Stochastic Environmental Resea (2004) 18: 67 – 78
DOI 10.1007/s00477-003-0130-z
© Springer-Verlag 2004

ORIGINAL PAPER

C. V. Chrysikopoulos · E. T. Vogler

Estimation of time dependent virus inactivation rates by geostatistical and resampling techniques: application to virus transport in porous media

Abstract A methodology is developed for estimating temporally variable virus inactivation rate coefficients from experimental virus inactivation data. The methodology consists of a technique for slope estimation of normalized virus inactivation data in conjunction with a resampling parameter estimation procedure. The slope estimation technique is based on a relatively flexible geostatistical method known as universal kriging. Drift coefficients are obtained by nonlinear fitting of bootstrap samples and the corresponding confidence intervals are obtained by bootstrap percentiles. The proposed methodology yields more accurate time dependent virus inactivation rate coefficients than those estimated by fitting virus inactivation data to a first-order inactivation model. The methodology is successfully applied to a set of poliovirus batch inactivation data. Furthermore, the importance of accurate inactivation rate coefficient determination on virus transport in water saturated porous media is demonstrated with model simulations.

List of symbols

$\mathbf{A}$	matrix $(n+p \times n+p)$ of the kriging system, defined in (18)
$\mathbf{b}$	vector of true model parameters
$\hat{\mathbf{b}}$	bootstrap vector of estimated model parameters
$\mathbf{b}_\beta$	vector with the mean values of multiple bootstrap parameter estimates, defined in (22)
B	number of resamples
C	concentration of viruses suspended in the liquid phase, M/L^3
C^*	sorbed virus concentration (virus mass/solid mass), M/M
C_g	concentration of virus directly in contact with solids, M/L^3
$C_\circ$	initial virus concentration, M/L^3
$\mathbf{d}$	vector $(n+p)$ of the kriging system, defined in (20)
D	hydrodynamic dispersion coefficient, L^2/t
$\mathbf{e}$	vector of random numbers with zero mean and known covariance matrix
$E[\,]$	expectation operator
f	known trial or base time-dependent functions
$\mathbf{g}$	vector of model simulated data
L	Lagrangian, defined in (14)
k	mass transfer rate constant, t^{-1}
K_d	partition coefficient, L^2/M
M	normalized virus log-concentration
$\bar{M}$	mean normalized virus log-concentration
MSE	mean square error of kriging estimator, defined in (21)
n	number of available observations or experimental data points
p	number of drift coefficients
R	covariance function
S	objective function
t	time, t
$\mathbf{u}$	vector of independent variables
U	average interstitial velocity, L/t
x	spatial coordinate in the direction of flow, L
$\mathbf{x}$	vector $(n+p)$ of unknowns of the kriging system, defined in (19)
y_i	observed data
$\mathbf{y}$	vector of observed data

Greek letters

α	resistivity coefficient of suspended viruses in the liquid phase, t^{-1}
α^*	resistivity coefficient of sorbed viruses, t^{-1}
β	deterministic but unknown drift coefficients

C. V. Chrysikopoulos (✉) · E. T. Vogler
Department of Civil and Environmental Engineering,
University of California,
Irvine, CA 92697, USA
E-mail: costas@eng.uci.edu

γ	semi-variogram, defined in (31)
ϵ	zero-mean stochastic process
ζ	semi-variogram model parameter, t
θ	porosity (liquid volume/porous medium volume), L^3/L^3.
κ	central confidence interval
λ	inactivation rate coefficient of suspended viruses in the liquid phase, t^{-1}
$\hat{\lambda}$	estimated inactivation rate coefficient of suspended viruses in the liquid phase, t^{-1}
λ^*	inactivation rate coefficient of sorbed viruses, t^{-1}
$\lambda_\circ$	initial inactivation rate coefficient of suspended viruses in the liquid phase, t^{-1}
$\lambda^*_\circ$	initial inactivation rate coefficient of sorbed viruses, t^{-1}
μ_ϵ	mean or expected value of $\epsilon(t)$
$\nu_1, \ldots, \nu_p$	1/2 Lagrange multipliers
$\xi_1, \ldots, \xi_n$	deterministic weight coefficients
ρ	bulk density of the solid matrix (solids mass/aquifer volume), M/L^3
σ^2	variance or sill of the semi-variogram, equal to $R(0)$
$\hat{\sigma}_\beta$	standard error of bootstrap parameter values, defined in (23)
τ	incremental time between measurements

Other signs

$\in$	an element of
$\notin$	not an element of

1 Introduction

Viruses are intracellular parasites with size ranging from 0.02 to 0.3 μm (Brock and Madigan, 1991) that may infiltrate into the subsurface from human and animal sewage through municipal wastewater discharges, septic tanks, sanitary landfills and agricultural practices (Keswick and Gerba, 1980; Chrysikopoulos, 1999; Schijven and Hassanizadeh, 2000). A virus contains a nucleic acid that is surrounded by a protein coat consisting of a number of protein molecules. Disruption of coat proteins and degradation of the nucleic acid is a complex process known as inactivation. In subsurface formations, viruses often migrate along groundwater flowpaths and their transport and fate are significantly affected by virus inactivation and sorption onto the solid matrix (Vilker, 1981; Chu et al., 2001; Redman et al., 2001).

There is a certain relationship between virus adsorption and inactivation. Available experimental observations indicate that the rate of virus inactivation is smaller for sorbed than liquid-phase viruses owing to the apparent protection, provided by the solid matrix, against the disruption of the coat protein and degradation of the nucleic acid (Hurst et al., 1980; Liew and Gerba, 1980; Gerba, 1984; Yates and Yates, 1988). Consequently, recent research efforts on virus adsorption and transport in porous media recognize that inactivation rates of liquid-phase and sorbed or attached viruses should not be assumed equal (Sim and Chrysikopoulos, 1999, 2000).

The majority of available mathematical models for the prediction of fate and transport of viruses in subsurface formations describe virus inactivation by a first-order rate expression with constant rate coefficient (e.g., Tim and Mostaghimi, 1991; Yates and Ouyang, 1992; Chrysikopoulos and Sim, 1996; Sim and Chrysikopoulos, 1995, 1998; Jin et al., 2000 to mention a few). However, several experimental studies suggest that virus inactivation rate coefficients exhibit temporal variability due to the existence of various virus subpopulations undergoing sequential inactivation with different inactivation rate coefficients (Parkinson and Huskey, 1971; Pollard and Solosko, 1971; Yamagishi and Ozeki, 1972; Grant et al., 1993). The mathematical complexity associated with the sequential inactivation that is described by several discrete first-order rate coefficients, each governing a different inactivation phase (Crane and Moore, 1986) is simplified by approximating the multiphasic sequential inactivation by a pseudo first-order expression with a time dependent inactivation rate coefficient determined from available experimental data (Sim and Chrysikopoulos, 1996).

In this work, an effective technique for the estimation of temporally variable virus inactivation rate coefficients from experimental data is developed. The technique consists of a geostatistical procedure called universal kriging in conjunction with the bootstrap resampling method and a nonlinear regression procedure.

2 Time-dependent inactivation

Experimental virus inactivation studies suggest that inactivation rate coefficients are time dependent (Hurst et al., 1980; Pollard and Solosko, 1971; Yamagishi and Ozeki, 1972; Crane and Moore, 1986). Furthermore, Sim and Chrysikopoulos (1996) have shown that experimental data from several batch virus inactivation studies can be reasonably described by a pseudo first-order approximation

$$\frac{\mathrm{d}C(t)}{\mathrm{d}t} = -\lambda(t)C(t) \ , \tag{1}$$

where C is the concentration of suspended viruses in the liquid phase; t is time; and λ is the time-dependent inactivation rate coefficient of suspended viruses given by

$$\lambda(t) = \lambda_\circ \mathrm{e}^{-\alpha t} \ , \tag{2}$$

where $\lambda_\circ$ is the initial inactivation rate coefficient and α is the resistivity coefficient. Substituting (2) into (1) and solving the resulting expression subject to the initial

condition $C(0) = C_\circ$, where $C_\circ$ is the initial virus concentration, yields

$$\ln\left[\frac{C(t)}{C_\circ}\right] = \frac{\lambda_\circ}{\alpha}\left[\exp(-\alpha t) - 1\right] \ . \tag{3}$$

To simplify the notation, the normalized virus log-concentration is expresses as

$$M(t) = \ln\left[\frac{C(t)}{C_\circ}\right] \ . \tag{4}$$

It should be noted that for an arbitrary time $t_\circ$, the slope $\mathrm{d}M(t_\circ)/\mathrm{d}t$ represents the inactivation rate coefficient evaluated at time $t_\circ$

$$\lambda(t_\circ) = -\frac{\mathrm{d}M(t_\circ)}{\mathrm{d}t} = -\lim_{\tau\to 0}\left[\frac{M(t_\circ+\tau) - M(t_\circ)}{\tau}\right] \ , \tag{5}$$

where τ is the incremental time between observations of M. Consequently, in view of (2) and (5), the accuracy of inactivation rate coefficient estimation is affected by the efficiency of the method employed for slope determination or evaluation of the parameters $\lambda_\circ$ and α. As illustrated in Fig. 1, the determination of λ at an arbitrary time $t_\circ$ is essentially equivalent to the slope estimation of the change of M at $t_\circ$.

3 Theory

3.1 Slope estimation using universal kriging

The slope estimation technique used in this work is based on a relatively flexible geostatistical method known as universal kriging often used in mining, hydrological, and geological applications (Journel and Huijbregts, 1978; Isaaks and Srivastava, 1989; Kitanidis, 1997). The kriging estimator is a weighted linear combination of the available experimental data with weights determined such that on the average the estimator error is zero (unbiasedness property) and the square estimation error is as small as possible (minimum variance property). The weights are obtained from the solution to a system of linear equations (kriging system).

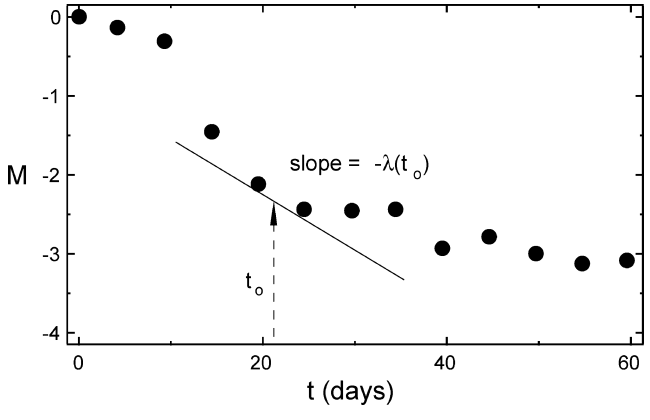

Fig. 1 Typical distribution of normalized virus log-concentration (solid circles) versus time. The slope of the change in normalized log-concentration at an arbitrary time $t_\circ$ is equal to $-\lambda(t_\circ)$

For a set of n available observations of the normalized virus log-concentration, $M(t_1), \ldots, M(t_n)$, the following linear estimator can be employed for the inactivation rate coefficient at any arbitrary time $t_\circ$

$$\hat{\lambda}(t_\circ) = \sum_{i=1}^{n} \xi_i M(t_i) \ , \tag{6}$$

where the hat signifies an estimated inactivation rate coefficient; and $\xi_1, \ldots, \xi_n$ are deterministic but unknown weight coefficients. The normalized virus log-concentration is assumed to have the form

$$M(t) = \sum_{k=1}^{p} \beta_k f_k(t) + \epsilon(t) \ , \tag{7}$$

where $\beta_1, \ldots, \beta_p$ are deterministic but unknown coefficients often referred to as drift coefficients; $f_1(t), \ldots, f_p(t)$ are known time-dependent functions, which are called trial or base functions; and $\epsilon(t)$ is a zero-mean stochastic process (residual). Consequently, the mean of the normalized, time-dependent virus log-concentration is given by

$$\bar{M}(t) = \sum_{k=1}^{p} \beta_k f_k(t) \ , \tag{8}$$

where $\bar{M}(t) = E[M(t)]$ is the mean normalized virus log-concentration (where $E[\ ]$ is the expectation operator). Note that the functions for the time-dependent, normalized virus log-concentration (7) and its mean (8) are both linear in the drift coefficients.

To satisfy the unbiasedness requirement, the weight coefficients $\xi_1, \ldots, \xi_n$ should be selected so that the average estimation error is zero,

$$E\left[\hat{\lambda}(t_\circ) - \lambda(t_\circ)\right] = 0 \ , \tag{9}$$

for any of the unknown drift coefficients $\beta_1, \ldots, \beta_p$. In view of (5), (6) and (8), the preceding expression can be written as

$$\sum_{k=1}^{p}\left\{\sum_{i=1}^{n} \xi_i f_k(t_i) + \lim_{\tau\to 0}\left[\frac{f_k(t_\circ+\tau) - f_k(t_\circ)}{\tau}\right]\right\}\beta_k = 0 \ . \tag{10}$$

For this condition to be valid for any drift coefficient $\beta_1, \ldots, \beta_p$ it is evident that

$$\sum_{i=1}^{n} \xi_i f_k(t_i) = -\lim_{\tau\to 0}\left[\frac{f_k(t_\circ+\tau) - f_k(t_\circ)}{\tau}\right] \ . \tag{11}$$

To satisfy the minimum variance requirement, the variance or mean square error $\mathrm{MSE} = E[(\hat{\lambda}(t_\circ) - \lambda(t_\circ))^2]$, should be as small as possible. In view of (5), (6) and (8), the variance of the linear estimator $\hat{\lambda}(t_\circ)$ can be expressed as

$$\begin{aligned}\mathrm{MSE} &= E\left[\left(\hat{\lambda}(t_\circ)-\lambda(t_\circ)\right)^2\right] \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}\xi_i\xi_j R(t_i,t_j) \\ &\quad + 2\sum_{i=1}^{n}\xi_i \lim_{\tau\to 0}\left[\frac{R(t_i,t_\circ+\tau)-R(t_i,t_\circ)}{\tau}\right] \\ &\quad + \lim_{\tau\to 0}\left[\frac{R(t_\circ+\tau,t_\circ+\tau)+R(t_\circ,t_\circ)-2R(t_\circ+\tau,t_\circ)}{\tau^2}\right],\end{aligned} \tag{12}$$

where

$$R(t_i,t_j) = E[(M(t_i)-\bar{M}(t_i))(M(t_j)-\bar{M}(t_j))] = E[\epsilon(t_i)\epsilon(t_j)] \tag{13}$$

is the covariance function representing the mutual variability between $M(t_i)$ and $M(t_j)$, or equivalently the mutual variability between $\epsilon(t_i)$ and $\epsilon(t_j)$. Minimization of the objective function (12) subject to the constraint (11) can be obtained by the Lagrange multipliers method (Kitanidis, 1997, p. 232). This method requires formation of the Lagrangian

$$\begin{aligned}&L(\xi_1,\ldots,\xi_n,v_1,\ldots,v_p) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}\xi_i\xi_j R(t_i,t_j) + 2\sum_{i=1}^{n}\xi_i\lim_{\tau\to 0}\left[\frac{R(t_i,t_\circ+\tau)-R(t_i,t_\circ)}{\tau}\right] \\ &\quad + \lim_{\tau\to 0}\left[\frac{R(t_\circ+\tau,t_\circ+\tau)+R(t_\circ,t_\circ)-2R(t_\circ+\tau,t_\circ)}{\tau^2}\right] \\ &\quad + 2\sum_{k=1}^{p}v_k\left\{\sum_{i=1}^{n}\xi_i f_k(t_i) + \lim_{\tau\to 0}\left[\frac{f_k(t_\circ+\tau)-f_k(t_\circ)}{\tau}\right]\right\},\end{aligned} \tag{14}$$

where $2v_1,\ldots,2v_p$ are the Lagrange multipliers (the 2 is used only for mathematical convenience). A system of $n+p$ linear equations is formed by taking the derivatives of $L(\xi_1,\ldots,\xi_n,v_1,\ldots,v_p)$ with respect to $\xi_1,\ldots,\xi_n,\ v_1,\ldots,v_p$ and setting them equal to zero, as follows:

$$\begin{aligned}&\sum_{j=1}^{n}\xi_j R(t_i,t_j) + \sum_{k=1}^{p}v_k f_k(t_i) + \lim_{\tau\to 0}\left[\frac{R(t_i,t_\circ+\tau)-R(t_i,t_\circ)}{\tau}\right] \\ &\quad = 0, \quad i = 1,2,\ldots,n\ ,\end{aligned} \tag{15}$$

$$\begin{aligned}&\sum_{i=1}^{n}\xi_i f_k(t_i) + \lim_{\tau\to 0}\left[\frac{f_k(t_\circ+\tau)-f_k(t_\circ)}{\tau}\right] = 0, \\ &\quad k = 1,2,\ldots,p\ .\end{aligned} \tag{16}$$

Equations (15) and (16) define a system of $n+p$ linear equations with $n+p$ unknowns that can be solved for the unknown coefficients $\xi_1,\ldots,\xi_n,v_1,\ldots,v_p$. This system of equations is the kriging system that can be expressed in matrix notation as

$$\mathbf{A}\mathbf{x} = \mathbf{d}\ , \tag{17}$$

where

$$\mathbf{A} = \begin{bmatrix} R(t_1,t_1) & \cdots & R(t_1,t_n) & f_1(t_1) & \cdots & f_p(t_1) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ R(t_n,t_1) & \cdots & R(t_n,t_n) & f_1(t_n) & \cdots & f_p(t_n) \\ f_1(t_1) & \cdots & f_1(t_n) & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ f_p(t_1) & \cdots & f_p(t_n) & 0 & \cdots & 0 \end{bmatrix}, \tag{18}$$

$$\mathbf{x} = \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \\ v_1 \\ \vdots \\ v_p \end{bmatrix}, \tag{19}$$

$$\mathbf{d} = -\begin{bmatrix} \left.\frac{R(t_1,t_\circ+\tau)-R(t_1,t_\circ)}{\tau}\right|_{\tau\to 0} \\ \vdots \\ \left.\frac{R(t_n,t_\circ+\tau)-R(t_n,t_\circ)}{\tau}\right|_{\tau\to 0} \\ \left.\frac{f_1(t_\circ+\tau)-f_1(t_\circ)}{\tau}\right|_{\tau\to 0} \\ \vdots \\ \left.\frac{f_p(t_\circ+\tau)-f_p(t_\circ)}{\tau}\right|_{\tau\to 0} \end{bmatrix}. \tag{20}$$

Note that all elements in vector $\mathbf{d}$ represent appropriate derivatives of the covariance function and the trial or base functions. Assuming that $R(t_i,t_j)$ is known or it can be determined from available experimental data, the coefficients $\xi_1,\ldots,\xi_n,v_1,\ldots,v_p$ are easily obtained by solving the kriging system (17). Subsequently, the estimate $\hat{\lambda}(t_\circ)$ is evaluated from (6). Note that neither $M(t_\circ)$ nor $M(t_\circ+\tau)$ need to be measured for the estimation of $\hat{\lambda}(t_\circ)$. Furthermore, in view of (15) and (16) the MSE expression (12) can be simplified as

$$\begin{aligned}\mathrm{MSE} &= \lim_{\tau\to 0}\left\{\sum_{i=1}^{n}\xi_i\left[\frac{R(t_i,t_\circ+\tau)-R(t_i,t_\circ)}{\tau}\right]\right. \\ &\quad \left. + \sum_{k=1}^{p}v_k\left[\frac{f_k(t_\circ+\tau)-f_k(t_\circ)}{\tau}\right]\right\} \\ &\quad + \lim_{\tau\to 0}\left[\frac{R(t_\circ+\tau,t_\circ+\tau)+R(t_\circ,t_\circ)-2R(t_\circ+\tau,t_\circ)}{\tau^2}\right].\end{aligned} \tag{21}$$

3.2 Parameter estimation by resampling techniques

The determination of the mean of the time-dependent, normalized virus log-concentration, $\bar{M}(t)$, involves the estimation of the drift coefficients, $\beta_1,\ldots,\beta_p$. These coefficients can be determined by fitting expression (3)

to available virus inactivation experimental data. It should be noted that point estimation of an unknown parameter does not provide any information about the variance of its error. Interval estimation can be used to define the confidence interval, which provides a measure of the certainty of this interval containing the true value of the parameter. The basic idea of interval estimation, developed by Neyman (1937), is intuitively simple but its application can lead to considerable difficulties. For models that are linear with respect to the parameters, there are several methods available for exact confidence interval estimation (intervals that maintain nominal coverage probability). However, for nonlinear models the most widely used techniques for interval estimation are linearization methods. Such methods assume that the nonlinear model may be approximated by a linear function throughout the section covered by the confidence interval. Consequently, linearization methods often provide poor approximate confidence intervals, i.e., they underestimate nominal coverage probability (Donaldson and Schnabel, 1987). On the other hand, nonlinear intervals constructed by resampling methods are shown to be relatively accurate (Duncan, 1973; Wu, 1986).

There are several resampling techniques available for dependable construction of confidence intervals. The jackknife, cross-validation, balanced repeated-replication, and bootstrap methods, are conceptually similar, and computationally intensive statistical methods that require very little modeling effort (Diaconis and Efron, 1983). Each of these methods generates numerous artificial data sets from the original experimental data and evaluates the variability of a statistical property of interest from its observed variability over all of the generated artificial data sets. A major advantage of the resampling techniques is that the error in the experimental data does not necessarily have to be homoscedastic or normally distributed. Among the available resampling techniques, bootstrap is considered more efficient for confidence interval estimation (Efron and Tibshirani, 1993). Consequently, in this work, only the bootstrap resampling technique is employed.

3.2.1 Bootstrapping

Since the bootstrap estimator was introduced by Efron (1979), the literature on the bootstrap method has grown rapidly (e.g., Efron, 1981; Bickel and Freedman, 1981; Singh, 1981; Efron and Gong, 1983; Wu, 1986; Hall, 1988; Efron and Tibshirani, 1993; Politis and Romano, 1994; Chernick, 1999; Politis et al., 1999 to mention a few). The concept of the bootstrap is conceptually simple and its theoretical foundations are described elsewhere (Efron, 1982). The method is outlined here in just a few steps. Consider a data set composed of n observations. Select a random sample with replacement of size n from the original data set. This random resample or bootstrap sample may contain observations more than once. Using the random resample, obtain a bootstrap vector of estimated model parameters $\hat{\mathbf{b}}$, by some parametric or nonparametric procedure. Repeat the resampling process a large number of times, B, and keep a record of the bootstrap parameter estimates $\hat{\mathbf{b}}_i$, where subscript i denotes bootstrap iteration ($1 \leq i \leq B$). The mean of multiple bootstrap estimated model parameters is the "best" bootstrap vector of parameter estimates

$$\mathbf{b}_\beta = \sum_{i=1}^{B} \frac{\hat{\mathbf{b}}_i}{B} \quad , \tag{22}$$

and the standard error is the square root of the sample variance of bootstrapped parameter values (Efron, 1981)

$$\hat{\sigma}_\beta = \left[\sum_{i=1}^{B} \frac{(\hat{\mathbf{b}}_i - \mathbf{b}_\beta)^T (\hat{\mathbf{b}}_i - \mathbf{b}_\beta)}{B-1} \right]^{1/2} \quad . \tag{23}$$

Efron (1982) has shown via Monte Carlo experiments that the bootstrap standard error is slightly downward biased; thus it is not conservative. For an illustration of the bootstrap procedure see Fig. 2.

3.2.2 Nonlinear fitting of bootstrap samples

The fitting of a mathematical model with nonlinear parameters to experimental data requires iterative methods. There are several approaches available for nonlinear parameter estimation. Here, the nonlinear least squares regression method is adopted. In general, the objective of the nonlinear least squares method is to obtain estimates of the model parameters that minimize the residual sum of squares between simulated and observed data. The objective function may be written as (Beck and Arnold, 1997; Chrysikopoulos et al., 1990)

$$S(\hat{\mathbf{b}}) = \left[\mathbf{y} - \mathbf{g}(\mathbf{u}, \hat{\mathbf{b}}) \right]^{\mathrm{T}} \left[\mathbf{y} - \mathbf{g}(\mathbf{u}, \hat{\mathbf{b}}) \right] \tag{24}$$

where $\mathbf{y} = \mathbf{g}(\mathbf{u}, \mathbf{b}) + \mathbf{e}$ is a vector of n observed data, $\mathbf{g}$ is a vector of n model simulated data, $\mathbf{u}$ is a vector of independent variables, and $\mathbf{e}$ is a vector of n random numbers with zero mean and known covariance matrix.

Minimization of the objective function is not a trivial task, because of the nonlinearities in $\mathbf{g}(\mathbf{u}, \mathbf{b})$. Several techniques have been developed for unconstrained nonlinear estimation. Simple iterative minimization algorithms, such as trial and error or exhaustive search, are seldom used due to their inefficiency. However, there is a wide selection of nonlinear estimation methods that can be used for the least-squares or maximum likelihood parametric estimation problem. The most advantageous methods can be classified in two major categories, the modified Newton and Gauss–Newton linearization approaches. The first approach to the nonlinear estimation problem uses a Taylor series expansion to linearize the

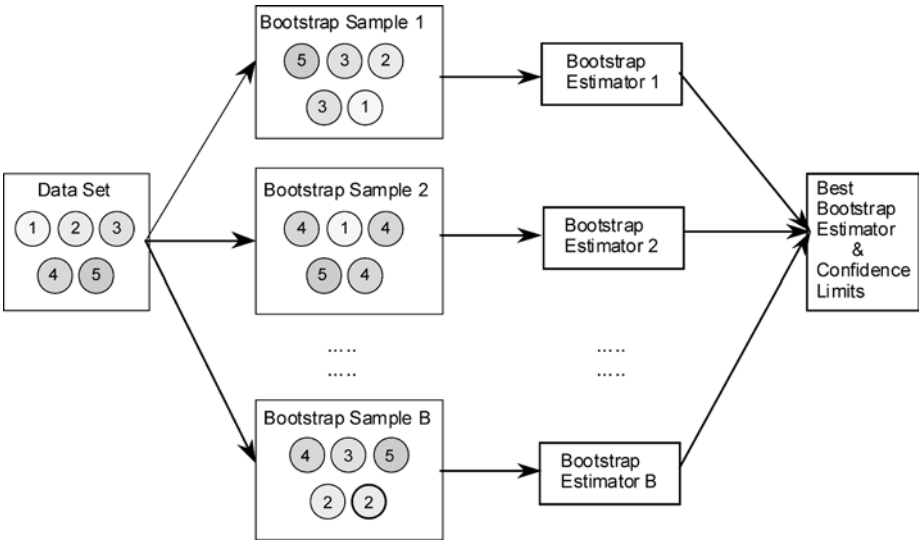

Fig. 2 Illustrative example of the bootstrap method for a set of five data points and B bootstrap replications (adopted from Chrysikopoulos et al., 2001)

objective function, whereas the second approach for the nonlinear estimation problem is to expand the nonlinear model in a Taylor series around the initial parameter estimates. An advantageous modification of Gauss-Newton based on the work of Levenberg (1944) and Marquardt (1963) eliminates potential numerical difficulties when a non-full column rank Jacobian matrix is encountered. Here, the Levenberg–Marquardt method is employed for the fitting of expression (3) to each bootstrap sample of the virus inactivation normalized, log-concentration data.

3.2.3 Confidence intervals by bootstrap percentiles

Although several bootstrap methods for obtaining confidence intervals are available (Efron and Tibshirani, 1993), in this study, the percentile method that makes use of the bootstrap distribution is employed. That is, the estimation of a confidence interval

$$\left[\hat{\mathbf{b}}_\ell(\kappa), \hat{\mathbf{b}}_\mathrm{u}(\kappa)\right] \ , \tag{25}$$

where subscripts ℓ and u denote the lower and upper limits of the vector of true model parameters $\mathbf{b}$, respectively, is approximated by the κ central confidence interval. The probability κ $(0 < \kappa < 1)$ indicates a $100\kappa\%$ confidence that $\mathbf{b} \in [\hat{\mathbf{b}}_\ell, \hat{\mathbf{b}}_\mathrm{u}]$ or a $100(1-\kappa)\%$ confidence that $\mathbf{b} \notin [\hat{\mathbf{b}}_\ell, \hat{\mathbf{b}}_\mathrm{u}]$. The larger the κ, the greater the chance that the unknown parameter is included in the confidence interval. For example, the 95% confidence limits for $\mathbf{b}$ based on 2000 bootstrap replications are given by $\hat{\mathbf{b}}_\ell = 50$th and $\hat{\mathbf{b}}_\mathrm{u} = 1950$th largest estimates of $\mathbf{b}$. Obviously, elements of the vector of model parameters are treated individually. For simple parameter estimation, approximately 100 bootstrap replications are sufficient. However, for confidence interval estimation the number of bootstrap iterations should be on the order of 1000 (Efron and Tibshirani, 1993).

4 Application to poliovirus inactivation

The experimental data shown in Fig. 1 for poliovirus type 1 (strain LSc) batch inactivation at 50 °C in the presence of 0.2 g of sediment, collected by Liew and Gerba (1980), were employed in this work to test the proposed bootstrap resampling technique in conjunction with a regression procedure based on the Levenberg–Marquardt method.

4.1 Bootstrap estimation of initial inactivation rate and resistivity coefficients

In view of Eqs. (3) and (8) it is evident that the drift coefficients can be expressed as:

$$\beta_1 = -\frac{\lambda_\circ}{\alpha} \ , \tag{26}$$

$$\beta_2 = \frac{\lambda_\circ}{\alpha} \ . \tag{27}$$

Certainly, evaluation of the two drift coefficients requires prior estimation of the initial inactivation rate, $\lambda_\circ$, and the resistivity coefficient, α, from the normalized virus log-concentration experimental data set.

A FORTRAN-90 program was developed in order to select a random sample (bootstrap sample) from the virus inactivation log-concentration data set, fit the bootstrap sample with expression (3), and estimate the corresponding $\lambda_\circ$ and α. A bootstrap sample was randomly selected with replacement from the original experimental data set. The bootstrap sample contains as many elements as the number of available experimental log-concentration data in the original data set. It should be noted that multiple repetitions of the same log-concentration may occur in a bootstrap sample. The subroutine *mrqmin* (Press et al., 1992) was used to obtain the desired bootstrap parameter estimates $\hat{\lambda}_\circ$ and $\hat{\alpha}$ by

fitting expression (3) to the normalized virus log-concentrations of the bootstrap sample.

The complete resampling procedure was repeated 2000 times. Consequently, two bootstrap vectors of 2000 estimated model parameters $\hat{\mathbf{b}}_{\lambda_\circ} = (\hat{\lambda}_{\circ_1}, \hat{\lambda}_{\circ_2}, \ldots, \hat{\lambda}_{\circ_{2000}})^{\mathrm{T}}$ and $\hat{\mathbf{b}}_{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_{2000})^{\mathrm{T}}$ are created. In view of (22), averaging the bootstrapped parameter values yields the "best" bootstrap estimates $\hat{\lambda}_{\circ_\beta}$ and $\hat{\alpha}_\beta$. In this work, determination of confidence limits was based on bootstrap percentiles. The 95% confidence limits of $\hat{\lambda}_{\circ_\beta}$ (95% confidence that $\hat{\lambda}_{\circ_\beta} \in [\hat{\lambda}_{\circ_\ell}, \hat{\lambda}_{\circ_u}]$) are given by $\hat{\lambda}_{\circ_\ell} = 50$th and $\hat{\lambda}_{\circ_u} = 1950$th largest bootstrap estimates of $\lambda_\circ$, respectively; and the 95% confidence limits of $\hat{\alpha}_\beta$ (95% confidence that $\hat{\alpha}_\beta \in [\hat{\alpha}_\ell, \hat{\alpha}_u]$) are given by $\hat{\alpha}_\ell = 50$th and $\hat{\alpha}_u = 1950$th largest bootstrap estimates of α, respectively.

For the experimental poliovirus inactivation data set considered here, the parameters $\hat{\lambda}_{\circ_\beta}$, and $\hat{\alpha}_\beta$ together with the corresponding confidence limits were determined and are listed in Table 1. Histograms of 2000 bootstrap replications of $\lambda_\circ$ and α were presented in Fig. 3. A dotted line drawn at the parameter estimate and dashed lines drawn at the two confidence limits were included with each histogram in Fig. 3. Both histograms are roughly Gaussian in shape suggesting that confidence interval evaluation based on bootstrap percentiles is a reasonable approach.

Figure 4a presents the normalized log-concentration poliovirus experimental data (solid circles) and simulated log-concentration history (solid curve) determined with the inactivation model (3) and the bootstrap estimates of $\lambda_\circ$ and α listed in Table 1. The residuals or detrended data (original data less the fitted drift) are presented in Figure 4b and suggest a relatively good agreement between the simulated and experimental data. Note that $\epsilon(t)$ oscillates about a zero value and exhibits a statistical dependence. Intuitively, it is expected that an improved estimation of the time dependent virus inactivation rate, $\lambda(t)$, can be obtained if the temporal correlation of $\epsilon(t)$ is incorporated in the estimation process.

Table 1 Bootstrap estimates of $\lambda_\circ$ and α

Parameter (days^{-1})	Estimate	95% Confidence Limits	
		Lower	Upper
$\hat{\alpha}_\beta$	0.035	0.020	0.049
$\hat{\lambda}_{\circ_\beta}$	0.128	0.097	0.167

4.2 Covariance function determination

A random field is partially described by second-order characteristics, e.g., its mean function or expected value, and covariance function (Christakos, 1992). Consequently, the stationary stochastic random field for the parameter $\epsilon(t)$ is characterized by the mean function:

$$\mu_\epsilon(t) = E[\epsilon(t)] = 0 \ , \tag{28}$$

indicating a zero-mean stochastic process, and by the covariance function:

$$R(t_i, t_j) = R(\tau) = E[\epsilon(t)\epsilon(t+\tau)] \ , \tag{29}$$

representing the mutual variability between $\epsilon(t_i)$ and $\epsilon(t_j)$ or equivalently the mutual variability between $\epsilon(t)$ and $\epsilon(t+\tau)$, where $\tau = |t_i - t_j|$. The covariance function of a stationary random field is related to the semi-variogram by (Journel and Huijbregts, 1978)

$$R(\tau) = R(0) - \gamma(\tau) \ , \tag{30}$$

where $R(0) = E\left[\epsilon^2(t)\right]$ is the variance of $\epsilon(t)$, representing the mean square deviation of $\epsilon(t)$ from its mean value, $\mu_\epsilon(t) = 0$; and $\gamma(\tau)$ is the semi-variogram defined as:

$$\gamma(\tau) = \frac{1}{2} E\left[\left(\epsilon(t) - \epsilon(t+\tau)\right)^2\right] \ . \tag{31}$$

To determine the covariance function $R(t_i, t_j)$ the detrended data, shown in Fig. 4b, were used for the construction of the appropriate raw semi-variogram or scatter plot. The raw semi-variogram is essentially a plot of the square difference $[\epsilon(t_i) - \epsilon(t_i + \tau)]^2/2$ as a function of the incremental time between measurements, τ. For n experimental data points, there are $n(n-1)/2$ such pairs that comprise the raw variogram. For the the poliovirus inactivation data set examined here, there are thirteen measurements that yield a cloud of 78 pairs, indicated by the solid circles in Fig. 5a. Dividing the axis of τ into seven consecutive intervals and by averaging the pairs of measurement in each interval, the experimental semi-variogram of the detrended data is constructed and is illustrated by the solid squares in Fig, 5b.

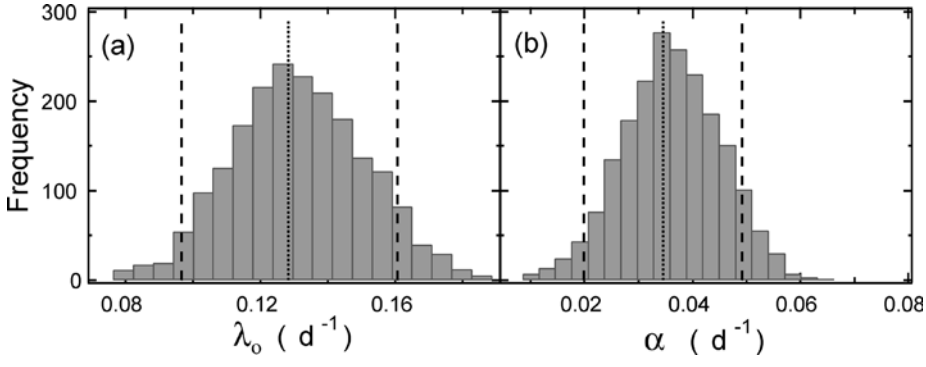

Fig. 3 Histograms of 2000 bootstrap estimates of **(a)** the initial inactivation rate coefficient, $\lambda_\circ$, and **(b)** the resisitivity coefficient, α, evaluated from the normalized log-concentration experimental data set presented in Fig. 1. The bars represent frequency. The best bootstrap estimator value $\hat{\lambda}_{\circ_\beta}$ and $\hat{\alpha}_\beta$ are indicated by dotted lines and their lower and upper 95% confidence limits $\hat{\lambda}_{\circ_\ell}$, $\hat{\alpha}_\ell$ and $\hat{\lambda}_{\circ_u}$, $\hat{\alpha}_u$, respectively, are represented by dashed lines

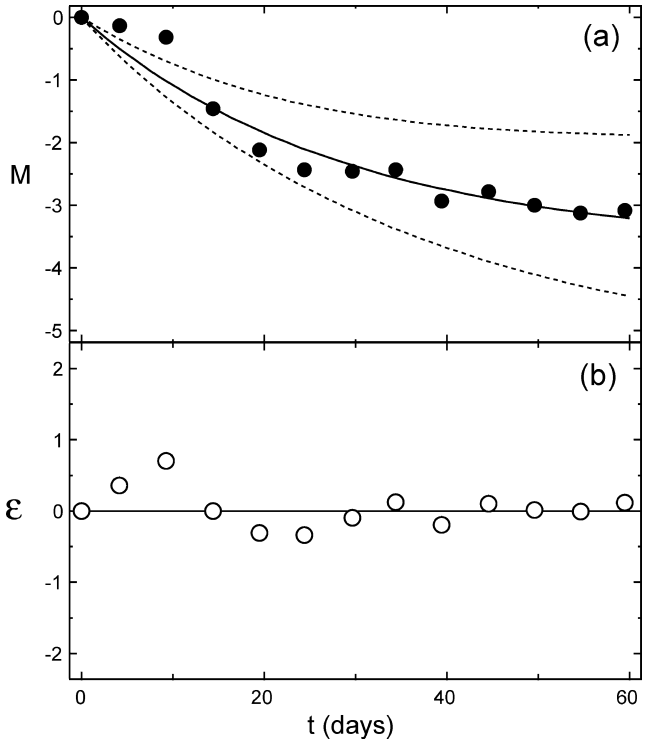

Fig. 4 a Normalized log-concentration experimental data (solid circles) of poliovirus batch inactivation in the presence of a sediment, collected by Liew and Gerba (1980), and simulated concentration history (solid curve) together with the appropriate lower and upper 95% confidence intervals (dashed curves) determined by the inactivation model (3) with the estimated parameter values listed in Table 1. **b** Residuals or detrended experimental data (open circles)

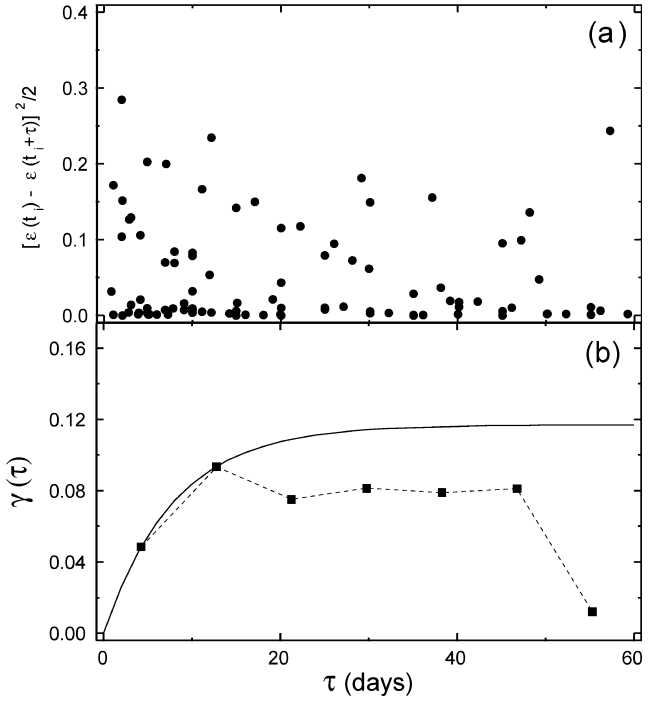

Fig. 5 a Scatter plot of the detrended experimental data. **b** Experimental semi-variogram of detrended data (solid squares) and fitted exponential model (solid curve)

It was assumed that the experimental semi-variogram can be fitted with the following exponential, theoretical semi-variogram model:

$$\gamma(\tau) = \sigma^2 \left[1 - \exp\left(-\frac{\tau}{\zeta}\right)\right], \quad \zeta > 0 \ , \tag{32}$$

where $\sigma^2 = R(0)$ is the variance or the sill of the semi-variogram, and ζ is a model parameter that determines how fast the semi-variogram increases to its sill value. For a value of $\tau = 3\zeta$ the semi-variogram is approximately equal to 95% of σ^2; this time lag is known as the range of the exponential semi-variogram model. The unknown model parameters of the semi-variogram are determined by a relatively subjective graphical fitting procedure that leads to estimates $\sigma^2 = 0.117$ and $\zeta = 7.93$ days. In Fig. 5b, the fitted model is indicated by the solid curve and is in agreement with the experimental semi-variogram. What is most important is the semi-variogram behavior at very small time scales, near the origin of the semi-variogram, where we can observe whether the random variable is differentiable, continuous non-differentiable, or discontinuous. The discrepancy between the experimental semi-variogram and theoretical model (32) at large time lags is of no major concern because the theoretical semi-variogram model can easily be modified to include a linear drift in order to more accurately reproduce the experimental semi-variogram at large incremental times; however, semi-variograms that differ by a quadratic function are essentially equivalent for purposes of kriging (Kitanidis, 1993). It should be noted that it is not important how close the theoretical semi-variogram model fits the sequence of points comprising the experimental semi-variogram, but the appropriateness of the semi-variogram function selected, based on the type of continuity and stationarity assumed for the random variable (Wackernagel, 1995).

Combining Eqs. (30) and (32) yields the desired covariance function:

$$\begin{aligned} R(t_i, t_j) = R(\tau) &= \sigma^2 \exp\left[-\frac{\tau}{\zeta}\right] \\ &= 0.117 \exp\left[-\frac{|t_i - t_j|}{7.93}\right] \ , \end{aligned} \tag{33}$$

where the latter formulation is the consequence of substitution of the fitted model parameters. The preceding expression is required by the proposed virus inactivation rate estimation procedure discussed in Sect. 3.1.

4.3 Estimation of inactivation rate coefficients

In view of Eqs. (3), (8), (26) and (27) it is evident that the appropriate trial or base functions are:

$$f_1(t) = 1 \ , \tag{34}$$

$$f_2(t) = \exp[-\alpha t] \ . \tag{35}$$

Furthermore, substitution of Eqs. (33) through (35) into (17) yields the following kriging system of linear equations:

$$\begin{bmatrix} \sigma^2 & \cdots & \sigma^2 \exp\left[-\frac{|t_1-t_n|}{\zeta}\right] & 1 & \exp[-\alpha t_1] \\ \sigma^2 \exp\left[-\frac{|t_2-t_1|}{\zeta}\right] & \cdots & \sigma^2 \exp\left[-\frac{|t_2-t_n|}{\zeta}\right] & 1 & \exp[-\alpha t_2] \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \sigma^2 \exp\left[-\frac{|t_n-t_1|}{\zeta}\right] & \cdots & \sigma^2 & 1 & \exp[-\alpha t_n] \\ 1 & \cdots & 1 & 0 & 0 \\ \exp[-\alpha t_1] & \cdots & \exp[-\alpha t_n] & 0 & 0 \end{bmatrix}$$

$$\times \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \\ \nu_1 \\ \nu_2 \end{bmatrix} = \begin{bmatrix} \frac{\sigma^2}{\zeta} \exp\left[-\frac{|t_1-t_\circ|}{\zeta}\right] \\ \frac{\sigma^2}{\zeta} \exp\left[-\frac{|t_2-t_\circ|}{\zeta}\right] \\ \vdots \\ \frac{\sigma^2}{\zeta} \exp\left[-\frac{|t_n-t_\circ|}{\zeta}\right] \\ 0 \\ \alpha \exp[-\alpha t_\circ] \end{bmatrix} . \quad (36)$$

The preceding set of linear equations is solved for the unknown coefficients $\xi_1, \ldots, \xi_{13}, \nu_1$, and ν_2, for 30 different $t_\circ$ values ranging from 0 to 60 days, using the poliovirus inactivation experimental data presented in Fig. 4a, and parameter values $\alpha = 0.035$ days^{-1}, $\sigma^2 = 0.117$, and $\zeta = 7.93$ days. For each $t_\circ$ considered the corresponding inactivation rate coefficient was determined by (6) and plotted in Fig. 6 (solid diamonds) together with the inactivation rate coefficients as predicted from (2) with the resistivity and initial inactivation rate coefficients (bootstrap estimates) listed in Table 1 (solid curve) as well as the associated lower and upper 95% confidence intervals (dashed curves).

Figure 6 indicates that the fluctuations of the estimated inactivation rate coefficient as a function of time obtained by the procedure developed in this work do follow the general trend predicted by the exponential decay model (2); however, they are in much better agreement with the fluctuations of the original poliovirus inactivation experimental data (compare the temporal structure of the solid circles in Fig. 4a with the solid diamonds in Fig. 6). The non-monotonic decrease of inactivation rates is attributed to the frequently observed multiphasic inactivation caused by the existence of several virus subpopulations within a virus population, undergoing sequential inactivation with different inactivation rate coefficients (Yamagishi and Ozeki, 1972; Grant et al., 1993).

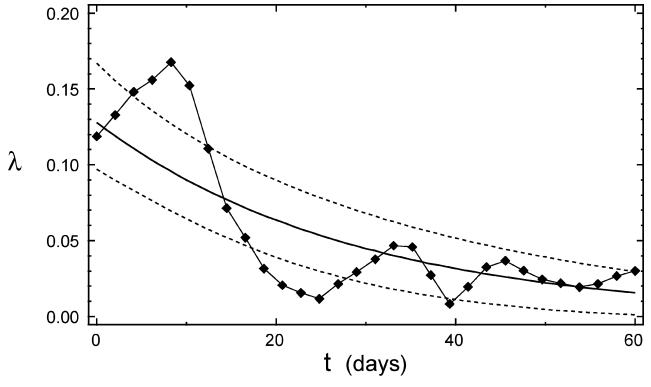

Fig. 6 Simulated behavior of time-dependent inactivation rate coefficient together with the appropriate lower and upper 95% confidence intervals (dashed curves), determined by expression (2) with the estimated parameter values listed in Table 1, and by the proposed slope estimation procedure (solid diamonds) (here $\alpha = 0.035$ days^{-1} and $\lambda_\circ = 0.128$ days^{-1})

5 Application to virus transport in porous media

5.1 Mathematical model

Transient virus transport through one-dimensional, homogeneous, saturated porous media, accounting for virus adsorption and inactivation, is governed by the following partial differential equation (Sim and Chrysikopoulos, 1995)

$$\frac{\partial C(t,x)}{\partial t} + \frac{\rho}{\theta}\frac{\partial C^*(t,x)}{\partial t} = D\frac{\partial^2 C(t,x)}{\partial x^2} - U\frac{\partial C(t,x)}{\partial x} - \lambda(t)C(t,x) - \lambda^*(t)\frac{\rho}{\theta}C^*(t,x) , \quad (37)$$

where C^* is the sorbed phase virus concentration (virus mass/solids mass); D is the hydrodynamic dispersion coefficient; U is the average interstitial velocity; λ^* is the time dependent inactivation rate coefficient of the sorbed phase viruses; ρ is the bulk density of the solid matrix; θ is the porosity of the medium (liquid volume/aquifer volume); and x is the spatial coordinate in the direction of flow. The left-hand side of the preceding equation consists of the accumulation terms, whereas the last two terms represent the inactivation of suspended and sorbed viruses, respectively.

Assuming that the adsorption process consists of virus diffusion to the outer layer of a solid particle by nonequilibrium mass transfer and virus immobilization onto the solid particle while in equilibrium with the liquid phase virus concentration in the outer layer, also accounting for inactivation of sorbed viruses, the accumulation of sorbed viruses can be represented by the following mass balance

$$\frac{\rho}{\theta}\frac{\partial C^*(t,x)}{\partial t} = k\left[C(t,x) - C_g(t,x)\right] - \lambda^*(t)\frac{\rho}{\theta}C^*(t,x) , \quad (38)$$

where k is the mass transfer rate constant; and C_g is the liquid phase virus concentration in direct contact with solids. Furthermore, it is assumed that the following linear equilibrium relationship is valid

$$C^*(t,x) = K_d C_g(t,x) , \quad (39)$$

where K_d is the partition or distribution coefficient (liquid volume/solids mass).

The inactivation rate coefficients are considered to be time dependent, and consequently the inactivation of viruses in the liquid phase is described by (1) and for the the inactivation of viruses in the solid phase is described by the following first-order rate expression:

$$\frac{dC^*(t)}{dt} = -\lambda^*(t)C^*(t) \ , \tag{40}$$

where the time dependent inactivation rate coefficients of suspended viruses in the liquid phase is described by (2) and the time dependent inactivation rate coefficients of sorbed viruses onto the solid phase is described by the following expression:

$$\lambda^*(t) = \lambda^*_\circ e^{-\alpha^* t} \ , \tag{41}$$

where $\lambda^*_\circ$ is the initial inactivation rate coefficient of sorbed viruses; and α^* is the resistivity coefficient of sorbed viruses. The magnitude of α is proportional to the resistivity of the dominant subpopulation, because the overall inactivation is controlled by the dominant subpopulation. The inactivation rate coefficients of viruses in the liquid phase are assumed to be twice as large as the coefficients of sorbed viruses (Reddy et al., 1981; Yates and Ouyang, 1992)

$$\lambda^*_\circ = \frac{\lambda_\circ}{2} \ . \tag{42}$$

Furthermore, the resistivity coefficient of sorbed viruses is considered to be equal to the resistivity coefficient of viruses in the liquid phase (Sim and Chrysikopoulos, 1996)

$$\alpha^* = \alpha \ . \tag{43}$$

The appropriate initial and boundary conditions for a semi-infinite, one-dimensional porous formation in the presence of a continuous source of viruses are:

$$C(0,x) = C^*(0,x) = 0 \ , \tag{44}$$

$$C(t,0) = C_\circ e^{-\lambda t} \ , \tag{45}$$

$$\frac{\partial C(t,\infty)}{\partial x} = 0 \ , \tag{46}$$

where $C_\circ$ is the source concentration. The condition (44) establishes that there is no initial liquid phase and adsorbed virus concentrations within the porous medium. The boundary condition (45) describes an exponentially decaying virus concentration at the inlet. The downstream boundary condition (46) preserves concentration continuity for a semi-infinite system. The governing virus transport equation (37) in conjunction with the relationships (2), (38), (39), and (41) is solved numerically subject to initial/boundary conditions (44)–(46). The numerical solution is obtained by the method of finite differences using the IMSL subroutine DLSARG (IMSL, 1991).

5.2 Model simulations

To demonstrate the effect of virus inactivation behavior on virus transport in water saturated porous media, the one-dimensional model presented in the previous section is employed to simulate the transport of poliovirus for the parameter values listed in Table 2. The virus inactivation rate coefficients are determined directly from the experimental data presented in Fig. 1. Both the exponentially decaying analytical expression (2) as well as the proposed slope estimation procedure are employed for the estimation of the inactivation rate coefficients. Snapshots of normalized poliovirus concentration suspended in the liquid phase for simulation times of 10, 20, and 40 days are presented in Fig. 7. The solid curves correspond the case where the inactivation rate coefficients are determined with analytical expression (2), whereas the dashed curves are constructed with inactivation rate coefficients determined by the slope estimation procedure developed in this work. It should be

Table 2 Model parameters for virus transport simulations

Parameter	Value	Reference
D	32.04 cm^2/h	Bales et al. (1991)
k	1.2 h^{-1}	Vilker and Burge (1980)
K_d	2.08×10^{-2} mL/mg	Vilker (1981)
U	5.04 cm/h	Bales et al. (1991)
α	0.035 days^{-1}	
ζ	7.93 days	
ρ	1.5 g/cm^3	
σ^2	0.117	
θ	0.25	

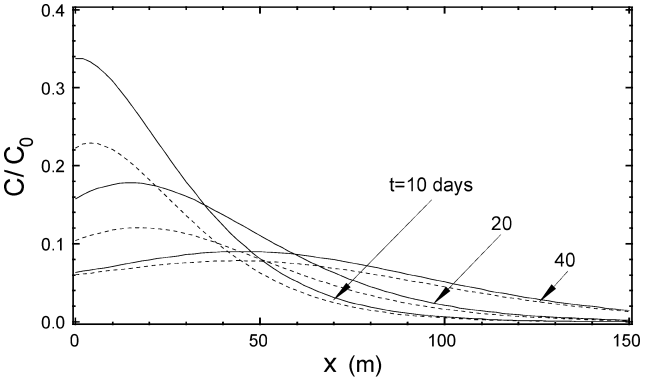

Fig. 7 Normalized poliovirus concentration snapshots simulated at various times with the transport parameter values listed in Table 2 accounting for the experimentally observed inactivation behavior (see data in Fig. 1). The solid curves correspond to a time dependent inactivation rate coefficient determined by expression (2), whereas the dashed curves correspond to an inactivation rate coefficient obtained by the proposed slope estimation procedure

noted that the normalized poliovirus concentration at the inlet ($x = 0$ m) is reduced with increasing time as imposed by the boundary condition (45). The transport simulations indicate that distributions of poliovirus concentrations with inactivation rate coefficients evaluated by the analytical expression (2) are overestimated compared to the concentrations predicted with inactivation rate coefficients determined by the proposed geostatistical procedure. The observed difference between the poliovirus snapshots simulated with the two different virus inactivation coefficient estimation procedures becomes progressively less pronounced with increasing time. However, this set of simulations clearly indicates that the evaluation procedure of the time dependent virus inactivation rate coefficients can significantly affect the predicted virus migration in porous media.

6 Summary

This paper introduces a new procedure for the determination of time dependent virus inactivation rate coefficients from existing virus inactivation experimental data. The procedure employs universal kriging for the slope estimation of the virus inactivation data and the bootstrap resampling technique for the estimation of the initial inactivation rate and resistivity coefficients. The inactivation rate coefficient estimation procedure is relatively laborious and computationally demanding because an experimental semi-variogram is necessary for the determination of the covariance function of the experimental virus inactivation data. Nevertheless, inactivation rate coefficients determined by the proposed procedure are in much better agreement with observed virus inactivation experimental data than the inactivation rate coefficients determined by the exponential decay model. Furthermore, simulations of virus transport in one-dimensional water saturated porous media suggest that the procedure employed for inactivation rate coefficient estimation can significantly influence the predicted migration behavior of viruses.

Acknowledgement The authors would like to thank Dr. Dionissios T. Hristopulos whose suggestions improved this work. Most of this research was performed while the first author was on sabbatical leave in the Environmental Engineering Department at the Technical University of Crete, Greece, where Prof. George Karatzas generously provided his computational facilities.

References

Bales RC, Hinkle SR, Kroeger TW, Stocking K (1991) Bacteriophage adsorption during transport through porous media: Chemical perturbations and reversibility. Environ. Sci Technol 25(12): 2088–2095

Beck JV, Arnold KJ (1977) Parameter Estimation in Engineering and Science. Wiley, New York, NY

Bickel P, Freedman DA (1981) Some asymptotic theory for the bootstrap. Ann Statist, 9(6): 1196–1217

Brock TD, Madigan MT (1991) Biology of Microorganisms, 6th edn., Prentice Hall, Englewood Cliffs, NJ

Chernick MR (1999) Bootstrap Methods: A Practitioner's Guide. Wiley, New York, NY

Christakos G (1992) Random Field Models in Earth Sciences, Academic Press, San Diego, CA

Chrysikopoulos CV (1999) Virus transport in the subsurface. In: Katsifarakis KL (ed), Groundwater Pollution Control. Vol. 2, Chap. 3, WIT Press, Ashurst, Southampton, UK, pp. 95–144

Chrysikopoulos CV, Sim Y (1996) One-dimensional virus transport in homogeneous porous media with time dependent distribution coefficient. J. Hydrol., 185: 199–219

Chrysikopoulos CV, Roberts PV, Kitanidis PK (1990) One-dimensional solute transport in porous media with partial well-to-well recirculation: Application to field experiments. Water Resour Res, 26(6): 1189–1195

Chrysikopoulos CV, Hsuan P-Y, Fyrillas MM (2002) Bootstrap estimation of the mass transfer coefficient for a dissolving nonaqueous phase liquid pool in porous media. Water Resour Res, 38(3): 10.1029/2001WR000661

Chu Y, Jin Y, Flury M, Yates MV (2001) Mechanisms of virus removal during transport in unsaturated porous media. Water Resour Res, 37(2): 253–263

Crane SR, Moore JA (1986) Modeling enteric bacterial die-off: a review. Water Air Soil Pollut, 27: 411–439

Diaconis P, Efron E (1983) Computer-intensive methods in statistics. Sci Am, 248(5): 116–129

Donaldson JR, Schnabel RB (1987) Computational experience with confidence regions and confidence intervals for nonlinear least squares. Technometrics, 29(1): 67–93

Duncan GT (1978) An empirical study of jackknife-constructed confidence regions in nonlinear regression. Technometrics, 20(2): 123–129

Efron B (1979) Bootstrap methods: another look at the jackknife. Ann Statist, 7(1): 1–26

Efron B (1981) Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods. Biometrica, 68(3): 589–599

Efron B (1982) The jackknife, the bootstrap and other resampling plans. SIAM, monograph 38, CBMS-NSF

Efron B (1987) Better bootstrap confidence intervals. Am Statist Ass., 82: 171–185

Efron B, Gong G (1983) A leisurely look at the bootstrap, the jackknife and cross-validation. Am Statistic, 37(1): 36–48

Efron B, Tibshirani R (1986) Bootstrap methods for the standard errors, confidence intervals, and other measures of statistical accuracy. Statist Sci, 1(1): 54–77

Freedman, DA (1981) Bootstraping regression models. Ann. Statist., 9(6): 1218–1228

Gerba, CP (1984) Applied and theoretical aspects of virus adsorption to surfaces. Adv. Appl. Microb., 30: 133–168

Grant SB, List EJ, Lidstrom ME (1993) Kinetic analysis of virus adsorption and inactivation in batch experiments; virus adsorption, virus inactivation, equilibrium isotherms. Water Resour Res, 29(7): 2067–2085

Hall P (1988) Theoretical comparison of bootstrap confidence intervals. Ann Statist, 16(3): 927–953

Hurst CJ, Gerba CP, Cech I (1980) Effects of environmental variables and soil characteristics on virus survival in soil. Appl Environ Microb, 40(6): 1067–1079

IMSL (1991) IMSL Math/Library User's Manual, ver. 2.0, IMSL Press, Houston, Texas

Isaaks EH, Srivastava RM (1989) An Introduction to Applied Geostatistics. Oxford University Press, New York, NY

Jin Y, Chu Y, Li Y (2000) Virus removal and transport in saturated and unsaturated sand columns. J. Contamin. Hydrol., 43: 111–128

Journel AG, Huijbregts CJ (1978) Mining Geostatistics. Academic Press, London, UK

Keswick BH, Gerba CP (1980) Viruses in groundwater. Environ. Sci. Technol., 14(11): 1290–1297

Kitanidis PK (1993) Generalized covariance functions in estimation. Math. Geol., 25(5): 525–540
Kitanidis PK (1997) Introduction to Geostatistics: Applications in Hydrogeology. Cambridge University Press, Cambridge, UK
Levenberg K (1944) A method for the solution of certain nonlinear problems in least squares. Quart Appl Math, 2: 164–168
Liew PF, Gerba CP (1980) Thermostabilization of enteroviruses by estuarine sediment. Appl Environ Microb, 40(2): 305–308
Marquardt DW (1963) An algorithm for least-squares estimation of nonlinear parameters. SIAM J. Appl. Math., 11(2): 431–441
Neyman, J (1937) Outline of a theory of statistical festimation based on classical theory of probability. Royal Soc London Phil Trans A, 236: 333–380
Parkinson JS, Huskey RJ (1971) Deletion mutants of bacteriophage lambda, 1, Isolation and initial characterization. J Mol Biol, 56: 369–384
Politis DN, Romano JP (1994) The stationary bootstrap. J Am Statist Assoc, 89: 1303–1313
Politis DN, Romano JP, Wolf M (1999) Subsampling. Springer Verlag
Pollard EC, Solosko W (1971) The thermal inactivation of T_4 and λ bacteriophage. Biophys J, 11: 66–74
Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1992) Numerical Recipes in Fortran 77: The Art of Scientific Computing, Cambridge University Press
Reddy KR, Khaleel R, Overcash MR (1981) Behavior and transport of microbial pathogens and indicator organisms in soils treated with organic wastes. J Environ Qual, 10(3): 255–266
Redman JA, Grant SB, Olson TM, Estes MK (2001) Pathogen filtration, heterogeneity, and the potable reuse of wastewater. Environ Sci Technol, 35: 1798–1805
Schijven JF, Hassanizadeh SM (2000) Removal of viruses by soil passage: overview of modelling processes and parameters. Critical Rev Environ Sci Technol, 30(1): 49–127
Sim Y, Chrysikopoulos CV (1995) Analytical models for one-dimensional virus transport in saturated porous media. Water Resour Res, 31(5): 1429–1437
Sim Y, Chrysikopoulos CV (1996) One-dimensional virus transport in porous media with time dependent inactivation rate coefficients. Water Resour Res, 32(8): 2607–2611
Sim Y, Chrysikopoulos CV (1998) Three-dimensional analytical models for virus transport in saturated porous media. Transp. Porous Media, 30(1): 87–112
Sim Y, Chrysikopoulos CV (1999) Analytical models for virus adsorption and inactivation in unsaturated porous media. Colloids Surfaces A: Physicochem. Eng Aspects, 155: 189–197
Sim Y, Chrysikopoulos CV (2000) Virus transport in unsaturated porous media. Water Resour Res, 36(1): 173–179
Singh, K (1981) On the asymptotic accuracy of Efron's bootstrap. Ann Statist, 9(6): 1187–1195
Sobsey MD, Dean CH, Knuckles ME, Wagner RA (1980) Interaction and survival of enteric viruses in soil materials. Appl Environ Microb, 40: 92–101
Tim US, Mostaghimi S (1991) Model for predicting virus movement through soils. Ground Water, 29(2): 251–259
Vilker VL (1981) Simulating virus movement in soils. In: Iskandar IK (ed.) Modeling Waste Renovation: Land Treatment, Wiley, New York, NY, pp. 223–253
Vilker VL, Burge WD (1980) Adsorption mass transfer model for virus transfer in soils. Water Res, 14: 783–790
Wackernagel H (1995) Multivariate geostatistics. Springer-Verlag, Berlin, Germany
Wu FJ (1986) Jackknife, bootstrap and other resampling methods in regression analysis. Ann Stat, 14(4): 1261–1295
Yamagishi H, Ozeki H (1972) Comparative study of thermal inactivation of phage $\phi 80$ and lambda. Virology, 48: 316–322
Yates MV, Yates SR (1988) Modeling microbial fate in the subsurface environment. Crit Rev Environ Control, 17(4): 307–344
Yates MV, Ouyang Y (1992) VIRTUS: a model of virus transport in unsaturated soils. Appl Environ Microb, 58(5): 1609–1616