Model Parameter Estimation for Particle Transport

By Andrew N. Ernest,[1] James S. Bonner,[2] Member, ASCE, and Robin L. Autenrieth,[3] Member, ASCE

Abstract: A generic parameter estimation algorithm was developed for the simultaneous extraction of multiple parameters from experimental or field observations. The basis of the method lies in the minimization of the variation between model predictions and observations via the iterative numerical determination of the functional minima of the model with respect to the parameters being determined. A multivariable Newton technique, and derivatives thereof, are used for this purpose. By relying on numerical differentiation in the iterative process, the technique is not limited by the model complexity, solution methodology, or scale of its domain. Model parameters have been estimated successfully for analytical and numerical solutions to dispersive and advective-dispersive particle transport model equations. The effect of sampling or measurement error and initial parameter estimates on algorithm performance has been evaluated. This approach has been successfully applied to determine data requirements and constraints (quantity, quality, and location) for measurement of hydrodynamic and transport characteristics of dye clouds and aquatic particles in laboratory-mixed settling columns.

Introduction

The affinity of many environmentally significant contaminants, including heavy metals and chlorinated hydrocarbons, for fine-grained sediments in aquatic systems is well-documented (Carpenter 1987; Elzerman and Coates 1987; Rodgers 1983). Adequate representation of particle transport is therefore necessary if the fate of such contaminants is to be evaluated (Lee et al. 1981; Lick and Kang 1987). The particle transport mechanisms that need to be characterized in natural systems include fluid convection, Stokes settling, Brownian diffusion, eddy dispersion, particle dispersion, flocculation, and differential settling (Hunt and Pandya 1984; Lick 1982; O'Melia 1972; Russel 1981). Contaminant interactions of import include sorption, desorption, and chemical reaction (Autenrieth 1986; Bonner et al. 1986).

The ultimate objective of this work is the development, testing, and application of a framework for the estimation of model parameters from experimental or field data. The work presented in this paper was conducted in conjunction with an ongoing experimental effort (Bonner et al. in press 1990; Ducharme 1989; Rorschach 1990; Sanders 1990; McCreary 1990) focusing on particle-mediated contaminant transport dynamics with emphasis on hydrodynamic transport, transport of heterogeneous particles, flocculation and contaminant sorption/desorption. To maintain applicability as the experimental and modeling efforts progressed, a generic modular framework was developed, within which widely differing predictive models could be

[1]Res. Asst., Dept. of Civ. Engrg., Texas A&M Univ., College Station, TX 77843.
[2]Asst. Prof., Dept. of Civ. Engrg., Texas A&M Univ., College Station, TX.
[3]Asst. Prof., Dept. of Civ. Engrg., Texas A&M Univ., College Station, TX.

Note. Discussion open until March 1, 1992. To extend the closing date one month, a written request must be filed with the ASCE Manager of Journals. The manuscript for this paper was submitted for review and possible publication on August 2, 1989. This paper is part of the ***Journal of Environmental Engineering***, Vol. 117, No. 5, September/October, 1991. ©ASCE, ISSN 0733-9372/91/0004-0573/$1.00 + $.15 per page. Paper No. 26274.

implemented. Of paramount interest is the ability to simultaneously extract multiple parameter values from existing models with minimal modifications or rederivations.

This paper focuses on the research conducted in the development of the basic framework, its application to hydrodynamic and particle transport models, and an evaluation of its effectiveness. The framework is evaluated purely in terms of its performance with respect to the applied models. General theoretical evaluation of parameter estimation with problem data can be found in Gawthrop (1984) and Stefanski (1985).

Approach

Dispersive Transport

The initial application of the parameter estimation framework was to a single-parameter, one-dimensional dispersive transport model. Corresponding to the first phase of the parallel experimental approach involving hydrodynamic characterization of a mixed settling column, the governing differential equation used is

$$\frac{\partial C}{\partial t} = D_z \frac{\partial^2 C}{\partial z^2} \quad \ldots\ldots (1)$$

where C = the concentration of the dispersing species at any given time t and distance z; D_z = the hydrodynamic dispersion coefficient, which may be viewed as the cumulative effect of Brownian diffusion, eddy diffusion, and mechanical dispersion. With respect to the experimental effort, D_z indicates the dispersive tendency of a neutrally buoyant dye under the following boundary conditions

$$D_z \left.\frac{\partial C}{\partial z}\right|_{z=0} = 0 \quad \ldots\ldots (2a)$$

$$D_z \left.\frac{\partial C}{\partial z}\right|_{z=h} = 0 \quad \ldots\ldots (2b)$$

where h = the distance between the two boundaries. This was selected as 2.0 m to conform with the aforementioned experiments. An impulse function was applied as the initial condition, to simulate actual experimental conditions

$$C(z,0) = C_{eq}\delta(z - 0) \quad \ldots\ldots (3)$$

where C_{eq} = the uniformly mixed concentration of the dispersing species.

Advective-Dispersive Transport

The one-dimensional advective-dispersive transport model was used as the basis for rigorous evaluation of the parameter estimation framework. It was also the model used in the preliminary analysis of data from the second phase of the corresponding experimental effort (Ducharme 1989; Sanders 1990; Bonner et al. 1990), particle transport studies in a mixed experimental settling column

$$\frac{\partial C}{\partial t} = D_z \frac{\partial^2 C}{\partial z^2} - v_z \frac{\partial C}{\partial z} \quad \ldots\ldots (4)$$

with experiment specific boundary conditions

$$D_z \left.\frac{\partial C}{\partial z}\right|_{z=0} = v_z C|_{z=0} \quad \ldots\ldots (5a)$$

$$D_z \left.\frac{\partial C}{\partial z}\right|_{z=h} = 0 \quad \ldots\ldots (5b)$$

and a uniform initial condition of

$$C(z,0) = C_{eq} \quad \ldots\ldots (6)$$

where v_z = the mean velocity of the dispersing species; C_{eq} = the uniform initial concentration; and h = the distance between the two boundaries. A useful nondimensional parameter indicating the relative dominance of advective and dispersive transport is the Peclet number, defined as

$$Pe = \frac{v_z h}{D_z} \quad \ldots\ldots (7)$$

Parameter Estimation

All gradient-based nonlinear parameter estimation techniques require an initial educated guess for the parameter vector. This parameter vector is then iteratively modified by a parameter step vector until a specified stopping criterion is achieved. It is in the evaluation of the parameter step vector that the parameter estimation techniques differ. Each of the classical parameter estimation techniques has its advantages over the others, generally in the area of convergence rates and requisites for convergence. It is not uncommon to incorporate several of the classical estimation techniques into one algorithm, to ensure convergence under as many conditions as possible. Srinivasan and Aiken (1986) show that convergence restrictions for stiff problems can be overcome in some cases by reparameterizing or restructuring the governing model.

Possibly the most mathematically elegant gradient-based method is the Newton method (Bard 1974). Many of the gradient estimators are in fact simplifications or approximations to the Newton method. The Newton method requires the evaluation of the Hessian of the likelihood function as an indication of the gradient and direction toward the optimum parameter set and is used in the evaluation of the parameter step vector.

The Gauss-Newton method is a simplification of the full Newton method, such that the second-order terms are neglected in the evaluation of the Hessian matrix. However, the Newton method is locally quadratically convergent, while the Gauss-Newton does so only linearly. Further, the Gauss-Newton method works best when the system of normal equations is quasi-linear and the minimum value of the residual function is small (Dennis and Schnabel 1983). In the case of highly nonlinear problems with large residuals, this method is likely not to converge.

Other popular methods include the damped Gauss-Newton, in which the parameter step vector $\boldsymbol{\delta}^k$ is scaled down sufficiently to ensure local convergence, even in highly nonlinear problems. This, however, decreases the rate of convergence even further. The Levenberg-Marquardt method is another adaptation of the Gauss-Newton method (Draper and Smith 1981), in which $\mathbf{G}$ is augmented to ensure better convergence in even large residual or highly nonlinear problems. The desired effect of the augmenting matrix in this

case is to replace the second derivative terms lost from the full Newton method. Although, with appropriate selection of the augmenting matrix, the Levenberg-Marquardt technique can approach the performance of the Newton method, algorithm control criteria can be computationally expensive.

The residual function is defined as the sum of the square of the differences between model predictions and the observations. For linear models, this function is then differentiated with respect to each of the free parameters, and the resulting set of equations, called normal equations, are explicitly solved in terms of the free parameters.

When applied to a nonlinear model, the explicit nature of the solution of the parameter set is lost. The resulting normal equations in this case are nonlinear. Further, depending on the model being fitted, it is often the case that the normal equations cannot be derived. To maintain the generality of the methodology, numerical techniques are used to both derive and solve the normal equations.

Derivation of the normal equations is most easily accomplished through the use of finite divided differences. The result can be expressed in functional form and implemented in an iterative root-finding scheme to locate the best estimate for the parameter set. Most iterative equation-solver techniques employ an initial estimate of the root, which is updated until a specified stopping criterion is reached. In the case of the dispersive transport model, only one parameter, D_z, is required to be estimated. Here, the experimental data can be expressed as functions of discrete space and time, corresponding to the sampling locations and frequency

$$C_{\text{obs}_i} = C_{\text{obs}}(z_i,t_i) \quad \text{(8)}$$

In addition, the mathematical model can be expressed similarly

$$C_{\text{pred}_i} = C_{\text{pred}}(z_i,t_i) \quad \text{(9)}$$

Both (8) and (9) are also functions of the free parameter D_z. The physical representation of D_z in the experimental data [(8)] may be related to the mixing intensity and represented mathematically as D_z in (9). These two functions can then be combined to create the residual function, which must then be a function of only D_z for any given observed data set

$$S_r(D_z) = \sum_{i=1}^{n} (C_{\text{pred}_i} - C_{\text{obs}_i})^2 \quad \text{(10)}$$

where n = the total number of observations. As there is only one parameter to be estimated, only one normal equation can be derived. This is accomplished by taking the derivative of the residual function with respect to D_z and setting it equal to zero

$$S_r'(D_z) = \frac{\partial S_r}{\partial D_z} = 2\sum_{i=1}^{n}\left[(C_{\text{pred}_i} - C_{\text{obs}_i})\frac{\partial C_{\text{pred}_i}}{\partial D_z}\right] = 0 \quad \text{(11)}$$

Eq. (11) constitutes the normal equation defining the optimum value of D_z. It is evident that only C_{pred_i} is a function of the free parameter, and as such is the only term for which the derivative is required. For more complex models, the functional expression for C_{pred_i} may not be analytically derivable, and may have to be treated as a black box. Therefore the method used to evaluate the derivative must be independent of the functional form. Finite

divided differences are used to this end. The derivative term is evaluated about the current parameter value by perturbing the function a value of $\pm\Delta D_z$

$$\left.\frac{\partial C_{\text{pred}_i}}{\partial D_z}\right|_{D_z^k} \approx \frac{C_{\text{pred}_i}(D_z^k + \Delta D_z) - C_{\text{pred}_i}(D_z^k - \Delta D_z)}{2\Delta D_z} \quad \ldots\ldots\ldots\ldots (12)$$

This classical central difference scheme would require three model evaluations at any given parameter value for the evaluation of the normal equation—one at the current parameter value and one on either side of it. The higher-order of accuracy of this scheme is desired over the reduced computational expense of forward or backward difference schemes.

Location of the root of the normal equation is accomplished iteratively, using either the Newton-Raphson or secant methods

$$D_z^{k+1} = D_z^k - \frac{S_r'(D_z^k)}{S_r''(D_z^k)} \quad \ldots\ldots\ldots\ldots (13)$$

$$D_z^{k+1} = D_z^k - S_r'(D_z^k)\left[\frac{D_z^k - D_z^{k-1}}{S_r'(D_z^k) - S_r'(D_z^{k-1})}\right] \quad \ldots\ldots\ldots\ldots (14)$$

Further simplification of the secant method to reduce the overall number of model runs per iteration was not implemented because both the possibility and rate of convergence would be excessively impaired, enough to preclude applicability to stiff problems.

To generalize this method to multiple parameters, it was applied to the advective-dispersive transport model. Here, the free parameters are v_z and D_z. The existing single-parameter framework was modified to incorporate an independent search (steepest ascent) algorithm. Here, each parameter was estimated successively, until they all met a preset stopping criterion. This algorithm provided a great deal of flexibility in independently controlling the rate of convergence for each parameter. Such flexibility is desired in situations where model solution stiffness varies widely with each parameter. Overall convergence in this method is generally slower than it is if all parameters are updated simultaneously. Simultaneous solution of systems of nonlinear equations is effectively achieved via the Newton approach

$$\mathbf{J}(P^k)\boldsymbol{\delta}^k = -[F(P^k)] \quad \ldots\ldots\ldots\ldots (15)$$

and

$$\mathbf{P}^{k+1} = \mathbf{P}^k + \boldsymbol{\delta}^k \quad \ldots\ldots\ldots\ldots (16)$$

where $[F(P^k)]$ represents the system of nonlinear equations being solved; $\mathbf{J}(P^k)$ = the matrix of partial derivative terms; $\mathbf{P}^k$ = the parameter vector, and $\boldsymbol{\delta}^k$ is the parameter step vector. The superscript k indicates the current iteration level.

Applying the generic parameter estimation methodology to a system of m parameters, the residual function can be expressed as

$$S_r(P_1,P_2,P_3,\ldots,P_m) = \sum_{i=1}^{n} (C_{\text{pred}_i} - C_{\text{obs}_i})^2 \quad \ldots\ldots\ldots\ldots (17)$$

where P_i represents the different parameters. From this, m normal equations can be derived

$$\mathbf{S}'_r = \begin{Bmatrix} \partial(S_r)/\partial(P_1) \\ \partial(S_r)/\partial(P_2) \\ \vdots \\ \vdots \\ \partial(S_r)/\partial(P_m) \end{Bmatrix} = \begin{Bmatrix} f_1(P_1,P_2,P_3, \ldots, P_m) \\ f_2(P_1,P_2,P_3, \ldots, P_m) \\ \vdots \\ \vdots \\ f_m(P_1,P_2,P_3, \ldots, P_m) \end{Bmatrix} = \mathbf{0} \quad \ldots\ldots\ldots (18)$$

It is then necessary to simultaneously solve this system of equations for the optimum parameter set. Combining (15), (16), and (18), the following formulation is derived

$$\mathbf{P}^{k+1} = \mathbf{P}^k + \mathbf{S}'^k_r(\mathbf{J}^k)^{-1} \quad \ldots\ldots\ldots (19)$$

where

$$\mathbf{J} = \begin{bmatrix} \partial^2 S_r/\partial P_1^2 & \partial^2 S_r/\partial P_1 \partial P_2 & \cdots & \partial^2 S_r/\partial P_1 \partial P_m \\ \partial^2 S_r/\partial P_2 \partial P_1 & \partial^2 S_r/\partial P_2^2 & \cdots & \partial^2 S_r/\partial P_2 \partial P_m \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ \partial^2 S_r/\partial P_m \partial P_1 & \partial^2 S_r/\partial P_m \partial P_2 & \cdots & \partial^2 S_r/\partial P_m^2 \end{bmatrix} \quad \ldots\ldots\ldots (20)$$

Eq. (20) shows that the Jacobian matrix $\mathbf{J}$, which holds the first derivatives of the normal equations (or the second derivatives of the residual function) is also the Hessian of the residual function. For any Newton iteration, therefore, it is necessary to evaluate the value of the residual function, and its first and second derivatives, at the current parameter set values.

Finite divided differences are again used in evaluating both the first- and second-derivative terms. Residual function values are obtained by executing the predictive model with parameter values perturbed from the current parameter set to provide information to generate the finite divided differences. For a two parameter model, the residual function needs to be evaluated with eight perturbed parameter sets, excluding the current set, at each Newton iteration. In general, for m parameters, $2m$ additional model evaluations are needed to compute the first derivatives of the residual function, and $2m(m - 1)$ further evaluations are necessary for the second derivatives. The total number of model runs required per iteration for a model with m parameters is given by X

$$X = 2m^2 + 1 \quad \ldots\ldots\ldots (21)$$

The Gauss-Newton (Chapra and Canale 1988; Dennis and Schnabel 1983) method is a simplification of the classical Newton technique. The Gauss-Newton analogy to (19) is

$$\mathbf{P}^{k+1} = \mathbf{P}^k + \mathbf{S}'^k_r(\mathbf{G}^k)^{-1} \quad \ldots\ldots\ldots (22)$$

The Jacobian matrix used in the full Newton method is replaced with a matrix **G**, which contains only the first derivative terms in (20)

$$\mathbf{G} = \begin{bmatrix} \sum_{i=1}^{n} (\partial C/\partial P_1)^2 & \sum_{i=1}^{n} \partial C/\partial P_1 \, \partial C/\partial P_2 & \cdots & \sum_{i=1}^{n} \partial C/\partial P_1 \, \partial C/\partial P_m \\ \sum_{i=1}^{n} \partial C/\partial P_2 \, \partial C/\partial P_1 & \sum_{i=1}^{n} (\partial C/\partial P_2)^2 & \cdots & \sum_{i=1}^{n} \partial C/\partial P_2 \, \partial C/\partial P_m \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ \sum_{i=1}^{n} \partial C/\partial P_m \, \partial C/\partial P_1 & \sum_{i=1}^{n} \partial C/\partial P_2 \, \partial C/\partial P_1 & \cdots & \sum_{i=1}^{n} (\partial C/\partial P_m)^2 \end{bmatrix} \quad \ldots (23)$$

and

$$\left.\frac{\partial C_{\text{pred}_i}}{\partial P_i}\right|_{\text{P}_i^k} \approx \frac{C_{\text{pred}_i}(P_i^k + \Delta P_i) - C_{\text{pred}_i}(P_i^k - \Delta P_i)}{2\Delta P_i} \quad \ldots (24)$$

The advantage of this method over that of Newton is that it requires fewer model evaluations per iteration

$$X = 2m + 1 \quad \ldots (25)$$

While the Newton method shows a quadratic increase in the required number of model runs per iteration, the computational effort increases only linearly in the Gauss-Newton method.

All the iterative techniques discussed herein operate by adding to an initial guess for the set of parameter values with a parameter step vector, based on some criteria for convergence. Differences between algorithms occur only in the evaluation of the parameter step vector. This inherent modularity permits easy substitution of the desired technique for solving the system of normal equations. It is equally uncomplicated to add modules to permit switching between solution techniques based on user-specified performance-based criteria. The parameter estimation algorithm used in this work included implementations of the independent search algorithm, the full Newton method, and a bounding algorithm to restrict parameter values within user specified limits.

Computer Experiments

To evaluate the performance of the parameter estimation algorithm with the different predictive models, a series of computer experiments was conducted on both a CRAY YMP/116 and a VAX 3100. An explicit finite difference (central) scheme was used to approximate the solution for both predictive models. A vertical settling column (Ducharme 1989) 197.5-cm tall was simulated with a fixed spatial grid and a dynamic time-step size to maintain solution stability. An analytical solution also was used for the dispersive transport model, but no perceptible difference was found in algorithm performance.

Box and Coutie (1956) proposed an approach to evaluating model behavior by relying on model-generated data to conduct sensitivity analyses on both model parameters and data reliability. A similar approach was taken

to evaluate the sensitivity of the coupled parameter estimation algorithm and predictive models to simulated problem data sets.

In the context of this paper, the estimated parameter set is the set of parameters that produces the minimum residual function value for the given data set. If the data set used is model generated, then the estimated parameter set should be identical to the parameter set (herein referred to as *true* or *modeled*) used to generate that data, unless the noncontinuous nature (sparcity) of the data has an effect on the estimation results.

The location of the true or modeled parameter set in the parameter domain will necessarily have an effect on the predictive model behavior. In the case of nonlinear terms, the effect is most likely to be in the stiffness of the solution surface. To characterize the resultant effect on the parameter estimation algorithm performance, the predictive model was used to generate several data sets over a range in the parameter sets that is considered environmentally significant. Residual function values were then determined in the vicinity of the true parameter set to provide a residual function surface. The gradient of the residual function surface at any prescribed parameter set is an indication of the driving force toward the true parameter set.

Sampling or measurement error was simulated by randomly altering the concentrations predicted by the advective-dispersive transport model over a range in parameter sets that is considered environmentally significant. The maximum amplitudes ($\pm 1\%$, $\pm 5\%$, $\pm 10\%$, $\pm 25\%$, and $\pm 40\%$) of the random noise applied to the uncorrupted data were selected as a percentage of the uniform initial concentration (C_{eq}). For example, a random noise level of $\pm 10\%$ would force the corrupted data $[\tilde{C}(z,t)]$ to fall within the limits imposed by (26)

$$\tilde{C}(z,t) = C(z,t) \pm 0.1 \times C_{eq} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (26)$$

Data sparcity was simulated by prescribing a data-sparcity level between 0% and 25%, where 0% would indicate that all the uncorrupted data were used, and 100% signifying that none were used. Each data point in the uncorrupted data set was evaluated in turn, by generating a random number between 0 and 1 and comparing it with the prescribed data-sparcity level. If the generated random number was greater than the data-sparcity fraction, the corresponding data point was used in the estimation algorithm. Otherwise, it was discarded. The uncorrupted data sets contained a total of 500 data points in each case, so a sparcity level of 25% would result in the random elimination of approximately 125 data points from the initially regular grid of observations. This may be interpreted as the fraction of data that either was not collected or had to be rejected from a regularly gridded field or laboratory sampling scheme.

It is evident that this method for characterizing data sparcity is heavily dependent on the sampling frequency and nature of the sampling grid for the uncorrupted data set. A more meaningful parameter is the resulting mean sampling frequency of the modified data set. In the case of a predictive model with one time and one space scale, this would be the average number of data points per unit time per unit length.

Results and Comments

Dispersive Transport

The residual function surface for the dispersive transport model [(1)] was evaluated at five different values of the modeled D_z, ranging over five orders

of magnitude. Fig. 1(a) shows the residual function approach zero at its corresponding true D_z. The convergence characteristics of the model at the specified D_z values (0.01,0.1,1.0,10.0, and 100.0 cm²/s) were then evaluated. This was accomplished by using the dispersive transport model to generate data at the specified D_z values. The parameter estimation technique [(14)] was then used to estimate the corresponding true D_z from initialized values of 10%, 50%, 75%, and 150% of the true parameter.

Fig. 1(b) shows the correlation between the estimated D_z and its corresponding modeled value. For the boundary and initial conditions specified in (2) and (3), large values of the D_z reduce the time to steady state for systems being modeled by (1). The effect of the increasing system dynamics

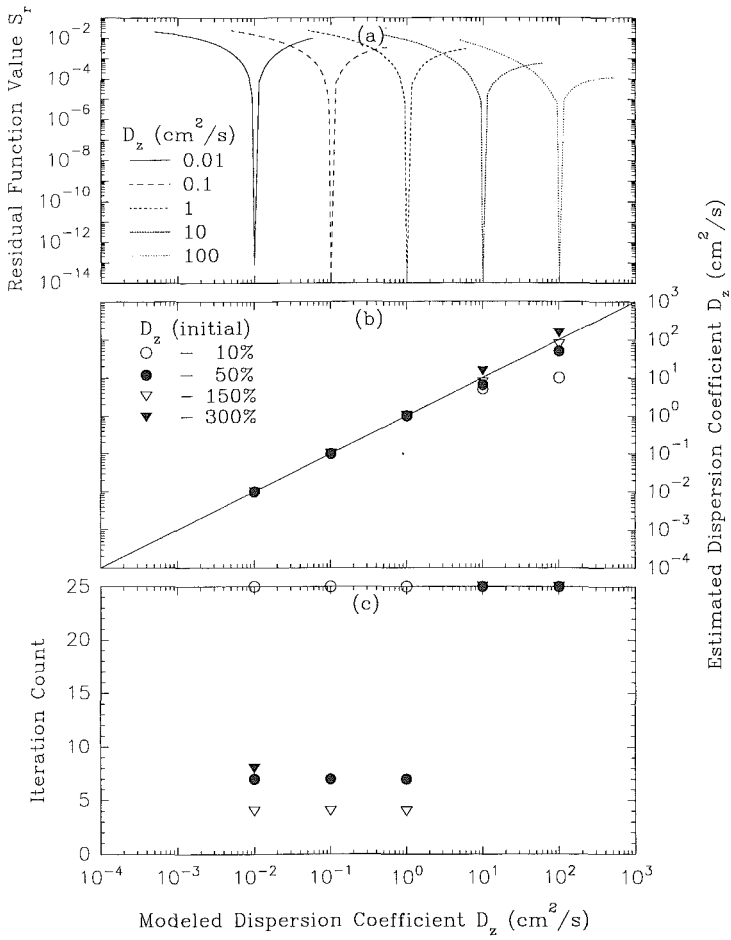

FIG. 1. Dispersive Transport Model: (a) Residual Function Value as Function of Dispersion Coefficient for Various Modeled Dispersion Coefficients; (b) Correlation between Model-Generated Data Set Parameters and Parameter Estimation Results for Various Initial Guesses; (c) Final Iteration Count for Estimated Parameter

to evaluate the sensitivity of the coupled parameter estimation algorithm and predictive models to simulated problem data sets.

In the context of this paper, the estimated parameter set is the set of parameters that produces the minimum residual function value for the given data set. If the data set used is model generated, then the estimated parameter set should be identical to the parameter set (herein referred to as *true* or *modeled*) used to generate that data, unless the noncontinuous nature (sparcity) of the data has an effect on the estimation results.

The location of the true or modeled parameter set in the parameter domain will necessarily have an effect on the predictive model behavior. In the case of nonlinear terms, the effect is most likely to be in the stiffness of the solution surface. To characterize the resultant effect on the parameter estimation algorithm performance, the predictive model was used to generate several data sets over a range in the parameter sets that is considered environmentally significant. Residual function values were then determined in the vicinity of the true parameter set to provide a residual function surface. The gradient of the residual function surface at any prescribed parameter set is an indication of the driving force toward the true parameter set.

Sampling or measurement error was simulated by randomly altering the concentrations predicted by the advective-dispersive transport model over a range in parameter sets that is considered environmentally significant. The maximum amplitudes ($\pm 1\%$, $\pm 5\%$, $\pm 10\%$, $\pm 25\%$, and $\pm 40\%$) of the random noise applied to the uncorrupted data were selected as a percentage of the uniform initial concentration (C_{eq}). For example, a random noise level of $\pm 10\%$ would force the corrupted data $[\tilde{C}(z,t)]$ to fall within the limits imposed by (26)

$$\tilde{C}(z,t) = C(z,t) \pm 0.1 \times C_{eq} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (26)$$

Data sparcity was simulated by prescribing a data-sparcity level between 0% and 25%, where 0% would indicate that all the uncorrupted data were used, and 100% signifying that none were used. Each data point in the uncorrupted data set was evaluated in turn, by generating a random number between 0 and 1 and comparing it with the prescribed data-sparcity level. If the generated random number was greater than the data-sparcity fraction, the corresponding data point was used in the estimation algorithm. Otherwise, it was discarded. The uncorrupted data sets contained a total of 500 data points in each case, so a sparcity level of 25% would result in the random elimination of approximately 125 data points from the initially regular grid of observations. This may be interpreted as the fraction of data that either was not collected or had to be rejected from a regularly gridded field or laboratory sampling scheme.

It is evident that this method for characterizing data sparcity is heavily dependent on the sampling frequency and nature of the sampling grid for the uncorrupted data set. A more meaningful parameter is the resulting mean sampling frequency of the modified data set. In the case of a predictive model with one time and one space scale, this would be the average number of data points per unit time per unit length.

Results and Comments

Dispersive Transport

The residual function surface for the dispersive transport model [(1)] was evaluated at five different values of the modeled D_z, ranging over five orders

of magnitude. Fig. 1(a) shows the residual function approach zero at its corresponding true D_z. The convergence characteristics of the model at the specified D_z values (0.01,0.1,1.0,10.0, and 100.0 cm²/s) were then evaluated. This was accomplished by using the dispersive transport model to generate data at the specified D_z values. The parameter estimation technique [(14)] was then used to estimate the corresponding true D_z from initialized values of 10%, 50%, 75%, and 150% of the true parameter.

Fig. 1(b) shows the correlation between the estimated D_z and its corresponding modeled value. For the boundary and initial conditions specified in (2) and (3), large values of the D_z reduce the time to steady state for systems being modeled by (1). The effect of the increasing system dynamics

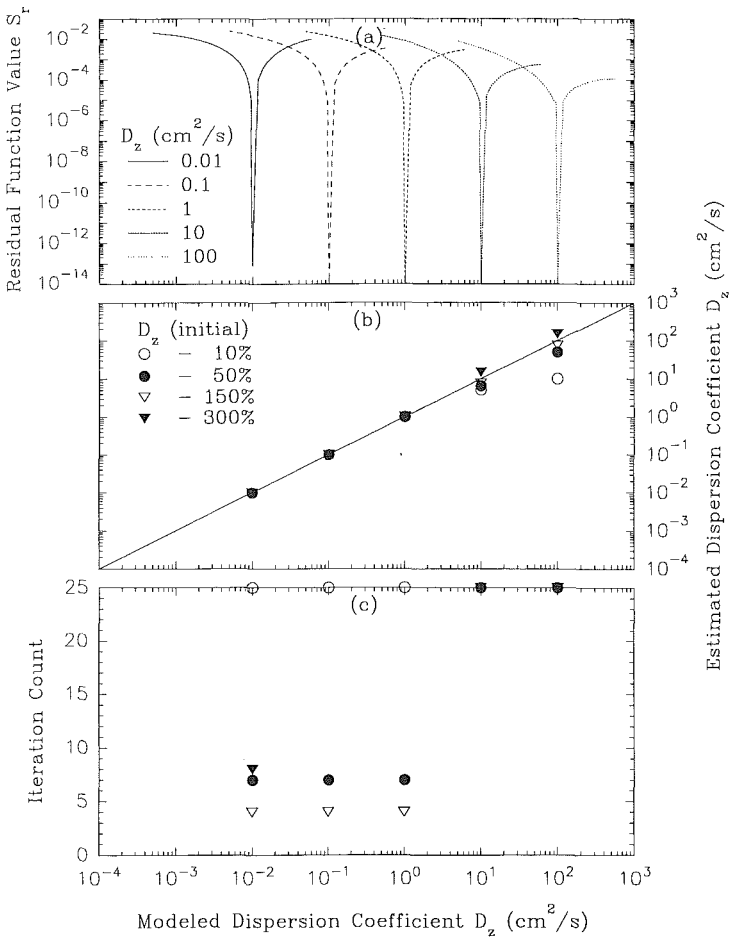

FIG. 1. Dispersive Transport Model: (a) Residual Function Value as Function of Dispersion Coefficient for Various Modeled Dispersion Coefficients; (b) Correlation between Model-Generated Data Set Parameters and Parameter Estimation Results for Various Initial Guesses; (c) Final Iteration Count for Estimated Parameter

can be seen in Figs. 1(b) and (c), where high values of D_z can be linked to poor correlation between the estimated and modeled D_zs [Fig. 1(b)], and an increase in the number of iterations required to achieve the tolerance level [Fig. 1(c)].

Further, as the modeled D_z increases in magnitude, the residual surface becomes less dynamic. It should be noted that both the X- and Y-axes in Fig. 1(a) are scaled logarithmically, indicating that while the shape of the curves are similar over the different modeled D_zs, the derivatives of the residual functions with respect to D_z will vary logarithmically. This indicates that rates of convergence should decrease dramatically for increasingly dispersive systems. Fig. 2 shows the estimated D_z and the corresponding residual function value at each iteration for each of the five modeled parameter

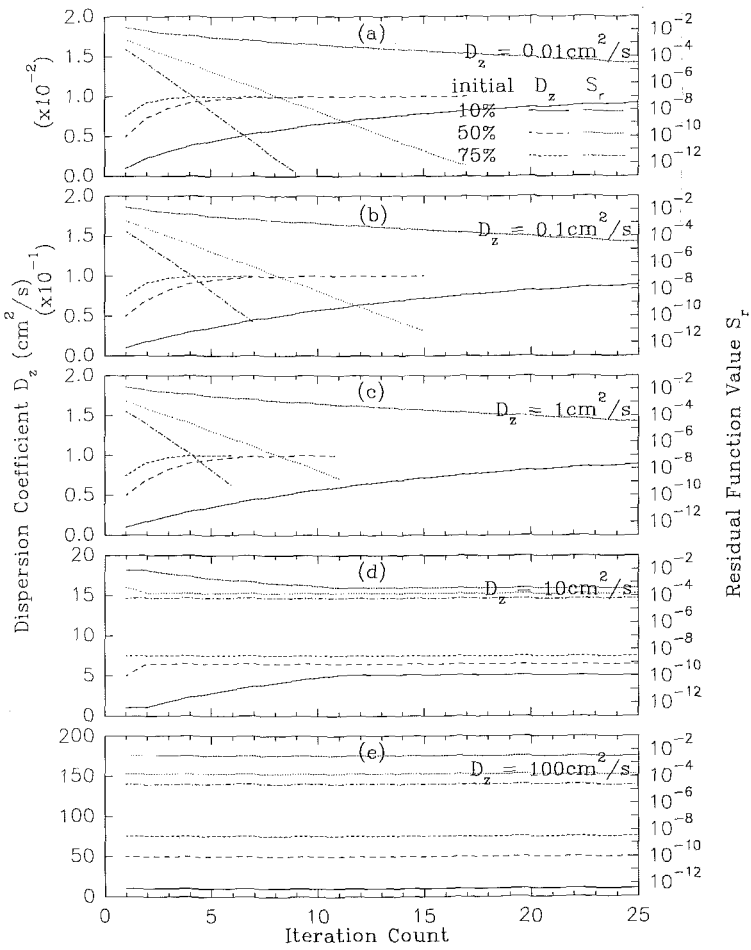

FIG. 2. Dispersive Transport Model. Rate of Convergence of Algorithm at Various Initial Guesses. Dispersion Coefficient: (a) 0.01; (b) 0.1; (c) 1; (d) 10; (e) 100 cm²/s

values. For modeled D_zs of 10 cm^2/s and more [Figs. 2(d) and (e)], the parameter estimation algorithm failed to reach the true D_z for all of the specified initial guesses. In fact, the normal equations are so nondynamic for these cases that virtually no change in either estimated D_z or residual function value can be perceived. Fig. 2(e) highlights this, in that the algorithm has reached a threshold at which successive iterations produce no visible change in D_z or the residual function value.

Convergence rates increased when the initial estimates given were increased from 10% to 75% of the true parameter. However, initial estimates of 150% resulted in algorithm divergence at all values of the modeled D_z, except the lowest (0.01 cm^2/s). Here, the parameter estimates were seen to undergo a damped oscillation around the true parameter before converging. The slope of the residual function surface is smaller at D_z values greater than the true D_z, than it is for smaller values [Fig. 1(a)]. This nonsymmetry can be seen to increase for increasing values of the modeled D_z. This is to be expected, insofar as increasing dispersion levels tend to damp out the observed with the predicted component in the residual function [(11)]. It should be noted that decreasing distance between the boundaries will have a similar effect on system dynamics, and therefore, on convergence.

For a D_z of zero, the predictive model will be at equilibrium at the initial condition for all time. As the estimated D_z decreases below the true D_z and approaches zero [Fig. 1(a)], the slope of each curve becomes increasingly shallow. This causes convergence to be much slower for initial guesses closer to zero and can be seen in Figs. 2(a), (b), and (c). For a zero slope, the algorithm will remain static at the corresponding parameter estimate.

Nonconvergence at higher dispersion values may be attributed to a combination of extreme solution stiffness of highly dispersive systems and the use of inadequate sampling schemes. The problem is exacerbated by the use of numerical approximations in the predictive model solution and the iterative estimation procedure. Under highly dispersive conditions, completely mixed conditions will be achieved almost instantaneously. As a result, the sampled data set may be excessively weighted with completely mixed concentration values, making the normal equation [(11)] virtually invariant in the vicinity of the true parameter value. The greater the amount of weight on steady-state conditions, the greater the region of damping in the normal equation.

Advective-Dispersive Transport

The model defined by (4), (5), and (6) was solved using a finite difference procedure. The time scale was set at the time required for traversal of the spatial domain due to the advective component alone, and discretized into 50 equal steps. The spatial domain was discretized into 10 elements. This resulted in a regular grid of 500 elements, at which model-generated observations could be provided for any combination of specified D_z and v_z.

Residual Function Behavior

To understand the relative importance of each parameter as a driving force toward a residual function minimum, the value of the residual function was evaluated over a range of D_z and v_z. The residual is a function of both the predictive model parameters (D_z and v_z) and the observed data. Each plot in Fig. 3 corresponds to a different observed data set, derived from predictive model runs at the stated values of D_z and v_z. For the ideal data

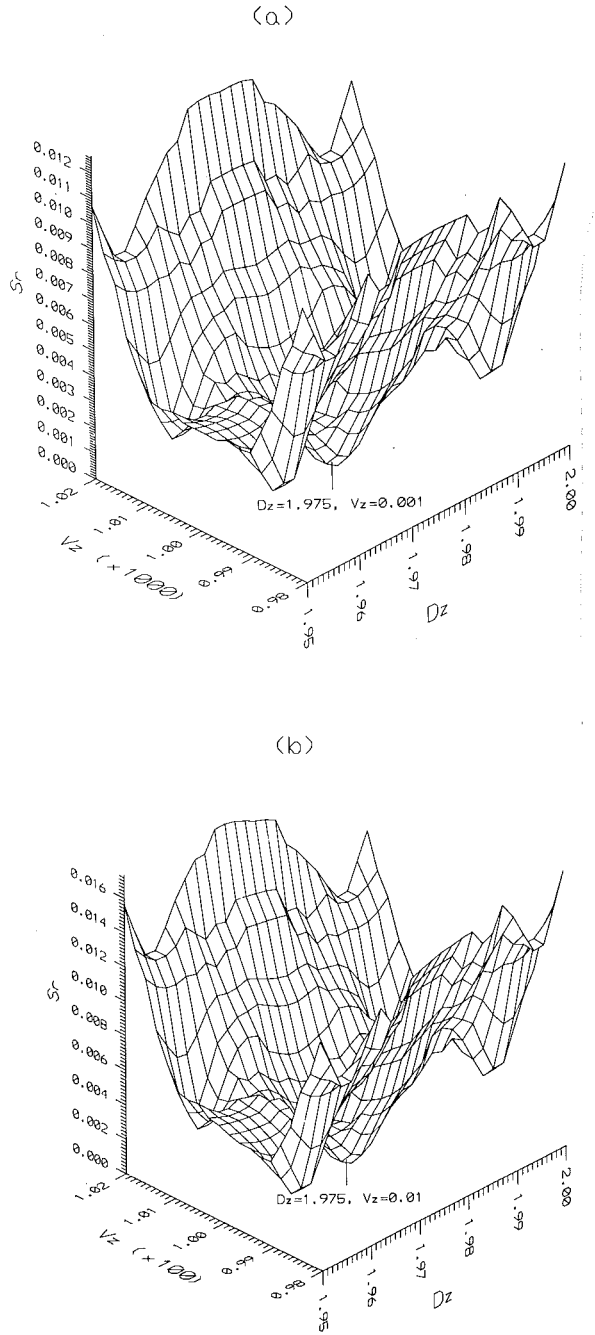

FIG. 3. Residual values as Function of Dispersion Coefficient and Velocity at Modeled Peclet Numbers of: (a) 0.1; (b) 1.0; (c) 10; (d) 100

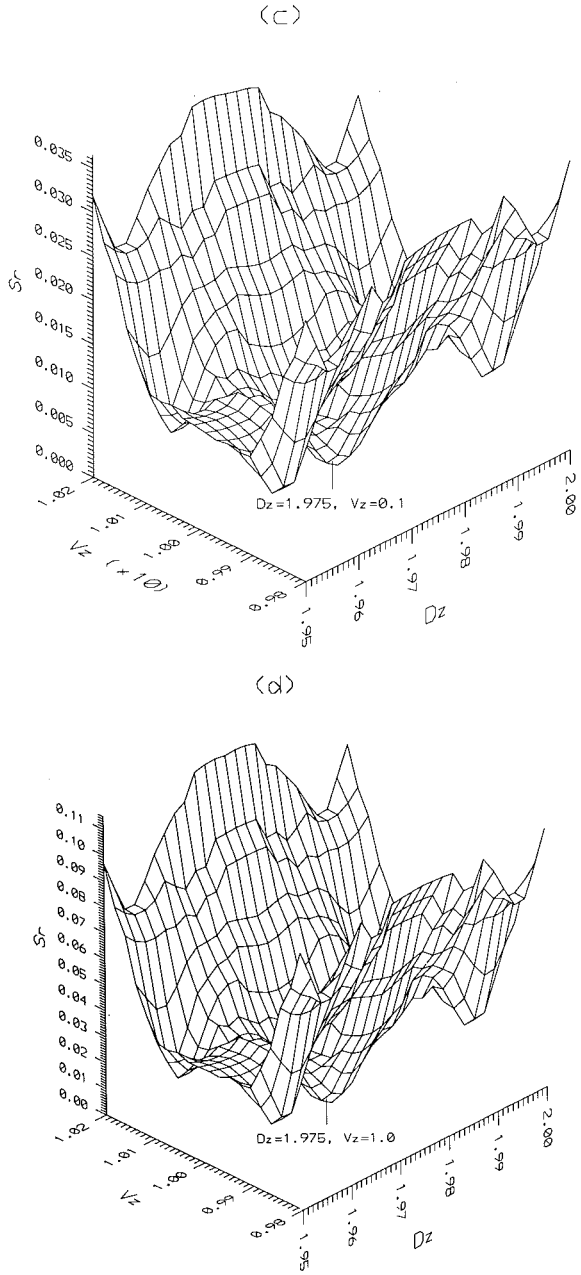

FIG. 3 (*Continued*)

sets presented, the residual function minimum should be zero at the stated values of D_z and v_z.

It can be seen in all plots that the true minimum is only a small depression in a valley running across the v_z axis. This valley indicates that the estimation procedure should converge close to the correct v_z, even if the estimated D_z is currently far from the correct value. This results in an observed inherent stability of the procedure with respect to v_z. In addition to the similarity in the plots, it can be seen that several local minima exist in the vicinity of the true solution.

The degree of exaggeration applied to the v_z axis decreases with increasing advective dominance (higher Peclet-number values) indicating that as the system becomes more and more advective dominant, the sides of the valley become more shallow. This decrease in slope, which is the driving force for the Newton-Raphson process, leads to a decrease in rate of convergence in the v_z direction. At Peclet numbers greater than 100, it can be seen that the value of the residual function may be almost invariant with changes in either parameter, making the system slowly convergent at best and highly unstable at worst. In dispersive dominant systems, the excessive slope of the valley walls appeared to cause the estimated v_z to oscillate about the correct solution, permitting D_z to converge more rapidly than otherwise.

Random-Noise Effect

Fig. 4 shows the rate at which the parameters D_z and v_z approach their optimum value under various random-noise levels and model-generated data sets. The tendency of the algorithm to become unstable and to oscillate at higher Peclet (Pe) numbers can be seen in Fig. 4(e). At relatively low random-noise levels ($< \pm 5\%$), the effect is negligible, especially at the lower Peclet numbers (<75). The general nature of convergence of the D_z and v_z with respect to each other is given in Fig. 4(e). Here, it is apparent that the direction taken by the estimation algorithm is dominated by the effect of the D_z term, while, at the same time, it is entirely possible to converge on a v_z value and either wildly oscillate around the optimum dispersion coefficient or diverge from it completely.

The ability of the procedure to converge to the correct parameter values is dependent on both the dynamics of the system and the amount of noise in the data. Fig. 5(a) shows the accuracy of the algorithm over a wide range of these two factors. The modeled Peclet number represents the parameter set used to generate the data sets, while the estimated parameters result from the application of the parameter estimation procedure on the data set. It should be noted that while the Peclet number is used here as a single representative parameter of the system, it is only indicative of the ratio of the advective to dispersive fluxes, it does not embody the system's dynamics. It is to be expected that as the overall magnitude of both flux terms increase, the dynamics of the system will increase.

The Peclet numbers shown in Fig. 5(a) incorporate a wide range of the respective fluxes. The parameter estimation procedure proved to be tolerant to a large range in system dynamics. However, at high Peclet numbers, at which the system is highly advective dominant, it is apparent that the ability of the method to accurately estimate both D_z and v_z becomes limited. This, however, is because of an inherent problem of excessive numerical dispersion associated with predictive numerical models of advective dominant systems. This numerical damping effect is exacerbated by the differencing strategy of the estimation procedure.

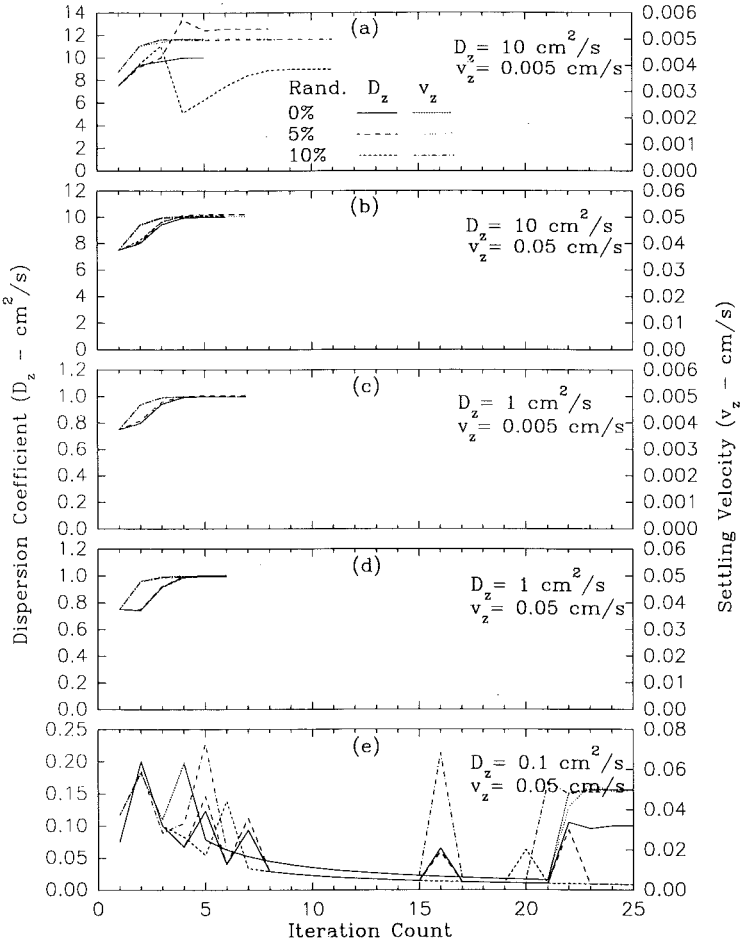

FIG. 4. Rate of Convergence of Algorithm at Various Random-Noise Levels. Modeled Peclet Numbers: (a) 0.1; (b) 1; (c) 1; (d) 10; (e) 100

Even at extremely high random-noise levels ($\pm 25\%$), the working range of the estimation procedure extends between Peclet numbers of 0.1 and 100, sufficient for most environmental problems. Both the performance and working Peclet range of the algorithm improve as the noise level decreases. The random-noise levels depicted involve noise on both sides of the correct modeled concentration levels and, as such, are representative of a random-noise band of twice the width of the value portrayed.

Comparison of Figs. 5(a) and 5(b) shows that convergence is reached within 10 iterations, when under optimum conditions, system Peclet numbers in the range of 10, and random-noise levels less than 25%. However, as the system Peclet number tends away from this range, toward 100 or less than 1, the required number of iterations increases. The further the system Peclet number outside this range, the larger the number of iterations re-

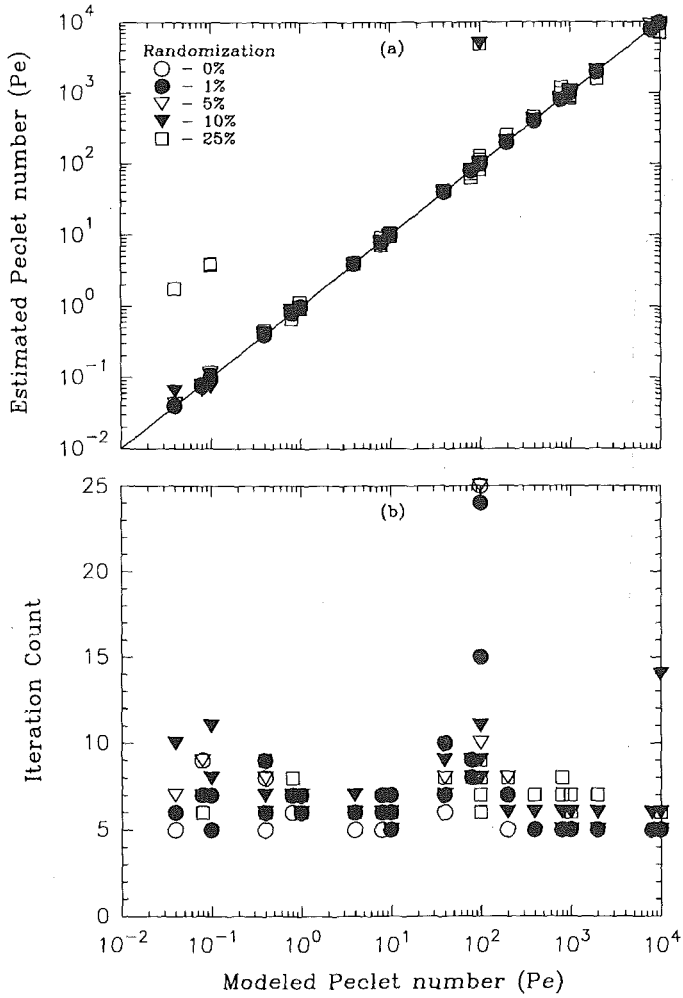

FIG. 5. Random-Noise Effect: (a) Correlation between Model-Generated Data Set Parameters and Parameter Estimation Results at Various Random-Noise Levels; (b) Final Iteration Count for Estimated Parameter

quired. It is evident that if the estimation procedure takes more than 15 iterations, it is more than likely diverging from the correct solution. However, an exception to this can be seen at the lower system Peclet numbers, of about 0.1. Here, after 30 iterations, the true parameter set is still being approached. This slow convergence can be attributed to the ±25% random-noise level. A similar phenomenon can be seen at a Peclet number of 100 and a random noise level of ±10%. However, it is important to note that, in these cases, the range of application can be extended by increasing the computational effort.

At Peclet numbers less the 0.01 and random-noise levels of ±25%, the

estimation procedure converges on a parameter set with a Peclet number of more than an order of magnitude greater than that of the true parameter set. This is the result of a combination of the high random-noise level and unbalanced weighting of data. At this low Peclet-number value, dispersion is the predominant transport mechanism. However, the time scale of the data used was obtained using the advective flux only. It is evident that this estimated time scale will exceed that of the actual problem. In this case, the dynamics of the dispersive system is accounted for by fewer data points than is desirable, and steady-state uniform concentrations predominate. The resulting spread in the data will accommodate a greater degree of advection than dispersion.

It should be noted that while the estimation algorithm has been tested rigorously over a wide range of conditions, no attempt has been made to provide a theoretical evaluation of its convergence criteria. A rigorous study of convergence criteria for parabolic partial differential equations can be found in Kunisch and White (1985).

Data-Sparcity Effect

The effect of data sparcity was evaluated in a similar fashion as random noise. Fig. 6 shows the rate of convergence of the estimation procedure under various levels of data sparcity. It can be seen that the performance of the estimation algorithm is affected more by sparse data than random noise at lower Peclet numbers. This is to be expected. At lower Peclet numbers, the unbalanced weighting of steady-state conditions will be exacerbated by decreasing the sampling frequency.

Two specific trends can be seen in Fig. 6 with regard to algorithm performance. First, increasing data sparcity tends to force the algorithm to estimate advective components that are higher than the true values. The same effect can be seen in the dispersive component in addition to making the convergence unstable. This can be related to the shape of the residual function surface (Fig. 3) in the direction of D_z.

Limits for convergence are clearly defined in Fig. 7. Under the conditions used to generate the data sets (time scales and spatial and temporal sampling frequency) convergence to a reasonable estimate of the true Peclet number can be expected for systems with Peclet numbers between 0.1 and 100.0 and at 375 observations (25% sparcity imposed on a 500-element data set). Algorithm performance may be improved by implementing a better sampling scheme. For example, the range of Peclet numbers over which convergence is ensured can be widened by increasing the sampling resolution when the system is more dynamic, and relaxing the resolution at other times. During the evaluation of the algorithm it was noted that considerably fewer observations were required for convergence if an accurate initial estimate of D_z was provided. It was found in general that the algorithm performance was more susceptible to small variations in D_z than is v_z.

It should be noted that while the algorithm may not converge for gross sample errors or data scarcity, under such conditions it is doubtful that manual model calibration would be feasible or practical. Further, the autocalibration provided by the parameter estimation algorithm can be used to determine data requirements.

Conclusions

A generic parameter estimation framework was developed and tested on model-generated data for an indication of the limits of applicability. A one-

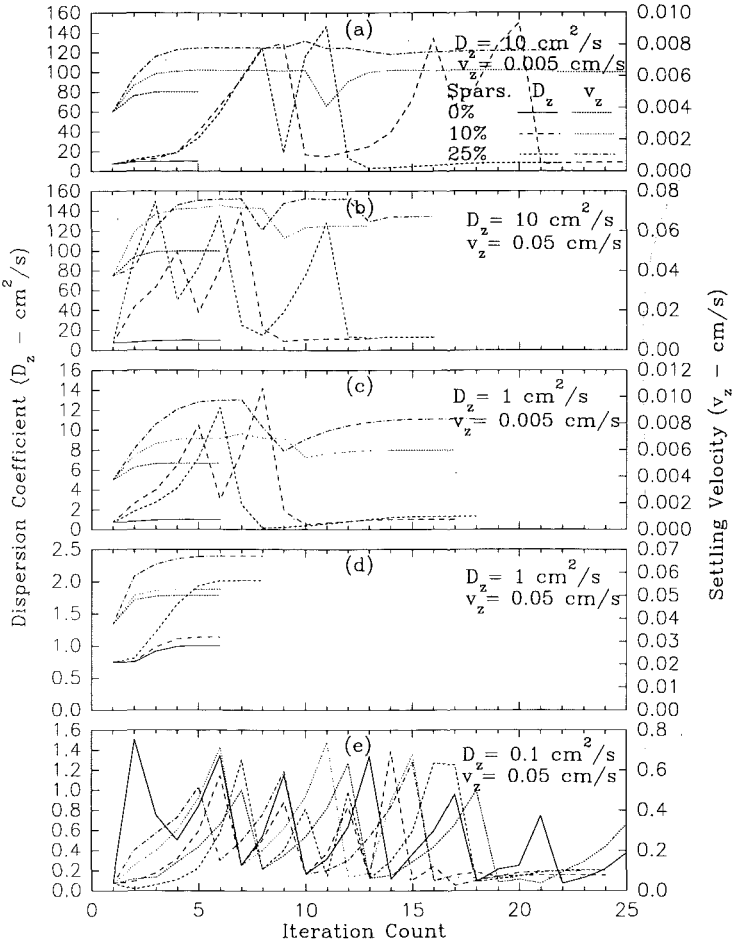

FIG. 6. Rate of Convergence of Algorithm at Various Data-Sparcity Levels. Modeled Peclet Numbers: (a) 0.1; (b) 1; (c) 1; (d) 10; (e) 100

parameter dispersive transport model and a two-parameter advective-dispersive transport model were used. The effect of sampling and measurement error on algorithm performance were evaluated. The following conclusions may be drawn from this study.

- The parameter estimation procedure was used successfully to parameterize a single-parameter dispersive transport model with a D_z range of 0.01 cm^2/s–100.0 cm^2/s.
- The parameter estimation procedure was used successfully to simultaneously extract D_z and v_z values from an advective-dispersive particle transport model with a Peclet-number range of 0.1–1,000.0.
- The rate, and whether convergence will occur, is strongly dependent on the slope of the residual function surface at the location of the initial

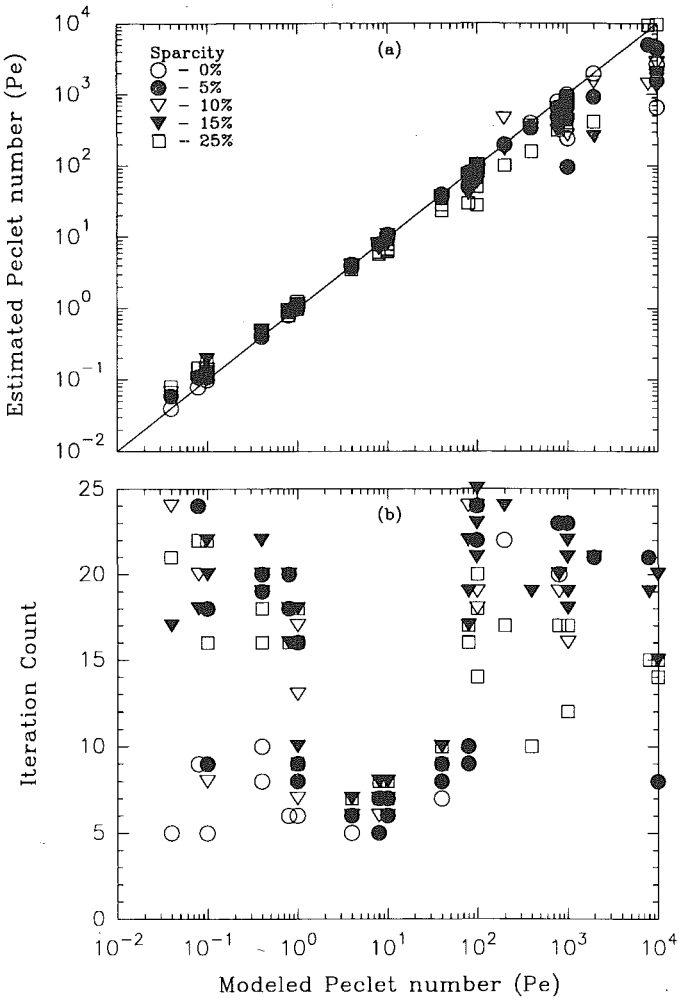

FIG. 7. Data-Sparcity Effect: (a) Correlation between Model-Generated Data Set Parameters and Parameter Estimation Results at Various Data-Sparcity Levels; (b) Final Iteration Count for Estimated Parameter

parameter estimates. The possibility of convergence improves if the initial parameter estimate is made on the side of the residual surface from the true solution that has the steepest slope.

- The rate of convergence and the accuracy of the resulting estimated parameters degrades with increasing random noise and sparcity in the observed data.
- Convergence is rapid within a Peclet range of 1–100 and random-noise level of less than 25%.
- The spatial and temporal location of observations may affect both convergence rates and the estimated parameter values, and must be co-

ordinated with stiff areas in the solution domain to ensure accurate parameter estimates.
- The technique can be used to determine predictive model inadequacies and data requirements (quantity, quality, and location) for model calibration.

The parameter estimation algorithm was found to be an effective tool in model calibration. The algorithm converged under conditions known to confound standard model-calibration methods, although algorithm performance was degraded.

Acknowledgments

This project was supported by the U.S. Environmental Protection Agency under cooperative agreement CR-814257-01, entitled "Study of Processes Affecting Particle-Mediated Transport of Hazardous Materials in Marine Systems." The contents do not necessarily reflect the views and policies of the U.S. EPA, nor does the mention of trade names of commercial products constitute endorsement or recommendation for use. We would like to thank our project officers J. Paul and Jerry Pesch, of the U.S. EPA Environmental Research Laboratory, Narragansett, Rhode Island, for technical support, and Bill Batchelor for his constructive comments.

Appendix I. References

Autenrieth, R. L. (1986). "Sorption and desorption partition characterization of selected organochlorine compounds with phytoplankton," thesis presented to Clarkson University, at Potsdam, New York, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Bard, Y. (1974). *Non-linear parameter estimation*. Academic Press, New York, N.Y.

Bonner, J. S., Hunt, C. D., Paul, J. F., and Biermam, V. J. (1986). "Prediction of vertical transport of low level radioactive middlesex soil at a deep ocean disposal site." *Tech. Report EPA 520/1-86-016*, Environmental Research Laboratory, U.S. Environmental Protection Agency, Narragansett, R.I.

Bonner, J. S., Ernest, A. N., Autenrieth, R. L., and Ducharme, S. L. (in press 1990). "Parameterizing models for contaminated sediment transport." *Transport and transformation of contaminants near the sediment-water interface*. J. V. DePinto, ed., Springer-Verlag, Berlin, Germany.

Box, G. E. P., and Coutie, G. A. (1956). "Application of digital computers in the exploration of functional relationships." Proc., Institute of Electrical Engineers, London, England, 103B (Suppl. 1) 100.

Carpenter, B. (1987). "Pcb sediment decontamination: Technical/economical assessment of selected alternative treatments." *Tech. Report EPA/600/S2-86/112*, U.S. Environmental Protection Agency, Washington, D.C.

Chapra, S. C., and Canale, R. P. (1988). *Numerical methods for engineers*. Second Ed., McGraw-Hill, New York, N.Y.

Dennis Jr., J., and Schnabel, R. (1983). *Numerical methods for unconstrained optimization and non-linear equations*. Prentice Hall, Englewood Cliffs, N.J.

Draper, N., and Smith, H. (1981). *Applied regression analysis*. John Wiley & Sons, New York, N.Y.

Ducharme, S. L. (1989). "Design and validation of a settling column for particle transport studies," thesis presented to Texas A&M University, at College Station, Texas, in partial fulfillment of the requirements for the degree of Master of Science.

Elzerman, A., and Coates, J. (1987). "Hydrophobic organic compounds on sediments: Equilibria and kinetics of sorption." *Envir. Sci. Tech.*, 8(1), 58–63.

Gawthrop, P. J. (1984). "Parameter estimation from noncontiguous data." *Proc.*, Institute of Electrical Engineers, London, England, 131(6), 261–266.
Hunt, J. R., and Pandya, J. (1984). "Sewage sludge coagulation and settling in seawater." *Envir. Sci. Tech.*, 18(2), 119–121.
Kunisch, K., and White, L. (1985). The parameter estimation problem for parabolic equations and discontinuous observation operators. *SIAM J. Control Optim.*, 23(6), 900–927.
Lee, D.-Y., Lick, W., and Kang, S. (1981). "The entrainment and deposition of fine-grained sediments in Lake Erie." *J. Great Lakes Res.*, 7(3), 224–233.
Lick, W., and Kang, S. (1987). "Entrainment of sediments and dredged materials in shallow lake waters." *J. Great Lakes Res.*, 13(4), 619–627.
Lick, W. (1982). "Entrainment, deposition and transport of fine grain sediments in lakes." *Hydrobiologia*, 91(1), 31–40.
McCreary, E. R. (1990). "Biodegradation of hazardous wastes using activated sludge microbes," thesis presented to Texas A&M University, at College Station, Texas, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.
O'Melia, C. R. (1972). "Coagulation and flocculation." *Physico-chemical processes for water quality control.* Wiley Interscience, New York, N.Y.
Rodgers, P. (1983). "Model simulation of pcb dynamics in Lake Michigan." *Physical behavior of PCB's in the Great Lakes*, D. Mackay, S. Paterson, S. Eisenich, and M. Simmons, eds., 311–328.
Rorschach, R. C. (1990). "Desorption of hexachlorobiphenol from selected particulate matters," thesis presented to Texas A&M University, at College Station, Texas, in partial fulfillment of the requirements for the degree of Master of Science.
Russel, W. (1981). "Brownian motion of the small particles suspended in liquids." *Ann. Rev. Fluid Mech.*, 13, 425–455.
Sanders, S. C. (1990). "Vertical transport and dynamic size distribution of new bedford harbor sediments," thesis presented to Texas A&M University, at College Station, Texas, in partial fulfillment of the requirements for the degree of Master of Science.
Srinivasan, V., and Aiken, R. C. (1986). "Stage-wise parameter estimation for stiff differential equations." *AIChE J.*, 32(2), 195–199.
Stefanski, L. A. (1985). "The effects of measurement error on parameter estimation." *Biometrika*, 72(3), 583–592.

APPENDIX II. NOTATION

The following symbols are used in this paper:

C = mass concentration;
$\tilde{C}$ = mass concentration corrupted by introducing random noise;
C_{eq} = uniformly mixed concentration of the transported species;
C_{obs} = functional form describing observed concentrations;
C_{obs_i} = mass concentration at observation i;
C_{pred} = functional form describing predictive mathematical model;
C_{pred_i} = mass concentration, corresponding to observation i, predicted by mathematical model;
D_z = dispersion coefficient;
D_{z_i} = effective dispersion coefficient for particle category i;
D_z^k = estimated dispersion coefficient at the kth iteration;
$\mathbf{F}$ = vector of normal equations;
f_i = normal equation for the ith parameter;
$\mathbf{G}$ = matrix of 1st derivatives for Gauss-Newton technique;
h = distance between boundaries;
$\mathbf{J}$ = Jacobian matrix;

m = number of parameters to be estimated;
n = total number of derivatives;
Pe = Peclet number;
P_i = parameter i;
$\mathbf{P}^k$ = vector of parameter values at iteration k;
S_r = residual function;
$\mathbf{S}'_r$ = vector of derivatives of residual function with respect to each parameter (normal equations);
$\mathbf{S}'^k_r$ = vector of derivatives of residual function with respect to each parameter (normal equations) at iteration i;
t = time;
t_i = time corresponding to observation i;
v_z = settling velocity;
X = number of predictive model evaluations required per parameter estimation iteration;
z = distance;
z_i = distance corresponding to observation i;
ΔD_z = finite difference step for evaluating all derivatives with respect to dispersion coefficient;
ΔP_i = finite difference step for evaluating all derivatives with respect to parameter i;
δ = impulse function;
$\boldsymbol{\delta}^k$ = vector of values to be added to parameter vector at kth iteration to obtain better parameter estimates;
μ = dynamic viscosity; and
$\mathbf{0}$ = $n \times 1$ zero vector.