Interpolation of Steady-State Concentration Data by Inverse Modeling

by Ronnie L. Schwede¹ and Olaf A. Cirpka²

Abstract

In most groundwater applications, measurements of concentration are limited in number and sparsely distributed within the domain of interest. Therefore, interpolation techniques are needed to obtain most likely values of concentration at locations where no measurements are available. For further processing, for example, in environmental risk analysis, interpolated values should be given with uncertainty bounds, so that a geostatistical framework is preferable. Linear interpolation of steady-state concentration measurements is problematic because the dependence of concentration on the primary uncertain material property, the hydraulic conductivity field, is highly nonlinear, suggesting that the statistical interrelationship between concentration values at different points is also nonlinear. We suggest interpolating steady-state concentration measurements by conditioning an ensemble of the underlying log-conductivity field on the available hydrological data in a conditional Monte Carlo approach. Flow and transport simulations for each conditional conductivity field must meet the measurements within their given uncertainty. The ensemble of transport simulations based on the conditional log-conductivity fields yields conditional statistical distributions of concentration at points between observation points. This method implicitly meets physical bounds of concentration values and non-Gaussianity of their statistical distributions and obeys the nonlinearity of the underlying processes. We validate our method by artificial test cases and compare the results to kriging estimates assuming different conditional statistical distributions of concentration. Assuming a beta distribution in kriging leads to estimates of concentration with zero probability of concentrations below zero or above the maximal possible value; however, the concentrations are not forced to meet the advection-dispersion equation.

Introduction

In subsurface hydrology, the number of direct measurements of quantities of interest typically are limited because they require drilling of expensive observation wells. Interpolation techniques are needed to estimate hydraulic heads, concentration, or other hydrological quantities at locations where no measurement is available. Due to subsurface heterogeneity and typical sparsity of data, deterministic interpolation is not practical and would not lead to satisfying results. For risk analysis and in other management applications, the uncertainty of the estimated quantity may be as important as the estimated value itself. In previous studies the importance of obtaining reliable uncertainty bounds has often not been sufficiently considered (e.g., Jones et al. 2003, Reed et al. 2004).

Geostatistical interpolation of point measured data has been used extensively in groundwater hydrology. Typically, some variant of kriging is used. Kriging, also known as best linear unbiased estimator (BLUE), yields the conditional mean and covariance of the interpolated quantity at any location in the domain, given measurements at observation locations, the prior mean, and covariance functions (Kitanidis 1997). Stationarity
of the prior statistics simplifies kriging, but is no prerequisite. When the estimated variable is a multi-Gaussian random space variable, the kriging estimate maximizes the posterior likelihood according to Bayes’ theorem (see Kitanidis 1986), but kriging does not require a Gaussian prior distribution. If the prior distribution is not assumed Gaussian, the posterior distribution would not be Gaussian either. The critical prerequisite of kriging (and cokriging) is that the relationship between the measured and interpolated variable is linear. Strong nonlinearity leads to erroneous best estimates and may yield bounded statistical distributions.

The present study deals with steady-state concentration data, for which interpolation may be particularly challenging. Quite obviously, concentrations cannot be Gaussian distributed because negative concentrations are impossible. Kitanidis and Shen (1996) power-transformed concentration values, assumed that the transformed variables were Gaussian distributed, interpolated them by kriging, and transformed the best estimates and their confidence intervals back to the concentration space. While the power-transformation guarantees non-negativity, it does not meet other constraints, such as a maximum concentration. Various recent studies indicate that unconditional statistical distributions of concentration may be approximated by a scaled beta distribution (Fiorotto and Caroni 2002, 2003; Caroni and Fiorotto 2005; Bellin and Tonina 2007; Schwede et al. 2008). The coefficients of the beta distribution, however, vary considerably in space (and in case of transient transport also in time), reflecting the strong nonstationarity of concentration. This hampers the application of simple transformation rules in the interpolation of concentration measurements.

The prior statistical distribution of steady-state concentrations may be obtained by unconditional Monte Carlo simulations. Here, multiple realizations of the hydraulic log-conductivity field are generated, and each realization is used for flow and transport simulations applying the same boundary conditions. The very fact that the resulting statistical distributions of concentration are beta-like even when multi-Gaussian statistics of the log-conductivity field has been assumed (see studies cited above), indicates a strong nonlinear dependence of steady-state concentration on the primary uncertain material property, namely hydraulic log-conductivity. We conjecture that the statistical interdependence of concentration at two locations is also nonlinear, which jeopardizes the applicability of kriging as an interpolation scheme for steady-state concentrations.

The distribution of concentration in an aquifer is the result of advective-dispersive transport and depends on the underlying flow field. We strongly believe that estimated concentration distributions should be meaningful in a hydraulic context. Interpolation of concentration measurements using only the concentration data does not enforce such behavior, which may result in concentration distributions that cannot be interpreted as the result of advective-dispersive transport.

In this article, we therefore present a rigorous statistical approach for the interpolation of steady-state concentration measurements by conditioning the underlying hydraulic log-conductivity field on the available hydrological data, using an inverse geostatistical approach. We generate multiple realizations of the hydraulic log-conductivity field and condition them on the measurements by the quasi-linear geostatistical approach of Kitanidis (1995). In the course of conditioning the log-conductivity fields on concentration measurements, we obtain one realization of the spatial concentration distribution for each conditional hydraulic log-conductivity field. Generating a large number of conditional fields of conductivity and corresponding concentration leads to conditional statistical distributions of concentration throughout the domain. From these distributions, we can evaluate statistical metrics such as the mean and variance, or particular percentiles of concentration at interpolation locations.

This approach has the advantage that we estimate a truly independent material property of the domain, namely the hydraulic conductivity field. Each realization of concentration meets the advection-dispersion equation, so that physically impossible results are excluded by construction. Effects of nonlinearities are fully accounted for. Finally, we can incorporate other hydrological measurements, provided they depend on hydraulic conductivity.

We compare the concentration distributions obtained by this approach to kriging estimates assuming either a Gaussian or a beta distribution of the interpolated concentration. The theoretical background of describing groundwater flow and transport, interpolation of steady-state concentrations by kriging, and interpolation by inverse modeling are described in the section Theory. The section Application to Hypothetical Test Problems includes applications to virtual two-dimensional (2D) and three-dimensional (3D) test cases, in which the results of kriging and the conditional Monte Carlo approach are compared and to the true values. Finally, we discuss our approach and draw some conclusions.

Theory

Governing Equations

We consider steady-state groundwater flow without internal sources or sinks given by:

\[-\nabla \cdot (K \nabla h) = 0 \text{ on } \Omega \quad (1)\]

in which \(K\) is the hydraulic conductivity, \(h\) is the hydraulic head, and \(\Omega\) is the domain. We set Dirichlet boundary conditions at the in- and outflow of the domain, \(\Gamma_{\text{in}}\) and \(\Gamma_{\text{out}}\), and no-flow Neumann conditions along the other boundaries of the domain, \(\Gamma_{\text{no}}\):

\[
\begin{align*}
    h &= h_{\text{in}} & \text{on } \Gamma_{\text{in}} \\
    h &= h_{\text{out}} & \text{on } \Gamma_{\text{out}} \\
    \mathbf{n} \cdot (K \nabla h) &= 0 & \text{on } \Gamma_{\text{no}}
\end{align*}
\quad (2)
\]
for given functions \( h_{in} \) and \( h_{out} \) along the respective boundaries. \( \Gamma = \Gamma_{in} \cup \Gamma_{out} \cup \Gamma_{no} \) is the boundary of the domain \( \Omega \) and normal vector \( \mathbf{n} \). Extension to fields with internal sources/sinks is straightforward, but will not be considered in the test cases.

The seepage velocity, \( v \), in a porous medium is:

\[
v = \frac{q}{\theta} = -\frac{K}{\theta} \nabla h \tag{3}
\]

with the porosity \( \theta \) and the specific discharge \( q \) given by Darcy’s law.

Steady-state transport of a conservative compound \( c \) is described by the advection-dispersion equation (with dispersion tensor \( D \) according to Scheidegger [1961]) in the following form:

\[
\nabla \cdot (v c - D \nabla c) = 0 \text{ on } \Omega \tag{4}
\]

subject to the boundary conditions:

\[
c = c_{in}(x) \text{ on } \Gamma_{in} \tag{5}
\]

\[
\mathbf{n} \cdot (D \nabla c) = 0 \text{ on } \Gamma_{out} \cup \Gamma_{no}
\]

in which the concentration distribution, \( c_{in}(x) \), on the inflow boundary, \( \Gamma_{in} \), is known.

**Prior Statistics of Concentration**

For the given geostatistical properties of the hydraulic log-conductivity field and given boundary conditions in flow and transport, statistical properties of concentration at points within the domain can be evaluated prior to any measurement. For the interpolation by kriging (cf. section Interpolation by Kriging), we need the prior mean, \( \overline{c}(x) \) and covariance function, \( R_{cc}(x, y) \):

\[
\overline{c}(x) = E[c(x)] \tag{6}
\]

\[
R_{cc}(x, y) = E[(c(x) - \overline{c}(x))(c(y) - \overline{c}(y))] \tag{7}
\]

in which \( E[\cdot] \) denotes the expected value, and \( x \) and \( y \) are the vectors of coordinates at two different locations.

In standard kriging applications, the geostatistical properties of the spatial variable to be interpolated were estimated from the measurements. Reliable estimates require many measurements, particularly for a nonstationary field. In the present study, we follow a different approach, in which we obtain the prior statistical properties of concentration from unconditional Monte Carlo simulations. We define a discretized domain, for which we assume the hydraulic log-conductivity field to be Gaussian distributed with known mean and covariance function. We generate unconditional realizations of the log-conductivity field using the spectral approach of Dietrich and Newsam (1993). Defining boundary conditions for flow and transport (cf. section Governing Equations) allows calculation of an unconditional realization of concentration by solving the forward model for each realization of the hydraulic log-conductivity field.

The expected value in Equations 6 and 7 is approximated by the arithmetic mean over all realizations. We obtain stable estimates of the nonstationary prior two-point statistics of steady-state concentration by considering a large number of unconditional realizations (in our case 15,000). The prior mean \( \overline{c}(x) \) and covariance function \( R_{cc}(x, y) \) computed by Monte Carlo simulations are considerably more accurate than first-order estimates based on perturbation methods (Pannone and Kitanidis 2001; Schwede et al. 2008).

**Interpolation by Kriging**

In the following, we denote measured concentration values with subscript \( m \) and concentrations at interpolation points with subscript \( 0 \). Also, bold variables are vectors, that is, \( c_m \) is the vector of all concentration measurements, and \( \overline{c}_m \) is the vector of corresponding prior expected values, obtained by unconditional Monte Carlo simulations as described in the section Prior Statistics of Concentration. We assume that the concentration measurements are unbiased but prone to an uncorrelated measurement error expressed by the variance \( \sigma_i^2 \) of measurement \( i \). This leads to a diagonal covariance matrix \( \mathbf{V} \) of measurement error made of the individual variances \( \sigma_i^2 \).

**Evaluation of the Best Estimate and the Estimation Variance**

The interpolated concentration, \( \hat{c}_0 \), at location, \( x_0 \), obtained by kriging is the best linear estimate \( \hat{c}_0 \):

\[
\hat{c}_0 = \overline{c}_0 + \xi \mathbf{R}_m \mathbf{m}_0 \tag{8}
\]

in which \( \mathbf{R}_m \) is the prior covariance vector relating the variations of concentration about the expected value at all measurement locations to those at the point of interpolation, \( \overline{c}_0 \) is the prior mean concentration at location \( x_0 \), and \( \xi \) can be obtained by solving the linear kriging system of equations:

\[
[\mathbf{R}_{mm} + \mathbf{V}] \xi = \mathbf{m}_m - \overline{c}_m \tag{9}
\]

in which \( \overline{c}_m \) is the vector of prior mean concentration values at the measurement locations and \( \mathbf{V} \) is the covariance matrix quantifying measurement and model errors.

The variance \( \hat{\sigma}_0^2 \) of the estimated concentration is:

\[
\hat{\sigma}_0^2 = R_{00} - R_{m0}^T [\mathbf{R}_{mm} + \mathbf{V}]^{-1} \mathbf{R}_{m0} \tag{10}
\]

in which \( R_{00} \) is the prior variance at location \( x_0 \).

It may be worth noting that this version of kriging does not require stationarity. That is, \( \overline{c}(x) \) may vary in space, and \( R_{cc}(x, y) \) may depend on the exact locations \( x \) and \( y \) rather than their distance \( x - y \).

**Assumed Posterior Concentration Distributions**

As stated in the introduction, the kriging estimate holds for any shape of the prior statistical distribution and makes no direct statements about the shape of the posterior distribution of the interpolated variable. With the given
best estimate, $\hat{c}_0$, and estimation variance, $\hat{\sigma}_0^2$, any two-parametric distribution could be defined. In the application discussed in the section Application to Hypothetical Test Problems, we test two distributions: Gaussian, which may be the most common, even though naïve guess, and the beta distribution, which has been shown appropriate for unconditional statistical distributions of concentration (Fiorotto and Caroni 2002, 2003; Caroni and Fiorotto 2005; Bellin and Tonina 2007; Schwede et al. 2008). These distributions have the following probability density functions $p_G(c)$ and $p_\beta(c)$, respectively:

$$p_G(c) = \frac{1}{\hat{\sigma}_0 \sqrt{2\pi}} \exp\left(-\frac{(c - \hat{c}_0)^2}{2 \hat{\sigma}_0^2}\right)$$

$$p_\beta(c) = \frac{1}{B(\alpha, \beta)} c^{\alpha-1} (1-c)^{\beta-1}$$

in which $B(\alpha, \beta) = \int_0^1 u^{\alpha-1} (1-u)^{\beta-1} du$ is the beta function with parameters $\alpha$ and $\beta$, which can be evaluated from the best estimate $\hat{c}_0$ and estimation variance $\hat{\sigma}_0^2$ by:

$$\alpha = \hat{c}_0 \left(\frac{\hat{c}_0(1-\hat{c}_0)}{\hat{\sigma}_0^2} - 1\right)$$

$$\beta = (1-\hat{c}_0) \left(\frac{\hat{c}_0(1-\hat{c}_0)}{\hat{\sigma}_0^2} - 1\right)$$

It should be noted that the beta distribution given in Equation 12 is defined on the interval $[0, 1]$. That is, the concentrations and their statistical metrics must be scaled by the maximal possible concentration value, $c_{\text{max}}$, which is the inflow concentration $c_{\text{in}}$ in the applications discussed in the section Application to Hypothetical Test Problems.

As characteristic value of the estimated concentration distribution at the interpolation locations, we use the median rather than the mean of the distribution. The median is much less affected by outliers than the mean value and is much closer to the most likely value when the distribution is nonsymmetric. This is of particular relevance for the beta distribution in case of a high variance, $\hat{\sigma}_0^2$, where the probability mass is clustered at values close to zero and close to the maximum concentration.

**Interpolation by Conditioning of Conductivity Fields**

**General Approach**

Rather than directly interpolating concentration measurements, we suggest putting the interpolation into an inverse modeling context. The general idea is to generate a large number, $N$, of conditional realizations of the hydraulic log-conductivity field. Each realization exhibits the full range of variability on all scales, and flow and transport simulations based on each realization lead to simulated concentration values at the measurement locations that agree with the measured values within their given uncertainty. The scheme requires a stable method of inferring hydraulic log-conductivity fields from measurements of steady-state concentrations, which was recently developed by Schwede and Cirpka (2009). Other types of measurements, such as of hydraulic heads or direct measurements of hydraulic conductivity, can easily be integrated into the conditioning procedure. That is, rather than directly interpolating concentration measurements, we seek hydraulic conductivity fields that can explain the observed values.

In the conditioning process, we simulate concentration values throughout the domain, so that we obtain an ensemble of possible concentration values at locations other than the measurement locations from the set of conditional log-conductivity fields. With a sufficiently large number of realizations, $N$, we obtain an empirical statistical distribution of steady-state concentration at each grid point of our discretized domain conditioned on the concentration values at the observation points and other measurements. This method may be addressed as conditional Monte Carlo (CMC) approach. Figure 1 schematically illustrates this approach.

This method is computationally quite demanding, but it has several advantages over direct interpolation methods: (1) we infer a truly independent material property of the domain, namely hydraulic conductivity; (2) the steady-state concentration depends in a strong nonlinear way on the hydraulic log-conductivity field and this nonlinearity is fully accounted for; (3) each realization of the concentration field meets the advection-dispersion equation. Physically impossible results are excluded by construction; (4) we can account for other types of measurements that also depend on the hydraulic conductivity field. Direct interpolation methods would require evaluating cross-covariance functions between those measurements and concentration and could hardly address nonlinearity in these interdependencies; and (5) our estimates of the concentration field are conditioned on particular boundary conditions. This is an advantage when the boundary conditions are known, but admittedly may pose a problem when they are uncertain.

The nonlinear dependence of steady-state concentration on hydraulic log-conductivity requires an

![CMC schematic](image-url)
iterative inverse modeling approach, such as the method of Kitanidis (1995). The nonlinearity also prohibits linearized uncertainty propagation as applied in cokriging. That is, even if a single smooth cokriging estimate of the log-conductivity field can be evaluated that meets all measurements within their respective uncertainties, the associated conditional covariance of log-conductivity would be biased, and the uncertainty of dependent concentration values could not well be approximated by conditional first-order second moment (FOSM) methods (e.g., Dettinger and Wilson 1981; Kunstmann et al. 2002; Cirpka and Nowak 2004). To address nonlinearity in uncertainty propagation, we generate an ensemble of equally likely log-conductivity fields exhibiting random small-scale fluctuations according to the underlying geostatistical model and meeting all measurements within their uncertainty. This is a standard approach in geostatistical inversion (e.g., RamaRao et al. 1995; Gómez-Hernández et al. 1997).

**Conditioning Method**

For generation of conditional hydraulic conductivity fields, we apply the method of smallest possible modification (e.g., Gutjahr et al. 1994). Our approach is based on the conditional realization variant of the quasi-linear geostatistical approach of inversion by Kitanidis (1995). The specific implementational issues of inverting steady-state concentration values have been discussed by Schwede and Cirpka (2009), although not in a conditional-simulation context. In this section, we give a brief overview of the conditioning method.

A conditional realization of the log-conductivity field, here denoted by \( Y_{\text{cond}}(x) \), has three contributors:

\[
Y_{\text{cond}}(x) = Y'_u(x) + Y'_c(x) + \sum X_i(x) \beta_i \tag{15}
\]

where \( Y'_u(x) \) is an unconditional field with zero mean and covariance function \( R_{Y'y}(h) \); \( Y'_c(x) \) is a smooth correction function with prior mean of zero and prior covariance function \( R_{Y'y}(h) \), \( X_i(x) \) is the ith deterministic trend function and \( \beta_i \) the corresponding trend coefficient. Discretizing the field by \( n \) elements, Equation 15 becomes:

\[
Y_{\text{cond}} = Y'_u + Y'_c + X\beta \tag{16}
\]

where \( Y_{\text{cond}}, Y'_u, \) and \( Y'_c \) are \( n \times 1 \) vectors, \( X \) is the \( n \times n_B \) matrix of discretized base functions, and \( \beta \) is the \( n_B \times 1 \) vector of trend coefficients. If the covariance function, \( R_{Y'y}(x,y) \) is stationary and the domain is discretized by a regular grid, \( Y'_u \) may be generated by efficient spectral methods (Dietrich and Newsam 1993). The prior mean of \( \beta \) is \( \beta^0 \), and its prior covariance matrix is \( R_{\beta\beta} \). \( Y'_c \) and \( \beta \) are obtained by minimizing the following objective function (e.g., Kitanidis 1995; Schwede and Cirpka 2009):

\[
L(Y'_c, \beta | c_{i,m}) = (c_{i,m}(Y_{\text{cond}}) - c_{i,m})^T V^{-1} (c_{i,m}(Y_{\text{cond}}) - c_{i,m}) \\
+ Y'_c^T R_{Y'y}^{-1} Y'_c + (\beta - \beta^0)^T R_{\beta\beta}^{-1} (\beta - \beta^0) \tag{17}
\]

where \( c_{i,m}(Y_{\text{cond}}) \) is the model outcome of the measured concentrations for the log-conductivity field \( Y_{\text{cond}} \) and \( c_{i,m} \) is the measurement vector used in the ith conditional Monte Carlo simulation:

\[
c_{i,m} = c_m + \varepsilon_i \tag{18}
\]

in which \( \varepsilon_i \) is a realization of the measurement error drawn from a Gaussian distribution with zero mean and covariance matrix \( V = \text{diag}(\sigma_1^2, \ldots, \sigma_m^2) (\varepsilon_i \sim \mathcal{N}(0,V)) \forall i = 1, \ldots, N \). This randomization is needed to account for the uncertainty of the measurements, expressed by the variances \( \sigma_1^2, \ldots, \sigma_m^2 \) related to the measuring process.

To minimize \( L(Y'_c, \beta | c_{i,m}) \), we linearize Equation 17 about the current estimate, apply matrix identities leading to cokriging-like equations (Kitanidis 1995), use the continuous adjoint state approach for the calculation of sensitivities (Sun and Yeh 1990), accelerate matrix-matrix multiplications by spectral methods (Nowak et al. 2003), stabilize the scheme by a modified Levenberg-Marquardt approach (Nowak and Cirpka 2004), and use adaptive enlargement of the sampling volume in the calculation of sensitivities (Schwede and Cirpka 2009).

In the examples given in the section Application to Hypothetical Test Problems, we use measurements of steady-state concentration, hydraulic head, and borehole-dilution test data for conditioning. In the Appendix, we explain how the characteristic residence time of a tracer within the borehole relates to the temporal moment generation equations. The set of measurements may be extended to pumping test and flowmeter test data (Li et al. 2007, 2008) or any other data depending on hydraulic conductivity.

**Analysis of Conditional Realizations for Interpolation of Concentration**

For each conditional realization of the hydraulic log-conductivity field, we obtain as a byproduct the concentration field by solving Equations 1 and 4 subject to Equations 2 and 5. The realizations of the concentration field exhibit the full range of spatial variability and meet the concentration measurements within the given measurement error at the measurement locations.

From the ensemble of all conditional realizations of the concentration field, we compute the estimated mean value and its uncertainty. As estimator we use the median (0.5 quantile) of the distribution, because it is less sensitive to outliers than the arithmetic mean, which is typically used as estimator for the statistical mean value. The conditional statistical distribution of concentration need not be symmetric. Therefore, we calculate the lower and upper uncertainty bounds at the 0.16 and 0.84 quantiles, respectively, which would correspond to the mean ±1 standard deviation if the distribution was Gaussian.

**Application to Hypothetical Test Problems**

In this section, we compare our approach of interpolation by inversion to kriging in hypothetical 2D and 3D confined aquifers with generated random log-hydraulic conductivity values.
conductivity fields. The heterogeneity of the aquifer is modeled by a Gaussian second-order stationary field with an exponential covariance model. We use the spectral methods of Dietrich and Newsam (1993, 1997) for the fast generation of these fields. The exact statistical and hydrogeological properties of the fields are listed in Table 1. All parameters listed in Table 1 are assumed to be known. In realistic field applications, the geostatistical parameters may be uncertain. Techniques of estimating geostatistical parameters of log-hydraulic conductivity from indirect measurements exist (e.g., Kitanidis and Vomvoris 1983; Li et al. 2007, 2008; Nowak 2009) and may be included in the inversion procedure. To obtain reliable estimates, however, more data points than those considered in the current example would be necessary (see some of the references given above). The purpose of the present contribution is to prove the principle of concentration interpolation by inverse modeling. For this purpose, we restrict ourselves to somewhat idealized conditions and reduced datasets. In field applications, all available data would be included and also the uncertainty about uncertainty had to be addressed.

Two-Dimensional Test Case

Figure 2 shows the true log-hydraulic conductivity field, the associated concentration distribution, and a concentration profile perpendicular to the direction of mean flow. The solute is introduced over a known section of the inflow boundary at the left-hand side. As indicated by Figure 2b, there are nine observation points for steady-state concentration and hydraulic heads are measured, and borehole-dilution tests are performed. The results of the interpolation by inversion (cf. section Interpolation by Conditioning of Conductivity Fields) are based on 2750 conditional realizations. Figure 3a shows the results for the interpolation by inverse modeling for the transects at $x = 1.5 \mathrm{~m}$ and $x = 2.5 \mathrm{~m}$. Figures 3b and 3c show the interpolated concentration profile based on kriging, assuming a beta (Figure 3b) and a Gaussian distribution (Figure 3c), respectively.

Obviously, the interpolation by inverse modeling (Figure 3a) fits the real spatial distribution of concentration quite well. The almost rectangular shape of the concentration profile is retained by the median of the estimate, and the position and width of the plume are captured at least qualitatively. We believe that these characteristics are enforced by making each realization of the concentration field fit the advection-dispersion equation with known boundary conditions. Even though the prior statistics used in kriging result from unconditional concentration simulations with the same boundary conditions, the direct dependence on transport is neglected in the interpolation by kriging. In the given problem, normalized concentration values smaller than zero or larger than unity are physically impossible. These bounds are naturally met with the conditional Monte Carlo approach (Figure 3a) and enforced by the kriging estimate assuming a beta distribution (Figure 3b). The latter estimate leads to a spatial concentration profile that is slightly too smooth and exhibits a

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definition of the domain</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Domain size</td>
<td>$\Omega$</td>
<td>2D: $5 \times 5 \mathrm{~m}$</td>
</tr>
<tr>
<td>Grid spacing</td>
<td></td>
<td>3D: $5 \times 5 \times 2 \mathrm{~m}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2D: $0.05 \times 0.05 \mathrm{~m}$</td>
</tr>
<tr>
<td>Log-conductivity $Y$</td>
<td></td>
<td>3D: $0.1 \times 0.1 \times 0.04 \mathrm{~m}$</td>
</tr>
<tr>
<td>Log-conductivity field variance</td>
<td>$\sigma_Y^2$</td>
<td>1 $\ln^2 \left( \mathrm{m/s} \right)$</td>
</tr>
<tr>
<td>Correlation length</td>
<td>$\lambda$</td>
<td>2D: $1.25 \times 1.25 \mathrm{~m}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3D: $1.25 \times 1.25 \times 0.5 \mathrm{~m}$</td>
</tr>
<tr>
<td>Covariance model</td>
<td>Exponential</td>
<td></td>
</tr>
<tr>
<td>Prior geometric mean of conductivity</td>
<td>$\exp(\beta^{*})$</td>
<td>$10^{-5} \mathrm{~m/s}$</td>
</tr>
<tr>
<td>Uncertainty of prior mean</td>
<td>$R_{\beta\beta}$</td>
<td>0.48 $\ln^2 \left( \mathrm{m/s} \right)$</td>
</tr>
<tr>
<td>Transport parameters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean velocity</td>
<td>$\mathbf{\tau}$</td>
<td>$1.16 \times 10^{-5} \mathrm{~m/s} \left( = 1 \mathrm{~m/d} \right)$</td>
</tr>
<tr>
<td>Porosity</td>
<td>$\theta$</td>
<td>0.4</td>
</tr>
<tr>
<td>Molecular diffusion coefficient</td>
<td>$D_m$</td>
<td>$10^{-9} \mathrm{~m}^2/\mathrm{s}$</td>
</tr>
<tr>
<td>Longitudinal dispersivity</td>
<td>$a_l$</td>
<td>$10^{-2} \mathrm{~m}$</td>
</tr>
<tr>
<td>Transverse dispersivity</td>
<td>$a_t$</td>
<td>$2 \times 10^{-3} \mathrm{~m}$</td>
</tr>
<tr>
<td>Measurement error</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hydraulic heads (absolute)</td>
<td></td>
<td>$3 \times 10^{-3} \mathrm{~m}$</td>
</tr>
<tr>
<td>Normalized concentration (relative)</td>
<td></td>
<td>1%</td>
</tr>
<tr>
<td>Borehole-dilution tests (relative)</td>
<td></td>
<td>1%</td>
</tr>
</tbody>
</table>
too small plume width. We attribute these shortcomings to the decoupling of transport and concentration estimation. The naïve interpretation of the kriging results assuming a Gaussian posterior statistical distribution of concentration (Figure 3c) leads to a finite probability of concentration outside the physically possible range, and at some locations even the estimated mean value of concentration is negative (cf. Figure 3c at, for example, \( y \approx 3.7 \) [m]). This confirms that the statistical distribution of steady-state concentration must not be assumed Gaussian.

The variance of the estimated concentration in the conditional Monte Carlo approach is smaller than the estimation variance of concentration in kriging. In Figure 3 the gray shaded areas are bounded by the 0.16 and 0.84 quantiles. These areas are much smaller in Figure 3a than in Figures 3b and 3c for both transects. This difference might be important in risk analysis, because a smaller uncertainty requires a smaller margin of safety in a technical design.

Figures 3b and 3c also include the prior mean, which is needed in the kriging approach. The prior mean concentration is the unconditional ensemble mean concentration, which has been intensively analyzed in macrodispersion studies (e.g., Gelhar and Axness 1983; Dagan 1984; Neuman et al. 1987). The profile is approximately Gaussian. The kriging estimate tries to keep as close to the prior mean as possible while meeting the measurements (see Equation 8). This explains why the kriging estimate assuming Gaussianity is systematically too low within the plume and too high at the right fringe. At the left fringe, the true concentration profile is outside the gray shaded uncertainty range in both kriging estimates (cf. Figures 3b and 3c).

Figure 4 shows the unconditional and conditional probability density functions (pdfs) obtained by the different interpolation methods at a particular location \( (x = 2.5 \, [m], \, y = 2 \, [m]) \). The prior pdf shown as dotted curve in Figure 4 exhibits the typical bimodal shape of the beta distribution. The probability mass is clustered at low and high concentration values. This behavior of unconditional concentration pdfs has been reported before (e.g., Caroni and Fiorotto 2005; Schwede et al. 2008). The true concentration value at this location is close to one (\( \approx 0.9995 \) [m]). The pdf according to the conditional Monte Carlo simulation shown as the black solid curve in Figure 4 has high probability at high concentrations, which is in agreement with the true concentration value. The conditional pdf shows very low probability of intermediate concentrations but again an increased probability of low concentration (but much less probability mass at low concentrations than the unconditional pdf). It may be worth noting that the arithmetic mean of the conditional concentration pdf is more strongly affected by the comparably high probability of very low concentration values than the median. This explains why the median may be a better metric for the estimated value than the arithmetic mean. The pdf of the kriging estimate assuming a Gaussian distribution of concentration shows a fairly large probability of concentration values outside the physically possible range. The pdf of the kriging estimate assuming a beta distribution of concentration meets the physical bounds of concentration, but it shows a fairly high probability of intermediate concentration values. We conclude that the conditional Monte Carlo approach estimates the conditional concentration distribution best.
Three-Dimensional Test Case

We have also performed a 3D study, the exact settings and parameters of this study can be found in Table 1. In this study, we used 27 steady-state concentration measurements, which are equidistantly distributed within the observed domain. For the interpolation approach by inverse modeling, we included three measurements of hydraulic heads and three dilution measurements located in the center of the domain at three different z locations. The results of the 3D study are quite similar to the 2D one, but the computational effort is much higher. We generated 12,500 unconditional Monte Carlo realizations
Conclusions

In this study, we have interpolated steady-state concentration data by inferring the underlying hydraulic conductivity field using conditional geostatistical simulations. With the ensemble of hydraulic log-conductivity fields we also obtain an ensemble of concentration fields, each of them meeting the measurements within their uncertainty bounds. Our approach of interpolating concentration by generating conditional realizations of hydraulic conductivity fields leads to reliable interpolation estimates including their uncertainty. The method overcomes two basic problems known for the geostatistical interpolation of steady-state concentration data. First, the method is not limited to a particular statistical distribution of the interpolated quantity, although it requires Gaussianity of the inverted quantity. Second, the method implicitly meets the physical bounds of the interpolated quantity, because the physical transport process is explicitly simulated. The nonlinear relationship between log-conductivity and concentration is accounted for by construction. Because of this nonlinearity, we evaluate the uncertainty of the estimated log-conductivity and concentration fields by conditional simulations rather than by cokriging plus linearized uncertainty propagation.

In comparison to direct linear geostatistical interpolation of the concentration measurements by kriging, the estimate by inversion is closer to the true value. We attribute the superiority of the conditional Monte Carlo method to the fact that each conditional realization meets the advection-dispersion equation with given boundary conditions. Kriging gives the best linear unbiased estimate, which is a conditional mean value, and the variance of estimation at all points of interpolation. Assuming a Gaussian conditional distribution of concentration yields a finite probability of concentration values outside the physically possible range. This confirms that concentrations should not be assumed to be Gaussian variables. Combining the kriging results with a beta distribution enforces the conditional concentration distribution to meet physical bounds, but systematic bias remains, presumably because the underlying relationships are nonlinear.

The computational expense of our approach is that the realizations of hydraulic log-conductivity and associated concentration fields can consistently be conditioned on all measurements that depend on hydraulic conductivity. To do the same in kriging would require estimating the cross-covariances between concentration and all types of measurements, and would significantly be complicated by nonlinearity in these relationships.

A particular advantage of our approach is that the realizations of hydraulic log-conductivity and concentration fields can consistently be conditioned on all measurements that depend on hydraulic conductivity. Monte Carlo simulations, however, can easily be parallelized. We cannot exclude that an adequate transformation of concentration would improve the kriging estimates. Considering the unconditional concentration statistics, a beta-to-Gaussian transformation would be suggestive. Unfortunately, the associated coefficients vary in space so that the presumably best mapping procedure is nontrivial.

Acknowledgments

This study has been funded by the Swiss National Science Foundation (SNF) under the grant 200020_118017. We thank Scott James and two anonymous reviewers for their constructive suggestions.

References


Appendix

Measurements of Borehole–Dilution Tests

It is nearly impossible to directly measure groundwater velocities. Velocity-like information in the subsurface may be obtained from a borehole-dilution test (Drost et al. 1968; Grisak et al. 1977; Palmer 1993; Novakowski et al. 2006; Pitak et al. 2007). Here, a conservative tracer is released into a borehole and the decrease of concentration in the borehole is monitored.
Adding the tracer mass, \( m_{\text{tracer}} \), at time zero into the borehole with volume, \( V \), at location, \( x_b \), and mixing the solution in the borehole results in the following concentration time curve:

\[
c(x_b, t) = \frac{m_{\text{tracer}}}{V} \exp \left( -\frac{Q}{V} t \right) \tag{A1}
\]

in which \( Q \) is the discharge passing through the borehole.

The \( k \)th temporal moment of transient concentration \( c(x, t) \) is defined as:

\[
m_k = \int_0^\infty t^k c(x, t) \, dt \tag{A2}
\]

Obviously, the characteristic time \( m_1/m_0 \) of concentration within the borehole is

\[
m_1(x_b) \over m_0(x_b) = \frac{V}{Q} \tag{A3}
\]

We can also calculate the moments in the complete domain by solving the moment generating equation (Harvey and Gorelick 1995; Cirpka and Kitanidis 2000). The advection-dispersion with tracer mass introduced into the borehole at time zero is

\[
\frac{\partial c}{\partial t} + \mathbf{v} \cdot \nabla c - \nabla \cdot (D \nabla c) = m_{\text{tracer}} \Pi(x, x_b, y_b, r_b) \delta(t) \tag{A4}
\]

subject to:

\[
c = 0(x) \text{ on } \Gamma_{\text{in}} \nabla \cdot (D \nabla c) = 0 \text{ on } \Gamma_{\text{out}} \cup \Gamma_{\text{no}} \tag{A5}
\]

in which \( x_b \) and \( y_b \) are the horizontal coordinates of the borehole, \( \delta(\cdot) \) is the Dirac delta function, and \( \Pi(x, x_b, y_b, r_b) \) is a function that is constant inside of the borehole with radius \( r_b \), zero outside the borehole, and that integrates to unity:

\[
\Pi(x, x_b, y_b, r_b) = \begin{cases} \frac{1}{\Delta z \pi r_b^2} & \text{if } (x - x_b)^2 + (y - y_b)^2 \leq r_b^2 \\ 0 & \text{otherwise} \end{cases} \tag{A6}
\]

where \( \Delta z \) is the thickness of the aquifer, and \( x, y \) are the horizontal coordinates.

Multiplication with \( t^k \) and integration over time yields (Harvey and Gorelick 1995; Cirpka and Kitanidis 2000):

\[
\nabla \cdot (\mathbf{v} m_k - D \nabla m_k) = k m_{k-1} + m_{\text{tracer}} \Pi(x, x_b, y_b, r_b) \delta_k \text{ on } \Omega \tag{A7}
\]

in which \( \delta_k \) is the Kronecker delta. Equation A7 is subject to the boundary conditions:

\[
m_k = 0 \text{ on } \Gamma_{\text{in}} \nabla \cdot (D \nabla m_k) = 0 \text{ on } \Gamma_{\text{out}} \cup \Gamma_{\text{no}} \tag{A8}
\]

Because we are interested in the characteristic time \( m_1/m_0 \) of concentration within the borehole rather than the actual moments, we can replace the internal flux-related boundary condition within the well volume by the internal Dirichlet boundary condition:

\[
m_0 = \begin{cases} 1 & \text{if } (x - x_b)^2 + (y - y_b)^2 \leq r_b^2 \\ 0 & \text{otherwise} \end{cases} \tag{A9}
\]

For the first moment we do not set an internal boundary condition. Then, the computed value \( m_1(x_b) \) corresponds to the characteristic time of the borehole-dilution test. The inversion of temporal moments of concentration has been described by Cirpka and Kitanidis (2000).