) = \begin{cases} \frac{1}{\sqrt{4\pi D(T_{i+1}-T_i)}} \left[\exp\left(-\frac{(s_{i+1}-s_i)^2}{4D(T_{i+1}-T_i)}\right) \right. \\ \quad \left. + \exp\left(-\frac{(s_{i+1}+s_i)^2}{4D(T_{i+1}-T_i)}\right) \right] & \text{for } s_{i+1} \geq 0 \\ 0 & \text{for } s_{i+1} < 0 \end{cases} \quad (12)$$

In this case, $i = 1, \dots, n-1$, where n is the total number of observations.

Based on this pdf and the observation data, we can identify the maximum likelihood estimate for D , again assuming conditional independence (which is dictated by the Brownian motion model). In other words, we seek the value of D that maximizes:

$$p(\mathbf{s} | D) = \prod_{i=1}^{n-1} p(s_{i+1} | s_i, D) \quad (13)$$

A search algorithm such as the Newton-Raphson or Secant method (see, for example, Gill et al. 1986) can be implemented to identify the mode of this pdf, corresponding to the maximum likelihood estimate of D . Although the uncertainty in parameter D could be taken into account by using a fully Bayesian approach, generating conditional realizations of the unknown function would then be significantly more computationally expensive relative to the algorithm that will be presented in the following section. Moreover, uncertainty in hyperparameters such as D has been shown to have a relatively minor effect in geostatistical interpolation (see, for example, Kitanidis 1986) for applications with sample sizes that warrant the use of geostatistical interpolation methods. As a result, a single value of hyperparameters is typically used in practical kriging applications, estimated using either an experimental variogram or statistical approaches such as the maximum likelihood approach described here.

5.3 Conditional realizations

The method developed for the generation of conditional realizations takes advantage of the Markovian properties of reflected Brownian motion, and is an exact method with computational time requirements that are comparable to those of generating conditional realizations using kriging with an unknown mean. The method does not require numerical integration of the pdf, nor does it rely on computationally expensive numerical methods such as MCMC methods for sampling the pdf.

The method involves using Bernoulli trials to generate a realization of observations composed of a combination of the actual observations and their reflections about the axis $s = 0$, as will be described in this section. Figure 3 shows measurements, as well as a sample

realization of the observations, as used for the conditional realization generation. A conditional unconstrained realization $\mathbf{s}_{cu,l}$ reproducing this modified set of observations is then generated using the linear variogram kriging model, where l denotes the l -th realization. A review of linear variogram kriging with an unknown mean is presented in the Appendix. The dotted line in Fig. 3 shows a conditional realization satisfying the observation realization. Finally, because Brownian motion is a Markov process, the probability of a given path is exactly equal to the probability of that path reflected against a boundary. Using this property, it is possible to reflect all negative components of the conditional realizations about the boundary $s = 0$, thereby generating a nonnegative realization $\mathbf{s}_{cc,l}$ that reproduces the original nonnegative set of observations. The solid line in Fig. 3 shows the final constrained conditional realization, satisfying the original set of observations.

We will now show that this algorithm is consistent with the statistical model presented in Eq. 10, and deal with the remaining question of how to generate a realization of observations (that are either positive, negative, or a combination thereof) in a manner consistent with the assumed statistical model.

The probability density function for the parameter value s_o at point t_o , defined in Eq. 10, can be rewritten as follows for $s_o \geq 0$:

$$p(s_o | s_i, s_{i+1}) = \frac{1}{\sqrt{2\pi\sigma_s^2}} \left\{ P_1 \left[\exp\left(-\frac{(s_o - \mu_{P,1})^2}{2\sigma_P^2}\right) + \exp\left(-\frac{(s_o + \mu_{P,1})^2}{2\sigma_P^2}\right) \right] + P_2 \left[\exp\left(-\frac{(s_o - \mu_{P,2})^2}{2\sigma_P^2}\right) + \exp\left(-\frac{(s_o + \mu_{P,2})^2}{2\sigma_P^2}\right) \right] \right\} \quad (14)$$

where

$$P_1 = \frac{\exp\left(-\frac{(s_{i+1} - s_i)^2}{4D(T_{i+1} - T_i)}\right)}{\exp\left(-\frac{(s_{i+1} - s_i)^2}{4D(T_{i+1} - T_i)}\right) + \exp\left(-\frac{(s_{i+1} + s_i)^2}{4D(T_{i+1} - T_i)}\right)} \quad (15)$$

$$P_2 = 1 - P_1 \quad (16)$$

$$\mu_{P,1} = \frac{s_{i+1}(t_o - T_i) + s_i(T_{i+1} - t_o)}{T_{i+1} - T_i} \quad (17)$$

$$\mu_{P,2} = \frac{s_{i+1}(t_o - T_i) - s_i(T_{i+1} - t_o)}{T_{i+1} - T_i} \quad (18)$$

$$\sigma_P^2 = 2D\left(\frac{(t_o - T_i)(T_{i+1} - t_o)}{T_{i+1} - T_i}\right) \quad (19)$$

By looking at the values of $\mu_{P,1}$ and $\mu_{P,2}$, the four terms in this probability density function can be

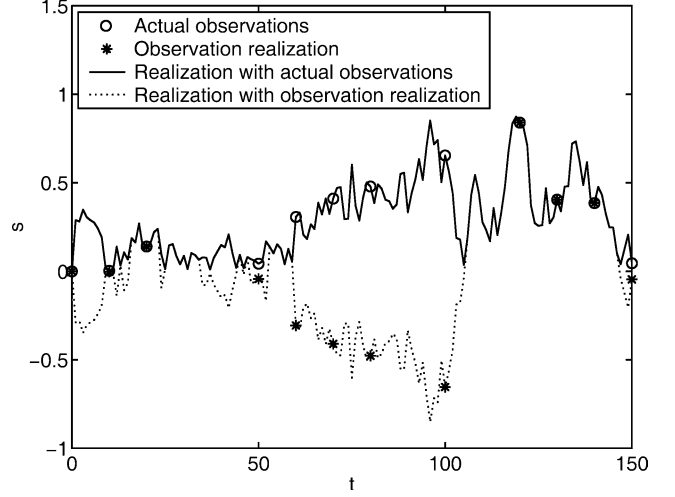


Fig. 3 Efficient conditional realization generated using non-negativity enforcing method. Markers indicate actual observations and observation realization as used for conditional generation with nonnegativity enforcing method. Dotted and solid lines represent conditional realization before and after enforcement of nonnegativity

interpreted as follows. If there was no constraint on s , there would be four possible scenarios for successive observations (s_{i+1} and s_i). The first possibility corresponds to both observations having positive values. The second possibility is the mirror image of the first, i.e., both observations have negative signs. From Eq. (14), the joint probability of these two possibilities is P_1 . The third possibility is for the first observation to be negative and the second positive. The fourth is the mirror image of the third, with the first observation being negative and the second positive. The joint probability of these two possibilities is P_2 .

Thus, for each conditional realization to be generated, where l denotes the l -th realization, the algorithm proceeds as follows:

1. Generate a realization of the observations, keeping magnitude constant, but randomly selecting a sign for each observation. Start with the original set of observations s_i , $i = 1, \dots, n$ (all nonnegative). Keep the sign of the first observation positive. For $i = 2, \dots, n$:
 - (a) Generate a uniformly distributed random number α in the range $[0, 1]$.
 - (b) If $\alpha \leq P_1$, $s_{i,l}$ has the same sign as $s_{i-1,l}$. Otherwise, $s_{i,l}$ has the opposite sign of $s_{i-1,l}$.
2. Generate an unconditional unconstrained realization $\mathbf{s}_{uu,l}$, and condition it on the modified observations according to the ordinary kriging algorithm with a linear variogram (see Appendix) yielding a conditional, unconstrained realization, $\mathbf{s}_{cu,l}$.
3. Take the absolute value of the realization to obtain a conditional constrained realization, $\mathbf{s}_{cc,l}$, of the unknown function.

6 Interpolation Equations – with Effect of Measurement Error

As in the previous case, we are interested in the interpolation of a given parameter to a certain set of points, given sparse measurements at other points. However, we are now taking into account the effect of imperfect measurements at the observation locations. Unlike in the previous case, the value of the function \mathbf{s} at each point will depend on its value at every other point, not simply at adjacent points.

6.1 Derivation of probability density function

Assuming a random, normally distributed, additive error, the model becomes:

$$y_i = s_i + \varepsilon_i \quad (20)$$

where s_i is the actual parameter value at the measurement time T_i , ε_i is a normally distributed random measurement error with mean 0 and variance σ_ε^2 , and y_i is the observed measurement at time T_i . The errors ε_i are independent of one-another and of the s values.

We are interested in deriving the probability density function of the function \mathbf{s} at times t_j where $j = 1, \dots, m$, given imperfect measurements y_i at times T_i , where $i = 1, \dots, n$, which can be expressed as $p(\mathbf{s}|\mathbf{y})$. If we discretize the function \mathbf{s} on a grid that includes the observation points, using standard manipulations and assuming conditional independence (i.e. $p(\mathbf{s}^{(y)}|\mathbf{s}^{(y)}, \mathbf{y}) = p(\mathbf{s}^{(y)}|\mathbf{s}^{(y)})$), it can be shown that

$$p(\mathbf{s}|\mathbf{y}) = p(\mathbf{s}^{(y)}, \mathbf{s}^{(y)}|\mathbf{y}) = p(\mathbf{s}^{(y)}|\mathbf{s}^{(y)})p(\mathbf{s}^{(y)}|\mathbf{y}) \quad (21)$$

where $\mathbf{s}^{(y)}$ are function values at discretization points that correspond to observation locations, and $\mathbf{s}^{(y)}$ are all other discretization points. Thus, the problem of finding the posterior of the complete \mathbf{s} is naturally broken into two parts: (i) find the posterior pdf of $\mathbf{s}^{(y)}$, and (ii) find the pdf of $\mathbf{s}^{(y)}$ given $\mathbf{s}^{(y)}$. This expression can be further manipulated to yield:

$$\begin{aligned} p(\mathbf{s}|\mathbf{y}) &\propto p(\mathbf{s}^{(y)}|\mathbf{s}^{(y)})p(\mathbf{y}|\mathbf{s}^{(y)})p(\mathbf{s}^{(y)}) \\ &\propto p(\mathbf{y}|\mathbf{s}^{(y)})p(\mathbf{s}) \end{aligned} \quad (22)$$

Therefore, the posterior distribution for the full vector \mathbf{s} is:

$$\begin{aligned} p(\mathbf{s}|\mathbf{y}) &\propto \prod_{i=1}^n \exp\left(-\frac{1}{2} \frac{(y_i - s_i^{(y)})^T (y_i - s_i^{(y)})}{\sigma_\varepsilon^2}\right) \\ &\cdot \prod_{j=1}^{m-1} \exp\left(-\frac{(s_{j+1} - s_j)^2}{4D(t_{j+1} - t_j)}\right) \end{aligned}$$

$$+ \exp\left(-\frac{(s_{j+1} + s_j)^2}{4D(t_{j+1} - t_j)}\right) \quad (23)$$

6.2 Determination of parameter D

We must modify the method used to determine the optimal value of parameter D because the available measurements no longer correspond to exact values of function \mathbf{s} . We now want to find the maximum likelihood estimate of D given the measurement vector \mathbf{y} . Estimating D directly would require an n -dimensional integration, which is not practical, especially when a large number of measurements is available. Therefore, an iterative Expectation-Maximization (EM) scheme is implemented to identify the optimal value of parameter D .

The EM approach is a general iterative method for computing the mode of the marginal pdf of a parameter D from the joint pdf of D and \mathbf{s} (Dempster et al. 1977; McLachlan and Krishnan 1997). The method proceeds by starting with an initial guess of D , and generating a large number N of conditional realizations of \mathbf{s} at the observation locations using the algorithm in Sect. 6.3. The k th estimate of D , $D^{(k)}$, is then updated to $D^{(k+1)}$ by finding the value of D that maximizes:

$$\varphi(D, D^{(k+1)}) = \frac{1}{N} \sum_{l=1}^N \ln f(\mathbf{s}_{cc,l}, D|\mathbf{y}) \quad (24)$$

where $f(\mathbf{s}_{cc,l}, D|\mathbf{y})$ is the posterior pdf of the l -th conditional realization up to a normalizing constant, as defined in Eq. 23. The procedure continues until the iterations converge on the maximum likelihood estimate of D . Note that only realizations at the observation locations are needed for the iterations of the EM algorithm because these are the points that define the variance of the process.

6.3 Conditional realizations

Unlike in the case with no measurement error, the value of the unknown function at each point is dependent on the values at all other points, and the method derived in Sect. 5.3 will not be directly applicable. However, a very efficient approximate approach as well as an efficient MCMC-based exact approach are possible. Ensemble properties of conditional realizations can then be used to define best estimates and confidence intervals.

Reflected Brownian motion approach

In this approximate approach, conditional realizations are generated in a method analogous to the reflected Brownian motion method used for generating conditional

realizations in the measurement-error-free case. A realization of the observations is generated identically as in the no-measurement-error case (see Sect. 5.3). An unconstrained conditional realization $\mathbf{s}_{cu,l}$ is generated using kriging with a linear variogram (with measurement error), as described in the Appendix. Finally, we reflect all negative components of the conditional realizations about the boundary $s = 0$, thereby generating nonnegative conditional realizations, $\mathbf{s}_{cc,l}$.

For the example presented in the application that follows, the best estimates and confidence intervals obtained from conditional realizations generated using this approximate approach were indistinguishable to the naked eye from those obtained using the MCMC approach which will be presented next. For very large measurement error variances (when the standard deviation of the error approaches the magnitude of the measurement values themselves), however, slight deviations in the confidence intervals can be observed. Therefore, for practical applications, the method developed in Sect. 5.3 can also be applied in cases where the observations have measurement error, taking care to incorporate the measurement error in the kriging approach when generating the conditional realizations.

MCMC approach

A Gibbs sampling algorithm (see, for example, Casella and George 1992) was implemented to obtain exact conditional realizations of the unknown function. The Gibbs algorithm samples the marginal posterior pdf's at successive discretization points of an unknown function. In order to improve on the efficiency of this technique, we take explicit advantage of the property derived in Sect. 6.1, namely the fact that the problem of finding the posterior of the complete \mathbf{s} is naturally broken into two parts: (i) find the posterior pdf of $\mathbf{s}^{(y)}$, and (ii) find the pdf of $\mathbf{s}^{(y)}$ given $\mathbf{s}^{(y)}$. We therefore use the Gibbs algorithm to obtain realizations at the observation locations, according to the posterior:

$$\begin{aligned}
 p(\mathbf{s}^{(y)}|\mathbf{y}) &\propto \prod_{i=1}^n \exp\left(-\frac{(y_i - s_i^{(y)})^2}{2\sigma_\varepsilon^2}\right) \\
 &\cdot \prod_{j=1}^{n-1} \left[\exp\left(-\frac{(s_{j+1}^{(y)} - s_j^{(y)})^2}{4D(T_{j+1} - T_j)}\right) \right. \\
 &\left. + \exp\left(-\frac{(s_{j+1}^{(y)} + s_j^{(y)})^2}{4D(T_{j+1} - T_j)}\right) \right] \quad (25)
 \end{aligned}$$

Once realizations at observation points have been obtained, a conditional realization at all discretization points is generated using the Reflected Brownian Motion approach outlined in Sect. 5.3. Because the conditional realization at the measurement points

already takes into account measurement error, measurement error need not be incorporated a second time, and the error-free implementation in Sect. 5.3 is therefore appropriate.

The marginal posterior of a single point $s_i^{(y)}$ in the nonnegative range ($s_i^{(y)} \geq 0$) can be rewritten as the sum of four Gaussian distributions:

$$p(s_i^{(y)}|\mathbf{s}^{(y)}, \mathbf{y}) \propto \sum_{j=1}^4 \frac{K_j}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \frac{(s_i^{(y)} - \mu_j)^2}{\sigma^2}\right) \quad (26)$$

where,

$$\sigma^2 = \frac{\sigma_P^2 \sigma_\varepsilon^2}{\sigma_P^2 + \sigma_\varepsilon^2} \quad (27)$$

$$\begin{aligned}
 K_1 &= P_1 \exp\left(-\frac{1}{2} \frac{(y_i - \mu_{P,1})^2}{\sigma_\varepsilon^2 + \sigma_P^2}\right) & \mu_1 &= \tau\left(\frac{\mu_{P,1}}{\sigma_P^2} + \frac{y_i}{\sigma_\varepsilon^2}\right) \\
 K_2 &= P_1 \exp\left(-\frac{1}{2} \frac{(y_i + \mu_{P,1})^2}{\sigma_\varepsilon^2 + \sigma_P^2}\right) & \mu_2 &= \tau\left(-\frac{\mu_{P,1}}{\sigma_P^2} + \frac{y_i}{\sigma_\varepsilon^2}\right) \\
 K_3 &= P_2 \exp\left(-\frac{1}{2} \frac{(y_i - \mu_{P,2})^2}{\sigma_\varepsilon^2 + \sigma_P^2}\right) & \mu_3 &= \tau\left(\frac{\mu_{P,2}}{\sigma_P^2} + \frac{y_i}{\sigma_\varepsilon^2}\right) \\
 K_4 &= P_2 \exp\left(-\frac{1}{2} \frac{(y_i + \mu_{P,2})^2}{\sigma_\varepsilon^2 + \sigma_P^2}\right) & \mu_4 &= \tau\left(-\frac{\mu_{P,2}}{\sigma_P^2} + \frac{y_i}{\sigma_\varepsilon^2}\right)
 \end{aligned} \quad (28)$$

where $P_1, P_2, \mu_{P,1}, \mu_{P,2}$, and σ_P^2 are as defined in (15)–(19).

This formulation suggests an efficient method for generating realizations from the marginal distribution of $s_i^{(y)}$. We know that the realization will be drawn from the nonnegative portion of one of these four Gaussian distributions. If there were no constraints, the probability of drawing from each Gaussian would be proportional to the value of its K_j . Therefore, a uniformly distributed random number α in the range $[0, 1]$ can be generated to choose a distribution, based on each Gaussian distribution's K_j value, normalized by the sum $\sum_{j=1}^4 K_j$. We are still only interested in sampling the nonnegative portion of this chosen Gaussian distribution. Therefore, we draw a number from this distribution, and, if it is nonnegative, we keep it. Otherwise, we draw another random number α , and select one of the Gaussian distributions anew. Once we obtain a nonnegative sample, this realization of $s_i^{(y)}$ is used as the next conditional constrained realization at point t_i , denoted $s_{cc,l}(t_i)$.

The overall algorithm proceeds as follows. We define the l -th conditional constrained realization as $\mathbf{s}_{cc,l}$. The chain can be initialized with any realization that has a non-zero posterior probability, $\mathbf{s}_{cc,0}$.

1. Initialize the iteration counter of the chain $l = 1$.
2. Obtain a new value $\mathbf{s}_{cc,l}(T) = (s_{cc,l}(T_1), s_{cc,l}(T_2), \dots, s_{cc,l}(T_n))'$ at the observation locations from $\mathbf{s}_{cc,l-1}$ through successive generation of values from the

marginal pdf at each observation point, using the most recent realization available at each other point. For example, noting the use of counter values l and $l - 1$:

$$\begin{aligned} p(s_{cc,l}(T_1)) &= p(s_{cc,l}(T_1)|s_{cc,l-1}(T_2), \dots, s_{cc,l-1}(T_n)) \\ p(s_{cc,l}(T_i)) &= p(s_{cc,l}(T_i)|s_{cc,l}(T_1), \dots, s_{cc,l}(T_{i-1}), s_{cc,l-1}(T_{i+1}), \dots, s_{cc,l-1}(T_n)) \\ p(s_{cc,l}(T_n)) &= p(s_{cc,l}(T_n)|s_{cc,l}(T_1), \dots, s_{cc,l}(T_{n-1})) \end{aligned}$$

This step proceeds as follows:

- (a) Generate a uniformly distributed random number α in the range $[0, 1]$.
 If $\alpha \sum_{j=1}^4 K_j < K_1$, $\mu = \mu_1$;
 If $K_1 < \alpha \sum_{j=1}^4 K_j < (K_1 + K_2)$, $\mu = \mu_2$;
 If $(K_1 + K_2) < \alpha \sum_{j=1}^4 K_j < (K_1 + K_2 + K_3)$,
 $\mu = \mu_3$;
 Otherwise, $\mu = \mu_4$.
 - (b) Generate a normally distributed random number γ with mean μ and variance τ (Eq. 27).
 - (c) If $\gamma < 0$, return to Step 2a. Otherwise,
 $s_{cc,l}(T_i) = \gamma$.
 - (d) Repeat this process for each observation point,
 $i = 1, \dots, n$.
3. Generate an unconditional realization $s_{uu,l}$ at all discretizations points t , and condition it on the values $s_{cc,l}(T)$ using the error-free Reflected Brownian Motion method described in Sect. 5.3 to obtain a full conditional constrained realization $s_{cc,l}$.
 4. Change counter l to $l + 1$ and return to step 2.

Because the chain is started with an arbitrary realization that is not necessarily representative of the intended pdf (Eq. 23), a block of the first generated realizations will need to be discarded before the realizations become equally likely samples from Eq. 23. Once convergence is reached, however, the resulting realization $s_{cc,l}$ and all subsequent realizations are equally likely samples from the full posterior pdf.

7 Example of Interpolation Application

A sample application of the developed methodology is now presented. The Humboldt River basin is located in North-Eastern Nevada, and its water resources are crucial to a variety of recreational and agricultural uses. The concentration history of dissolved arsenic in the North Fork of the Humboldt river and the total dissolved arsenic load supplied to downstream locations are derived based on sparse concentration measurements. The statistical confidence intervals about the best estimate of the concentration history and cumulative contaminant load are also obtained. Arsenic in the Humboldt River basin results in large part from mining operations in the area. Arsenic and other contaminants are released during normal mining practices when mineralized rock is crushed and exposed to oxygen and water.

The new methodology is compared with results obtained using kriging with a linear variogram, because the partial differential equation governing the prob-

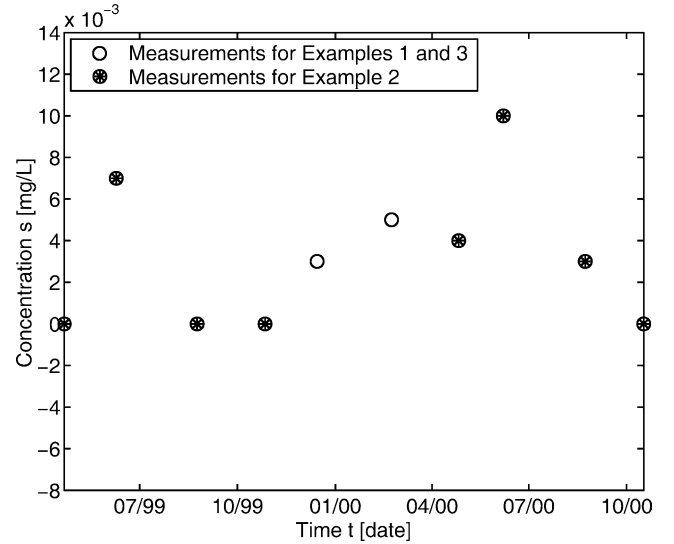


Fig. 4 Concentration measurements for North Fork of Humboldt River at North Fork Ranch, Elko county, Nevada

ability density is the same for these two models, as described in Eq. 3. The concentration data were obtained from the EPA STORET database and are plotted in Fig. 4 (Wilson, E. EPA Region 9. Dissolved arsenic data for North Fork Humboldt @ North Fork Ranch 21NEV-1. Personal communication. 2001.). In order to obtain total load estimates, flowrate data were also needed. River flow data for the equivalent time period, however, were not available. Therefore, flows for January 1, 1971, through December 31, 1981, were averaged to obtain a representative hydrograph for the stream (USGS 2001). These daily average flows were used to estimate the flowrate history for the period of April 21, 1999, through October 17, 2000, by assigning to each day a flowrate equivalent to the average flow for that calendar day. These flows are presented in Fig. 5.

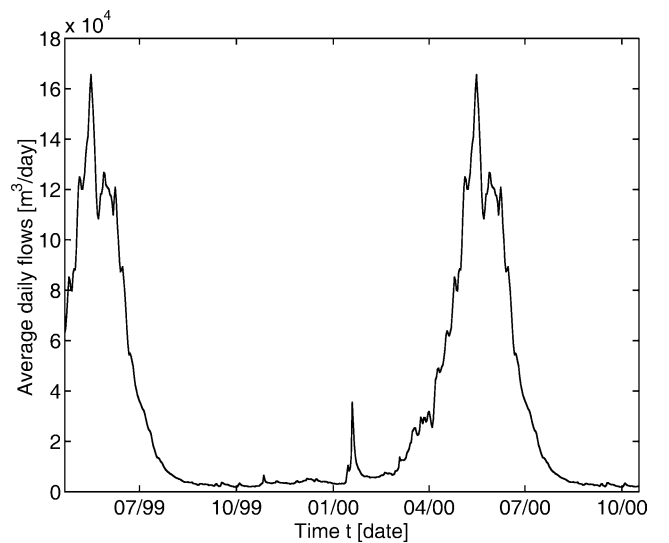


Fig. 5 Averaged flow data at one-day intervals for North Fork of Humboldt River, Elko county, Nevada

We discretize the concentration history into one-day intervals, with time zero starting on the day of the first measurement, April 21, 1999. Therefore, $t_j = j - 1$ days, where $j = 1, \dots, m$, and $m = 546$.

Three examples are presented. In the first example, all available data are used, starting at time $T_1 = 0$ and ending at $T_n = 546$ days, yielding $n = 10$ observations. For this example, it is assumed that there is no measurement error in the sampled dissolved arsenic concentrations. In the second example, we assume that data for two sampling times ($T_i = (238, 308)$ corresponding to December 15, 1999, and February 23, 2000) are not available to illustrate the effect of data that is more sparse and irregular on the confidence intervals associated with estimates (see Fig. 4). Finally, the third example uses all the measurements but it is assumed that the measurements include a random, normally distributed sampling error with a known variance of $\sigma_e^2 = 10^{-6} \text{ mg/L}^2$. This last example demonstrates the applicability of the method when observations are known not to be error-free, and illustrates the effect of this additional uncertainty on the analysis.

7.1 Example 1 results – Regular sampling with no measurement error

For the application of the nonnegativity enforcing methodology, the coefficient D that maximizes the likelihood of the observations was determined according to Eq. 13 to be $2.4 \times 10^{-7} (\text{mg/L})^2/\text{day}$. The likelihood of the observations is plotted as a function of parameter D in Fig. 6. A similar method was used to determine the optimal D for the linear variogram, substituting the pdf described in Eq. 5 for Eq. 7 in Eq. 13:

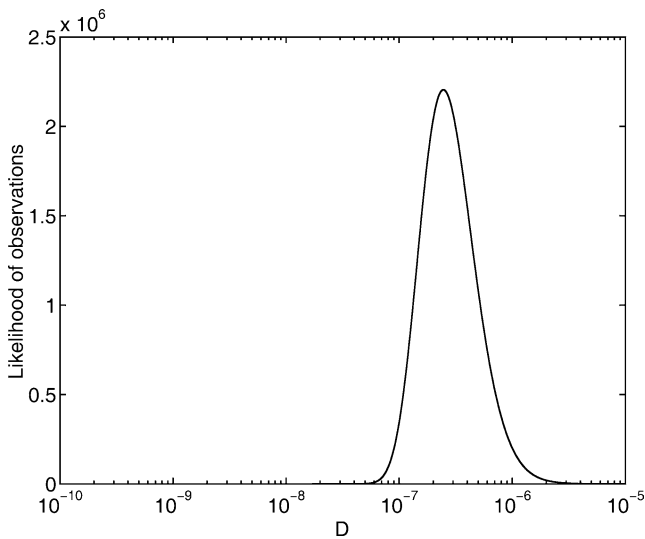


Fig. 6 Likelihood of measurements for nonnegativity enforcing method for Example 1 as a function of parameter D

$$p(\mathbf{s}|D) = \prod_{i=1}^{n-1} p(s_{i+1}|s_i, D) \quad (29)$$

$$p(s_{i+1}|s_i, D) = \frac{1}{\sqrt{4\pi D(T_{i+1} - T_i)}} \exp\left(-\frac{(s_{i+1} - s_i)^2}{4D(T_{i+1} - T_i)}\right) \quad (30)$$

Using this method, the optimal D for the kriging model with linear variogram was found to be $2.0 \times 10^{-7} (\text{mg/L})^2/\text{day}$.

For these examples, the median of the probability density function of the concentration at each time was selected to represent the best estimate. For the new method, the median and 95% confidence intervals of the probability density functions of concentration values at one day intervals were determined based on ensemble properties of conditional realizations generated using the method described in Sect. 5.3. These values are plotted in Fig. 7a. The equivalent plot for the linear kriging estimation is presented in Fig. 7b. As can be seen in Fig. 7, the nonnegativity enforcing method behaves essentially like a linear interpolator for high concentration values (i.e. behaves similarly to linear variogram kriging), but the best estimate in low concentration regions tends to be deflected away from $s = 0$.

The estimate for the total load was based on 1000 conditional realizations of possible concentration histories. The cumulative integrals of these realizations,

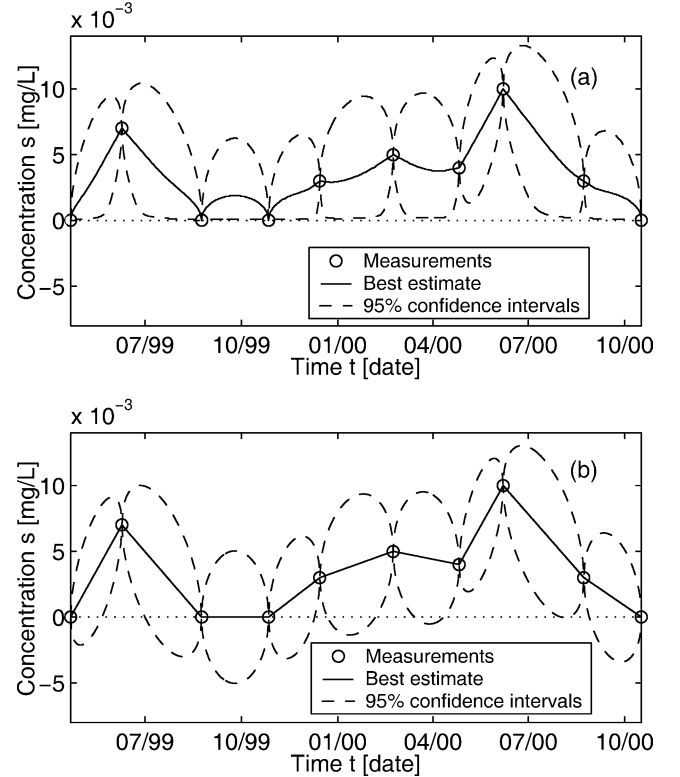


Fig. 7 Best estimate and 95% confidence intervals for concentration as a function of time for Example 1. **a** Nonnegativity enforcing method; **b** Linear variogram kriging method

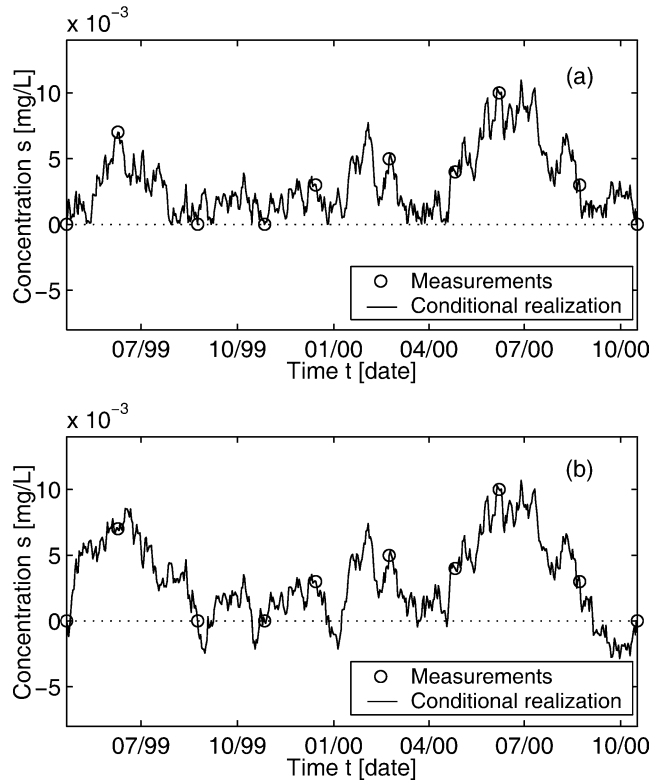


Fig. 8 Sample conditional realizations for concentration history for Example 1. **a** Nonnegativity enforcing method; **b** Linear variogram kriging method

weighted by the daily flowrate, were analyzed in order to determine the load best estimate (defined as the median of the distribution) and the 95% confidence intervals. For the nonnegativity enforcing method, conditional realizations were generated according to the Reflected Brownian Motion approach presented in Sect. 5.3. One such realization is plotted in Fig. 8a along with the observation points. As can be seen in this figure, the conditional realization honors the observations and is everywhere nonnegative. For the linear variogram kriging, the methodology outlined in the Appendix was followed, and a sample conditional realization is presented in Fig. 8b. Note that the two conditional realizations presented in Fig. 8 were generated using the same random number seed. The total contaminant load best estimate and 95% confidence intervals based on 1000 Monte Carlo simulations are presented in Fig. 9 for both the nonnegativity enforcing method and the linear variogram kriging method.

7.2 Example 2 results – Irregular sampling with no measurement error

The elimination of certain measurement times affects not only the calculation of the best estimates and their associated confidence intervals, but can also affect the optimal value of the parameter D . For this case, the

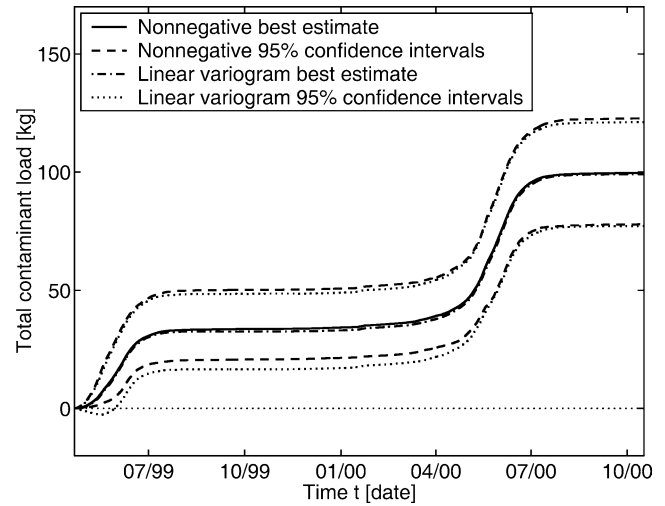


Fig. 9 Cumulative arsenic load best estimate and 95% confidence intervals for Example 1 using nonnegativity enforcing method and linear variogram kriging method

optimal D value was found to be $2.7 \times 10^{-7} (\text{mg/L})^2/\text{day}$ for the nonnegativity enforcing method, and $2.4 \times 10^{-7} (\text{mg/L})^2/\text{day}$ for the kriging method. For the new method, the resulting median and 95% confidence intervals of the probability distribution functions of concentration values at one day time intervals are plotted in Fig. 10a. The equivalent plot for the linear kriging

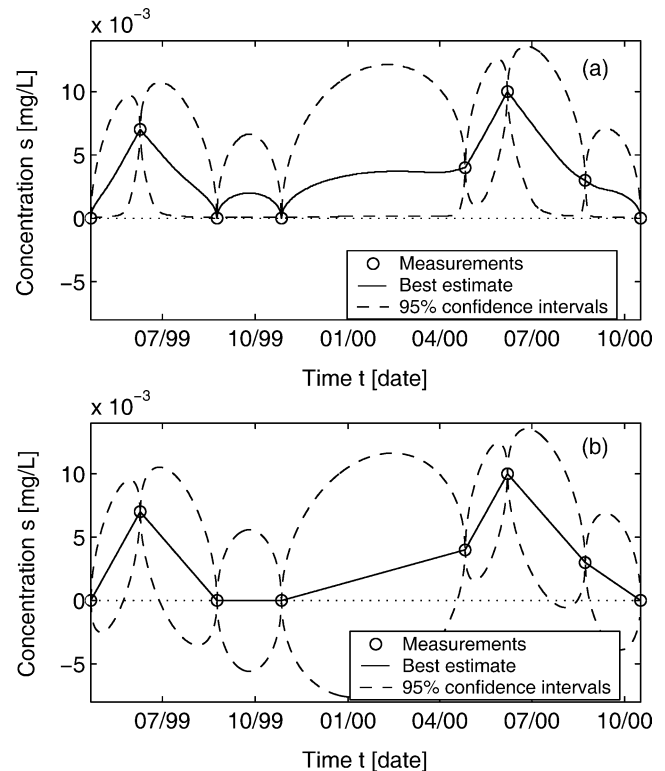


Fig. 10 Best estimate and 95% confidence intervals for concentration as a function of time for Example 2. **a** Nonnegativity enforcing method; **b** Linear variogram kriging method

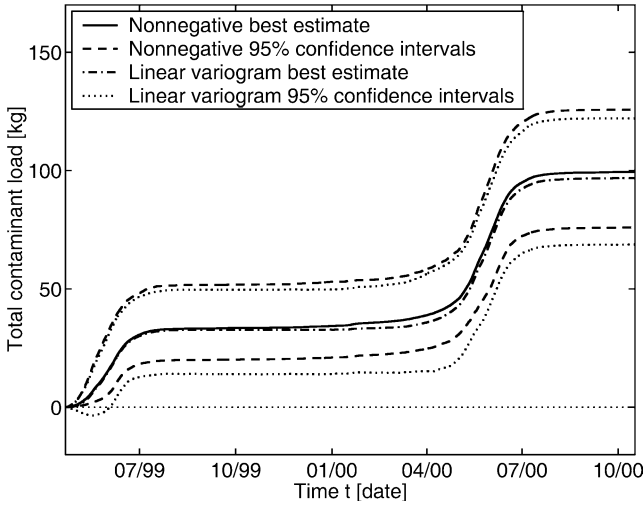


Fig. 11 Cumulative arsenic load best estimate and 95% confidence intervals for Example 2 using nonnegativity enforcing method and linear variogram kriging method

estimation is presented in Fig. 10b. The total contaminant load best estimate and 95% confidence intervals based on 1000 Monte Carlo simulations are presented in Fig. 11 for both the nonnegativity enforcing method and the linear variogram kriging method.

7.3 Example 3 results – Regular sampling with measurement error

To determine the effect of measurement error, the measurements were assumed to contain error with a variance of $\sigma_e^2 = 1 \times 10^{-6} \text{ (mg/L)}^2$. The optimal values for D , obtained using the EM algorithm, were $2.1 \times 10^{-7} \text{ (mg/L)}^2/\text{day}$ and $1.7 \times 10^{-7} \text{ (mg/L)}^2/\text{day}$ for the nonnegativity enforcing method and linear kriging, respectively. For the new method, the resulting median and 95% confidence intervals of the probability distribution functions of concentrations values at one day time intervals are plotted in Fig. 12a. The equivalent plot for the linear kriging estimation is presented in Fig. 12b. The total contaminant load best estimate and 95% confidence intervals based on 2000 Monte Carlo simulations are presented in Fig. 13 for both the nonnegativity enforcing method and the linear variogram kriging method.

7.4 Discussion

As can be seen from Figs. 7a and 8a, for example, the new methodology is effective at enforcing parameter nonnegativity for all points. The new methodology behaves similarly to linear variogram kriging in high concentration regions, which it is designed to do. Traditional kriging, on the other hand, leads to conditional realizations and confidence intervals reaching into the negative parameter range. These negative values have no

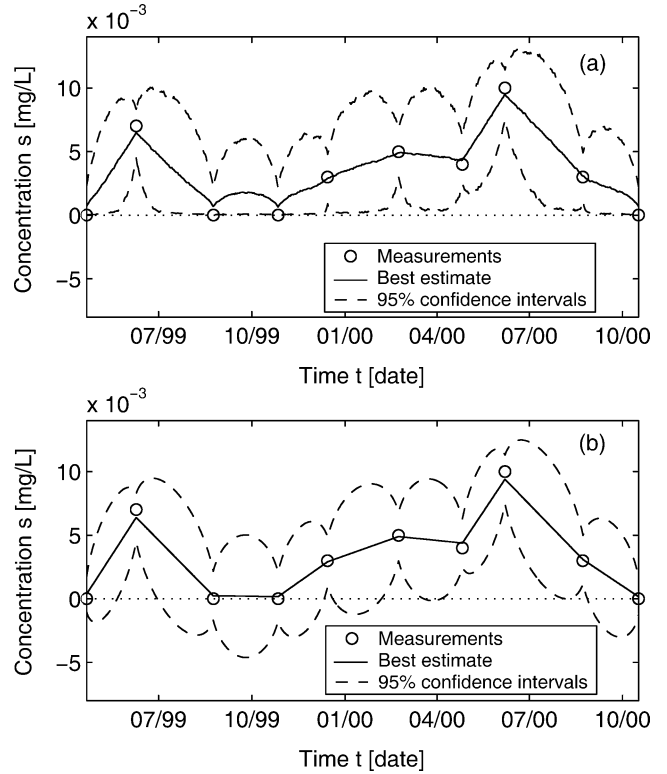


Fig. 12 Best estimate and 95% confidence intervals for concentration as a function of time for Example 3. **a** Nonnegativity enforcing method; **b** Linear variogram kriging method

physical significance, and can be misleading. Furthermore, even the total load estimate confidence intervals reach into the negative range. Such results are not physical, and can undermine the credibility of the analysis as a whole.

As expected, the removal of certain data points leads to wider confidence intervals about best estimates of the concentration history. This effect is more pronounced

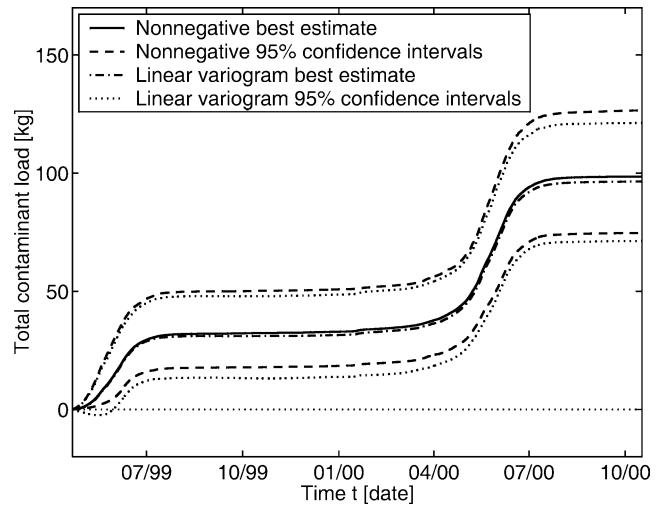


Fig. 13 Cumulative arsenic load best estimate and 95% confidence intervals for Example 3 using nonnegativity enforcing method and linear variogram kriging method

for traditional kriging, because the additional information provided by the nonnegativity constraint is not available.

The addition of measurement error results in wider confidence intervals close to measurement locations for both methods, but slightly narrower confidence intervals away from observations. This is due to the fact that some of the variability between the observation values is attributed to measurement error, resulting in a lower estimate for the variance parameter D .

The application of the new methodology has greatest impact in periods of low concentration because, by design, the method behaves similarly to linear kriging with a linear variogram for concentrations far from the constraint boundary. In the examples presented, periods of high concentration happen to correspond to high flow periods. Therefore, although the effect of the new method can be clearly seen in the figures illustrating the best estimates and statistical confidence intervals for the dissolved arsenic concentration history, the effect on the estimated total load is relatively minor. This can be seen in Figs. 9, 11 and 13, where the total loads are similar for the two methods. This is due to the fact that, when the concentration history is integrated over time, weighted by the daily flows, the effect of the lower-concentration periods is diminished, because they are weighted by lower flows. This indicates that the effect of the proposed algorithm on load calculations will be relatively minor for applications where high function values receive higher weights. The effect on the function values themselves (Figs. 7, 10 and 12), however, is strong.

In order to illustrate the effect of the new methodology in a case where there is no direct correlation between the function values and their assigned weights, the analysis performed in Example 3 was repeated using a constant flowrate for the entire sampling period. This flowrate was set equal to the mean flowrate over the sampling period, or $3338 \text{ m}^3/\text{day}$. The change in assumed flowrate does not affect concentration history estimates that were presented in Fig. 12 for Example 3. Figure 14 presents the estimated total dissolved arsenic loads for the sampling period, based on the use of a constant flowrate. As can be seen from this figure, the total load estimated using linear geostatistical inverse modeling is significantly lower than that estimated using the new methodology. The same applies to the upper and lower 95% confidence intervals. A similar effect would have been observed had the analyses performed in Examples 1 and 2 been repeated using the constant flowrate. In fact, if one examines the total load results obtained using the variable flowrate in Figs. 9, 11 and 13, one can see that the total load estimate is consistently slightly lower using kriging relative to the new method.

The higher total load, as estimated using the new methodology, is more representative of the possible concentration history scenarios, and therefore possible total contaminant loads being delivered to downstream

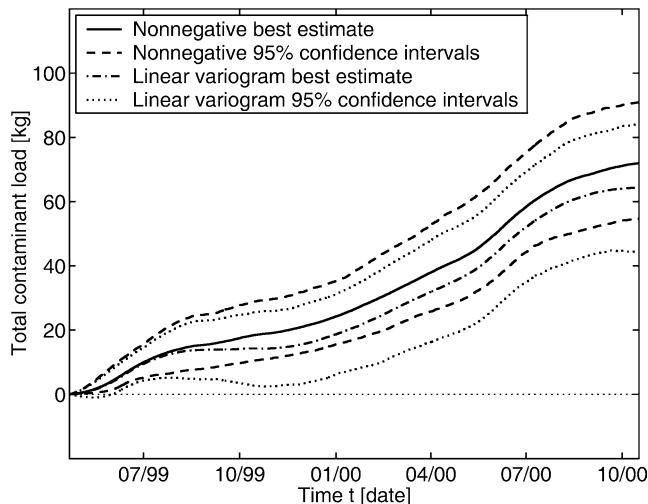


Fig. 14 Cumulative arsenic load best estimate and 95% confidence intervals for Example 3 with constant flowrate using nonnegativity enforcing method and linear variogram kriging method

locations. The conditional realizations, or concentration history scenarios, generated using the new methodology are designed to have equivalent characteristics to those generated using the linear geostatistical approach with a linear variogram, in the sense that the temporal variability of the unknown function is described by a Brownian motion. However, the realizations generated using the new methodology incorporate the additional information that concentrations cannot be negative.

The significant difference observed between the load predictions resulting from the kriging approach relative to the new methodology are critical when evaluating environmental risk based on interpolated load estimates. Using a linear kriging methodology that does not incorporate nonnegativity would consistently lead to underestimates of total loads, whether these are represented by the best estimate or by the corresponding confidence intervals. Such inaccuracy could have significant detrimental effects on management decisions made on the basis of interpolated estimates.

8 Conclusions

The method developed in this work provides an effective and statistically rigorous kriging-based stochastic interpolation tool for nonnegative functions. The main advantage of this method is that it preserves much of the computational efficiency associated with working in a multi-Gaussian setting (such as is the case with linear geostatistical approaches), while at the same time offering a method for enforcing nonnegativity that does not impose undesirable characteristics on the solution. The method is particularly applicable to situations where a Gaussian model would be deemed appropriate except for the fact that a nonnegativity constraint needs

to be enforced. The interpolation of chemical concentrations is one such application.

The method allows for the generation of conditional realizations of the unknown function, which can then be used to identify a best estimate of, as well as statistical confidence intervals about, the unknown function. The effect of measurement error can be incorporated without a significant increase in computational requirements. The effectiveness of the method was demonstrated for the case of sparse dissolved arsenic concentration measurements being used to estimate the total contaminant load over a given time interval.

The method offers several advantages relative to unconstrained kriging:

- Confidence intervals about the best estimates remain in the nonnegative range, by defining these intervals using a pdf that is non-zero only in the nonnegative range;
- The method allows for the generation of physically reasonable conditional realizations;
- The method allows for the incorporation of additional information, resulting in narrower confidence intervals relative to kriging, especially in areas close to the constraint boundary;
- Because the method incorporates additional information, it results in better estimates of total loads, as defined by the weighted integral of the estimated unknown function.

The computational cost associated with the new methodology is very manageable. For example, the generation of the 1000 nonnegative conditional realizations used for Fig. 9 took slightly over a minute on a 2.0 GHz machine, including the estimation of parameter D , and weighted integration and sorting of the realizations. The Reflected Brownian Motion method used for generating conditional realizations (which is exact when measurement error is not considered and a good approximation otherwise) avoids numerical integration and computationally expensive numerical sampling techniques. When the exact Gibbs sampling algorithm is used in the case where measurement error is taken into account, the computational cost is minimized by only applying this algorithm at observation locations, and obtaining conditional realizations at other discretization points using the Reflection Brownian Motion approach. In addition, the iterative EM algorithm used for estimating the parameter D can be implemented using a relatively small number of realizations per iteration. Therefore, for the examples presented here, the computational cost of the proposed methods was not significantly greater than that of generating the equivalent number of realizations following an unconstrained linear variogram model. Also, the proposed methods all involve generating a conditional realization first at the observation locations, then filling in the remaining points using an unconstrained linear variogram realization, and finally reflecting the resulting realizations about the constraint boundary. Therefore, the cost of

the proposed methods relative to that of generating unconstrained realizations will increase only with the number of observations, not the number of nodes at which the function is to be estimated.

The methods presented in this paper are based on the property that, in one dimension, the value of a function represented by Brownian is only dependent on its values at adjacent points. As such, these methods are currently only applicable to cases where the modeled parameters are a function of a single variable (e.g. time). Also, the Brownian motion formulation means that the method corresponds to a linear variogram in the case where nonnegativity is not enforced. The linear variogram makes a minimum of assumptions regarding the spatial structure of the unknown function (for example, it does not assume any large scale correlations) and is applicable to a wide range of problems. A method directly applicable in multiple dimensions and with a variety of covariance structures will be the topic of upcoming work.

Finally, there are various possible extensions to the method presented in this paper:

- The method of images could just as easily be applied to other inequality constraints, such as arbitrary minimum or maximum values on the parameter to be identified. Such minimum or maximum values could be identified based on prior information available from other sources (such as, for example, solubility limits of compounds).
- The method of images could be used to develop probability density functions for parameters with both a lower and an upper bound. Such an application would require the consideration of multiple images. However, only a manageably small number of these would have a physically significant effect on the solution (see, for example, Bostock 1971).
- The method could be applied in a case where measurement error does not follow a Gaussian distribution. The MCMC approach for generating conditional realizations may have to be modified, however, if the measurement error pdf results in a marginal posterior pdf from which it is not easy to sample directly.

Appendix

Review of Kriging Methodology

Kriging with a linear variogram and a constant, unknown mean is used as a component of the method developed in this work for generating conditional realizations of interpolated data. Kriging is also compared to the newly developed methodology for the presented application. Where used, linear variogram kriging is performed according to a standard procedure which is briefly described herein. The reader is referred to standard texts on geostatistics (e.g. Kitanidis 1997) for additional details and background.

The general system of equations that is solved to obtain a best estimate and variance is:

$$\sum_{j=1}^n \lambda_j [-D|T_i - T_j| - \sigma_\varepsilon^2(1 - \delta_{ij})] + v = -D|T_i - t_o| - \sigma_\varepsilon^2, \quad i = 1, \dots, n \quad (\text{A1})$$

$$\sum_{j=1}^n \lambda_j = 1 \quad (\text{A2})$$

where n is the number of observations, T are points at which measurements are available, t_o is the point at which an estimate is sought, σ_ε^2 is the variance of the measurement error, and the λ and v coefficients are found from the system. If we are not considering measurement error, $\sigma_\varepsilon^2 = 0$.

Once the system of equations has been solved, the best estimate of the function at points t_o is defined as:

$$\hat{s}_o = \sum_{j=1}^n \lambda_j y_j \quad (\text{A3})$$

where y are the observation values, the mean squared error of this estimate is:

$$\sigma_{s_o}^2 = -v + \sum_{j=1}^n \lambda_j D|T_i - t_o| \quad (\text{A4})$$

and the 95% confidence intervals about the best estimate are $\hat{s}_o \pm 2\sigma_{s_o}$.

The general system of equations that is solved to obtain conditional realizations is:

$$\sum_{j=1}^n [-D|T_i - T_j| - \sigma_\varepsilon^2(1 - \delta_{ij})] \xi_j + \hat{\beta} = s(T_i) + \eta\sigma_\varepsilon - \mathbf{s}_{uu,l}(T_i), \quad i = 1, \dots, n \quad (\text{A5})$$

$$\sum_{j=1}^n \xi_j = 0 \quad (\text{A6})$$

where l denotes that we are dealing with the l -th conditional realization, η is a $n \times 1$ vector of normally distributed random numbers with zero mean and unit variance, $\mathbf{s}_{uu,l}(T_i)$ is an unconditional unconstrained realization of a Brownian motion with generalized covariance function $-D|t_i - t_j|$ (see Gutjahr et al. 1994; Kitanidis 1995), and the ξ and $\hat{\beta}$ coefficients are found from the system. If we are not considering measurement error, $\sigma_\varepsilon = 0$ and η is not needed.

Once the system of equations has been solved, the conditional realization is defined as:

$$\mathbf{s}_{cu,l}(t_o) = -\sum_{j=1}^n D|t_o - T_j| \xi_j + \hat{\beta} + \mathbf{s}_{uu,l}(t_o) \quad (\text{A7})$$

Acknowledgements This research was partially funded by the Natural and Accelerated Bioremediation Research (NABIR) program, Biological and Environmental Research (BER), U.S. Department of Energy (grant #DE-FG03-00ER63046). We would

like to thank Dr. Eric Wilson, U.S. Environmental Protection Agency Region 9, for helping us obtain the Humboldt River data used in this study, and Prof. Peter W. Glynn for his valuable input.

References

1. Barnes RJ, You KH (1992) Adding Bounds to Kriging. *Mathem Geol* 24(2):171–176
2. Besag J, York J, Mollie A (1991) Bayesian image restoration with two applications in spatial statistics. *Ann Inst Statist Math* 43(1):1–59
3. Besag J, Green P, Higdon D, Mengersen K (1995) Bayesian computation and stochastic systems. *Stat Sci* 10(1):3–66
4. Bostock CA (1971) Estimating truncation error in image well theory. *Water Resour Res* 7(6):1658–1660
5. Casella G, George EI (1992) Explaining the Gibbs Sampler. *Am Statist* 46(3):167–174
6. Cressie NAC (1991) *Statistics for Spatial Data*, Wiley-Interscience, New York
7. Dempster AP et al (1977) Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Roy Statist Soc B* 39:1–38
8. Diggle PJ, Tawn JA, Moyeed RA (1998) Model-based geostatistics (with discussion). *Appl Statist* 47(3):299–350
9. Eubank RL (1988) *Spline smoothing and nonparametric regression*. M. Dekker, New York
10. Gill PE, Murray W, Wright MH (1986) *Practical Optimization*, Academic Press, San Diego
11. Gutjahr A, Bullard B et al (1994) Joint conditional simulations and the spectral approach for flow modeling. *Stochastic Hydrol Hydraul* 8(1):79–108
12. Handcock MS, Stein ML (1993) A Bayesian analysis of kriging. *Technometrics* 35(4):403–410
13. Hardle W (1990) *Applied nonparametric regression*, Cambridge University Press, New York
14. Hoeksema RJ, Kitanidis PK (1985) Analysis of spatial structure of properties of selected aquifers. *Water Resour Res* 21(9):563–572
15. Karlin S, Taylor HM (1975) *A First Course in Stochastic Processes*, Academic Press, New York
16. Kauffmann C, Kinzelbach W (1989) Parameter estimation in contaminant transport modelling. In: Kinzelbach KA (ed) *Contaminant transport in groundwater*, Balkema, Rotterdam
17. Kitanidis PK (1986) Parameter uncertainty in estimation of spatial functions: Bayesian analysis. *Water Resour Res* 22(4):499–507
18. Kitanidis PK (1995) Quasi-linear geostatistical theory for inverting. *Water Resour Res* 31(10):2411–2419
19. Kitanidis PK, Shen K-F (1996) Geostatistical interpolation of chemical concentration. *Adv Water Resour* 19(6):369–378
20. Kitanidis PK (1997) *Introduction to Geostatistics Applications in Hydrogeology*, Cambridge University Press, New York
21. McLachlan GJ, Krishnan T (1997) *The EM Algorithm and Extensions*, Wiley, NY
22. Michalak AM, Kitanidis PK (2003) A method for enforcing parameter nonnegativity in Bayesian inverse problems with an application to contaminant source identification. *Water Resour Res* 39(2):1033, doi:10.1029/2002WR001480
23. Saito H, Goovaerts P (2000) Geostatistical interpolation of positively skewed and censored data in a dioxin-contaminated site. *Environ Sci Tech* 34(19):4228–4235
24. Schimek MG (2000) *Smoothing and Regression: Approaches, Computation, and Application*, Wiley, New York
25. Silverman BW (1985) Some aspects of the spline smoothing approach to nonparametric regression curve fitting. *J Roy Statist Soc B* 47:1–52
26. Snodgrass MF, Kitanidis PK (1997) A geostatistical approach to contaminant source identification. *Water Resour Res* 33(4):537–546

27. USGS (2001) Daily Streamflow for the Nation USGS 10317400 N F Humboldt R NR N Fork, NV. water.usgs.gov/nwis/discharge/?site_no10317400. Accessed 11/01/01
28. Vardi Y, Lee D (1993) From image deblurring to optimal investments: maximum likelihood solutions for positive inverse problems. *J Roy Statist Soc B (Methodological)* 55(3):569–612
29. Waller LA, Carlin BP, Xia H, Gelfand v (1997) Hierarchical spatio-temporal mapping of disease rates. *J Am Statis Assoc*, 92(438): 607–617
30. Weir IA (1997) Fully Bayesian reconstructions from single-photon emission computed tomography data. *J Am Statist Assoc* 92(437):49–60
31. Yakowitz SJ, Szidarovsky F (1985) A comparison of geostatistical and nonparametric regression methods. *J Multivar Anal* 16(1):21–53
32. Zauderer E (1989) *Partial Differential Equations of Applied Mathematics*, Wiley-Interscience, New York