A Gibbs sampler for inequality-constrained geostatistical interpolation and inverse modeling

Anna M. Michalak[1]

Received 8 November 2007; revised 11 April 2008; accepted 12 May 2008; published 30 September 2008.

[1] Interpolation and inverse modeling problems are ubiquitous in environmental sciences. In many applications, the parameters being estimated or mapped have physical constraints, such as nonnegativity (e.g. concentration, hydraulic conductivity), solubility limits, censored data (e.g. due to dry wells or detection limits), and other physical boundaries or missing data. Geostatistical interpolation and inverse modeling techniques have often been applied for estimating such parameters, but these methods typically cannot enforce physical constraints. This paper describes a statistically rigorous and computationally efficient Gibbs sampler, a Markov chain Monte Carlo technique, based on an a priori truncated Gaussian distribution model, which allows for multiple and variable physical constraints to be enforced within a geostatistical framework. Sample interpolation and inverse modeling applications confirm that estimates, uncertainty bounds and conditional simulations reflect the specified constraints, leading to conclusions that are more consistent with the underlying conceptual model, and provide a more accurate measure of the posterior uncertainty of the parameters being estimated. In addition, especially in inverse modeling applications, a posteriori confidence bounds are narrower even in areas where constraints are not imposed. The method is applicable in multiple dimensions, for data with or without measurement error, and with any variogram model.

Citation: Michalak, A. M. (2008), A Gibbs sampler for inequality-constrained geostatistical interpolation and inverse modeling, *Water Resour. Res.*, *44*, W09437, doi:10.1029/2007WR006645.

1. Introduction

[2] Interpolation and inverse modeling problems are ubiquitous in environmental sciences. Methods based on geostatistical approaches have been used extensively to address these problems in surface and groundwater hydrology. In many applications, the parameters being estimated have physical constraints. Examples from environmental applications include solubility limits, nonnegativity constraints on hydraulic conductivity or species concentrations, concentration ranges for contaminants below an instrument's detection limit, maximum hydraulic head constraints when sampling wells are not sufficiently deep to locate the water table, and/or other physical boundaries. Enforcing known constraints on parameter values would allow the resulting parameter estimates to be consistent with the physical structure of the system. In addition, enforcing constraints can remove biases introduced when these constraints are ignored, and can decrease the uncertainty of estimates. As described in further detail in section 2, traditional geostatistical approaches are limited in their ability to enforce such constraints.

[3] The objective of this paper is to present a statistically rigorous, geostatistically-based, interpolation and inverse modeling approach applicable to inequality-constrained data and parameters. The approach is to be applicable with any variogram or covariance model, in multiple dimensions, and with multiple, and possibly variable, constraints. The proposed approach is conceptually related to the approach developed by *Michalak and Kitanidis* [2003, 2005], which uses a linear variogram model, and is applicable to a single constant constraint, for a system where the modeled parameter varies as a function of a single variable (e.g. time). Whereas this existing approach is based on Reflected Brownian Motion [e.g. *Karlin and Taylor*, 1975], the approach presented here is a generalization that makes use of truncated Gaussian distributions.

2. Background

2.1. Constrained Interpolation and Inverse Modeling

[4] In environmental applications, methods for incorporating inequality constraints into interpolation and inverse problems have been explored for a few problems, most commonly to enforce nonnegativity for estimated parameters. Within the geostatistical literature, these developed methods have relied on (1) constraining kriging weights to be nonnegative, (2) using Lagrange multipliers or related tools to enforce inequality constraints on the estimates, and (3) applying data transformations. These approaches are briefly summarized below.

[5] *Szidarovszky et al.* [1987] and *Deutsch* [1996] presented approaches that constrain kriging weights to be nonnegative, which, by default, yields nonnegative estimates given nonnegative observations. In general, however, nonnegative kriging weights are not necessary for producing nonnegative estimates. In addition, these approaches are

[1]Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, Michigan, USA.

0043-1397/08/2007WR006645

not applicable for multiple or variable constraints or censored data. *Barnes and You* [1992] present an alternative to this approach, which does not constrain the kriging weights, but only the estimates themselves. At locations where the estimated value would otherwise violate an inequality constraint, the estimate is set to the inequality bound, and an adjustment is calculated to the estimation variance. Both in this approach and those presented by *Szidarovszky et al.* [1987] and *Deutsch* [1996], if this variance is used to define symmetric confidence intervals for the estimated field, then these confidence intervals may not honor the inequality constraints. In addition, the approaches have not been extended for application to inverse problems.

[6] *Journel* [1986] proposed a soft kriging approach based on indicator kriging which can satisfy multiple constraints but requires the formulation of a set of internally consistent indicator covariance models, which can be difficult to derive in practice [*Yoo and Kyriakidis*, 2006]. The final uncertainties are approximated using a series of indicator values. The approach has also not been extended to inverse problems.

[7] Lagrange multipliers have been used extensively in estimation problems for equality and inequality constrained parameters. This method amounts to restricting a common Gaussian process by replacing the original objective function $f(\mathbf{s})$ by the Lagrange function

$$h(\mathbf{s},\boldsymbol{\nu}) = f(\mathbf{s}) - \sum_{i=1}^{k} \nu_i[g_i(\mathbf{s}) - b_i] \tag{1}$$

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 $\boldsymbol{\nu} = (\nu_1, \nu_2,\ldots,\nu_k)$ are Lagrange multipliers (see *Michalak and Kitanidis* [2003] for a more thorough discussion). This approach is effective at bounding best estimates, can be applied to multiple and variable constraints, and is valid for both interpolation and inverse problems. However, this approach is not easily suited to assessing uncertainty bounds for the estimated parameters. This is because, if this approach is used in generating conditional realizations of the modeled process, the ensemble of realizations yields a finite probability for elements of $\mathbf{s}$ being equal to the constraint value, which is contrary to the definition of continuous random variables. A related approach, based on spline formalism, was presented in *Dubrule and Kostov* [1986].

[8] Data transformations are another popular method that has been applied to geostatistical and other estimation problems, especially for enforcing parameter nonnegativity. The approach involves using a mapping function to transform the original variable. Such methods have been applied both to geostatistical interpolation and inverse problems (e.g. *Kitanidis and Shen* [1996], *Kitanidis* [1997, p.70], *Chilès and Delfiner* [1999], *Saito and Goovaerts* [2000], *Leuangthong and Deutsch* [2004], and *Diggle and Ribeiro* [2007]). The most common method is the power transformation, which is defined as

$$\widetilde{\mathbf{s}} = \begin{cases} (\mathbf{s}^{\kappa} - 1)/\kappa & \kappa > 0 \\ ln(\mathbf{s}) & \kappa = 0 \end{cases} \tag{2}$$

where $\mathbf{s}$ is the vector of values in the original domain, $\widetilde{\mathbf{s}}$ 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 at the limit of κ tending to zero. In general, data transformations lead to highly non-symmetric probability density functions for the estimated parameter values in the untransformed space, make linear inverse problems nonlinear, and present difficulties for measurements of zero (e.g. non-detects). These and other difficulties have been documented in *Snodgrass and Kitanidis* [1997], *Walvoort and de Gruijter* [2001], and *Michalak and Kitanidis* [2003], among others. In addition, traditional data transformations are only valid for enforcing a single upper or lower bound that is constant for all estimation times or locations.

[9] In a few instances, numerical sampling approaches have been proposed to address specific problems in spatial interpolation. *Militino and Ugarte* [1999] used expectation maximization to analyze censored spatial data. The approach involves "whitening" the available data, and using these transformed observations to estimate the censored values. *Abrahamsen and Benth* [2001] used a Monte Carlo approach together with data augmentation and prior information on the drift parameters describing the trend in the modeled data to apply inequality constraints in a kriging context. Both of these approaches assumed exact measurements, and are further reviewed in *De Oliveira* [2005]. *De Oliveira* [2005] developed an MCMC Bayesian model for modeling the spatial distribution of a parameter with censored observations, using a Metropolis-Hastings algorithm [*Hastings*, 1970]. *Fridley and Dixon* [2007] recently developed a similar algorithm focusing specifically on estimating observations that are below detection limits, and making use of both a Gibbs sampler and a Metropolis-Hastings step. Some of these sampling approaches assume error-free measurements, while others are designed for specific constraints on data values, or do not fully account for the additional uncertainty associated with censored measurements. In addition, none of these approaches have been extended for application to inverse problems.

[10] Recently, *Michalak and Kitanidis* [2003, 2005] developed a nonnegativity-enforcing interpolation and inverse modeling approach based on a Gibbs sampling algorithm. This approach provides nonnegative best estimates as well as confidence intervals. The method is described in more detail in section 2.2.

2.2. MCMC Methods in Hydrology

[11] Markov chain Monte Carlo (MCMC) approaches have been applied in hydrology to problems such as hydrologic model parameter optimization [e.g. *Vrugt et al.*, 2003a, 2003b] and parameterization of rainfall-runoff models [e.g. *Bates and Campbell*, 2001; *Campbell et al.*, 1999; *Marshall et al.*, 2004]. The majority of these approaches have been based on the Metropolis or Metropolis-Hastings algorithms. The Metropolis algorithm involves the sequential generation of correlated conditional realizations, which are accepted or rejected based on their a posteriori probability density relative to that of the previous accepted realization. The Metropolis-Hastings algorithm is an extension of this approach for the case where the transition probability between successive realizations is not symmetric. A useful introduction to these approaches is

presented in *Chib and Greenberg* [1995]. In hydrology, these approaches have primarily been applied to the estimation of a relatively small number of distinct variables representing different components of an examined system. A few other applications, such as parameter optimization for groundwater contaminant fate and transport models [e.g. *Balakrishnan et al.*, 2003] and zonation estimation [*Chen et al.*, 2006], have also been developed.

[12] The applications that are of interest in this paper involve the estimation of a spatially or temporally autocorrelated parameter field that can be described using a geostatistical variogram or covariance model. In this case, the state space, which includes each of the parameters to be estimated, is much larger relative to the applications listed above, with each location in discretized time or space represented by a random variable, yielding hundreds to millions of random variables to be characterized. A few applications of MCMC approaches to spatial processes have been presented in the literature. For example, *Chen et al.* [2004] presented an application of a Gibbs sampler for estimating the spatial distribution of iron concentrations using GPR tomographic data and borehole lithofacies logs. A Gibbs sampler is an approach that relies on sequentially sampling the marginal distribution of individual random variables. An overview of this general approach can be found in *Casella and George* [1992].

[13] *Michalak and Kitanidis* [2002, 2004a] developed a Metropolis-Hastings algorithm for enforcing nonnegativity in geostatistical inverse modeling. *Michalak and Kitanidis* [2003, 2005] developed a Gibbs sampling approach for enforcing nonnegativity using a prior probability density function based on the method of images applied to reflected Brownian motion [*Karlin and Taylor*, 1975]. This approach yields a prior pdf that is the sum of a Gaussian distribution and its reflection about the constraint boundary. This method eliminates some approximations necessary in the earlier approach, and the method was applied in an interpolation context to contaminant load estimation in the North Fork of the Humboldt River, and in an inverse modeling context to contaminant source identification at the Dover Air Force Base. The approach uses the method of images to define a marginal probability density function that is defined only for the nonnegative portion of the parameter space. This approach provides a rigorous statistical framework for defining a probability density function that constrains parameter values to be nonnegative at each estimation point, yielding nonnegative best estimates as well as confidence intervals. Given its reliance on the method of images, the approach is applicable for parameters that vary as a function of a single variable (e.g. time). Furthermore, the Brownian motion model implies that the modeled process can be represented using a linear variogram. *Fienen et al.* [2004] applied the approach to vertical deconvolution of hydraulic conductivities.

2.3. Use of Truncated Normal Distribution for Statistical Estimation

[14] Applications involving constrained estimation, including nonnegativity, other threshold constraints, and censored data abound in the statistical literature. One of the most common approaches to enforcing inequality constraints has been through the implementation of truncated distributions. Several of the implemented algorithms make use of the Gibbs sampling algorithm, described in section 2.2. *Smith and Roberts* [1993] provide a review of such applications, with those most relevant to the current work involving constrained parameter models, missing data, and censored data problems. *Gelfand et al.* [1992] also describe Gibbs sampling for constrained parameter and truncated data problems, specifically in the context of Bayesian analysis. In the context of mining and petroleum engineering, a Gibbs sampler applied with a truncated Gaussian distribution has been applied for generating individual simulations of sampled constrained processes [*Freulon and de Fouquet*, 1993; *Freulon*, 1994], and this approach has also been described in *Lantuéjoul* [2002] and *Chilès and Delfiner* [1999]. More recently, this approach has been used to delineate geologic facies [*Armstrong et al.*, 2003]. In these applications, the emphasis is on generating a single representative map of the sampled process, and not on assessing the uncertainty associated with the estimated process, or exploring the impact of constraints in inverse modeling applications.

[15] Overall, although the use of the Gibbs sampler for sampling truncated distributions as a solution to constrained problems is well established in the statistical literature, the applicability of such a method for estimating uncertainty in both geostatistical interpolation and inverse modeling is presented for the first time in this work.

3. Model Development

[16] Bayes' rule states that the posterior pdf of a state vector $\mathbf{s}$ given an observation vector $\mathbf{z}$ is proportional to the likelihood of the data given the state, times the prior pdf of the state. Symbolically:

$$p''(\mathbf{s}) = \frac{p(\mathbf{z}|\mathbf{s})p'(\mathbf{s})}{\int p(\mathbf{z}|\mathbf{s})p'(\mathbf{s})d\mathbf{s}} \quad (3)$$

In this context, prior and posterior are with respect to using the data $\mathbf{z}$, and the state vector $\mathbf{s}$ represents the parameter field under estimation. If we want to enforce constraints of the form $\mathbf{g}(\mathbf{s}) \geq \mathbf{b}$, we modify the probability distribution by multiplying it by heaviside functions corresponding to the desired constraints, effectively truncating the distribution at the constraint boundary, and rescaling the remaining portion of the probability density function to integrate to unity:

$$p''(\mathbf{s}) = \frac{p(\mathbf{z}|\mathbf{s})p'(\mathbf{s})\prod_{l=1}^{k}\mathcal{H}(\mathbf{g}_l(\mathbf{s}) - \mathbf{b}_l)}{\int \left(p(\mathbf{z}|\mathbf{s})p'(\mathbf{s})\prod_{l=1}^{k}\mathcal{H}(\mathbf{g}_l(\mathbf{s}) - \mathbf{b}_l)\right)d\mathbf{s}} \quad (4)$$

where

$$\mathcal{H}(\mathbf{g}_l(\mathbf{s}) - \mathbf{b}_l) = \begin{cases} 1 & \mathbf{g}_l(\mathbf{s}) > \mathbf{b}_l \\ 1/2 & \mathbf{g}_l(\mathbf{s}) = \mathbf{b}_l \\ 0 & \mathbf{g}_l(\mathbf{s}) < \mathbf{b}_l \end{cases} \quad (5)$$

The inclusion of the heaviside function in the integral in the denominator of the pdf ensures that, at each point in the discretized parameter field $\mathbf{s}$, the pdf still integrates to unity, which amounts to truncating the prior pdf at the inequality constraints. This approach is theoretically applicable with any probability density function, but is used here to truncate a Gaussian prior. Note that this setup can be used for specifying both lower and upper limits on parameter values. In addition, because both $\mathbf{g}$ and $\mathbf{b}$ are vectors, different upper and/or lower bounds can be specified for each component of the vector $\mathbf{s}$, such that the constraints can vary in space or in time. For example, for the simple case of a nonnegative parameter with an upper limit $s_{\max}$ (e.g. a solubility limit), $k = 2$ and

$$\begin{array}{ll} \mathbf{g}_1(\mathbf{s}) = \mathbf{s} & \mathbf{b}_1 = 0 \\ \mathbf{g}_2(\mathbf{s}) = -\mathbf{s} & \mathbf{b}_2 = s_{\max} \end{array} \tag{6}$$

More complex examples of constraints are described in the applications presented in section 5.

[17] In geostatistical approaches, the prior represents the deviation of the unknown vector from a model of the trend. As such, if we assume a truncated multi-Gaussian distribution, we define the unconstrained portion of the prior probability of the parameter values as

$$p'(\mathbf{s}) = \frac{1}{\sqrt{(2\pi)^m |\mathbf{Q}|}} \exp\left[-\frac{1}{2}(\mathbf{s} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{Q}^{-1}(\mathbf{s} - \mathbf{X}\boldsymbol{\beta})\right] \tag{7}$$

where $\mathbf{s}$ is an $m \times 1$ state vector obtained from the discretization of the unknown parameter distribution that we wish to estimate, $\mathbf{X}\boldsymbol{\beta}$ is the model of the trend, where $\mathbf{X}$ is a $m \times p$ matrix of drift functions and $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown drift coefficients, $\mathbf{Q}$ is the $m \times m$ prior covariance matrix based on the geostatistical model of the parameter distribution, and $|\ |$ denotes matrix determinant. In the case of inverse modeling, $\mathbf{Q}$ is the prior covariance matrix of the estimated state vector. In the case of interpolation, $\mathbf{Q}$ represents the covariance among all estimation and measurement locations. Given that the method, as presented in this work, is based on a prior defined as a truncated Gaussian distribution, the method should only be applied for datasets where a multivariate Gaussian distribution would be appropriate, were it not for the presence of constraints.

[18] The likelihood of the observations can be expressed as:

$$p(\mathbf{z}|\mathbf{s}) = \frac{1}{\sqrt{(2\pi)^n |\mathbf{R}|}} \exp\left[-\frac{1}{2}(\mathbf{z} - \mathbf{h}(\mathbf{s}, \boldsymbol{\theta}))^T \mathbf{R}^{-1}(\mathbf{z} - \mathbf{h}(\mathbf{s}, \boldsymbol{\theta}))\right] \tag{8}$$

where $\mathbf{z}$ is an $n \times 1$ vector of observations. When the system is underdetermined, which is the case of most interest for the application of geostatistical approaches, $m > n$. The vector $\boldsymbol{\theta}$ contains other parameters needed by the model function $\mathbf{h}$ ($\mathbf{s}$, $\boldsymbol{\theta}$). We assume that the measurement error ε has zero mean and known covariance matrix $\mathbf{R}$. The measurement errors are typically assumed to be uncorrelated, yielding a diagonal matrix $\mathbf{R}$, with the measurement error variances $\sigma^2_{R,1}$ to $\sigma^2_{R,n}$ on the diagonal. Note that the measurement error encompasses both the actual observation error when data are collected, and any inaccuracies inherent in the physical or conceptual model used to represent the problem. If the dependence of observations is linear in the unknown parameter $\mathbf{s}$, then:

$$p(\mathbf{z}|\mathbf{s}) = \frac{1}{\sqrt{(2\pi)^n |\mathbf{R}|}} \exp\left[-\frac{1}{2}(\mathbf{z} - \mathbf{H}\mathbf{s})^T \mathbf{R}^{-1}(\mathbf{z} - \mathbf{H}\mathbf{s})\right] \tag{9}$$

where $\mathbf{H}$ is an $n \times m$ sensitivity matrix, where $H_{ij} = \partial z_i/\partial s_j$. For interpolation applications, $\mathbf{H}$ simply maps the measurement locations onto the unknown state vector, and is therefore made up of ones and zeros. For inverse modeling application, $\mathbf{H}$ represents the sensitivity of each observation to each element of the state vector, as determined by a physical model of the system (e.g. a groundwater transport model). The method, as presented here for inverse modeling applications, assumes that the forward model is indeed linear. Although the presented approach could also be implemented within an iterative solution to a non-linear inverse problem, the computational feasibility of such an application will be the topic of future research.

4. Conditional Realizations and Estimation of Uncertainty

[19] One of the advantages of using a stochastic approach to interpolation and inverse problems is that physically significant confidence intervals and conditional realizations can be obtained in addition to a best estimate of the parameter values. In the case of classical linear geostatistical methods, the posterior pdf of the unknown parameters is a multivariate Gaussian distribution, and can therefore easily be sampled.

[20] The constrained posterior pdf in equation (4) is not Gaussian, and therefore does not lend itself to straightforward computation of confidence intervals or generation of conditional realizations. Therefore, a Markov chain Monte Carlo (MCMC) method is used to obtain conditional realizations from the posterior pdf. Ensemble properties of conditional realizations can then be used to infer other statistics of the estimated parameters, such as a measure of central tendency and percentiles, which can be used as a measure of the uncertainty associated with the estimate. In the discussion that follows, the median of the conditional realizations is used as the measure of central tendency, because it represents the estimated value for which there is an equal probability of the true parameter value being above or below the estimate. The mean can clearly also be used as a measure of central tendency, depending on the application.

[21] MCMC methods allow for the sampling of probability density functions in multiple dimensions with computational effort that is manageable relative to performing the multi-dimensional integrations that would otherwise be required. The dimensionality of the posterior pdf is equal to the number of points in the discretized parameter field, and can therefore easily be on the order of thousands to millions.

[22] One of the methods falling into the MCMC category is the Gibbs sampling algorithm. In this approach, conditional realizations are generated by sequentially sampling the marginal (i.e. 1-dimensional) probability density function at each point in the discretized parameter field, while holding the values at all other points constant. This marginal

pdf is defined using the most recent conditional realization available at each of the other points in the discretized parameter space. It can be shown that, once the chain has converged, the realizations resulting from this process are equally likely realizations from the full multi-dimensional posterior pdf (see, for example, *Casella and George* [1992]).

4.1. Description of Constrained Geostatistical Gibbs Sampling (CGGS) Algorithm

[23] This section describes the Gibbs sampling algorithm as implemented in the constrained interpolation and inverse modeling method. We define the l-th conditional constrained realization as $s_{cc,l}$. In this context, a conditional realization is one that has been conditioned on the data $\mathbf{z}$. The chain can be initialized with any realization that has a non-zero posterior probability. The Gibbs sampler proceeds as follows (see, for example, *Casella and George* [1992] for a more detailed discussion):

1. Set initial values $s_{cc,0} = (s_{uc,0}(x_1), s_{uc,0}(x_2),\ldots, s_{uc,0}(x_m))^T$ and initialize the iteration counter of the chain $l = 1$.

2. Obtain a new conditional constrained realization $s_{cc,l} = (s_{cc,l}(x_1), s_{cc,l}(x_2),\ldots, s_{cc,l}(x_m))^T$ from $s_{cc,l-1}$ through successive generation of values from the marginal pdf at each point (note the use of counter values l and $l-1$):

$$
\begin{aligned}
p(s_{cc,l}(x_1)) &= p(s_{cc,l}(x_1)|s_{cc,l-1}(x_2),\ldots,s_{cc,l-1}(x_m))\\
p(s_{cc,l}(x_2)) &= p(s_{cc,l}(x_2)|s_{cc,l}(x_1),s_{cc,l-1}(x_3)\ldots,s_{cc,l-1}(x_m))\\
p(s_{cc,l}(x_i)) &= p(s_{cc,l}(x_i)|s_{cc,l}(x_1),\ldots,s_{cc,l}(x_{i-1}),s_{cc,l-1}(x_{i+1}),\ldots,\\
&\qquad s_{cc,l-1}(x_m))\\
p(s_{cc,l}(x_m)) &= p(s_{cc,l}(x_m)|s_{cc,l}(x_1),\ldots,s_{cc,l}(x_{m-1}))
\end{aligned}
$$

3. Change counter l to $l+1$ and return to step 2 until convergence is reached.

[24] When convergence is reached, the resulting realization $s_{cc,l}$ is a conditional constrained realization from the full posterior pdf. Steps 2 and 3 can then be repeated to obtain additional conditional realizations. In the presented applications, convergence is evaluated by tracking the values of the two components (likelihood and prior) of the posterior probability of the realizations. When the running average of both components stabilizes, convergence has been reached. The number of realizations discarded during this ramp-up portion of the Markov chain is listed in section 5 for each of the presented applications. Once the realizations are representative of the probability space of the estimated parameters, the chain is run until the probability space has been appropriately sampled. The starting realization and the length of the Markov Chain are also listed in section 5 for each of the presented applications. The reader is referred to the original work by *Geman and Geman* [1984], and subsequent publications by *Casella and George* [1992], *Gamerman* [1997], *Gelman et al.* [1995], and *Robert and Casella* [2004] for a more thorough discussion of convergence in the context of the Gibbs sampler.

4.2. Derivation of Marginal Posterior Probability Density Function

[25] In order to apply the Gibbs sampling algorithm as described in the previous section, a marginal pdf from which it is possible to draw samples of $\mathbf{s}$ at a single point x_i is needed. The problem is broken into two components, first presenting the marginal likelihood, then the marginal prior.

4.2.1. Marginal Likelihood of s_i

[26] The likelihood of the measurements, defined in equation (9), can be expressed as [*Michalak and Kitanidis*, 2003]:

$$p(\mathbf{z}|\mathbf{s}) = (2\pi)^{-n/2}\sigma_R^{-n}\exp\left[\sum_{j=1}^{n}\left(-\frac{1}{2\sigma_R^2}\left(z_j-\sum_{k=1}^{m}H_{j,k}s_k\right)^2\right)\right] \tag{10}$$

Therefore, the marginal likelihood for a single point in the discretized parameter field $\mathbf{s}$, if values at all other points are held constant, is:

$$
\begin{aligned}
p(\mathbf{z}|s_i) &= (2\pi)^{-n/2}\sigma_R^{-n}\\
&\quad\cdot\exp\left[\sum_{j=1}^{n}\left\{-\frac{1}{2\sigma_R^2}\left(z_j-\sum_{k=1,k\neq i}^{m}H_{j,k}s_k-H_{j,i}s_i\right)^2\right\}\right]\\
&= (2\pi)^{-n/2}\sigma_R^{-n}\prod_{j=1}^{n}\exp\left[-\frac{H_{j,i}^2}{2\sigma_R^2}\left(s_i-\frac{z_j-\sum_{k=1,k\neq i}^{m}H_{j,k}s_k}{H_{j,i}}\right)^2\right]
\end{aligned} \tag{11}
$$

combining the n Gaussian distributions, we obtain

$$p(\mathbf{z}|s_i) = \frac{1}{\sqrt{2\pi\tau_L}}\exp\left(-\frac{1}{2}\frac{(s_i-\mu_L)^2}{\tau_L}\right) \tag{12}$$

where

$$
\begin{aligned}
\mu_L &= \frac{\mathbf{a}^T\mathbf{b}}{\mathbf{b}^T\mathbf{b}} \qquad \tau_L = \frac{\sigma_R^2}{\mathbf{b}^T\mathbf{b}}\\
a_j &= z_j-\sum_{k=1,k\neq i}^{m}H_{j,k}s_k, \qquad j=1,\ldots,n\\
b_j &= H_{j,i}, \qquad j=1,\ldots,n
\end{aligned} \tag{13}
$$

[27] In the case of interpolation

$$
\begin{aligned}
H_{j,i} &= 1 \quad \text{if} \quad x_{z,j} = x_{s,i}\\
H_{j,i} &= 0 \quad \text{otherwise}
\end{aligned} \tag{14}
$$

yielding $\mu_L = z_j$ if the ith estimation point $x_{s,i}$ is at the jth measurement point $x_{z,j}$, and $p(\mathbf{z}|s_i) \propto c$, where c is a constant, otherwise. For the special case of interpolation with no measurement error, $p(\mathbf{z}|s_i) = \delta_{s_i=z_j}$ at a measurement point; in other words, the estimate at a measurement location is fixed at the measurement value.

[28] Note that the equations presented in this section have been written under the common assumption that measurement errors are independent, as described in section 3. The presented approach can also be applied with correlated errors, modifying the presented equations to reflect the resulting off-diagonal terms in $\mathbf{R}$, in a manner analogous to the way in which the off-diagonal terms in $\mathbf{Q}$ are treated in the following section.

4.2.2. Marginal Prior of s_i

[29] The prior probability density function of the discretized parameter field $\mathbf{s}$ was defined in equations (4) and (7). The unconstrained portion of the marginal prior pdf of a single point in the discretized parameter field is therefore $p'(s_i|\mathbf{s})$, where, in this case, $\mathbf{s} = \{s_1,\ldots,s_{i-1},s_{i+1},\ldots,s_m\}$. This portion of the marginal prior can be rearranged to be

$$p'(s_i|\mathbf{s}) \propto \exp\left[-\frac{1}{2}(\mathbf{s}-\mathbf{X}\boldsymbol{\beta})^T\mathbf{Q}^{-1}(\mathbf{s}-\mathbf{X}\boldsymbol{\beta})\right] \propto \exp\left[-\frac{1}{2}\sum_{j=1}^{m}\sum_{k=1}^{m}\left(s_j-\sum_{l=1}^{p}X_{j,l}\beta_l\right) \cdot \left(Q^{-1}\right)_{j,k}\left(s_k-\sum_{i=1}^{p}X_{j,l}\beta_l\right)\right] \tag{15}$$

where $(Q^{-1})_{j,k}$ refers to the j,k-th component of $\mathbf{Q}^{-1}$, not the inverse of the j,k-th component of $\mathbf{Q}$. This expression is the probability density function of s_i, given known values for all other elements of $\mathbf{s}$. This is equivalent to a standard kriging problem where all elements of $\mathbf{s}$ except for s_i are treated as the measurement locations, and s_i is the point at which an estimate is needed. The mean and variance of the distribution of s_i define a normal distribution:

$$p'(s_i|\mathbf{s}) = \frac{1}{\sqrt{2\pi\tau_P}}\exp\left(-\frac{1}{2}\frac{(s_i-\mu_P)^2}{\tau_P}\right) \tag{16}$$

where the mean μ_P and variance τ_P are obtained as in the solution of a regular universal kriging system of equations, holding all measurement and other estimation points constant, and estimating at a single point:

$$\begin{bmatrix} \mathbf{Q}_{k\neq i,k\neq i} & \mathbf{X}_{k\neq i} \\ \mathbf{X}_{k\neq i}^T & 0 \end{bmatrix}\begin{bmatrix} \boldsymbol{\lambda} \\ -\boldsymbol{\nu} \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_{k\neq i,i} \\ \mathbf{X}_i^T \end{bmatrix} \tag{17}$$

which yields the kriging weights $\boldsymbol{\lambda}$ and the Lagrange multipliers $\boldsymbol{\nu}$. These parameters define the mean and variance of the estimate of s_i as:

$$\mu_P = \sum_{j=1}^{m-1}\lambda_j s_j$$
$$\tau_P = Q_{i,i} - \sum_{j=1}^{m-1}\lambda_j Q_{j,i} + \sum_{l=1}^{p}X_{i,l}\nu_l \tag{18}$$

where $\mathbf{s}_j = \mathbf{s}_{k\neq i}$ includes all the most recent sampled values at all other estimation points, and $Q_{j,i} = Q_{k\neq i,i}$ is the covariance between the current estimation point, and all other estimation points. For interpolation, the measurement locations are included as estimation points in $\mathbf{Q}$. Equivalently, the influence of the trend $\mathbf{X}\beta$ can first be removed, and these statistics can then instead be represented as:

$$\mu_P = -\frac{\sum_{k=1,k\neq i}^{m} G_{i,k}s_k}{G_{i,i}} \qquad \tau_P = \frac{1}{G_{i,i}} \tag{19}$$

where

$$\mathbf{G} = \mathbf{Q}^{-1} - \mathbf{Q}^{-1}\mathbf{X}\left(\mathbf{X}^T\mathbf{Q}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Q}^{-1} \tag{20}$$

[30] Note that obtaining μ_P and τ_P at each point requires the solution of m systems of m linear equations, which are then used for each realization. However, the kriging weights and Lagrange multipliers are only a function of the arrangement of the measurements and estimation locations, not of the values of the samples or realizations at these locations (see equation (17)). Therefore, these systems of equations need only be solved a single time, and not for every realization, yielding considerable computational savings.

4.2.3. Marginal Posterior of s_i

[31] Combining the marginal likelihood and prior, the full marginal posterior pdf of s_i can be expressed, to within a normalizing constant, as:

$$p(s_i|\mathbf{s},\mathbf{z}) \propto \prod_{l=1}^{k}\mathcal{H}(g_l(s_i)-b_l)\frac{1}{\sqrt{2\pi\tau_L}} \cdot \exp\left(-\frac{1}{2}\frac{(s_i-\mu_L)^2}{\tau_L}\right)\frac{1}{\sqrt{2\pi\tau_P}}\exp\left(-\frac{1}{2}\frac{(s_i-\mu_P)^2}{\tau_P}\right) \tag{21}$$

which can be written as:

$$p(s_i|\mathbf{s},\mathbf{z}) \propto \frac{1}{\sqrt{2\pi\tau}}\exp\left(-\frac{1}{2}\frac{(s_i-\mu)^2}{\tau}\right)\prod_{l=1}^{k}\mathcal{H}(g_l(s_i)-b_l) \tag{22}$$

where,

$$\tau = \frac{\tau_P\tau_L}{\tau_P+\tau_L} \qquad \mu = \tau\left(\frac{\mu_P}{\tau_P}+\frac{\mu_L}{\tau_L}\right) \tag{23}$$

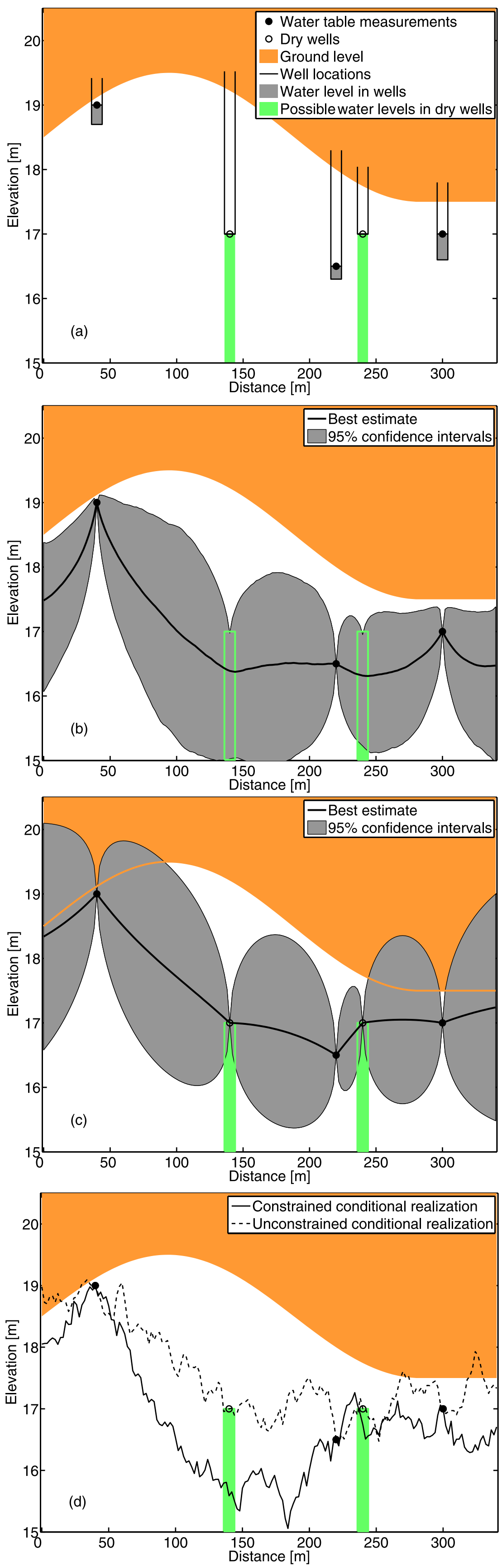
This general formulation can be applied to include constraints in kriging with no measurement error (e.g. ordinary kriging, universal kriging, kriging with a trend, etc.), kriging with measurement error (a.k.a. continuous-part kriging), and linear geostatistical inverse modeling.

4.2.4. Sampling of Marginal Posterior of s_i

[32] We are interested in sampling from a normal distribution that has been truncated at the inequality constraints. As such, we simply draw a random sample from the distribution $N(\mu,\tau)$, until we obtain one that is in the allowable range for s_i. Once we obtain an admissible sample, this realization of s_i is used as the next conditional constrained realization at point x_i, denoted $s_{cc,l}(x_i)$. The computational efficiency of this sampling approach can further be improved by implementing a more efficient algorithm for obtaining a sample from a one-dimensional truncated normal distribution. *Robert* [1995], for example, presents an efficient method for sampling the portion of a normal distribution lying within the constraint boundaries, based on an acceptance-rejection scheme using an exponential distribution.

5. Applications

[33] The following section includes three sample applications of the CGGS approach. The first application represents a hypothetical case of interpolating hydraulic head in



the presence of censored data and variable topography. The second application involves the estimation of contaminant load in a river with censored data and a nonnegativity constraint. The third application is a hypothetical inverse modeling example involving the estimation of the historical distribution of a contaminant in an aquifer. These sample applications were selected to be both simple and illustrative of a variety of applications for the presented approach, with the recognition that field applications will often have higher dimensionality relative to the presented examples.

5.1. Example 1: Hydraulic Head Mapping

[34] The first application involves the estimation of a hypothetical hydraulic head distribution in an unconfined aquifer. Five wells are present in the aquifer, three of which measure the hydraulic head, and two of which are dry. The two dry wells represent censored data, or inequality constrained data, because we know that the water table is below the bottom of the well, but we do not know its exact location. We wish to estimate the water table depth at all unsampled locations, including the two dry wells. As an additional constraint, we know that the groundwater table cannot be above the land surface, and all estimation locations can also therefore be thought of either as inequality constrained parameters, or as censored data. For simplicity, we assume that the measurements in the wells are exact (i.e. no measurement error). An example where this assumption does not hold is presented in the second application.

[35] Figure 1a presents the available data and estimation constraints. We assume that the unconstrained spatial covariance of the water table distribution can be represented using an exponential covariance function:

$$Q_{ij} = \sigma^2 \exp(-h_{ij}/l) \tag{24}$$

with parameters $\sigma^2 = 1.0$ m^2, $l = 600$ m, where h_{ij} is the separation distance between estimation points i and j. In a practical setting, this covariance information would be derived from additional sampling or previous experience. The hydraulic head is estimated at 1 m intervals.

[36] Figure 1c presents the estimated water table distribution using ordinary kriging. Because ordinary kriging cannot directly handle the inequality constraints imposed by the dry wells, these measurement location either have to be ignored, or the water table depth has to be set to a fixed value. In this case, ignoring the dry wells would yield best estimates at the dry well locations above the bottom of one of these wells, a physically inconsistent result. Therefore, as is often done in practice, the water table in the dry wells is assumed to be at the bottom of the well. This approach underestimates the uncertainty in the groundwater level at and around the dry well locations. In addition, ordinary

Figure 1. (a) Hypothetical hydraulic head measurements. (b) Best estimate and 95% confidence intervals for hydraulic head distribution using constrained estimation method. (c) Best estimate and 95% confidence intervals for hydraulic head distribution using unconstrained estimation method. (d) Sample conditional realizations for hydraulic head distribution using constrained and unconstrained estimation methods.

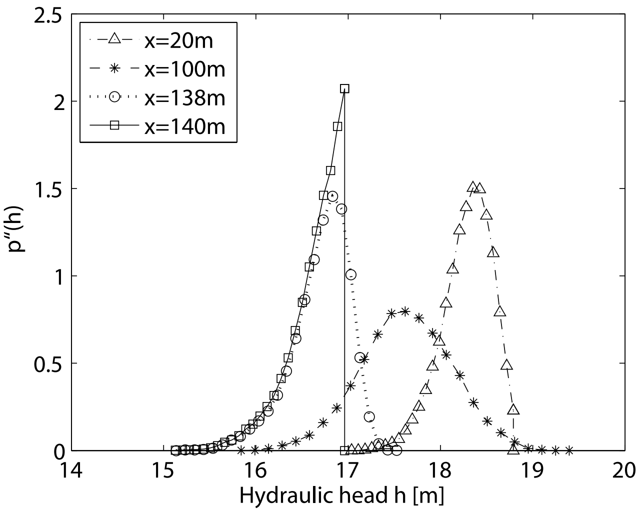


Figure 2. A posteriori hydraulic head probability density function at four sample points.

kriging cannot incorporate the ground level constraint on the water table depth, and the uncertainty bounds of the estimate extend above ground. This is again a physically inconsistent result stemming from the Gaussian assumptions common to estimating uncertainty bounds using the kriging variance.

[37] Figure 1b presents the estimated water table distribution using the CGGS algorithm. No measured water table depth is specified at the dry well locations, but the groundwater level is constrained to be below the bottom of the well. In addition, the realizations are everywhere constrained to be below the ground surface. The Markov chain was initialized with the ordinary kriging best estimate of the hydraulic head distribution, because this was a realistic starting point that also did not violate the constraints for this application. The chain is run for a total of 50,000 realizations, because the large distance between measurements requires a relatively large number of realizations in order to effectively sample the uncertainty space. Because the starting point for the Markov chain is very good in this case, only the first 100 realization of the chain are discarded in the analysis. Contrary to the ordinary kriging result, none of the estimates or their uncertainty bounds violate the physical constraints imposed by the dry wells or ground surface. In addition, the uncertainty associated with the groundwater level at the dry wells is realistically represented, with wide uncertainty on the water table depth at these locations. Interestingly, for both dry wells, the best estimate of the groundwater depth is significantly below the bottoms of the wells for this example. Finally, the uncertainty intervals on the water table depth are not symmetric close to constraint boundaries, which is realistic for this situation and results from the use of a non-Gaussian a priori pdf.

[38] Figure 1d presents a sample conditional realization generated using a multi-Gaussian assumption as is representative of a traditional kriging sampling approach, and a sample conditional realization generated using the proposed constrained algorithm. The two realizations exhibit a similar degree of spatial variability, because they are both based on the same prior covariance model. As was already demonstrated in Figure 1b, however, the realization generated using the CGGS algorithm does not reach above ground level, and is better reflective of the uncertainty in groundwater levels at the locations of the dry wells.

[39] Note that the marginal pdf at each point for each conditional realization is modeled as a truncated Gaussian distribution. However, the overall pdf describing the uncertainty across the ensemble of realizations can take on a variety of forms (Figure 2). Away from constraint boundaries (e.g. x = 100 m), the pdf is close to Gaussian and similar to that which would be obtained using a traditional kriging setup. At locations where constraints have a significant impact on the estimates (e.g. x = 140 m), the final pdf looks like the truncated Gaussian pdf used in the sampling procedure. Near such boundaries (e.g. x = 138 m, x = 20 m), the distribution is skewed. These pdfs reflect the spread in the ensemble of conditional realizations used to characterize the uncertainty associated with parameter values at unsampled locations. Note that the product of a multidimensional truncated-Gaussian prior with a Gaussian likelihood function does not point-wise yield a truncated Gaussian distribution.

5.2. Example 2: Contaminant Load Estimation

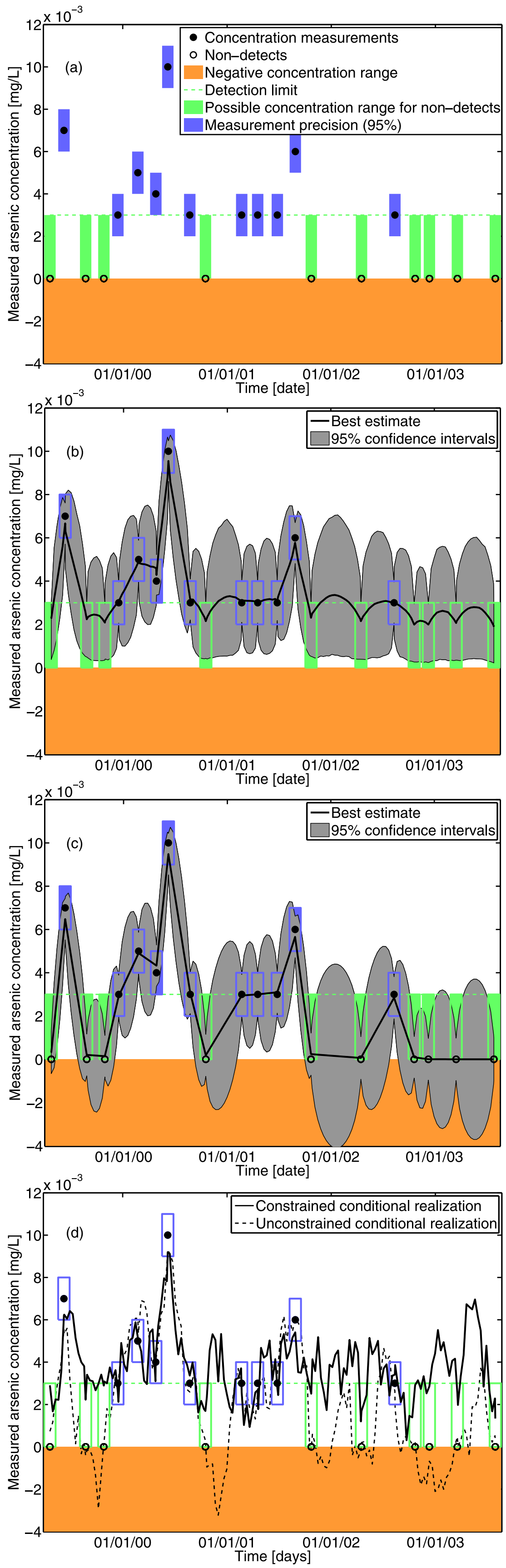
[40] The Humboldt River basin is located in North-Eastern Nevada in the United States, and its water resources have a variety of recreational and agricultural uses. The Humboldt River contains arsenic which results in large part from mining practices when mineralized rock is crushed and exposed to oxygen and water. The concentration history of dissolved arsenic in the North Fork of the Humboldt river and the total dissolved arsenic load supplied to downstream locations were previously estimated using the nonnegativity-enforcing Gibbs sampling approach of *Michalak and Kitanidis* [2005].

[41] The CGGS approach not only allows for this lower concentration bound to be enforced and measurement errors to be taken into account, but also provides a statistically rigorous methods for accounting for censored data where the measured concentration was below the detection limit. The new approach is compared with results obtained using ordinary kriging with a linear variogram, chosen based on an examination of the experimental variogram of available data and for easy comparison to Example 3 in *Michalak and Kitanidis* [2005]. The covariance structure of the contaminant concentrations is therefore modeled using a linear generalized covariance function:

$$K_{ij} = -\theta h_{ij} \tag{25}$$

where $\theta = 10^{-8}$ (μg/l) ^{2}day^{-1}, h_{ij} is the time lag between the i and j estimation times, and the generalized covariance takes the place of the covariance Q_{ij} used for stationary parameter distributions.

[42] The concentration data were obtained from the EPA STORET database [*EPA*, 2003] and are plotted in Figure 3a. All measured and estimated concentrations must be nonnegative. Based on documentation from EPA, the detection limit is 3 μg/l. The measurement error is assumed to be normally-distributed with a variance of 0.25(μg/l)^2, yielding 95% confidence bounds of ±1 μg/l, corresponding to the reported data precision.



[43] We discretize the concentration history into ten-day intervals, augmented by the times at which measurements were actually taken, with time zero starting on the day of the first measurement, 21 April 1999, yielding a total of m = 173 estimation times. Note that although we are interested in the variability of the arsenic concentration in time rather than in space, the approach presented in sections 3 and 4 is directly applicable, simply by substituting temporal coordinates t for spatial coordinates x in the algorithm presented in section 4.1.

[44] For this application, the median and 95% confidence intervals of the probability density functions of concentration values at ten day intervals are determined based on ensemble properties of conditional realizations generated using the method described in section 4. The Markov chain is initialized with the ordinary kriging best estimate of the concentration history, because this is a realistic starting point that also does not violate the constraints for this application. The chain is run for a total of 10,000 realizations. Because the starting point for the Markov chain is very good in this case, only the first 100 realization of the chain are discarded in the analysis. Results are plotted in Figure 3b. The equivalent plot using kriging with a linear variogram is presented in Figure 3c. As can be seen in these figures, the proposed approach behaves similarly to the kriging interpolation away from constraints, but the best estimate near constraints deviates from the kriging estimates. For the kriging application, the non-detect points are assumed to have no arsenic (0 μg/l). An alternative approach which is sometimes used is to set non-detects at half the detection limit (1.5 μg/l in this case).

[45] As can be seen from Figure 3b, the new methodology is effective at enforcing parameter nonnegativity, and constraining non-detects to below the detection limit without specifying a prior estimate of concentrations at those times. The measurement uncertainty is also reflected in the estimates. Note that the measurement error is modeled through the likelihood term (section 4.2), whereas the non-detects are modeled as an interval constraint in the range of zero to the detection limit. By design, the new methodology behaves similarly to kriging with a linear variogram in high concentration regions.

[46] Traditional geostatistical simulation, on the other hand, leads to conditional realizations and confidence intervals reaching into the negative parameter range, which have no physical significance, and can be misleading. In addition, kriging requires explicit assumptions about the concentration for non-detect samples, and cannot account for the finite uncertainty range for these measurement times.

[47] Figure 3d presents a sample conditional realization generated using a multi-Gaussian assumption as is repre-

Figure 3. (a) Concentration measurements for North Fork of Humboldt River at North Fork Ranch, Elko county, Nevada. (b) Best estimate and 95% confidence intervals for concentration as a function of time using constrained estimation method. (c) Best estimate and 95% confidence intervals for concentration as a function of time using unconstrained estimation method. (d) Sample conditional realizations for concentration history using constrained and unconstrained estimation methods.

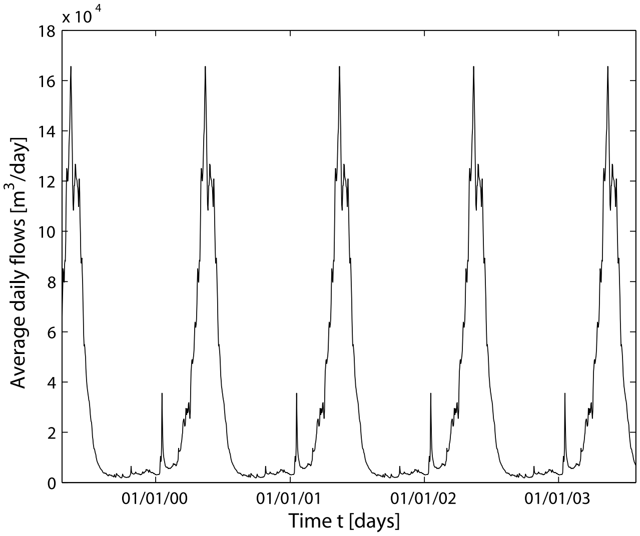


Figure 4. Averaged flow data at one-day intervals for North Fork of Humboldt River, Elko county, Nevada.

sentative of a traditional kriging setup, and a sample conditional realization generated using the CGGS algorithm. The two realizations exhibit similar degrees of spatial variability, because they are both based on the same covariance model. As was already demonstrated in Figure 3c, however, the realization generated using the proposed algorithm does not violate the nonnegativity constraint, and is better reflective of the uncertainty in concentrations at times of non-detect samples.

[48] The obtained estimates can also be used in conjunction with flowrate information to estimate total contaminant load, as was presented in *Michalak and Kitanidis* [2005]. River flow data for the equivalent time period, however, were not available. Therefore, flows for 1 January 1971, through 31 December 1981, are averaged to obtain a representative hydrograph for the stream [*USGS*, 2001]. These daily average flows are used to estimate the flowrate history for the period of 21 April 1999, through 30 July 2003, by assigning to each day a flowrate equivalent to the average flow for that calendar day. These flows are presented in Figure 4.

[49] To estimate the total contaminant load, individual conditional realizations are weighted using river flows, yielding an ensemble of contaminant loads that can be used to describe the uncertainty associated with this quantity. The total loads are presented in Figure 5 for the kriging and constrained approaches. As discussed in *Michalak and Kitanidis* [2005], for this river, high concentration events are associated with high flows. Therefore, because the proposed method has a stronger impact close to constraint boundaries, the impact of the constraints on total load is relatively limited. However, Figure 5 shows that the kriging approach consistently underestimates contaminant loads relative to the constrained approach, which is more consistent with the physical bounds on parameter values. This effect is more pronounced than that discussed in *Michalak and Kitanidis* [2005], because the current approach is able to provide a better representation of observations with non-detect concentrations. At the end of the four-year period, the new approach estimates a statistically significantly higher contaminant load relative to the kriging approach, at the 0.05 confidence level. This effect is important because it implies that methods that do not account for physical constraints on parameter values, especially with regard to nonnegativity and representation of non-detect values, can lead to strong underestimation of contamination.

5.3. Example 3: Estimation of Historical Contaminant Distribution

[50] The final example application involves the identification of the historical distribution of a contaminant in a two-dimensional aquifer, and is modeled after the heterogeneous example presented in *Michalak and Kitanidis* [2004b]. The two-dimensional distribution at time T_a is estimated based on downgradient concentration measurements taken at time $T_b = T_a + 2000$ days. The affected aquifer is represented as having a deterministically heterogeneous hydraulic conductivity field.

[51] The domain is finite, measuring 1024 m and 512 m in the x_1 and x_2 directions, respectively. It is discretized into 128×64 nodes in the x_1 and x_2 directions, respectively, resulting in an 8 m $\times$ 8 m grid. No-flux boundary conditions are applied at the top and bottom boundaries for both flow and transport. The left-hand side and right-hand side boundaries have prescribed constant heads, resulting in a mean gradient of 3.472×10^{-2} m/m. Details regarding the aquifer heterogeneity are available in *Michalak and Kitanidis* [2004b]. The flow solution is obtained using MODFLOW [*McDonald and Harbaugh*, 1988; *Harbaugh and McDonald*, 1996].

[52] The actual contaminant distribution at time T_a used in this example is presented in Figure 6a. The plume profile at time T_b is obtained using MT3DMS [*Zheng*, 1990; *Zheng and Wang*, 1999]. The boundary conditions used to solve the forward problem are:

$$\begin{aligned} C(\mathbf{x},t) = 0 \quad & x_1 = [0, 1024];\ x_2 \in [0, 512] \\ \left(\eta v_i C - \eta D_{ij}\frac{\partial C}{\partial x_j}\right) n_i = 0 \quad & x_1 \in [0, 1024];\ x_2 = [0, 512] \end{aligned} \tag{26}$$

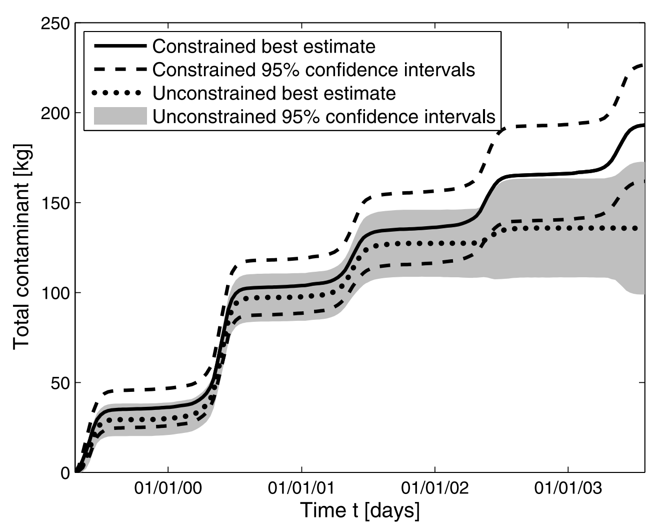


Figure 5. Cumulative arsenic load best estimate and 95% confidence intervals for constrained and unconstrained estimation methods.

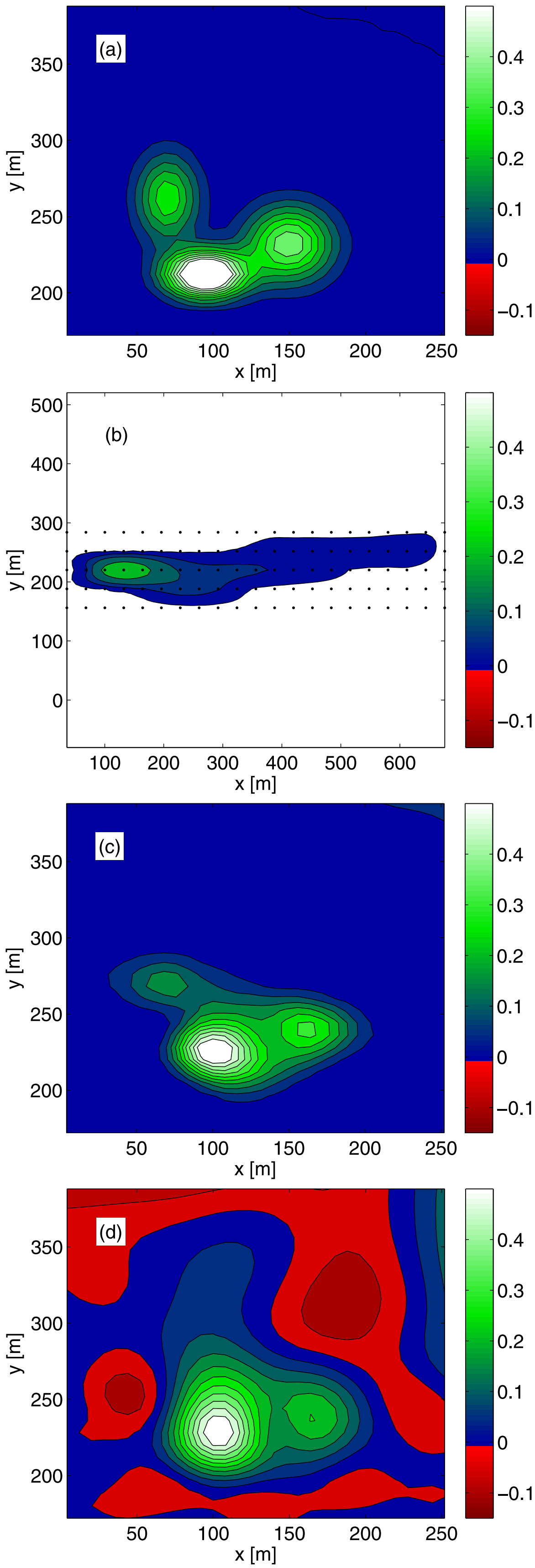


Figure 6. (a) Actual contaminant distribution at time T_a. (b) Actual contaminant distribution at time T_b and measurement locations for inverse modeling application. (c) Recovered contaminant distribution for time T_a using constrained estimation method. (d) Recovered contaminant distribution for time T_a using unconstrained estimation method. Concentration units are mg/l.

where t is time, x_i are the spatial directions (i = 1, 2), $\mathbf{x} = (x_1, x_2)$, C is resident concentration, η is porosity, D_{ij} is the i,jth entry of the dispersion tensor, and v_i is fluid velocity in the direction x_i. Uncorrelated random error with a standard deviation of 1×10^{-3} mg/l is added to the observations to represent measurement error.

[53] The distribution at time T_b is presented in Figure 6b, along with sampling locations. The sampling is conducted on a 32 m × 32 m grid, yielding a total of 105 observation locations. We recover the contaminant distribution in the region $\Omega_a = \{\mathbf{x}\colon x_1 \in (0,256),\ x_2 \in (168, 392)\}$. For the purpose of solving the inverse problem, this area is discretized into 8 m intervals, yielding 896 points at which the concentration at time T_a is to be estimated. This represents a strongly underdetermined problem. The adjoint approach of *Michalak and Kitanidis* [2004b] is applied to define the sensitivity matrix **H** needed to solve the inverse problem, which defines the sensitivity of each observation at time T_b to a historical concentration at each location in the domain Ω_a. The covariance of the concentration distribution at time T_a is taken from this earlier work, where it was estimated using a Restricted Maximum Likelihood approach. A cubic generalized covariance model was used:

$$K_{ij} = \theta h_{ij}^3 \tag{27}$$

where h_{ij} is the physical separation distance between the ith and jth locations at which the contaminant distribution is to be estimated, and $\theta = 10^{-6}$ (mg/l)^2 m^{-3} [*Michalak and Kitanidis*, 2004b].

[54] Figure 6d presents the recovered historical distribution using a linear geostatistical inverse modeling approach, analogous to the method presented in *Michalak and Kitanidis* [2004b]. Although the overall distribution at time T_a is recovered reasonably well, the best estimate includes areas with negative concentrations. Even in locations where the best estimate itself is positive, the uncertainty bounds can encompass negative values, as seen in the one-dimensional slice presented in Figure 7b. Figure 6c presents the recovered historical distribution using the proposed CGGS approach. The Markov chain is initialized with the absolute value of the best estimate obtained using linear inverse modeling, because this is a starting point that does not violate the constraints for this application. The chain is run for a total of 1000 realizations. As in the two previous applications, the first 100 realization of the chain are discarded in the analysis. The applied constraint in this case is nonnegativity within the entire domain. As can be seen from this figure, the best estimate is indeed everywhere nonnegative. In addition, as seen in Figure 7a, the entire probability density function at each point is constrained to be nonnegative, yielding positive uncertainty bounds.

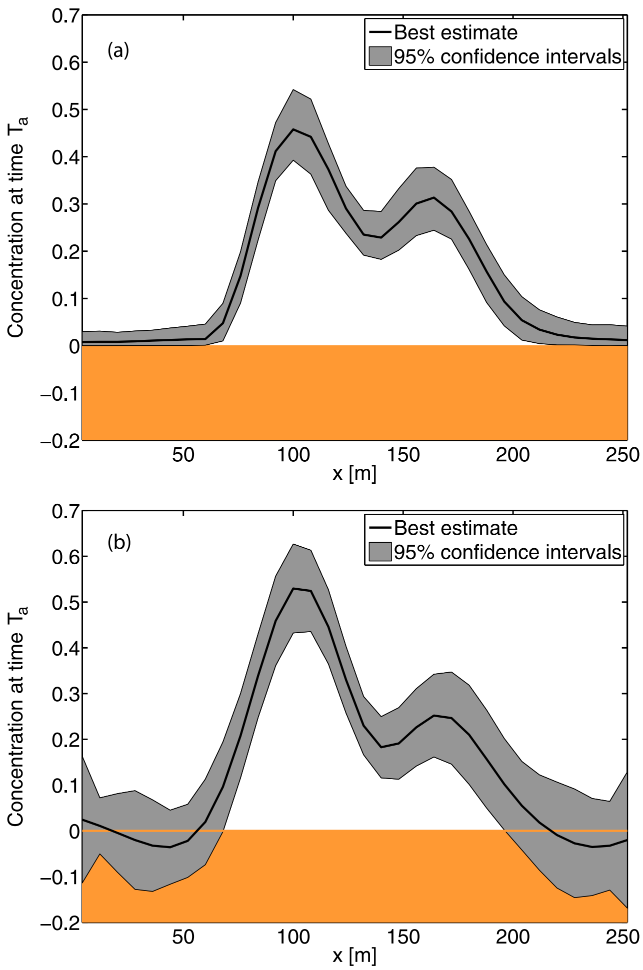


Figure 7. Cross-section at y = 236 m of recovered contaminant distribution for time T_a. (a) Using constrained estimation method. (b) Using unconstrained estimation methods. Concentration units are mg/l.

[55] Furthermore, the accuracy and precision of the estimates are improved by the addition of the constraint. First, the third peak in the historical contaminant distribution, which is absent in the estimate obtained using the linear approach, is correctly inferred in the best estimates of the constrained approach. Second, because of the strong implicit constraint on total contaminant mass offered by the plume measurements, the addition of the nonnegativity constraint decreases the uncertainty throughout the domain. Conceptually, through mass conservation, by eliminating the possibility for negative concentrations, the possibility for some large positive concentrations is eliminated as well. This effect can be seen clearly by comparing Figures 7a and 7b, where the uncertainty bounds for the constrained approach are everywhere narrower relative to the linear approach. This effect is especially pronounced in areas of low concentration. Overall, the new approach successfully enforces physical constraints in an inverse modeling setup, while improving the precision and accuracy of the obtained estimates. Note that the inverse modeling approach used for the solution of the solute inverse problem assumes a known transport model, parameterized in **H**. The current literature on stochastic methods for solving solute transport inverse problems in groundwater hydrology does not consider uncertainties in transport parameters (see, e.g. review in *Michalak and Kitanidis* [2004b]), although a few recent works have considered such uncertainty in a deterministic context [e.g. *Sun et al.*, 2006; *Sun*, 2007]. Relaxing this assumption within a probabilistic framework is the topic of ongoing research. The innovation presented in the current paper, however, focuses specifically on the treatment of the constraints within the solution space.

6. Conclusions

[56] The presented approach provides a statistically-rigorous methodology for geostatistical interpolation and inverse modeling, subject to multiple and spatially-variable inequality constraints. The approach uses a Gibbs sampler to characterize the marginal probability distribution at each estimation point, using a truncated Gaussian prior probability distribution. As presented in the current work, the method is applicable to cases where a truncated Gaussian distribution is a good representation of the a priori uncertainty on the parameter distribution, and would therefore be less applicable to highly skewed distributions. In addition, for applications to inverse problems, the method assumes a linear and known forward model.

[57] From a methodological perspective, the three presented applications demonstrate the applicability of the proposed approach to a wide range of problems and constraints. Censored data with an upper bound were used in Example 1, whereas interval censored data were used in Example 2. Nonnegativity was enforced in Examples 2 and 3. Example 1 also demonstrated the applicability of the proposed approach to data with a spatially-variable inequality constraint. Applications to interpolation were presented in Examples 1 and 2, whereas an application to geostatistical inverse modeling was shown in Example 3. Example 1 presented an application in one spatial dimension, Example 2 presented a temporal problem, and Example 3 represented an application in two spatial dimensions. Finally the three presented examples also applied three different covariance structures (exponential covariance, linear generalized covariance, cubic generalized covariance).

[58] Scientifically, the presented examples demonstrate the broad need for constrained interpolation and inverse modeling approaches, with sample applications to groundwater flow, surface water quality, and groundwater solute transport. The presented approach can also be applied to a wide range of other data types and problems.

[59] Overall, the geostatistical Gibbs sampler using a truncated Gaussian prior distribution offers an effective, simple, computationally efficient and universally applicable approach for spatial and temporal interpolation and inverse modeling problems in environmental applications where inequality constraints are present. In addition, the method yields estimates with high precision and accuracy relative to classical geostatistical methods.

[60] **Acknowledgments.** This material is based upon work supported by the National Science Foundation under grant 0644648 "Development of Geostatistical Data Assimilation Tools for Water Quality Monitoring." Olaf Cirpka, Jinsong Chen and one anonymous reviewer contributed significantly to the final version of this paper.

References

Abrahamsen, P., and F. E. Benth (2001), Kriging with inequality constraints, *Math. Geol.*, *33*, 719–744.
Armstrong, M., A. G. Galli, G. Le Loc'h, F. Geffroy, and R. Eschard (2003), *Plurigaussian Simulations in Geosciences*, Springer-Verlag, Berlin.
Balakrishnan, S., A. Roy, M. G. Ierapetritou, G. P. Flach, and P. G. Georgopoulos (2003), Uncertainty reduction and characterization for complex environmental fate and transport models: An empirical Bayesian framework incorporating the stochastic response surface method, *Water Resour. Res.*, *39*(12), 1350, doi:10.1029/2002WR001810.
Barnes, R. J., and K. H. You (1992), Adding bounds to kriging, *Math. Geol.*, *24*, 171–176.
Bates, B. C., and E. P. Campbell (2001), A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling, *Water Resour. Res.*, *37*, 937–947.
Campbell, E. P., et al. (1999), A Bayesian approach to parameter estimation and pooling in nonlinear flood event models, *Water Resour. Res.*, *35*, 211–220.
Casella, G., and E. I. George (1992), Explaining the Gibbs sampler, *Am. Stat.*, *46*, 167–174.
Chen, J. S., S. Hubbard, Y. Rubin, C. Murray, E. Roden, and E. Majer (2004), Geochemical characterization using geophysical data and Markov Chain Monte Carlo methods: A case study at the South Oyster bacterial transport site in Virginia, *Water Resour. Res.*, *40*(12), W12412, doi:10.1029/2003WR002883.
Chen, J., S. Hubbard, J. Peterson, K. Williams, M. Fienen, P. Jardine, and D. Watson (2006), Development of a joint hydrogeophysical inversion approach and application to a contaminated fractured aquifer, *Water Resour. Res.*, *42*(6), W06425, doi:10.1029/2005WR004694.
Chib, S., and E. Greenberg (1995), Understanding the Metropolis-Hastings Algorithm, *Am. Stat.*, *49*, 327–335.
Chilès, J. P., and P. Delfiner (1999), *Geostatistics: Modeling Spatial Uncertainty*, John Wiley, New York.
De Oliveira, V. (2005), Bayesian inference and prediction of Gaussian random fields based on censored data, *J. Comput. Graph. Stat.*, *14*, 95–115.
Deutsch, C. V. (1996), Correcting for negative weights in ordinary kriging, *Comput. Geosci.*, *22*, 765–773.
Diggle, P. J., and P. J. Ribeiro Jr. (2007), *Model-Based Geostatistics*, Springer-Verlag, New York.
Dubrule, O., and C. Kostov (1986), An interpolation method taking into account inequality constraints. 1. Methodology, *Math. Geol.*, *18*, 33–51.
Environmental Protection Agency (2003), Downloading Data from the STORET Warehouse: An Exercise. (Available at http://www.epa.gov/storet/updates.html)
Fienen, M. N., P. K. Kitanidis, D. Watson, and P. Jardine (2004), An application of Bayesian inverse methods to vertical deconvolution of hydraulic conductivity in a heterogeneous aquifer at Oak Ridge National Laboratory, *Math. Geol.*, *36*, 101–126.
Freulon, X. (1994), Conditional simulation of a Gaussian random vector with nonlinear and/or noisy observations, in *Geostatistical Simulations*, pp. 57–71, Kluwer Acad. Publishers, Dordrecht, Netherlands.
Freulon, X., and C. de Fouquet (1993). Conditioning a Gaussian model with inequalities, in *Proceedings of Geostatistics Troia'92*, vol. 1, pp. 201–212, Kluwer Acad. Publishers, Dordrecht, Netherlands.
Fridley, B. L., and P. Dixon (2007), Data augmentation for a Bayesian spatial model involving censored observations, *Environmetrics*, *18*, 107–123.
Gamerman, D. (1997), *Markov Chain Monte Carlo*, Chapman and Hall, London.
Gelfand, A. E., A. F. M. Smith, and T. M. Lee (1992), Bayesian-analysis of constrained parameter and truncated data problems using Gibbs sampling, *J. Am. Stat. Assoc.*, *87*, 523–532.
Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (1995), *Bayesian Data Analysis*, Chapman and Hall, London.
Geman, S., and D. Geman (1984), Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE T. Pattern Anal.*, *6*, 721–741.
Harbaugh, A. W., and M. G. McDonald (1996), User's documentation for MODFLOW-96, an update to the U.S. Geological Survey modular finite-difference ground-water flow model, U.S. Geological Survey Open-File Report 96–485, 56 pp.
Hastings, W. K. (1970), Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, *57*(1), 97–109.
Journel, A. G. (1986), Constrained interpolation and qualitative information - the soft kriging approach, *Math. Geol.*, *18*, 269–305.
Karlin, S., and H. M. Taylor (1975), *A First Course in Stochastic Processes*, Academic Press, San Diego, Calif.
Kitanidis, P. K. (1997), *Introduction to Geostatistics Applications in Hydrogeology*, 249 pp., Cambridge Univ. Press, New York.
Kitanidis, P. K., and K. F. Shen (1996), Geostatistical interpolation of chemical concentration, *Adv. Water Resour.*, *19*, 369–378.
Lantuéjoul, C. (2002), *Geostatistical Simulation: Models and Algorithms*, Springer-Verlag, Berlin.
Leuangthong, O., and C. V. Deutsch (2004), Transformation of residuals to avoid artifacts in geostatistical modelling with a trend, *Math. Geol.*, *36*, 287–305.
Marshall, L., D. Nott, and A. Sharma (2004), A comparative study of Markov chain Monte Carlo methods for conceptual rainfall-runoff modeling, *Water Resour. Res.*, *40*(2), W02501, doi:10.1029/2003WR002378.
McDonald, M. G., and A. W. Harbaugh (1988), *A Modular Three-Dimensional Finite-Difference Ground-Water Flow Model*, U.S. Geological Survey Techniques of Water-Resources Investigations, book 6, chap. A1, 586 pp.
Michalak, A. M., and P. K. Kitanidis (2002), Application of Bayesian inference methods to inverse modeling for contaminant source identification at Gloucester Landfill, Canada, in *Computational Methods in Water Resources XIV*, vol. 2, edited by S. M. Hassanizadeh et al., pp. 1259–1266, Elsevier, Amsterdam, Netherlands.
Michalak, A. M., and P. K. Kitanidis (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.
Michalak, A. M., and P. K. Kitanidis (2004a), Application of geostatistical inverse modeling to contaminant source identification at Dover AFB, Delaware, *J. Hydraul. Res.*, *42*, 9–18.
Michalak, A. M., and P. K. Kitanidis (2004b), Estimation of historical groundwater contaminant distribution using the adjoint state method applied to geostatistical inverse modeling, *Water Resour. Res.*, *40*(8), W08302, doi:10.1029/2004WR003214.
Michalak, A. M., and P. K. Kitanidis (2005), A method for the interpolation of nonnegative functions with an application to contaminant load estimation, *Stoch. Env. Res. Risk A*, *19*, 8–23, doi:10.1007/s00477-004-0189-1.
Militino, A. F., and M. D. Ugarte (1999), Analyzing censored spatial data, *Math. Geol.*, *31*, 551–561.
Robert, C. P. (1995), Simulation of truncated normal variables, *Stat. Comput.*, *5*, 121–125.
Robert, C. P., and G. Casella (2004), *Monte Carlo Statistical Methods*, 2nd ed., Springer-Verlag, New York.
Saito, H., and P. Goovaerts (2000), Geostatistical interpolation of positively skewed and censored data in a dioxin-contaminated site, *Environ. Sci. Technol.*, *34*, 4228–4235.
Smith, A. F. M., and G. O. Roberts (1993), Bayesian computation via the Gibbs sampler and related Markov-chain Monte Carlo methods, *J. R. Stat. Soc., B*, *55*, 3–23.
Snodgrass, M. F., and P. K. Kitanidis (1997), A geostatistical approach to contaminant source identification, *Water Resour. Res.*, *33*, 537–546.
Sun, A. Y. (2007), A robust geostatistical approach to contaminant source identification, *Water Resour. Res.*, *43*(2), W02418, doi:10.1029/2006WR005106.
Sun, A. Y., S. L. Painter, and G. W. Wittmeyer (2006), A constrained robust least squares approach for contaminant release history identification, *Water Resour. Res.*, *42*(4), W04414, doi:10.1029/2005WR004312.
Szidarovszky, F., E. Y. Baafi, and Y. C. Kim (1987), Kriging without negative weights, *Math. Geol.*, *19*, 549–559.
USGS (2001), Daily Streamflow for the Nation USGS 10317400 N F Humboldt R NR N Fork, NV. (Available at http://www.water.usgs.gov/nwis/discharge/?site_ no10317400, accessed 11/01/01)
Vrugt, J. A., H. V. Gupta, L. A. Bastidas, W. Bouten, and S. Sorooshian (2003a), Effective and efficient algorithm for multiobjective optimization of hydrologic models, *Water Resour. Res.*, *39*(8), 1214, doi:10.1029/2002WR001746.
Vrugt, J. A., H. V. Gupta, W. Bouten, and S. Sorooshian (2003b), A shuffled complex evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, *Water Resour. Res.*, *39*(8), 1201, doi:10.1029/2002WR001642.
Walvoort, D. J. J., and J. J. de Gruijter (2001), Compositional kriging: A spatial interpolation method for compositional data, *Math. Geol.*, *33*, 951–966.

Yoo, E.-H., and P. C. Kyriakidis (2006), Area-to-point kriging with inequality-type data, *J. Geograph. Syst.*, *8*, 357–390, doi:10.1007/s10109-006-0036-7.

Zheng, C. (1990), MT3D, A modular three-dimensional transport model for simulation of advection, dispersion, and chemical reactions of contaminants in groundwater systems, Report to the Kerr Environmental Research Laboratory, US Environmental Protection Agency, Ada, Oklahoma.

Zheng, C., and P. P. Wang (1999), MT3DMS: A modular three-dimensional multispecies model for simulation of advection, dispersion and chemical reactions of contaminants in groundwater systems; Documentation and User's Guide, Contract Report SERDP-99-1, U. S. Army Engineer Research and Development Center, Vicksburg, MS.

A. M. Michalak, Department of Civil and Environmental Engineering, University of Michigan, 183 EWRE Building, 1351 Beal Avenue, Ann Arbor, MI 48109-2125, USA. (amichala@umich.edu)