A method for estimating distributions of mass transfer rate coefficients with application to purging and batch experiments

K.J. Hollenbeck a,*, C.F. Harvey b, R. Haggerty c, C.J. Werth d

a Department of Hydrodynamics and Water Resources, Technical University of Denmark, Building 115, 2800 Lyngby, Denmark
b Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
c Department of Geosciences, 104 Wilkinson Hall, Oregon State University, Corvallis, OR 97331, USA
d Department of Civil Engineering, University of Illinois at Urbana-Champaign, Urbana-Champaign, USA

Received 12 February 1998; received in revised form 20 November 1998; accepted 20 November 1998

Abstract

Mass transfer between aquifer material and groundwater is often modeled as first-order rate-limited sorption or diffusive exchange between mobile zones and immobile zones with idealized geometries. Recent improvements in experimental techniques and advances in our understanding of pore-scale heterogeneity demonstrate that two or even a few rate coefficients are insufficient in many cases. Here, we investigate a piece-wise linear model for a continuous distribution of rate coefficients, that has several advantages over previously used 'statistical' distribution models (with functional form from gamma or lognormal PDF's): (1) distributions of arbitrary, even bimodal, shapes can be represented; (2) linear estimation methods can be applied to determine the distribution from experimental data; (3) the uncertainty in the distribution can be determined for each of its sections; and (4) the relationship between the time scales of available data and those of estimatable mass transfer processes can be investigated. A statistical model refinement algorithm is presented that reduces the number of parameters sections of the piece-wise linear model to the admissible minimum. We show that purging experiments allow estimation of a wider zone of the rate distribution than do batch experiments, and hence will provide predictions that are accurate over a wider range of time scales. Finally, in an application to TCE gas-purging desorption data, the piece-wise linear rate-distribution model has a higher
probability of being adequate than those using a gamma or lognormal distribution or a single rate coefficient. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Mass transfer; Diffusion; Sorption; Estimation

1. Introduction

Accurate models of mobile–immobile domain solute mass transfer are important for predicting solute migration, spreading, dilution and reaction. Sorption and diffusive mass transfer of contaminants commonly limit the rate of aquifer or soil remediation. Determining the rate coefficients of mass transfer (or equivalently their time scales) is often difficult because a variety of diffusive and sorptive processes contribute to mass transfer. The properties controlling the rate of mass transfer include: the types of minerals and their spatial distribution; the types and amounts of natural organic matter; the volume, size and geometry of macropores, mesopores, and micropores within grains and aggregates; the geometry, chemistry and mineralogy of coatings on the surfaces; and the chemistry of the water and the solute (Rao et al., 1980; Valocchi, 1989; Cvetkovic and Shapiro, 1990; Wood et al., 1990; Ball and Roberts, 1991a,b; Weber et al., 1991; Barber et al., 1992; Harmon et al., 1992; Farrell and Reinhard, 1994; Harmon and Roberts, 1994; Wilson, 1995; Pignatello and Xing, 1996; Weber and Huang, 1996; Wels et al., 1996; Holmen and Gschwend, 1997). A growing body of research demonstrates that mass transfer in natural porous media occurs over multiple time scales (e.g., Smith et al., 1990; Pavlostathis and Mathavan, 1992; Grathwohl and Reinhard, 1993; Holmewand and Gladdan, 1995; Pignatello and Xing, 1996). In other words, mass transfer models may require a wide range of rate coefficients to describe the data.

Transport models accounting for multiple time scales of mass transfer have been in use for almost 30 years (e.g., Ruthven and Loughlin, 1971; Rao et al., 1982; Rasmuson, 1985; Pedit and Miller, 1994). Among the more recent contributors to this field, Haggerty and Gorelick (1995) and Culver et al. (1997) use series of discrete rate coefficients, while Chen and Wagenet (1995) and Cunningham et al. (1997) use a continuous distribution of rate coefficients. Several recent papers (Culver et al., 1997; Werth et al., 1997; Haggerty and Gorelick, 1998) estimate a distribution of rate coefficients by fitting laboratory and field data. They assume functional forms of the rate distribution that are scaled versions of the gamma or lognormal distribution functions known from statistics. Here, we introduce a mass transfer model with a continuous piece-wise linear distribution of rate coefficients as a tool for examining the estimation problem.

The piece-wise linear model has several advantages over distribution models borrowed from statistics. First, it can be used to model distributions that have arbitrary, even multimodal, shape. Second, analytical solutions of the transport equation for some boundary/initial conditions can be found in the Laplace domain, with stable numerical inversion. Third, multivariate linear statistics can be used to formulate uncertainty in rate distribution estimates. Fourth, the mass transfer response observed during a particular time window can be linked to the corresponding piece of the rate distribution, and thus studied in detail.
This last advantage means that we can use the piece-wise linear model to determine the time scales over which predictions can be made using parameters estimated from a particular experiment. With the piece-wise linear model we can estimate the capacity associated with different sections of a distribution of rate coefficients, and we can also estimate the uncertainty associated with each of these sections. We can then use these estimated uncertainties to determine the time scales over which we can accurately predict solute behavior. In contrast, the parameters of ‘statistical’ distribution models affect the shape of the entire distribution, so the uncertainty in a parameter estimate cannot be translated directly into the uncertainty for a time window of mass transfer processes. Typically, uncertainty increases for time scales much shorter and longer than the experimental time scale. The piece-wise linear model provides a useful tool for answering questions such as: how much does our uncertainty increase for time scales longer than the experimental time scale? How does this uncertainty differ for various experimental designs, and how can we design experiments which minimize the uncertainty of capacities associated with very slow rates?

This paper is organized in three sections. In Section 2, the mass transfer model is derived. It is shown that for any continuous distribution of diffusion coefficients, an equivalent continuous distribution of first-order rates exists. In Section 3, the methodology of rate distribution estimation is developed. Solutions of the transport equation for two common experiment types are presented, followed by statistical theory for finding the best model. In Section 4, the methodology is applied to evaluate experimental designs, using both synthetic and actual data.

2. Model derivation

2.1. Background: discrete rate coefficients

Our derivation starts from the advection–dispersion equation for the common two-region model of chemical transport in groundwater

$$\frac{\partial c_m}{\partial t} + \beta \frac{\partial c_{im}}{\partial t} = D \frac{\partial^2 c_m}{\partial x^2} - \nu \frac{\partial c_m}{\partial x} = AD(c_m)$$

(1)

where $c_m$, $c_{im}$ [M/L$^3$] are the mobile and immobile concentrations within an elementary volume, and $t$ is time. The capacity coefficient $\beta$ is dimensionless. For the purposes of this study, it suffices to formulate Eq. (1) in one dimension and for a homogeneous medium, but the operator $AD(\cdot)$ describing advection ($\nu$ [L/T]: groundwater velocity) and dispersion $D$ [L$^2$/T]: dispersion coefficient) can be extended to the general case. The above formulation is equivalent to a variety of formulations of rate-limited mass transfer (Haggerty and Gorelick, 1995, Fig. 1), and the expression ‘mobile/immobile’ can be understood in a wide sense, covering for example also sorption (Reichle et al., 1998). For example, Eq. (1) is equivalent to the conservation equation for a common formulation of the mobile–immobile domain model when $\beta = (\theta_m + \rho_b K_d)/\theta_m$, with $\theta_m$, $\theta_{im}$ [1] the volumetric fractions of immobile and mobile zones, $\rho_b$ [M/L$^3$] the bulk density, and $K_d$ [L$^2$/M] the distribution coefficient. In practice, $\beta$ can be viewed as a
A lumped parameter to be estimated from data rather than computed from its often unknown constituent parameters.

If exchange between the mobile and immobile phases is rate-limited, the kinetics of the mass transfer process must be formulated. A first-order rate law is often assumed for lack of better knowledge and for simplicity. The discrete ‘multirate’ model of Haggerty and Gorelick (1995) is built on the traditional two-site description of mass transfer, but has \( M \) instead of two sites, each with a capacity coefficient and a rate parameter, \( \alpha [T^{-1}] \). There are thus \( M + 1 \) differential equations, which can be expressed as a linear system.

\[
\frac{\partial c_m}{\partial t} + \sum_{i=1}^{M} \beta_i \frac{\partial c_{im}}{\partial t} = AD(c_m) \tag{2}
\]

\[
\frac{\partial c_{im}}{\partial t} = \alpha_i (c_m - c_{im}) \quad i = 1 \ldots M \tag{3}
\]

For any particular type of experiment, the mobile concentration depends on the mass transfer properties of all \( M \) immobile zones. Accurate prediction of removal of residual contaminant or long-term solute transport hinges on accounting for multiple time scale of mass transfer, which is difficult with a standard single-rate modeling approach.
The first-order mass transfer model is generally a simplified representation of reality. First, the model is an approximation to diffusion, which usually controls the rate of mass transfer. Second, the model does not represent sorption hysteresis or limits on the number of adsorption sites (Pignatello and Xing, 1996). Pedit and Miller (1994) developed multiple-rate models that include non-linear isotherms, but conclude that for practical application, ‘time and effort involved in such detailed work would be prohibitive’. Therefore, even though in principle our mathematical derivation could be extended to non-linear isotherms, we assume linear ones, as did Brusseau et al. (1989) in their multiprocess model. The important advantage of a linear isotherm formulation is that we are able to use Laplace transforms in the mathematical manipulations.

2.2. Continuous distribution of rate coefficients

While an improvement over the simple two-site model, the discrete multirate model still assumes the existence of a countable number of rates rather than a continuous distribution. Culver et al. (1997) show that model predictions converge as the number of sites is increased. However, the solution of the corresponding high-dimensional system of differential equations can become vulnerable to numerical error, both for finite-difference time-stepping schemes (Culver et al., 1997) and a matrix exponential solution (Haggerty and Gorelick, 1995). In their review article, Pignatello and Xing (1996) summarize a large body of research that conceptualize natural materials as having a distribution of sorption sites with a distribution of energies and rate coefficients. To represent this distribution, we extend the mass transfer model of Eqs. (2) and (3) to a truly continuous distribution of rate coefficients:

\[
\frac{\partial c_m}{\partial t} + \int_0^\infty b_\alpha(\alpha) \frac{\partial c_{im}(\alpha)}{\partial t} d\alpha = \Delta D(\alpha)
\]  

(4)

where immobile concentrations are now associated with sites featuring a characteristic rate coefficient. Eq. (4) is formulated without a retardation factor, as equilibrium mass transfer can be described by a rate distribution extending to very ‘fast’ sites (those with large \( \alpha \)).

The function \( b_\alpha(\alpha) \) in Eq. (4) describes how capacity density is distributed over all sites (note that \( b \) has units of [T], whereas \( \beta \) is dimensionless). The cumulative capacity for any interval \([\alpha_1, \alpha_2]\) and hence the total capacity, \( \beta_{\text{tot}} \), can be found by integrating the capacity density

\[
\beta_{[\alpha_1, \alpha_2]} = \int_{\alpha_1}^{\alpha_2} b_\alpha(\alpha) d\alpha \to \beta_{\text{tot}} = \int_0^\infty b_\alpha(\alpha) d\alpha
\]  

(6)

Several papers on multiple-rate mass transfer have borrowed the functional form of common statistical probability density functions (PDF’s, e.g., gamma, lognormal) to describe \( b_\alpha(\alpha) \). To avoid confusion with terminology used in statistics, we point out that \( b_\alpha(\alpha) \) is not a PDF of a random variable \( \alpha \). Likewise, \( \beta \) is not a cumulative density
function (CDF). As seen in Eq. (6), the integral \( \int_{b}^{a} B(\alpha) \, d\alpha \) is not 1, but \( B_{\text{int}} \), the scaling factor also used in the literature for a gamma and lognormal "rate PDF".

Solutions to the transport problem can be found by taking the Laplace transform of Eqs. (4) and (5) with regard to time, and combining them to eliminate the immobile concentration. This technique has been widely applied to one-site and dual porosity models (Sudicky, 1990; Harvey and Gorelick, 1995). The Laplace domain equation for mobile concentration subject to a continuous distribution of rate coefficients is:

\[
\tilde{s} \tilde{c}_m - c_{m,0} + \tilde{c}_m \int_{0}^{\infty} \frac{\alpha b(\alpha)}{s + \alpha} \, d\alpha - \int_{0}^{\infty} c_{\text{im},0}(\alpha) \frac{\alpha b(\alpha)}{s + \alpha} \, d\alpha = AD(\tilde{c}_m)
\]

where the tilde denotes a variable in the Laplace domain, \( s \) the Laplace parameter, and the subscript 0 the initial value. As can be seen in Eq. (7), the mobile concentration is dependent on the particular distribution of capacity in the immobile zones, expressed in the first integral in Eq. (7), denoted \( \mathcal{J}_1 \) in the following. The mobile concentration is also dependent on the initial concentration in the distribution of immobile sites as expressed by the second integral above, denoted \( \mathcal{J}_2 \) in the following. If this initial distribution is uniform over all sites (i.e., equilibrium), \( \mathcal{J}_2 \) can be simplified by taking the initial immobile concentration out of the integral, such that \( \mathcal{J}_2 = c_{\text{im},0} \mathcal{J}_1 \). Explicit expressions for \( \mathcal{J}_1 \) for two types of rate distributions discussed below are given in Appendix A. Particular expression for mobile concentration in simple transport problems will be given below.

2.3. Piece-wise linear distribution of rate coefficients

The mathematical development to this point is applicable to any distribution of rate coefficients, whether first-order or diffusion. For the purpose of our paper, analysis of the estimation problem is facilitated by assuming a piece-wise linear shape for the capacity density distribution \( b(\alpha) \), an example illustrated in Fig. 1. (Note that since \( B_{[\alpha_1, \alpha_2]} = \int_{\alpha_1}^{\alpha_2} b(\alpha) \, d\alpha \), for the range of small \( \alpha \) to contain noticeable capacity, the corresponding \( b \) must be very large, therefore a logarithmic \( b \)-axis in Fig. 1.) The piece-wise linear distribution model gives us the flexibility to approximate many true rate distributions, but it does not in itself represent physical process knowledge (for that perspective, the reader is directed to the extensive article by Fesch et al. (1998)). The sections in the piece-wise linear rate distribution are delimited by a vector of ‘anchor’ rate coefficients, \( \alpha_s \), briefly ‘anchors’. The optimal capacity densities at the given anchors, \( \tilde{b}(\alpha_s) \), are found in the estimation. The entire continuous distribution of rate coefficients is defined from the linear interpolation between the elements of \( \tilde{b}(\alpha_s) \).

Thus, with a potentially small number of parameters to be estimated, a continuous distribution over all rates can be represented.

Our working hypothesis is that the range of rates, \( \alpha \), for which capacity densities can be estimated, is related to the time range of available data. To investigate this relationship, we will plot rate distributions as a function of the time scale of mass transfer \( t_\alpha = 1/\alpha \) and we will look at the cumulative capacity \( \beta \). Looking at cumulative capacity in the proposed fashion has the advantage of a better visualization for the rates’ wide spread in magnitudes (Fig. 1), but also the uncertainty analysis described below is easier for the cumulative capacity (even although we are estimating the individual
densities at the anchors). We expect that the estimation procedure will indicate that cumulative capacities have large uncertainties for mass transfer time scales that are either much smaller or much larger than the time scale over which data are collected. This is illustrated for an example case in Fig. 1. For common types of experiments, our goal is to compare the width of the estimatable zone with that of the available data, and hence to compare the experiments themselves with respect to their potential for estimation of continuous rate distributions.

The reasons for uncertainty in capacity are different for the early and late times. For mass transfer that occurs over time scales \( t_a \) much smaller than the time of our first data point \( t_1 \), we cannot estimate how the capacity density is distributed, but we can determine the total capacity. This explains why the uncertainty in the cumulative capacity drops as we consider time-scales on the order of our data. For mass transfer that occurs over time scales much longer than the time scale of our experiment, we simply cannot estimate the capacity. If one nevertheless introduces anchors with large \( t_a \) in the vector \( \mathbf{b}(\mathbf{a}_j) \) to be estimated, the uncertainty in cumulative capacity will continue to grow larger as \( t_a \) gets larger than the time scale of the experiment. In the extreme, the estimation problem will become ill-posed.

2.4. Mathematical equivalence with diffusion model

There is an ongoing debate in the literature whether first-order or diffusive models are more appropriate (e.g., Wu and Gschwend, 1986; Ball and Roberts, 1991b; Werth and Reinhard, 1997). We will briefly demonstrate that for distributions of rate and diffusion coefficients, both mass transfer models are mathematically equivalent. Haggerty and Gorelick (1995) showed that a discrete multirate model as formulated in Eqs. (2) and (3) is exactly equivalent to a diffusion model for a single diffusion coefficient. For example, for spherical grains, the equivalent first-order multirate parameters are determined as

\[
\alpha_j = j^2 \pi^2 \alpha_d \tag{8}
\]

\[
\beta_j = \frac{6}{j^2 \pi^2} \beta_{\text{tot}} \tag{9}
\]

where the subscript ‘d’ stands for diffusion and \( \alpha_d = D_a/a^2 \) is the given (single) diffusion rate coefficient. The expansion is for \( j = 1 \) to \( \infty \). An example of a single diffusion rate and the equivalent sequence of first-order rate coefficients is shown in the top row of Fig. 2.

Now assume that the diffusion rate coefficient \( \alpha_d \) has a continuous distribution with capacity density distribution \( b_d \). Re-arranging Eq. (8) lets one express \( \alpha_d \) as a function \( h(\alpha_j) = \alpha_j/(j^2 \pi^2) \) for each \( j \). Using the rule for a derived density function, one can find the terms summing up to the equivalent first-order capacity density distribution (note the change in the argument of \( b_d \)) as:

\[
b_{1o,j}(\alpha) = b_d(h(\alpha_j)) \left| \frac{\partial h(\alpha_j)}{\partial \alpha_j} \right| \tag{10}
\]
where the subscript ‘fo’ stands for first-order. The total first-order capacity density is then found as the infinite series

\[ b_{fo}(\alpha) = \sum_{j=1}^{\infty} b_{fo,j}(\alpha) \beta_j \]  

with ‘weights’ \( \beta_j \) from Eq. (9).

In the context of solutions to the transport equation (e.g., Eq. (7)), the first-order distribution \( b_{fo}(\alpha) \) is mathematically equivalent to the diffusion distribution \( b_f(\alpha) \). An example showing an equivalent pair of continuous first-order and diffusion rate capacity densities is shown in the bottom row of Fig. 2. To compute the first-order curve in the figure, the series in Eq. (11) was expanded to 50 terms, at which point it had converged. Although the shapes of the distributions differ, the area under both curves (total capacity) is the same.

3. Methods: estimating rate distributions

Typically, mass transfer parameters are estimated from batch sorption experiments and gas-purging desorption experiments using analytical solutions for the traditional two-site model (Brusseau and Rao, 1989). To extend this methodology for continuous rate distributions, we developed Laplace-domain expressions for the batch and purging experiments. We use the quotient–difference inversion algorithm of de Hoog et al. (1982), which features double-accelerated convergence (our code splits the time vector in segments by order of magnitude as prompted by the argument of Gambolati et al. (1997)). This section concludes by elaborating on the inverse problem and presenting some estimation theory.
3.1. Modeling batch adsorption experiments

In a batch adsorption experiment, there is no immobile mass initially, whereas the mobile phase has a known initial concentration and \( c_{m,0} \); thus, \( \mathcal{F}_2 = 0 \). The advection and dispersion terms of Eq. (7) can be dropped for a perfectly-mixed batch reactor, as there are no spatial gradients of mobile concentration. The solution for the Laplace-domain mobile concentration is then

\[
\bar{c}_m = \frac{c_{m,0}}{s + s \mathcal{F}_1}
\]  

(12)

As can be seen from Eq. (5), the equilibrium concentrations in a batch experiment are \( c_{m,e} = c_{m,e}(\alpha) = c_a \). For mass balance \( \theta_m c_{m,0} = \theta_m c_a + \theta_m \int_0^\alpha b(\alpha) c_d d\alpha \), where \( \theta_m \) is the volume fraction of the mobile phase. It follows that \( c_a = c_{m,0}/(1 + \beta_m) \).

Taking the Laplace transform of Eq. (5) and using Eq. (12), the Laplace-domain expression for immobile concentration is

\[
\bar{c}_{im}(\alpha) = \frac{\alpha \bar{c}_m}{s + \alpha} = \frac{\alpha c_{m,0}}{(s + s \mathcal{F}_1)(s + \alpha)}
\]  

(13)

By inverting Eq. (13), one can assess how rapidly the equilibrium conditions are attained over the range of immobile sites. An example is shown in Fig. 3. The figure also shows that mobile concentration in a batch adsorption experiment can decrease in stages separated by periods of apparent equilibrium. This seems to be the case when the underlying cumulative capacity distribution has 'steps', as is the case in the example (see the inset in Fig. 3). Hence, there is a danger of assuming equilibrium conditions prematurely. From our experience of exploring the continuous rate distribution model

![Fig. 3. Hypothetical batch adsorption experiment: time-evolution of mobile (shown as top plane for comparison), and immobile concentrations over sites associated with a continuous distribution of rates (behavior at the anchors highlighted). Inset: corresponding cumulative capacity distribution.](image)
and as seen in Fig. 3, it appears that the time to equilibrium for a particular rate \( \alpha \) is about \( 10 \alpha^{-1} \). However, without analytical expressions for concentrations in the time domain, we cannot prove this conjecture.

### 3.2. Modeling purging desorption experiments

A purging experiment is a particular kind of column experiment. The material inside the column is first allowed to equilibrate with a contaminant. Then, a carrier gas is used to purge the contaminant from the column, with the eluent concentration being monitored (Werth and Reinhard, 1997). The exact mathematical description of the purging experiment has to be based on the full advection–dispersion Eq. (1) in one dimension, where in addition we assume a homogeneous dispersion coefficient. The initial and boundary conditions are

\[
\begin{align*}
    c_m(x,t=0) &= c_m(x,t=0) = c_m \\
    c_m(x,t \to \infty) &= c_m(x,t \to \infty) = 0 \\
    \left( c_m - \frac{D}{v} \frac{\partial c_m}{\partial x} \right)_{x=0,t} &= 0
\end{align*}
\]

Because of the second initial condition, \( \mathcal{F}_2 = c_m \mathcal{F}_1 \), so Eq. (7) becomes

\[
( s \bar{c}_m - c_m ) (\mathcal{F}_1 + 1) = D \frac{\partial^2 \bar{c}_m}{\partial x^2} - v \frac{\partial \bar{c}_m}{\partial x}
\]

The solution for \( \bar{c}_m \) is (cf., Jury and Roth, 1990, [solution to prob. 3.7])

\[
\bar{c}_m = \frac{c_m(\mathcal{F}_1 + 1)}{D(r_2 - r_1)} \left[ e^{r_1 x} \left( \frac{1}{r_1} - \frac{1}{r_2} \right) - \frac{1}{r_2} + \frac{1}{r_1} \right]
\]

where the two roots \( r_1 \) and \( r_2 \) are ‘+’ for \( r_1 \), ‘−’ for \( r_2 \):

\[
r_{1,2} = \frac{v}{2D} \left( 1 \pm \sqrt{1 + \frac{4D(s + s \mathcal{F}_1)}{v^2}} \right)
\]

The observations in the purging experiment are those of flux concentration, \( \bar{c}_m^t = c_m - (D/v)(\partial c_m/\partial x) \); thus likewise in Laplace space \( \bar{c}_m^t = \bar{c}_m - (D/v)(\partial \bar{c}_m/\partial x) \). The derivative can be computed from Eq. (15). A solution analogous to Eq. (15) can also be found for non-equilibrium initial conditions, but it requires rather extensive numerical integration. Solutions for other mass transfer models have been presented by Haggerty and Gorelick (1998).

### 3.3. Estimation of anchor capacities

The goal of our estimation method is to find the capacity densities at specified anchor rate coefficients, i.e., \( P \) parameters grouped in a vector \( b(\alpha) \), and their joint uncertainties. In other words, we fix the position of the anchors along the \( \alpha \)-axis and estimate the
Given $N$ measurements with independently distributed errors, $b(\alpha_s)$ can be found by minimizing the objective function

$$O(b(\alpha_s)) = \sum_{i=1}^{N} \frac{1}{\sigma_{x,i}^2} \left[ x_{\text{meas},i} - x_i(b(\alpha_s)) \right]^2$$

where $x_{\text{meas}}$ is the measured variable and $x_i(b(\alpha_s))$ the corresponding model prediction. The $\sigma_{x,i}^{-2}$ are weights, reflecting the precision of the measurements (the reciprocals of their error variances). These weights serve to normalize each squared deviation to unit variance. For a batch experiment, the measured variable $x$ is the mobile concentration; in purging experiments, $x$ is the flux concentration.

The optimal (minimal) sum-of-squares $O_{\min}$ is a random variable that follows a $\chi^2$ distribution with $v = N - P$ degrees of freedom (Press et al., 1992, Chap. 15.1):

$$O_{\min}(b(\alpha_s)) \sim \chi_{N-P}^2$$

Statistically, the probability of model adequacy can thus be defined as

$$P_{\text{adeq}} = 1 - Q(O_{\min}(b(\alpha_s)),N-P)$$

where $Q(O_{\min},v)$ is the $\chi^2$ cumulative density function. The smaller the optimal objective function value, the more likely adequate the model—for an adequate model, any realization of measurements is unlikely to result in a large $O_{\min}$ by chance. Here, ‘model’ is to be understood as the combination of the possibly simplified advection–dispersion equation and the optimal rate distribution parameters $b(\alpha_s)$, or other parameters for other rate distribution models.

### 3.4. Estimation uncertainty

Uncertainty in parameter estimates is represented in the estimation covariance matrix $\Sigma_{b(\alpha_s)}$. Its lower bound can be found from the Cramer–Rao theorem (Schweppe, 1973, Chap. 12) as:

$$\Sigma_{b(\alpha_s)} \geq H^{-1}$$

where $H$ is the Hessian matrix, which can be computed from the sensitivity of the model to its parameters (the Jacobian), weighted by the precisions in the observations (Press et al., 1992, Chap. 15.5):

$$H_{kj} = \sum_{i=1}^{N} \frac{1}{\sigma_{x,i}^2} \left[ \frac{\partial x_i(b(\alpha_s))}{\partial b(\alpha_{s,k})} \frac{\partial x_i(b(\alpha_s))}{\partial b(\alpha_{s,j})} \right]$$

In our application, the derivatives were found by numerical approximation. Eq. (20) gives an exact representation of the Hessian only if the model is correctly specified. This was assumed in the uncertainty analyses presented below. We verified the absence of a bias via a Monte Carlo study (i.e., by repeatedly estimating the rate distribution from synthetic data).

Estimates of the capacities at the anchors are correlated, and interpretation of multivariate estimates is conceptually difficult. With the piece-wise linear distribution model, it is possible to analyze estimation results for the single variable $\beta$. Cumulative
capacity from the ‘fastest’ (largest) anchor, $\alpha_{a,p}$ down to the $k$-th anchor, $\beta_k = \int_{a_{a,p}}^{a_{a,k}} b(\alpha) d\alpha$, can be expressed as a linear combination of the individual capacity densities $b(\alpha_a)$:

$$
\beta_k = \frac{1}{2} \begin{bmatrix}
\alpha_{a,p-1} - \alpha_{a,p} \\
\alpha_{a,p-2} - \alpha_{a,p} \\
\vdots \\
\alpha_{a,k+1} - \alpha_{a,k+2} \\
\alpha_{a,k} - \alpha_{a,k+2} \\
\alpha_{a,k} - \alpha_{a,k+1}
\end{bmatrix}^T 
\begin{bmatrix}
b(\alpha_{a,p}) \\
b(\alpha_{a,p-1}) \\
\vdots \\
b(\alpha_{a,k+2}) \\
b(\alpha_{a,k+1}) \\
b(\alpha_{a,k})
\end{bmatrix} = I_k^T b(\alpha_{a,p} \ldots k) \quad (21)
$$

Notice the reverse sequence in $\alpha (P > k)$, which gives an increasing sequence in $t_a = 1/\alpha$. It is more insightful to interpret uncertainty in cumulative capacity along the axis of mass transfer time scales, because it can then be related to data time scales.

The estimation variance for the linear combination $\beta_k = b(\alpha_{a,p} \ldots k)$ is

$$
\sigma_{\beta_k}^2 = I_k^T \Sigma_{b(\alpha_a)} I_k \quad (22)
$$

The above formulation takes account of the correlation in the estimates $b(\alpha_a)$ as contained in the covariance matrix $\Sigma_{b(\alpha_a)}$. Picking out only part of the covariance matrix $\Sigma_{b(\alpha_a)}$ when computing intermediate variances of $\beta_k (k \neq 1)$ does however represent an approximation. Still, we feel the analytical formulation in Eq. (22) can elucidate the general nature of the uncertainty structure better than an alternative procedure based entirely on Monte Carlo trials (we verified that the results are similar). For appreciating estimation results, recall that even though $\beta$ is a linear combination of the individual estimated densities $b(\alpha_a)$, these themselves are found in a nonlinear estimation problem (mobile concentration is a nonlinear function of $b(\alpha_a)$, cf., Eqs. (12a) and (15)). The true confidence ranges at the anchor capacities may thus be asymmetric. Still, for distributions other than piece-wise linear, cumulative capacity is not a linear combination of the distribution parameters, thus no linear statistics can be used to express estimation uncertainty at all.

### 3.5. Model refinement algorithm

Model refinement for the piece-wise linear rate distribution can be achieved through elimination of redundant anchors. We developed an algorithm that successively removes anchors for which the estimated $b(\alpha_a)$ is closest to the value that follows from interpolation between neighboring points of $b(\alpha_a)$. Recalling that the continuous distribution is already defined by linear interpolation between the anchors’ capacities, such elimination has least impact on the shape of the distribution. For the reduced set of anchors $\alpha'_a$, a new optimal set of $b(\alpha'_a)$ is found. The reduced model (one anchor fewer) is tested based on the $F$-distribution of the ratio of the corresponding minimal objective function values:

$$
\frac{\left[ O_{\min}(b(\alpha'_a)) - O_{\min}(b(\alpha_a)) \right]}{O_{\min}(b(\alpha_a))/[N - P]} \sim F_{1, N - P} \quad (23)
$$
In practice, one will decide on a significance level \( p_{\text{improve}} \) for a threshold value of the \( F \)-statistic. If the ratio of the objective function values is larger than the threshold, the reduced model is rejected because it is not significantly better than the model with one anchor more. The reduced model must also not be over-simplified, i.e., statistically inadequate, as tested using Eq. (18). In this work, the significance levels for accepting model improvement and model adequacy (\( p_{\text{improve}} \) and \( p_{\text{adeq}} \)) were chosen to be the same (0.025).

4. Results: estimation examples

4.1. Model refinement algorithm: eliminating redundant anchors

A hypothetical batch experiment can serve to demonstrate the performance of the model refinement algorithm (Fig. 4). The assumed true rate distribution has four anchors; the three sections contain approximately the same capacity. Synthetic data were
generated by spoiling the predicted concentration histories for the true model; these data were then used in the estimation procedure for the following rate distribution models. The model with too many anchors is the one with an additional anchor at $t_a = 3.3 \times 10^2$ whose capacity density follows from linear interpolation of the densities at the neighboring anchors. Because the linear interpolation is already inherent by definition in the true model, the additional anchor is redundant—the predicted concentration histories are identical. Indeed, the redundant anchor is detected and removed by the refinement algorithm. The now refined model again has the same set of anchors as the true model, and hence the concentration histories are of course again the same for the refined and the true model. Taking the refinement one step further by removing an additional anchor at $t_a = 10^1$ (now the one closest to the line between its neighbors) however results in an over-simplified model. The predicted concentration history is on average outside the measurement error range, so the probability of model adequacy $\rho_{adeq}$ (Eq. 18) drops below the threshold of 0.025. Therefore, the refinement algorithm would terminate at the previous step, outputting as result the refined model in Fig. 4.

The example also shows that finding a rate distribution requires some manual interaction in the definition of a reasonable initial guess of the number and locations of anchors. In principle, one can start with a large number of them, and let the refinement algorithm find the essential ones. As shown below for some actual data, this approach is practical for about 10 anchors, but for larger numbers, the estimation can become unstable.

4.2. Which experiment type is better, batch or purging?

Here we consider whether batch or purge experiments provide better estimates of the distribution of rate coefficients, by comparing estimated distributions from both hypothetical batch and purging experiments simulated with the same ‘true’ rate coefficient distribution. In particular we focused on the zone of mass transfer time scales for which capacity could be estimated. To compare the two different experiments, we assumed hypothetical data to be of a quality attainable with state-of-the-art techniques. Hypothetical measurement times were taken as those from the actual purging experiment to be analyzed at the end of this section. The relative error of flux concentration was $k = 0.15$ among two replicates; in our experience, batch experiments can achieve a value of $k = 0.05$ for mobile concentration. The weight for each datum $c_i$ follows as $\sigma_i^2 = (kc)_i^{-2}$. For the purging case, flow velocity, dispersion coefficient, and column length were taken from the actual experiment. Given relative errors, the Hessian matrix is independent of $c_0$ (batch) and $c_\infty$ (purging), because the Laplace-domain and thus the time-domain expressions of concentrations are linear with respect to initial concentrations. Because the equation for the Hessian, Eq. (20), is only exact for the correct model parameters, we computed estimation uncertainty around the true rate distribution, in effect examining the best-case scenario.

Purging experiments outperform batch experiments, especially when the data only cover a limited window of time (Fig. 5), even though the purging data were perturbed with a greater relative error. Capacity densities for rates with $t_a < t_1$, with $t_1$ the time of the first datum, cannot be identified individually for either type of experiment. Their
Fig. 5. Batch vs. purging experiments: estimating a rate distribution from data covering only a portion of the desired mass transfer time scales. Shown are the estimated cumulative distributions along with 90% confidence intervals at the anchors for hypothetical experiments. The advective time in the purging experiment (PV = pore volume) is also marked; \( t_1 \) is the time of the first datum.

integral, cumulative capacity up to \( t_1 \), is defined better. (Note that uncertainty for a section is shown for the large \( t_\alpha \)-end, so at the first anchor, there is zero uncertainty in \( \beta \).) For both types of experiments, uncertainty in capacity densities grows for rate coefficients with \( t_\alpha \) larger than the time of the last datum, but the effect is much less pronounced for the purging experiment. The error bars grow also for the purging case when appending anchors with even larger \( t_\alpha \) to the vector \( b(\alpha) \) to be estimated (not graphed). Still, the estimatable zone in the rate distribution extends towards large mass transfer times, in our experience by about one order of magnitude further with purging experiments than it does with batch experiments. As can be seen in the batch case, the lower capacity uncertainty limit can drop even below that achieved at with fewer anchors. This behavior shows that including too many anchors with too large \( t_\alpha \) makes the estimation problem ill-posed. The computation of capacity variance (Eqs. (19)–(22)) becomes numerically unstable, because an ill-conditioned matrix \( H \) has to be inverted.

Two advantages seem to make the purging experiment superior, both evident from a plot of the sensitivity of the concentration observations to capacity densities at the anchors. Fig. 6 shows the terms whose sum forms the \( k \)-th diagonal element of the Hessian (cf., Eq. (20)). These individual terms, denoted \( s_k(t_i) \), represent the pseudo-univariate sensitivity at the \( k \)-th anchor (with characteristic time \( t_\alpha,k \)) at the time of the \( i \)-th datum:

\[
s_k(t_i) = \frac{1}{\alpha_k^2} \left[ \frac{\partial c(t;b(\alpha))}{\partial b(\alpha_k)} \right]^2 \quad k = 1 \ldots P
\]

As seen in Fig. 6, for the purging experiment, periods of relatively high sensitivity are differentiated much more between parameters than for the batch experiment. In other
words, the purging has the advantage that different data time windows improve capacity density estimates for different anchors. Reflecting the smaller correlation in the estimates, the condition number of $\Sigma_{\alpha_i \alpha_j}$, the estimates’ covariance matrix, is two orders of magnitude smaller for the purging case (but still very large: $2 \times 10^3$). Sensitivity ‘sequencing’ seems to outweigh the almost always higher absolute sensitivities for the batch experiment. At late times, however, a second advantage of the purging experiment comes into play. Concentration levels are so low then that for a constant relative error, the weights $\sigma_{\alpha_i}^2$ in Eq. (24) become very large. Hence, absolute sensitivities for anchors with large $t_\alpha$ grow to be larger in purging than in batch experiments.

We conducted some additional analyses to examine the generality of the finding that purging experiments provide better estimates. One of their potential drawbacks is that if not measured independently, the dispersion coefficient $D$ has to be estimated along with the rate distribution. In finding the uncertainty in cumulative capacity $\beta$, the estimation covariance between $D$ and the capacity density estimates $b(\alpha_i)$ had to be neglected, because no appropriate modification of the relevant Eq. (22) is possible. We repeated the exercises of finding the estimatable zone (Fig. 5) for different values of $D$, and found that parameter’s impact to be small. Only very strong dispersion causes individual sections of the rate distribution to become less identifiable. With respect to advection, we noticed that it becomes difficult to distinguish among capacities for mass transfer time scales smaller than the time scale of advection (one pore volume). If many additional late-time data are available, total capacity $\beta_{tot}$ can also be estimated ac-cu-
rately from batch experiments, as long as \( b_{\text{tot}} \) is not too large to cause mobile concentration to drop below detection limit. This is so because the equilibrium (final) concentration is \( c_{e} = c_{f}/(1 + b_{\text{tot}}) \), and the estimation variance of the average of many measurements of \( c_{e} \) becomes small even though each individual measurement has a rather high error. Finally, we found that an incorrect assumption of initial equilibrium before purging results in underestimates of the capacity of the slow sites. None of the conditions that can deteriorate the performance of purging experiments were present in the actual case to be analyzed below, and it should generally be possible to avoid them in practice.

Apart from the arguments following from the above statistical analyses, also practical considerations come into play when designing experiments for rate distribution estimation. Purging experiments are ideally first started after equilibration during an adsorption (uptake) phase, so they take twice as long as batch adsorption experiments, where the sample is generally known to be initially ‘clean’. Practical difficulties with purging experiments are keeping columns tightly sealed and obtaining accurate flow measurements. On the other hand, the purging experiment might be easier to conduct for actual aquifer material, because the sample does not have to be diluted and mixed continually. Pulse breakthrough curves will likely not be suitable for determination of rate coefficients, because the chemical’s residence time in the column is too short for slow mass transfer to be revealed.

4.3. Application to purging data with Santa Clara sediment

As an example, we applied the continuous, piece-wise linear rate distribution model to TCE desorption data from Werth and Reinhard (1997). TCE had been sorbed to Santa Clara sediment contained in columns (the TCE isotherm had been found to be non-linear with a Freundlich exponent of 0.45). After an equilibration period, the columns were purged and the corresponding flux values were measured over time. The flux detection limit was approximately \( 10^{-13} \) mol/min. Down to this value the error in flux measurements was relatively constant. Once the flux approached the detection limit, the columns were disconnected and baked to remove the residual mass of TCE. Two replicates with equilibration periods of 1 and 8 months, respectively, yielded essentially the same purging observations, indicating equilibrium had prevailed initially. There were \( N = 74 \) desorption data spanning a time period of approximately 1 week (10^4 min), of which one pair was identified as outliers and removed from the set. The relative error in flux concentration was 0.15 among two replicates, practically constant over all orders of magnitude in measured \( c_{f} \).

For the flux concentration solution, the piece-wise linear rate distribution gave the best fit to the observations (Fig. 7). The optimal rate distribution based on the gamma probability density function (with three parameters \( b_{\text{tot}}, \lambda, \eta \), and \( h(\alpha) = b_{\text{tot}} \alpha^{\lambda-1}/[\exp(\alpha/\eta)\eta^\lambda(\lambda)] \)) resulted in a poorer fit, and a single-rate model in an even worse one. The probability of the piece-wise linear model being adequate (cf., Eq. (18)) was \( p_{\text{adeq}} = 0.52 \). For the gamma-type distribution and single-rate models, \( p_{\text{adeq}} < 10^{-15} \), a consequence of the corresponding concentration predictions lying outside the 90% confidence range of many data points (Fig. 7). Because only two replicates were
available to compute measurement error and thus weights in the objective function (Eq. (16)) the probability levels are likely not exact. Given the much larger $P_{\text{sel}}$ for the piece-wise linear model, the conclusion that it is more adequate is unaffected. We also attempted fitting a lognormal type of a rate distribution to the observations expressed as mass fraction remaining $(m(t)/m_0)$, and obtained results no better than for the gamma one. (Obtaining the solution for $c^I$ for the lognormal model was thus not worth the necessary cpu-intensive numerical integration of $F_I$.)

Apparently, a finer resolution of the rate distribution than provided by ‘statistical’ functions (i.e., gamma, lognormal) is needed. The estimated cumulative capacity distribution has a plateau in the range of $0.25 < \tau < 1$ [min] (Fig. 7b). In all, the capacity density distribution (not shown separately) has three modes at $t_\alpha = 0.2$, 1.66, and $10^3$ min. The plateau causes a ‘biphasic’ mass release (Pavlostathis and Mathavan, 1992; Cunningham et al., 1997) with the second ‘phase’ corresponding to sites associ-
ated with a wide range of rate coefficients. The TCE data in our Fig. 7 had also been analyzed by Werth et al. (1997), but even though their model properly represented isotherm non-linearity, it could not describe the concentration ‘tail’ as well as our rate-distribution model. Other authors working with rate-distribution models, but assuming a gamma rate distribution model, reported poorer fits for other data, although they did not evaluate statistical adequacy. Culver et al. (1997) [Tables 4 and 5] reported a 20 to 30% error for column experiments. The fitted model prediction is outside the measurement variability range in parts of the curves shown by Pedit and Miller (1994). Connaughton et al. (1993) expressed $1 - m(t)/m_0$ for a gamma distribution of rate coefficients, and their fit was poor in some cases (see their Fig. 4a).

Some readers might object to the higher number of parameters in the piece-wise linear distribution model compared to the gamma one. However, statistical adequacy is an objective criterion for model evaluation. Given 72 data with relative error of 0.15, nine anchors cannot be called over-parameterization. A large number of parameters could only be a practical problem for estimation, but certainly not for forward modeling. As we have shown in this example, however, it is perfectly possible to estimate the $b$ coefficients for about 10 anchors. It is also easy to do so, because of the close correspondence between time windows in the concentration history and $t_a$ windows in the distribution plot. If concentration changes noticeably in a certain time window, the corresponding part in the distribution has to be resolved with more anchors. In practice, one can thus find an adequate rate distribution by first successively adding some anchors, then using the refinement algorithm to remove redundant ones.

5. Conclusions

- The advection–dispersion equation with rate-limited mass transfer for a truly continuous distribution of rate coefficients can be formulated in the Laplace domain, with numerical inversion for solution. For a continuous distribution of diffusion coefficients, an equivalent continuous distribution of first-order rate coefficients can be found.
- The piece-wise linear description of the rate distribution has several advantages: (1) generality, (2) stable Laplace inversion, (3) applicability of linear statistics for evaluation of estimation uncertainty in cumulative capacity, and (4) correspondence between time windows in observations and sections in the rate distribution. The fourth advantage simplifies rate distribution estimation in practice. The model refinement algorithm allows identification and removal of redundant section delimiters (‘anchor’ rate coefficients).
- The estimatable zone of the rate distribution is closely linked to the time window of available data. For sites with mass transfer time scales smaller than the experimental one, only a lumped (cumulative) capacity can be estimated. Uncertainty decreases over the data time zone; beyond the data time zone, it grows again, but much less in purging than in batch experiments.
- Purging experiments outperform batch experiments for estimating rate distributions for two reasons. (1) Estimation sensitivities to capacities at individual anchor rate coefficients are more distinctly sequenced with time, giving less correlated estimates. (2)
Smaller late-time measurement error yields an extended zone of estimatable rate coefficients towards large mass transfer time scales.

- The piece-wise linear rate distribution model yielded a much more statistically adequate fit to TCE gas-purging data than did models with ‘statistical’ distributions of rate coefficients (gamma, lognormal) or a single rate coefficient.

Acknowledgements

We are grateful for helpful comments from Mark Goltz and two anonymous reviewers. We thank the US EPA for supporting this work through NCERQA Grant R825825.

Appendix A. Rate distribution dependent expressions

Expressions for the integral $J_1$ for the piece-wise linear and gamma rate distributions are given in Table 1. For the single-rate model, $J_1$ is not an integral, but a simple expression following from a derivation analogous to that of Eq. (7) as $J_1 = \beta / (s + \alpha)$.

Note that the integral for the rate distribution based on the gamma PDF contains the factor $\exp(s/\eta)$. In practice, the rate distribution will commonly have low mean and variance, thus the PDF will be characterized by a small $\nu$ (the mean is proportional to $\eta$, the variance to $\eta^2$). Hence numerically, there is an overflow problem for large Laplace parameters $s$, which makes it difficult to obtain early-time solutions through inversion (keeping in mind that $c_m(0) = \lim_{x \to +0} \pi \delta (s)(s)$). The problem can be reduced, but not fully overcome, by computing the logarithms of the r.h.s. of Eq. (A2) and then exponentiating the sum of all terms. A first guess of $\lambda$ can be taken from the large-time slope of log-log-plotted $c^2(t)$, which is $-1 - \lambda$.

Table 1

<table>
<thead>
<tr>
<th>$J_1$ evaluated for the piece-wise linear (Eq. (A1)) and gamma rate distributions (Eq. (A2))</th>
</tr>
</thead>
<tbody>
<tr>
<td>$J_1 = \sum_{i=0}^{N} \left[ (b_{i+1} - b_i) \left( \frac{a_i^2 - a_{i+1}^2}{2(a_{i+1} - a_{i})} + \frac{a_i b_{i+1} - a_{i+1} b_i + s(b_{i+1} - b_i)}{a_{i+1} - a_i} \right) \times \left( -s \log(\alpha_{i+1} + s) + s \log(\alpha_i + s) + \alpha_{i+1} - \alpha_i \right) \right]$ (A1)</td>
</tr>
<tr>
<td>$J_1 = -\beta \int_{0}^{\infty} \frac{e^{-x/\eta} \sin(\lambda \pi)}{\eta \Gamma(\lambda)} (\pi + \lambda \Gamma(\lambda) \Gamma(-\lambda, 0, s/\eta) \sin(\lambda \pi))$ (A2)</td>
</tr>
</tbody>
</table>

The extended incomplete gamma function is defined as $\Gamma(x, z_0, z_1) = \int_{z_0}^{z_1} e^{-t} (t^{z_1 - 1} + t^{z_1 - 1}) dt$, or, with the incomplete gamma function $P$ defined as in the paper of Abramowitz and Stegun (1965) [Eq. 6.5.1], $\Gamma(x, z_0, z_1) = \Gamma(x, P(z_1, s) - P(z_0, s))$. 

References


