Probability density function of non-reactive solute concentration in heterogeneous porous formations

Alberto Bellin [a],*, Daniele Tonina [b,c]

[a] *Dipartimento di Ingegneria Civile e Ambientale, Università di Trento, via Mesiano 77, I-38050 Trento, Italy*
[b] *Earth and Planetary Science Department, University of California, Berkeley, USA*
[c] *USFS, Rocky Mountain Research Station 322 East Front Street, Suite 401 Boise, ID, 83702, USA*

Abstract

Available models of solute transport in heterogeneous formations lack in providing complete characterization of the predicted concentration. This is a serious drawback especially in risk analysis where confidence intervals and probability of exceeding threshold values are required. Our contribution to fill this gap of knowledge is a probability distribution model for the local concentration of conservative tracers migrating in heterogeneous aquifers. Our model accounts for dilution, mechanical mixing within the sampling volume and spreading due to formation heterogeneity. It is developed by modeling local concentration dynamics with an Ito Stochastic Differential Equation (SDE) that under the hypothesis of statistical stationarity leads to the Beta probability distribution function (pdf) for the solute concentration. This model shows large flexibility in capturing the smoothing effect of the sampling volume and the associated reduction of the probability of exceeding large concentrations. Furthermore, it is fully characterized by the first two moments of the solute concentration, and these are the same pieces of information required for standard geostatistical techniques employing Normal or Log-Normal distributions. Additionally, we show that in the absence of pore-scale dispersion and for point concentrations the pdf model converges to the binary distribution of [Dagan, G., 1982. Stochastic modeling of groundwater flow by unconditional and conditional probabilities, 2, The solute transport. Water Resour. Res. 18 (4), 835–848.], while it approaches the Normal distribution for sampling volumes much larger than the characteristic scale of the aquifer heterogeneity. Furthermore, we demonstrate that the same model with the spatial moments replacing the statistical moments can be applied to estimate the proportion of the plume volume where solute concentrations are above or below critical thresholds. Application of this model to point and vertically averaged bromide concentrations from the first Cape Cod tracer test and to a set of numerical simulations confirms the above findings and for the first time it shows the superiority of the Beta model to both Normal and Log-Normal models in interpreting field data. Furthermore, we show that assuming a-priori that local concentrations are normally or log-normally distributed may result in a severe underestimate of the probability of exceeding large concentrations.
© 2007 Elsevier B.V. All rights reserved.

Keywords: Concentration pdf; Heterogeneous geological formations; Ito's Stochastic Differential Equation; Beta distribution

1. Introduction

Densely monitored field-scale tracer tests and studies on contaminated sites revealed large variations of solute concentrations over short distances. This complexity is due to flow non-uniformity generated by spatial variability of the hydraulic conductivity K. Solute concentrations are attenuated by dilution associated with pore-scale dispersion (PSD) and their measurements are further affected by mixing within the sampling volume. In fact, solute concentration is defined as the mass of solute dissolved

---

in a unit volume of water: $C_{\Delta V}(\boldsymbol{x}, t) = M(x, t)/(n \Delta V)$, where n is the formation porosity and M is the mass of solute that at time t is within the sampling volume ΔV centered at $\boldsymbol{x}$, such that the larger ΔV the more it acts as an attenuation factor for the spatial variability of solute concentration.

Point concentrations are obtained operationally when the size of the sampling volume is much smaller than the characteristic scale of variability of K, such that one can safely assume that $\Delta V \rightarrow 0$. In such a situation, the only hydrodynamical process attenuating solute concentrations is PSD. However, in all other cases, solute concentration fluctuations that are perceived from measurements – or that should be considered in risk analysis, when exposures are evaluated at extraction wells – are attenuated by the combined effect of PSD and mixing within the sampling volume.

Since detailed mapping of hydraulic conductivity in the subsurface is unfeasible, solute transport is typically modeled within a stochastic framework by considering an ensemble of realizations of the K field. Most of the existing stochastic studies analyze the evolution of the ensemble mean $\langle C \rangle$ and variance σ_C^2 of point concentration (e.g., Kapoor and Gelhar, 1994; Zhang and Neuman, 1996; Dagan and Fiori, 1997; Kapoor and Kitanidis, 1998; Pannone and Kitanidis, 1999; Fiori and Dagan, 2000; Fiori, 2002; Fiorotto and Caroni, 2002; Caroni and Fiorotto, 2005). Others have specifically addressed the combined effect of pore-scale dispersion and sampling volume on solute concentration moments (Andricevic, 1998) and flux statistics at a compliance plane (Fiori et al., 2002). These studies concur in concluding that ergodic conditions are hardly obtained in applications, such that $\langle C \rangle$ does not suffice to describe the concentration spatial variability (Kitanidis, 1994; Fiori and Dagan, 1999), and a more effective representation should be used, for example the local probability distribution (Goovaerts, 1997, ch. 7).

Despite its importance in applications, the probability density function (pdf) of solute concentrations in heterogeneous aquifers has been considered in few studies. Dagan (1982) showed that in the absence of pore-scale dispersion point concentration at a given (fixed) location is at any time either the initial concentration C_0 or zero, and the resulting Cumulative Distribution Function (CDF) is binary with probability $P = \langle C \rangle / C_0$ of observing C_0 and $1 - P$ of observing zero concentration. Successively, Bellin et al. (1994) showed that although the concentration CDF is close to bimodality for small sampling volumes, it approaches the Normal distribution as the sampling volume grows large. However, all these studies neglected pore-scale dispersion, which for point concentrations, has a strong impact on σ_C^2 (Kapoor and Kitanidis, 1998; Fiori and Dagan, 2000), and thus on the concentration pdf.

More recently, Fiori (2001) argued that the Beta distribution can be used to represent the variability of local concentrations, because it is bounded between a minimum and a maximum and is flexible enough to comply with the above properties. Other evidence in support of the Beta distribution, involving the relationship between the kurtosis and the skewness, has been discussed by Chatwin and Sullivan (1990) and Chatwin et al. (1995) for atmospheric turbulent transport.

The effect of pore-scale dispersion on point concentration CDF was investigated numerically by Fiorotto and Caroni (2002). They showed a good fit of the Beta distribution with data obtained numerically by simulating a tracer test in a two-dimensional heterogeneous formation. In a companion paper, Caroni and Fiorotto (2005) proposed an integral expression of the concentration pdf, whose validity is limited to weakly heterogeneous formations, and showed numerically that the Beta distribution fits the experimental concentration CDF well for values of the log-conductivity variance up to $\sigma_Y^2 = 2$.

Our work differs from that of Caroni and Fiorotto (2005) in the following two aspects: (1) In Caroni and Fiorotto's work the Beta distribution is assumed a-priori, while we look for a physical interpretation leading to a theoretical justification of this choice; (2) they considered only concentration data obtained from numerical simulations, whereas in addition to that we confronted our model to field data.

In this work, we seek to answer the following questions: Why is the Beta distribution a better model of local uncertainty obtained from Monte Carlo numerical experiments than the Normal and Log-Normal distributions? Is it an artifact of numerical simulations, and in particular of the geostatistical model of the spatial variability of K, or a property of solute concentrations of passive tracers? And furthermore, is the Beta distribution also a good model for the global probability distribution, which describes the probability of exceeding a given concentration threshold irrespective of the location within the plume? We address these questions first developing a theoretical model for the concentration CDF and then applying it to concentration data from the Cape Cod tracer test (Leblanc et al., 1991) and from two-dimensional numerical simulations.

2. Mathematical statement of the problem

2.1. The general expression of the concentration pdf

Let us consider the instantaneous release of a conservative tracer with concentration C_0 within the source

volume V_0 in a heterogeneous aquifer of spatially variable K and constant porosity n. The evolution of the solute concentration can be described by the following Stochastic Differential Equation (SDE) (Cassiani et al., 2005):

$$dC_{\Delta V} = F(C_{\Delta V})dt + G(C_{\Delta V})dW(t) \qquad (1)$$

where $W(t)$ is the Wiener process with zero mean and dispersion coefficient equal to 1/2 (Gardiner, 1985, p. 66), and F and G are functions of the concentration, but not explicitly of time (autonomous SDE). Hereafter to simplify notation we omit the indication that $C_{\Delta V}$ varies in time and space, which is therefore considered as implicit. More specifically, F is the expected rate of change of $C_{\Delta V}$ and is referred to in genetics as drift function, whereas G, called the diffusion function, defines the disturbance to the expected rate of change caused by formation heterogeneity. Eq. (1), which has been shown consistent with the transport equation, was extensively applied to model solute transport in atmospheric turbulent flow and turbulent combustion processes (see e.g. Wandel et al., 2003; Cassiani et al., 2005).

One important result of stochastic theories is that solute concentrations evolve toward a maximum entropy state, corresponding to a well mixed Gaussian plume, spontaneously in heterogeneous aquifers (Kitanidis, 1994; Kapoor and Kitanidis, 1998). The rate of convergence of C toward $\langle C \rangle$ increases with pore-scale dispersion and sampling volume. A simple model describing such a behavior assumes that the expected rate of change of local concentrations is proportional to its "distance" from equilibrium:

$$F = \kappa[\langle C \rangle - C_{\Delta V}] \qquad (2)$$

where $\kappa > 0$ is the rate of convergence of $C_{\Delta V}$ to $\langle C \rangle$ due to dilution and mixing. The parameter κ is influenced by pore-scale dispersion and mixing in such a way that an increase in pore-scale dispersion coefficients or ΔV causes a faster convergence of $C_{\Delta V}$ to $\langle C \rangle$. This approach, which supposes that mixing and pore scale dispersion homogenize the concentration field, corresponds to the Interaction by Exchange with the Mean (IEM) (Villermaux and Devillon, 1972) also known as Linear Mean Square Estimation model (Dopazo and O'Brian, 1974) utilized in combustion turbulence and chemical engineering (Cassiani et al., 2005).

The otherwise smooth transition of $C_{\Delta V}$ to $\langle C \rangle$ described by the first right hand term of Eq. (1) is perturbed by flow non-uniformity, which introduces randomness into the system. This disturbance is described by the function G in the second right hand term of Eq. (1),

and is not uniform within the plume. In fact, Bellin et al. (1994) observed that for a fixed time the coefficient of variation of the solute concentration $CV[C_{\Delta V}]$ is small at locations close to the center of mass, and increases sharply as one approaches the fringes of the plume. This behavior of $CV[C_{\Delta V}]$ can be explained by observing that in a Monte Carlo framework randomness is reflected in variations of plume position and shape between independent realizations of the velocity field, and that the expected variations between any two independent realizations are larger at the fringes of the plume than close to the center of mass (e.g. Fisher et al., 1979; Kitanidis, 1988; Dagan, 1991; Gelhar, 1993).

These differences in randomness can be modeled by assigning the following form to the function G in the second right hand term of Eq. (1):

$$G = \sqrt{\epsilon\, C_{\Delta V}(C_0 - C_{\Delta V})} \qquad (3)$$

where $\epsilon > 0$ is a coefficient proportional to the formation heterogeneity. In Eq. (3), the location with respect to the center of mass is epitomized by $C_{\Delta V}$, with $C_{\Delta V} \approx C_0$ and $C_{\Delta V} \approx 0$ representing locations close to the center of mass and outside the ensemble plume, respectively. Consistently with the above reasoning, in Eq. (1) G is small compared to F when $C_{\Delta V}$ is either high, i.e. close to C_0, or small, and it is maximum for $C_{\Delta V} = C_0/2$.

After these preparatory steps the SDE (1) assumes the following form:

$$dC_{\Delta V} = \kappa[\langle C \rangle - C_{\Delta V}]dt + \sqrt{\epsilon\, C_{\Delta V}(C_0 - C_{\Delta V})}dW(t), \qquad (4)$$

which can be written in a more convenient form for mathematical treatment by dividing both sides of the equation by C_0 to obtain:

$$dZ = \kappa[\langle Z \rangle - Z]dt + \sqrt{\epsilon\, Z(1 - Z)}dW(t) \qquad (5)$$

where $Z = C_{\Delta V}/C_0$ and $\langle Z \rangle = \langle C \rangle/C_0$ are dimensionless solute concentrations, which vary between 0 and 1. Note that to simplify the notation we omitted to indicate the dependence of Z on ΔV, which therefore is considered as implicit.

The pdf of the normalized concentration described by Eq. (5) satisfies the following Kolmogorov forward equation (see e.g. Gardiner, 1985, ch. 4.3.6):

$$\frac{\partial f(z,t)}{\partial t} = \frac{\partial}{\partial z}[F(z)f(z,t)] + \frac{1}{2}\frac{\partial}{\partial z}\left\{G(z)\frac{\partial}{\partial z}[G(z)f(z,t)]\right\}. \qquad (6)$$

Under stationary conditions, when the SDE (1) reaches equilibrium and $\partial f/\partial t \rightarrow 0$, which is possible for autonomous diffusion processes (Pope, 1965), the solution of the Eq. (6) is given by (Cobb, 1981, p. 51):

$$f(z) = \frac{\Gamma[p+q]}{\Gamma[p]\Gamma[q]} z^{-1+p}(1-z)^{-1+q} \qquad (7)$$

where $\Gamma[x] = \int_0^\infty \tau^{x-1}e^{-\tau}d\tau$ is the Euler gamma function (x stands for p, q, or $p+q$ and τ is the integral dummy variable), $p = \langle Z \rangle \kappa/\epsilon = \langle Z \rangle/\beta$, and $q = (1-\langle Z \rangle)\kappa/\epsilon = (1-\langle Z \rangle)/\beta$. In these last two expressions, the parameter β assumes the following form:

$$\beta = \frac{\epsilon}{\kappa} = \frac{\sigma_Z^2}{\langle Z \rangle[1-\langle Z \rangle] - \sigma_Z^2} \qquad (8)$$

Note that the model (7) is the familiar Beta density function, with the polarization ratio β controlling the shape of the pdf.

The stationary hypothesis introduced in order to solve Eq. (6) does not imply that the solute concentration should reach a steady state condition, but rather that at each time step the population ensemble of the concentration has reached statistical equilibrium. In other words, the concentration pdf depends on time though the solute concentration, but not directly.

In the Eq. (8), β varies between zero and infinity conferring to the Beta density function a large flexibility. For example, Fig. 1 shows that when β is larger than both $\langle Z \rangle$ and $1-\langle Z \rangle$ the pdf is U-shaped with polarization to $z=0$ and $z=1$. On the other hand, when β is smaller than both $\langle Z \rangle$ and $1 - \langle Z \rangle$ the pdf is unimodal, and when β is between $\langle Z \rangle$ and $1-\langle Z \rangle$ the pdf has a J-shape with a
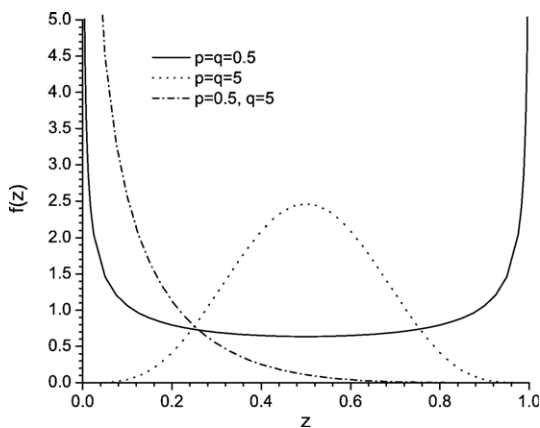

Fig. 1. Typical shapes of the beta distribution model for β larger than both $\langle Z \rangle$ and $1-\langle Z \rangle$ (i.e. p, $q<1$; solid line), smaller than both $\langle Z \rangle$ and $1-\langle Z \rangle$ (i.e. p, $q>1$; dotted line), and for $\beta>\langle Z \rangle$ but $\beta<1-\langle Z \rangle$ (i.e. $p<1$, $q>1$; dashed-dot line).

singular point at $z \rightarrow 0$ (see also Fig. 3 of (Cobb, 1981, p. 49)). Note that σ_Z^2 plays a fundamental role in controlling the shape of the pdf.

In the next section, we discuss two known cases for the pdf model: point solute concentration in the absence of pore-scale dispersion and the purely diffusive case in a uniform flow field, showing that our model (7) resembles these cases when PSD is set to zero and when $\sigma_Y^2 \rightarrow 0$ (homogeneous formation), respectively.

2.2. Two known cases of the concentration pdf

In an early work, Dagan (1982) showed that in the absence of pore-scale dispersion the point concentration of a conservative tracer is either $C=C_0$ ($Z=1$) or $C=0$ ($Z=0$), depending on whether the sampler is internal or external to the solute body, and the plume does not dilute. In this idealized situation $\sigma_Z^2 = \langle Z \rangle [1-\langle Z \rangle]$ and the pdf of Z assumes the following form:

$$f(z) = \langle Z \rangle \delta[1-z] + [1-\langle Z \rangle]\delta[z] \qquad (9)$$

where δ is the Dirac delta function. Hence, in this case knowing $\langle Z \rangle$ suffices to fully characterize the spatial random function Z regardless of the model of spatial variability. In Appendix A, we show that for $\beta \rightarrow \infty$, which results from setting $\sigma_Z^2 = \langle Z \rangle[1-\langle Z \rangle]$ into Eq. (8), the Beta model (7) reduces to the binary distribution (9) as required. Notice that the first right hand term of the SDE (5) vanishes for $\beta \rightarrow \infty$, thereby validating the assumption introduced in Section 2.1 that the second right hand term mimics solute spreading due to flow non-uniformities associated to formation heterogeneity.

Another interesting result is obtained when dilution, or mixing associated with ΔV, dominates over solute spreading caused by flow non-uniformities. Increases in pore-scale dispersion coefficients, or in ΔV, lead to reductions of σ_Z^2 (see the works by Kapoor and Kitanidis (1998) and Fiori and Dagan (2000)) and β, which converges linearly to zero as $\sigma_Z^2 \rightarrow 0$.

The limit of the model (7) for small β values is conveniently analyzed by examining the behavior of the central moments of Z:

$$\mu_\nu = \int_0^1 (z - \langle Z \rangle)^\nu f(z)dz$$
$$= \left[-\frac{p}{p+q}\right]^\nu {}_2F_1\left[-\nu, p, p+q, \frac{p+q}{p}\right] \qquad (10)$$

where $\nu>1$ is the order of the moment and ${}_2F_1$ is the hypergeometric function (Gradshteyn and Ryzhik, 1980, p. 1045). In view of the ensuing analysis, it is

convenient to write the moments (10) in the following dimensionless form:

$$\mu'_\nu \frac{\mu_\nu}{(\mu_2)^{\nu/2}} = \left[-\frac{p}{p+q}\right]^\nu \left[\frac{pq}{(p+q)^2(1+p+q)}\right]^{\nu/2}$$

$$\times {}_2F_1\left[-\nu, p, p+q, \frac{p+q}{p}\right]. \tag{11}$$

The careful inspection of Eq. (11) for $\beta \to 0$ conducted in Appendix B shows that $\mu'_\nu = 0$ when ν is odd, and $\mu'_\nu = 1 \cdot 3 \cdot 5 \cdots (\nu-1) = (\nu-1)!!$, that is the double factorial of $\nu-1$ (Arfken, 1985, p. 548), when ν is even. This property characterizes a Normal distribution with mean $\langle Z \rangle$ and variance σ_Z^2. Furthermore, the convergence of the moments Eq. (11) to those of the Normal distribution is proportional to β and $\sqrt{\beta}$ for ν even and odd, respectively. Note that small β values are obtained for $k \gg \epsilon$ in the SDE (5) such that Z converges rapidly to $\langle Z \rangle$, leading to an ergodic plume. Furthermore, the condition $\beta \to 0$ is obtained for $\sigma_Z^2 \to 0$, such that the Normal distribution degenerates to the Dirac Delta distribution $f(z) = \delta[z - \langle Z \rangle]$, a solution consistent with the fact that for an ergodic plume $C(\mathbf{x}, t) = \langle C(\mathbf{x}, t) \rangle$.

We emphasize that the concentration pdf can be seen as a measure of the uncertainty associated to the estimation of the local concentration by the ensemble mean, and most importantly it can be used to assess the probability that the actual (measurable) concentration exceeds a given threshold at a given location as required in risk assessment. As in Normal and Log-Normal distributions, note that the Beta model is fully characterized by the ensemble mean and variance of the concentration, which in the last decades have been the object of several investigations leading to a number of closed form solutions (see e.g. Dagan, 1982; Kapoor and Gelhar, 1994; Kapoor and Kitanidis, 1998; Fiori and Dagan, 2000), but at the same time enjoying a much larger flexibility with respect to these models.

3. Numerical validation of the concentration pdf model

In Section 2 we showed, by using a physically based stochastic model, that the concentration pdf of a passive solute is the Beta model (7). This result was anticipated heuristically by Fiorotto and Caroni (2002) and Caroni and Fiorotto (2005), who observed that the model (7) provides a good fit of the sample Cumulative Frequency Distribution (CFD) of concentrations obtained by means of a numerical Monte Carlo experiment (see e.g. Figs. 8 and 9 of Caroni and Fiorotto (2005)). In the ensuing section we validate the model (7) against numerical

simulations for both small and large sampling volumes and for different levels of formation heterogeneity.

3.1. Numerical experiment

We considered a two-dimensional isotropic formation with spatial variability of the log-transmissivity field, $Y = \ln(T_K)$, described by the isotropic exponential covariance model:

$$C_Y(r) = \sigma_Y^2 \exp[-r/I_Y], \tag{12}$$

where $T_K(x_1, x_2) = \int_0^b K(x_1, x_2, x_3)dx_3$ is the hydraulic transmissivity, r is the two-point separation distance and I_Y is the log-transmissivity integral scale.

A single realization of the transmissivity field was generated using HYDRO_GEN (Bellin and Rubin, 1996; Rubin and Bellin, 1998), and the flow equation was solved using the Galerkin finite element code developed by Bellin et al. (1992). Simulations were conducted in a domain $100 I_Y$ long and $70 I_Y$ wide with the average flow, U, in the x_1 direction. A unitary mean head gradient was imposed by assigning constant piezometric heads along the short sides of the domain and impervious boundaries along the remaining two sides.

Transport was solved by forward or backward particle tracking with pore-scale dispersion modeled as a Brownian motion with constant longitudinal and transverse dispersion coefficients $D_{d,L} = UI_Y/200$ and $D_{d,T} = UI_Y/2000$, respectively. The resulting Peclet numbers are $P_{eL} = 200$ and $P_{eT} = 2000$ in longitudinal and transverse directions, respectively.

In the forward particle tracking scheme (FPT), which we adopted in estimating the spatial distribution of the solute concentration, a total mass $M_0 = nC_0V_0$ of solute was instantaneously released within the initial volume V_0 with a square horizontal projection area A_0 of side $L = 1.5 I_Y$, and thickness equal to the aquifer depth b. The mass of solute was divided into $N_p = 90,000$ non-interacting particles of volume $v_p = nV_0/N_p = 2.5 \ 10^{-5} n \ I_Y^2 \ b$. The particles were tracked forward in time by utilizing the following tracking algorithm:

$$X_1(t + \Delta t) = X_1(t) + V_1[X_1(t), X_2(t)]\Delta t + \sqrt{2D_{d,L}\Delta t}\epsilon_1 \tag{13}$$

$$X_2(t + \Delta t) = X_2(t) + V_2[X_1(t), X_2(t)]\Delta t + \sqrt{2D_{d,T}\Delta t}\epsilon_2 \tag{14}$$

where $\mathbf{V} = (V_1, V_2)$ is the Eulerian velocity, $\mathbf{X} = (X_1, X_2)$ is the particle's trajectory, and ϵ_1, ϵ_2 are two independent random numbers both normally distributed with zero

mean and unitary variance. The time step Δt is selected such that the maximum step resulting from the application of Eqs. (13) and (14) to the computed velocity field is smaller than the grid's size.

Consistent with particle tracking the solute concentration was computed as follows:

$$C_{\Delta V}(t, \boldsymbol{x}) = \frac{M_{\Delta V}(\boldsymbol{x}, t)}{n \Delta V} = \frac{n_p(\boldsymbol{x}, t) V_0}{N_p \Delta V} C_0, \qquad (15)$$

where $M_{\Delta V}$ is the solute mass that at time t is within the sampling volume ΔV with horizontal square projection area ΔA of side Δ and thickness b centered at $\boldsymbol{x} = (x_1, x_2)$. Note that $M_{\Delta V}$ is obtained by counting the number $n_p(t, \boldsymbol{x})$ of particles, whose centers of mass are within ΔV at time t. Simulations were conducted with $\Delta = 0.2 \, I_Y$ and I_Y, such that $C_{\Delta V}$ mimics the vertically averaged concentration measured in a monitoring well fully penetrating the aquifer, in the first case, and a larger sampling volume in the second.

Numerical errors in the above forward particle tracking scheme (FPT) arise from neglecting particle deformation and assuming that the contribution of each particle to $M_{\Delta V}$ is either $\Delta m = v_p C_0 = M_0 / N_p$ or zero, depending on whether its center of mass is within or outside ΔV. The latter assumption leads to the following lower detection limit for the simulated concentration: $C_{\Delta V,\min} = C_0 V_0 / (N_p \Delta V)$, which is obtained when only one particle is found within ΔV. At this point, we introduce the relative sensitivity of the FPT as the ratio between the solute concentration obtained when only one particle is within ΔV and C_0:

$$S_c = \frac{C_{\Delta V,\min}}{C_0} = \frac{V_0}{N_p \Delta V}. \qquad (16)$$

Simulations were conducted with $S_c = 6.25 \times 10^{-4}$, meaning that concentrations lower than $S_c C_0$ cannot be detected.

Eq. (16) shows that in order to maintain the same detection limit and accuracy, a reduction of the sampling volume requires an increase in the number of particles. This would lead to an exceedingly large computational burden when $\Delta V \ll V_0$. To alleviate the computational burden and increase accuracy, we shifted from FTP to the backward particle tracking (BPT) approach when calculating the local concentration over many Monte Carlo realizations. This methodology consists in releasing n_f particles within ΔV and tracking them back in time towards the source (V_0) (Vanderborght, 2001). Similarly to FPT, the particle's movement is split into two steps: a Brownian jump, modeled as:

$$X_j^*(t - \Delta t; \boldsymbol{x}) = X_{t,j}(t; \boldsymbol{x}) - X_{b,j} \qquad j = 1, 2 \qquad (17)$$

where $X_{b,1}$ and $X_{b,2}$ assume the expression in the last right hand term of Eqs. (13) and (14), respectively and $X_{t,j}(t; \boldsymbol{x})$ is the j-th component of the particle's position at time t. At the second step, the particle is advected back in time according to the Eulerian velocity $\mathbf{V}$ at the position $\boldsymbol{x} = \mathbf{X}^*(t - \Delta t; \boldsymbol{x})$:

$$\begin{aligned} X_{t,j}(t - \Delta t, \boldsymbol{x}) = X_j^*(t - \Delta t; \boldsymbol{x}) \\ - V_j[\mathbf{X}^*(t - \Delta t, \boldsymbol{x})] \Delta t \qquad j = 1, 2. \end{aligned} \qquad (18)$$

where V_j is the j-th component of $\mathbf{V}$. Eqs. (17) and (18) are applied repeatedly until $t = 0$, when the concentration is computed as follows:

$$C_{\Delta V}(\boldsymbol{x}, t) = \frac{n_p(V_0, t)}{n_f} C_0 \qquad (19)$$

where $n_p(V_0, t)$ is the number of particles found within the source volume. In this case, the minimum detectable concentration is given by: $\Delta C_{\Delta V,\min} = C_0 / n_f$, corresponding to the case of only one particle reaching the source at time $t = 0$. In our simulation we used $n_f = 64{,}000$ and $80{,}000$ for the control volume with size $\Delta V = 0.2 \, I_Y$ and I_Y, respectively.

3.2. Analysis of the numerical results

The need to quantify the probability of exceeding a given concentration for regulatory limitations justifies the interest in studying the CDF of solute concentrations. More specifically, because of its importance in environmental analysis where samplers of different dimensions are used, we extended our analysis to the effects of sampling dimensions and distance from or time since injection on the CDF of $C_{\Delta C} / C_0$. Simulations were conducted in weakly (i.e. $\sigma_Y^2 = 0.2$) and moderately (i.e., $\sigma_Y^2 = 1.2$) heterogeneous two-dimensional formations with a square source area of side $L = 1.5 \, I_Y$.

Fig. 2a, and b show the CFDs of $Z = C_{\Delta V} / C_0$ for $\sigma_Y^2 = 1.2$ at the center of mass of the mean plume at times $t = 4 \, I_Y / U$ and $55 \, I_Y / U$, respectively. The CFDs are obtained numerically with 2000 Monte Carlo simulations for several ΔV and are assumed to resemble the corresponding CDFs. In the same figures the CFDs – note that according to the Bernoulli statistical theorem the sample CFD converges to the CDF of the population as the sample's size grows to infinity – are compared to the CDF of our Beta model, which is obtained by replacing the first two statistical moments of Z in the Eq. (A.1) with the sample mean and variance obtained from the Monte Carlo simulation. We utilized this procedure
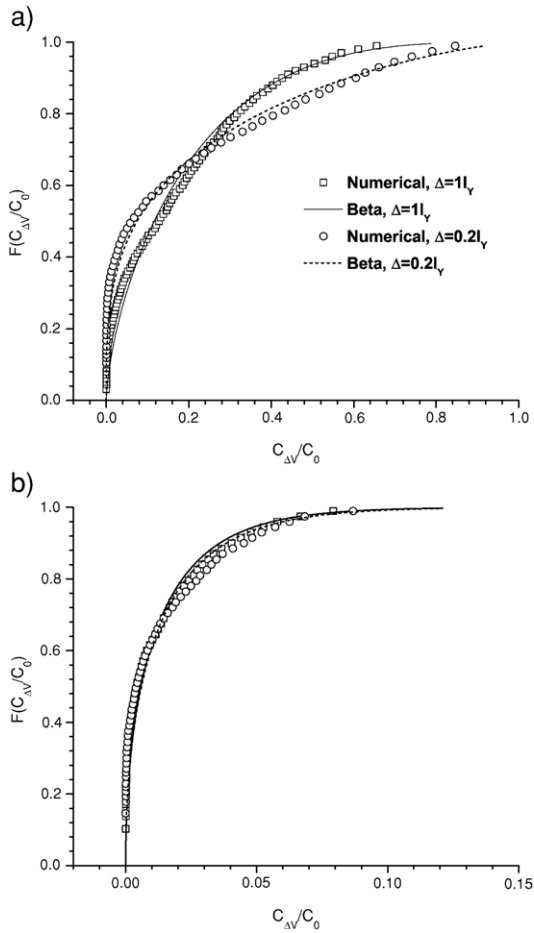
Fig. 2. Comparison between Beta model and the CDF of the solute concentration obtained numerically by Monte Carlo simulations for $\Delta=0.2I_Y$ and I_Y at a) $x=4I_Y$ and $t=4I_Y/U$ and b) at $x=55I_Y$ and $t=55I_Y/U$. In all cases $\sigma_Y^2=1.2$, $Pe_L=200$ and $Pe_T=2000$, and the source area is a square of side $L=1.5\ I_Y$.

rather than the least-square fitting or maximum likelihood estimation, because they require measured concentration data, which are often rare or absent in applications. Whereas the first two concentration moments, which are the only pieces of information required by the moment method, can be obtained from transport models.

Examination of Fig. 2a reveals that $C_{\Delta V}/C_0$ is adequately described by the Beta model, which shows a great flexibility in capturing modifications of the CFD induced by variations of the sampling volume. Moreover, numerical results of Fig. 2b show a declining sensitivity of CFD to the sampling volume at large times since injection, and this is accurately captured by the Beta model. On the other hand, Normal and Log-Normal models, often used in geostatistical studies,

resulted in a poor adaptation to the same set of data (not shown in the figures).

In Fig. 3, we report the results of simulations conducted to compute the concentration CDF at positions other than the center of mass of the mean plume. In all cases the Beta model (obtained as before by replacing in Eq. (A.1) the statistical moments with the sample's moments) matches the numerical CDF and performs much better than Normal and Log-normal models. A first comparison is between the CDFs at the position $x=(4\ I_Y,\ 0)$ for $t=4\ I_Y/U$ and $t=8\ I_Y/U$, when the center of mass of the mean plume is on the sampling point and ahead of it, respectively. Besides the good match of the Beta model with numerical simulations, we observe that the probability of exceeding moderate and high concentrations is greater at the center of mass of the mean plume, and this is consistent with the structure we assumed for the "noise" term G (Eq. (3)) in Eq. (1).

At a larger distance from the solute source, for example $x=(55I_Y,\ 0)$ and time $t=45\ I_Y/U$ (i.e., when concentrations are measured at a location ahead of the center of mass of the mean plume), the effect is more evident. However at large distances, because of solute spreading, the zone around the center of mass with small variations of the solute concentration increases in size and the sampling within this zone provides similar CDF regardless of the position. This effect is the same that reduces the importance of sampling dimension at large distances from the source or at greater time since injection (see Figs. 2b and 3).

Simulations conducted with $\sigma_Y^2=0.2$, and not shown here, produced similar results. Notice that since our
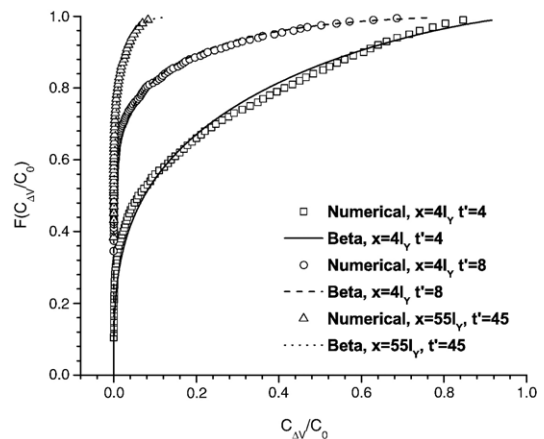

Fig. 3. Comparison between Beta model and the CDF of the solute concentration obtained numerically by Monte Carlo simulations at $x=4I_Y$ for $t=4I_Y/U$ and $8I_Y/U$ and at $x=55I_Y$ for $t=45I_Y/U$. Furthermore, $\Delta=0.2I_Y$, $\sigma_Y^2=1.2$, $Pe_L=200$ and $Pe_T=2000$, and the source area is a square of side $L=1.5I_Y$.

model (7) is not limited to weakly heterogeneous formations it comes as no surprise that it performs well also in moderately heterogeneous formations and it can be expected to perform well at higher heterogeneity of the flow field. Therefore, the model provides good predictions of the solute concentration CDFs for different heterogeneity of the formation, at different locations within the plume, and for different sampling dimensions.

4. Applications

The above model can be used in risk analysis and in all the applications requiring the probability that a threshold concentration is exceeded. In envisioning possible applications, one should consider that this model requires reliable estimates of $\langle C_{\Delta V} \rangle$ and $\sigma_{C_{\Delta V}}^2$ by either theoretical closed form solutions (see e.g., Kapoor and Gelhar, 1994; Zhang and Neuman, 1996; Dagan and Fiori, 1997; Kapoor and Kitanidis, 1999; Fiori and Dagan, 2000; Fiori, 2002), or numerical Monte Carlo simulations, which may be conditional to available measurements, thus alleviating the computational burden associated with the alternative approach of computing the concentration CDF directly from a large sample of Monte Carlo realizations. In fact, when the first two concentration moments are computed through Monte Carlo simulations the gain is in the smaller number of realizations needed to stabilize them with respect to the sample CFD.

In geostatistics several strategies have been elaborated in order to cope with the limited information provided by the measurements. For example, in the Simple Indicator Kriging (SIK) approach the prior information provided by the sample CFD is updated by a linear combination of indicators obtained from the measurements (see e.g. Goovaerts, 1997, p. 293). In doing that one assumes implicitly that the prior concentration CDF is statistically stationary with the consequence that the probability of exceeding a given threshold is assumed the same within the plume (Goovaerts, 1997; Goovaerts et al., 2005). Substituting the sample CFD with the Beta model of the concentration CDF may alleviate this problem providing more consistent prior information.

Apart from these applications sometimes it is useful to know the proportion of the plume volume where solute concentrations are above a given threshold, for example the solute concentration above which a remedial action should be taken. The Cumulative Frequency Distribution (CFD) obtained from concentration measurements is the simplest way of obtaining such information. However, a set of coarse measurements is typically available in

applications, such that the CFD is poorly characterized, in particular when the threshold is higher than the mean concentration. In order to gain confidence in the concentration CFD a suitable probability distribution is often adapted to the experimental data.

Furthermore, in the Bayesian approach to inverse problems one seeks for a set of model parameters **m** that maximizes their conditional distribution $p(\mathbf{m}|\mathbf{d})$ given the data **d**. According to the Bayes' theorem $p(\mathbf{m}|\mathbf{d})$ depends on the marginal distribution of the data $h(\mathbf{d})$, through the following expression: $p(\mathbf{m}|\mathbf{d}) = \mathfrak{T}(\mathbf{d}|\mathbf{m})\rho(m)/h(d)$, where $\mathfrak{T}$ is the conditional distribution of the data given the parameters and ρ is the marginal prior distribution of the parameters. For example, one may want to infer the most likely distribution of hydraulic conductivity from the information provided by concentration measurements, in which case $h(\mathbf{d})$ can be estimated by our model by placing $h(\mathbf{d})=f(Z)$.

In the remainder of this Section we show how the Beta model can be used to obtain the probability distribution of a sample of concentration measurements taken from the plume at a given time, i.e. a snapshot, that is used in both applications described above. In order to do that, we assume that the sample made of concentrations taken from the plume at a given time obeys the following SDE:

$$\mathrm{d}C_{\Delta V} = \kappa[\overline{C_{\Delta V}(t)} - C_{\Delta V}]\mathrm{d}t \\ + \sqrt{\epsilon C_{\Delta V}(C_0 - C_{\Delta V})}\mathrm{d}W(t), \qquad (20)$$

which is similar to the SDE (4) utilized for modeling the local concentration with the difference that the spatial average $\overline{C_{\Delta V}(t)}$ of the solute concentration replaces the ensemble average. We consider now that by definition $\overline{C_{\Delta V}(t)} = M(t)/V(t)$, where $M(t)$ is the total mass of solute and $V(t)$ is the volume occupied by the plume. For a plume larger than the scale of spatial variability of K, $V(t)$ varies slightly between independent realizations of K, and if, in addition to that, the plume is also well mixed the following approximation can be applied:

$$\overline{C_{\Delta V}(t)} \simeq \frac{1}{V_G(t)} \int\limits_{V_G(t)} \langle C_{\Delta V}(\boldsymbol{x}, t) \rangle \mathrm{d}\boldsymbol{x} \qquad (21)$$

and $V_G(t)$ is the volume occupied by a Gaussian plume with the same spatial moments of the actual plume. An indirect confirmation that the approximation introduced in Eq. (21) holds can be found in the paper by Fitts (1996), who showed that the spatial average of point concentrations above background concentration at both Cape Cod and Border tracer experiments is well approximated by the spatial mean of the solute

concentration obtained by solving the ADE with the mean velocity and macrodispersion coefficients inferred from concentration data (see the column "Mean" in Tables 3 and 4 of Fitts (1996)).

As in Eq. (4) the first right hand term of the SDE (20) epitomizes the effects of PSD and mechanical mixing within the sampling volume, which act to reduce spatial gradients of the concentration and thus drive the system toward a spatially constant concentration. It should be noted, however, that this is an approximation since strictly speaking the system evolves toward a uniform concentration only in a bounded domain, i.e. when the solute disperses within a constant volume (Kitanidis, 1994).

Under statistically stationary conditions the solution of the SDE (20) is given by the model (7) with the statistical moments replaced by the corresponding spatial moments, i.e. $\langle C_{\Delta V} (\boldsymbol{x}, t) \rangle$ with $\bar{C}_{\Delta V}(t) = \sum_{i=1}^{n_d} C_{\Delta V} (\boldsymbol{x}_i, t)/n_d$ and $\sigma_{\Delta V}^2(\boldsymbol{x}, t)$ with $S_{\Delta V}^2(t) = \sum_{i=1}^{n_d} [C_{\Delta V}(\boldsymbol{x}_i, t) - \bar{C}_{\Delta V}(t)]^2/(n_d - 1)$, where n_d is the number of measurements.

4.1. Analysis of the Cape Cod solute concentration data

The Cape Cod tracer experiment (Leblanc et al., 1991; Hess et al., 1992) provides a data set of solute concentrations, which is large enough for an experimental validation of the theoretical results presented in the previous section. A controlled amount of non-reactive solute, bromide, was released for 17 hours within a volume with dimensions of $1.2 \times 4 \times 4$ m, resulting in an initial concentration of $C_0 = 640$ mg/l. The thickness of the injection volume does not exceed 30 vertical integral scale of the hydraulic log-conductivity, I_{Y_v}, which has been estimated by Hess et al. (1992) in the range between 13 and 39 cm, such that ergodic conditions are only partially satisfied (Quinodoz and Valocchi, 1990; Dagan, 1991). Furthermore, Fitts (1996) showed that concentrations predicted by solving the classic ADE were inaccurate, although the plume thickness spans several vertical integral scales and the macrodispersion coefficients were obtained with high accuracy from the concentration measurements at a dense monitoring network of 656 multilevel samplers (MLS) (Leblanc et al., 1991, Figs. 8 and 9).

At Cape Cod, the background concentration of Bromide was estimated less than $C_{lim} = 0.1$ mg/l (Leblanc et al., 1991), such that concentrations smaller than this limit are removed from the sample because of the impossibility to distinguish between the two sources of solute mass when concentrations fall below this limit. With the lower limit of the population larger than zero, the variable Z should be redefined as follows: $Z = (C - C_{lim})/(C_0 - C_{lim})$, such that Z varies between 0 and 1 with zero

probability of observing solute concentrations smaller than C_{lim}.

4.1.1. Vertically averaged concentrations

First we analyze the vertically averaged concentration because it represents the type of information that most commonly is available in applications where sampling is performed in wells screened over the entire plume depth rather than in multilevel samplers as in the Cape Cod experiment.

Fig. 4a–e show the CFD of $Z(x_1, x_2) = (C_m(x_1, x_2) - C_{lim})/(C_0 - C_{lim})$ at 33, 83, 237, 426, and 511 days since injection, where $C_m(x_1, x_2)$ is the vertically averaged Bromide concentration at the monitoring well location $\boldsymbol{x} = (x_1, x_2)$. In all figures, the solid line shows the Beta CDF of Eq. (A.1) with the parameter β computed through Eq. (8), where $\langle Z \rangle$ and σ_Z^2 are approximated by the sample mean and variance, respectively. In each figure, the inset depicts the results in a semi-logarithmic scale that better evidences the discrepancy between the models and the experimental data for small concentrations. Furthermore, the dashed line shows the Log-Normal CDF,

$$P(Z) = \int_0^{\xi} \frac{1}{\varepsilon \sqrt{2\pi\sigma_\xi^2}} \exp\left[-\frac{(\varepsilon - \langle \xi \rangle)^2}{2\sigma_\xi^2} \right] d\varepsilon, \qquad (22)$$

where $\xi = \ln Z$. In Eq. (22)

$$\langle \xi \rangle = \exp\left[\langle Z \rangle + \sigma_Z^2/2 \right] \qquad (23)$$

and

$$\sigma_\xi^2 = \exp\left[2\langle Z \rangle + \sigma_Z^2 \right] \{ \exp[\sigma_Z^2] - 1 \} \qquad (24)$$

are the expected value and variance of the transformed variable written as a function of the first two statistical moments of Z. Likewise the Beta model, the Log-Normal CDF has been obtained by replacing the first two statistical moments of Z in Eqs. (23) and (24) with the sample mean and variance. Although unbounded, the Log-Normal distribution has been selected as a possible alternative to the Beta distribution because it is widely used in geostatistics for skewed random functions. Here, we do not consider the Normal distribution as a possible alternative to the Beta model because it resulted in a much poorer fit with the sample than the Log-Normal distribution.

In addition, we performed the best fitting of Beta CDF model (A.1) with the least-square method to the sample CFD obtaining the dot-dashed curve shown in Fig. 4a–e. With respect to matching the moments (moments method) the adaptation is better at low
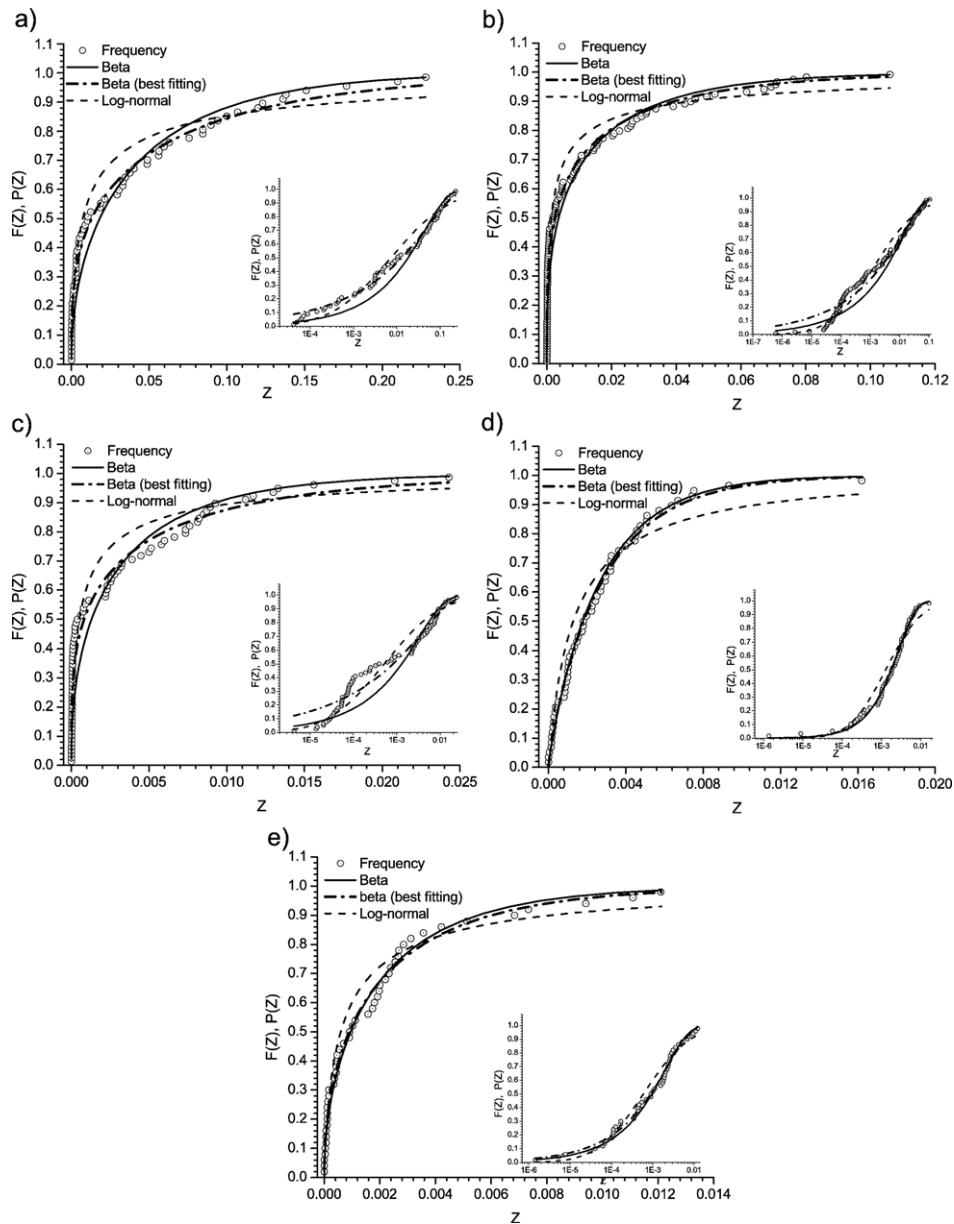
Fig. 4. Cumulative Frequency Distribution of the vertically averaged Bromide concentrations reported from the first Cape Cod tracer test at a) $t=33$, b) $t=83$, c) $t=237$, d) $t=426$, and e) $t=511$ days since injection. The Cumulative Frequency Distribution is compared with the Beta and Log-Normal CDF models. The dot-dashed line shows the best fitting of the Beta model with the sample CFD.

concentrations and poorer at large concentrations. The difference may be attributed to the more weight given by the least-square method to low concentration measurements, which outnumber large concentration measurements.

We applied the chi-square goodness-of-fit test to both Beta and Log-Normal distributions by using nine equiprobable classes and level of significance set at

0.05 (Conover, 1999, pg. 239), resulting in an upper limit of $\chi^2=12.59$ for the test statistic:

$$T = \sum_{j=1}^{nc} \frac{(O_j - E_j)^2}{E_j} \tag{25}$$

where nc$=9$ is the number of classes, O_j is the number of observations within class j, and E_j is the expected

number of observations within the same class. The results of the test are shown in columns three and seven of Table 1 for the Beta and Log-Normal distributions, respectively, and $C_{\lim}=0.1$ mg/l. At 203 days, none of the observations falls within the fourth class, which is then merged with the fifth class, reducing the total number of classes to 8, and χ^2 to 11.07. We found that the null hypothesis that Z is Log-Normally distributed is accepted at day 384, and rejected in all the other cases. On the other hand, the null hypothesis that Z is distributed according to the Beta model of Eq. (A.1) is accepted at day 33, 203, 273, 384, 426 and 511, and rejected in the remaining cases. With the only exception of day 461, note that when the null hypothesis is rejected for both distributions, the smallest T is obtained with the Beta distribution, which shows a better adaptation to the experimental data with respect to the Log-Normal model. The same test applied to the best fitting curve produced similar results with the only difference that the null hypothesis is accepted at the same days of the previous case plus day 111. This slightly better performance of the best fitting in term of test statistic is counterbalanced by a loss of accuracy in the probability of exceeding large concentrations, as shown in Fig. 4a–e. Therefore, subsequently we focus on the results obtained by

matching the first two moments of the probability model and the sample.

A closer inspection of Fig. 4b and c reveals that when the null hypothesis is rejected for the Beta model the mismatch between the model and the data is mostly concentrated at low Z values (see the inset of the figures). Consistently with our finding there are evidences that small concentration measurements may be inaccurate at Cape Cod. In a recent analysis of the Cape Cod data set, Fitts (1993) eliminated Bromide concentrations less than 0.3 mg/l with the justification that they were "substantially inaccurate due to chemical analysis method limitations", while Thierrin and Kitanidis (1994) eliminated all the concentrations smaller than 0.5 mg/l in their analysis of dilution.

By eliminating from the sample concentration measurements lower than $C_{\lim}=0.3$ mg/l because unreliable, as Fitts (1996) and Fiori and Dagan (1999) did in their studies, the null hypothesis that Z is Beta distributed is accepted in all cases it was accepted for $C_{\lim}=0.1$ mg/l plus days 83, 111, 139, 315, and 461 (the T statistic are reported in columns 6 and 10 of Table 1 for Beta and Log-Normal distributions, respectively). Increasing $C_{\lim}$ had a beneficial effect also on the Log-Normal distribution, but the null hypothesis was verified for a smaller number of days with respect to the

Table 1
Chi-square goodness-of-fit test for the beta and Log-Normal probability distribution models of the vertically averaged bromide concentration at Cape Cod

Test statistic, T									
Day	$F (C \leq$ 0.3 mg/l)	Beta distribution				Log-Normal distribution			
		T	T_1	$T_{1,R}$	T_{bf}	T	T_1	$T_{1,R}$	T_{bf}
33	0.15	11.45	6.06	0.53	9.25	19.64	0.97	0.05	12.79
55	0.20	15.42	3.29	0.21	14.63	16.26	1.51	0.09	13.79
83	0.35	33.63	25.15	0.75	7.97	26.00	15.61	0.60	12.41
111	0.31	21.75	12.62	0.58	8.68	19.75	2.50	0.13	14.08
139	0.50	57.37	32.63	0.57	12.46	56.33	22.31	0.40	20.42
174	0.27	27.56	14.44	0.52	21.64	40.53	5.59	0.14	23.08
203 [a]	0.18	10.78	4.08	0.38	9.95	14.63	0.04	0.00	18.05
237	0.48	37.20	25.39	0.68	13.55	33.70	17.84	0.53	15.8
273	0.12	5.39	2.29	0.43	7.02	15.68	0.01	0.00	13.63
315	0.37	28.90	13.89	0.48	10.48	37.45	5.14	0.14	18.76
349	0.16	12.62	4.02	0.32	12.73	26.79	0.73	0.03	27.96
384	0.13	5.48	1.45	0.26	7.61	8.09	0.23	0.03	7.9
426	0.17	2.84	0.44	0.15	3.36	16.42	0.30	0.02	7.96
461	0.20	14.36	2.32	0.16	12.46	12.16	3.38	0.28	6.46
511	0.31	9.59	6.03	0.63	8.88	11.43	1.58	0.14	10.47

T_1 is the contribution to the T statistic of concentrations smaller than 0.3 mg/l for $C_{\lim}=0.1$ mg/l, and T_{bf} is the T statistic for $C_{\lim}=0.3$ mg/l. Furthermore, $T_{1,R}=T_1/T$ is the relative contribution of concentrations smaller than 0.3 mg/l to the T statistic, and $F (C \leq 0.3$ mg/l) is the sample CFD for $C=0.3$ mg/l.
[a] Obtained by merging the classes number 4 and 5 in one class, such that the number of classes reduces to 8 and consequently $\chi^2=11.07$.

Beta distribution and, however, in cases when it was accepted for both distributions the smallest T statistic was that of the Beta model (see Table 1). In none of the days the null hypothesis was accepted for Log-Normal distribution and rejected for the Beta distribution.

In order to evaluate the impact of measurement errors further on the concentration pdf, we have set the threshold between doubtful and reliable concentrations

at $C_1 = 0.3$ mg/l, and computed the relative contribution to T of classes below C_1:

$$T_{1,R} = \frac{T_1}{T} = \frac{1}{T} \sum_{j=1}^{nc_1} \frac{(O_j - E_j)^2}{E_j} \qquad (26)$$

where nc_1 is the number of classes with the upper bound smaller than $Z_1 = (C_1 - C_{lim})/(C_0 - C_{lim}) = 0.0003125$,
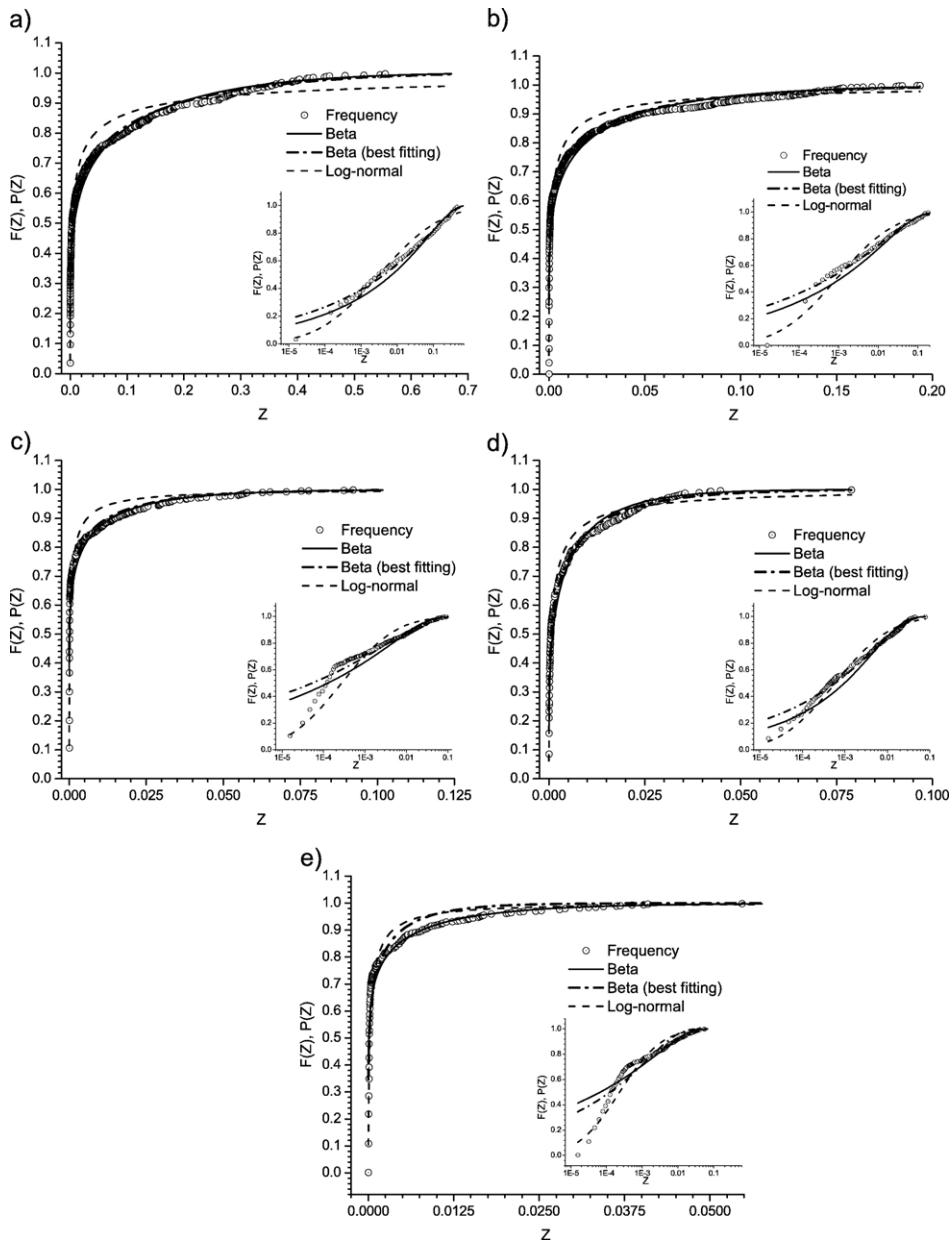

Fig. 5. Cumulative frequency distribution of the point concentration of Bromide from the first Cape Cod tracer test at a) $t=33$, b) $t=83$, c) $t=237$, d) $t=426$ and, e) $t=511$ days since injection. The Cumulative Frequency Distribution is compared with the Beta and Log-Normal CDF models. The dot-dashed line shows the best fitting of the Beta model with the sample CFD.

incremented by one if Z_1 is located in the upper half of the class containing it.

The results of this analysis are shown in columns five and nine of Table 1 for the Beta and Log-Normal models, respectively. Additionally, columns four and eight show the corresponding values of T_1. With the exception of day 461, when both models are rejected and the smallest T is that of the Log-Normal model, the largest contribution to T of classes $Z \le Z_1$ is that of the Beta model, and in 7 out of 15 cases this contribution exceeds 50% of T. In other words, when the null hypothesis is rejected for the Beta model, it is because of the poor adaptation to concentrations smaller than 0.3 mg/l, which are negatively impacted by measurement and analytical errors. A treatment of the measurement error, although theoretically possible, is beyond the scope of this work, suffice here to know that it impacts concentrations smaller than 0.3 mg/l negatively as evidenced in other studies.

We conclude that the Beta model of Eq. (A.1) is superior to the Log-Normal model of Eq. (22) in describing the probability distribution of vertically averaged Bromide concentrations at Cape Cod, and that its performance improves significantly at high concentrations. On the other hand, small concentrations are described better by the Log-Normal distribution, but this may be a consequence of the larger relative impact of measurement errors, which obscure the underlying probability distribution.

4.1.2. Point concentration

We repeated the analysis of the previous section for Bromide concentrations measured at the ports of the multilevel samplers. Each sampling port consists of a polyethylene tube with internal diameter of 0.47 cm (Leblanc et al., 1991). Since the sampler's size is one order of magnitude smaller than I_{Y_s}, concentration measurements at the sampling ports can be thought as point concentrations.

Fig. 5a–e show the CFD of $Z(x_1, x_2, x_3) = (C(x_1, x_2, x_3) - C_{lim})/(C_0 - C_{lim})$, where $C(x_1, x_2, x_3)$ is the Bromide concentration at the port's location (x_1, x_2, x_3) and $C_{lim} = 0.1$ mg/l. Although, both Beta and Log-Normal models fail the χ^2 goodness-of-fit test with χ^2 exceeding the 95th percentile, the Beta distribution follows the CFD closely at moderate to large concentrations, while as in the vertically averaged concentrations the Log-Normal distribution seems preferable, yet not ideal, at small concentrations (see the inset of Fig. 5a–e).

By assuming $C_{lim} = 0.3$ mg/l, the null hypothesis that point concentrations are distributed as the Beta model was accepted for days 203, 237, 273, 384 and 511, while

in all other cases the T statistic was much smaller than that obtained with $C_{lim} = 0.1$ mg/l. Contrarily, the null hypothesis that the concentration is Log-Normally distributed was accepted only for day 111.

In the light of these results, we conclude that the Beta distribution is a good model of the spatial variability for the Bromide concentration observed at the first Cape Cod tracer test. In particular, the Beta distribution models the experimental CFD accurately at high concentrations, but it is less satisfactory at low concentrations. However, chemical analysis errors have been reported by Hess et al. (1992) and Fitts (1996) to influence low concentration measurements negatively, such that the mismatch between CFD and Beta distribution function (Eq. (A.1)) should not be entirely attributed to our model's limitations. Conversely, the Log-Normal model performs better, yet not satisfyingly, at low concentrations, but poorly at high concentrations. The Normal model performs poorly in all cases. Overall, the Beta model is then preferable over the alternative Log-Normal model.

4.2. Numerical validation of the spatial CDF

Simulations are performed in a two-dimensional formation by using a single realization of the transmissivity field, which we assumed as the "real" field. The resulting concentration distributions at selected times since injection (snapshots), obtained by solving solute transport with the FPT scheme of Eqs. (13) and (14), are compared with the Beta model (7). Concentration measurements composing the sample are taken at the center of a grid with horizontal dimension equal to the
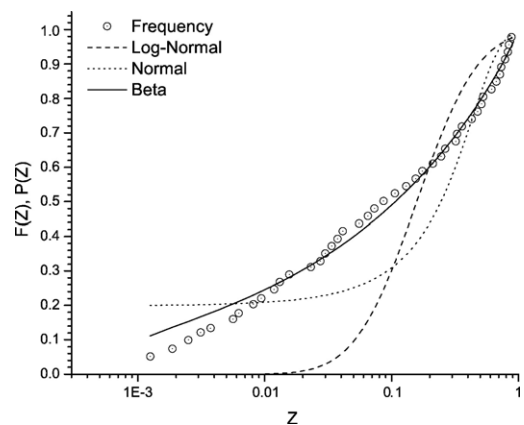

Fig. 6. Cumulative frequency distribution of solute concentrations obtained from numerical simulations of a field-scale tracer test, at $t = 4I_Y/U$ since injection. The Cumulative Frequency Distribution is compared with the following three CDF models: Beta, Normal and Log-Normal.

horizontal projection of ΔV. We considered two snapshots: one at an early time since injection, and the other at a later time.

Fig. 6 compares the experimental CFD of $Z=(C_{\Delta V}-C_{\Delta V,\text{min}})/(C_0-C_{\Delta V,\text{min}})$ at time $t=4I_Y/U$, obtained from the numerical simulation of the tracer experiment with sampling size $\Delta=0.2I_Y$, with the Beta model (A.1) and two additional models widely used in the theory of random functions: the Log-Normal model (22) and the Gaussian (Normal) model. Fig. 7 repeats Fig. 6 for $t=55I_Y/U$. For each model the parameters are computed by matching the first two statistical moments with the corresponding moments of the sample. For the Beta model the parameter β is computed by substituting the sample mean and variance of Z into Eq. (8).

Likewise the Cape Cod data analysis presented in Section 3, the Beta model (A.1) performs much better than the other two models. Capitalizing on the information provided by the first two moments of the solute concentration, the model (A.1) provides an accurate description of the concentration CFD. At low concentrations, the Beta model performs much better with the data produced by the simulated tracer experiment than with those obtained in the field, suggesting that the latter may be negatively affected by the chemical analysis error reported by Hess et al. (1992) and Fitts (1996), as we argued in the previous section. Furthermore, both Normal and Log-Normal models underestimate the probability of exceeding high concentrations, which may impact risk assessment negatively.

We performed the chi-square and Kolmogorov–Smirnov goodness-of-fit tests (Conover, 1999, pg.428) for the three models shown in Figs. 6 and 7, and for the snapshot at $t=28I_Y/U$ not shown in the figures.
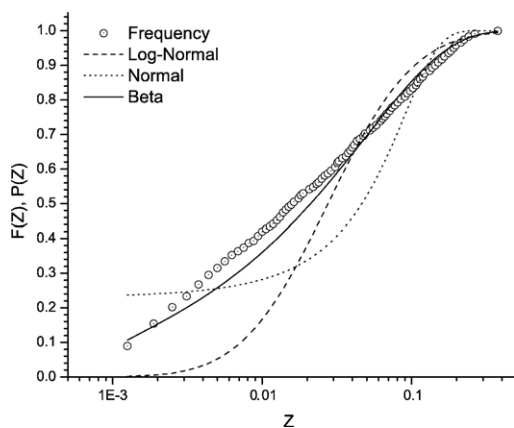

Fig. 7. Cumulative Frequency Distribution of solute concentrations obtained from numerical simulations of a field-scale tracer test, at $t=55I_Y/U$ since injection. The Cumulative Frequency Distribution is compared with the following three models: Beta, Normal and Log-Normal.

Although, strictly speaking all the above models are rejected by the chi-square goodness of fit test at 0.05 level of significance, the Beta model performs much better than both Log-Normal and Normal models, confirming what was already evident from the visual inspection of Figs. 6 and 7. Similar considerations are suggested by the Kolmogorov–Smirnov test with the only difference that the null hypothesis that the concentration CDF follows the Beta model is in this case accepted for $t=4I_Y/U$ and rejected for both Normal and Log-Normal models.

Simulations with a larger sampling volume of side $\Delta=1I_Y$ have similar results with the Beta distribution outperforming the Normal and Log-Normal models. However, in this case the chi-square goodness of fit test is met by the Beta model at both $t=4\,I_Y/U$ and $t=55I_Y/U$ and rejected at $t=28I_Y/U$, whereas Kolmogorov–Smirnov test is always satisfied by the Beta, but not by the Normal and Log-Normal distributions with the exception of the Normal model at time $t=4I_Y/U$.

5. Conclusions

We developed a new model for the concentration pdf of a conservative tracer migrating in a heterogeneous aquifer. The model is based on a Stochastic Differential Equation describing the evolution of the plume subjected to spreading, pore-scale dispersion, and mechanical mixing associated with the sampler volume. Under statistical stationarity, we show that the local solute concentration is distributed according to the Beta distribution, whose parameters depend on the first two statistical moments of the concentration. Similarly, the probability distribution of the solute concentration in a single realization is well described by the Beta distribution with the first two statistical moments replaced by the corresponding spatial moments. The former is a model of local uncertainty, i.e. it describes the variability of the solute concentration at a given location, the latter can be seen as a model of global uncertainty describing the probability of exceeding a given concentration irrespective of the location within the plume.

Analysis of the Bromide concentration measurements from the first Cape Cod tracer experiment confirmed that our model is a valuable tool for interpreting concentration data, as well as to quantify the uncertainty resulting from modeling solute transport with a limited amount of information on the spatial distribution of the hydraulic conductivity. In particular, our model captures the smoothing effect of the sampling volume accurately predicting the reduction of the probability of exceeding high concentrations, when larger sampling volumes are adopted. The relatively poorer accuracy in modeling the

spatial distribution of small concentrations may be either a model limitation, or a direct consequence of chemical analysis errors in the concentration data. The Log-Normal model performs slightly better than the Beta model for low concentrations for Cape Cod data, but at price of much poorer performance at high concentration levels with a significant negative impact on risk assessment.

Numerical simulations confirmed that our model is superior to both Log-Normal and Normal models in representing local and global uncertainty arising from the incomplete knowledge of spatial variability of the permeability. Similarly to Cape Cod data, the probability of exceeding high concentrations is well predicted, while the accuracy deteriorates slightly at low concentrations, with the limit between high and low concentrations at $(C-C_{lim})/(C_0-C_{lim})=0.01$ and 0.03 for $t=4I_Y/U$ and $55I_Y/U$, respectively. Thus, the range of concentrations for which our model shows high accuracy in predicting the probability of exceeding a given threshold reduces with the elapsed time since injection.

Acknowledgments

Appendix A. Solution of the concentration pdf for $\beta \rightarrow \infty$

Let us consider the limit of the concentration pdf (7) for $\beta \rightarrow \infty$. To handle this case properly, it is convenient to consider the CDF of the solute concentration:

$$
\begin{aligned}
F(z) = \mathrm{Prob}[Z \leq z] &= \int_0^z f[\zeta]\mathrm{d}\zeta \\
&= \frac{\Gamma[\omega]}{\Gamma[\omega\langle Z\rangle]\Gamma[\omega(1-\langle Z\rangle)]} B_z[\omega\langle Z\rangle, \omega(1-\langle Z\rangle)]
\end{aligned}
$$
(A.1)

where $B_z[a, b] = \int_0^z t^{a-1}(1-t)^{b-1}\mathrm{d}t$ is the incomplete Beta function (Abramowitz and Stegun, 1972, p. 258), and $\omega = 1/\beta$.

For $\omega \rightarrow 0$ the function Eq. (A.1) is discontinuous at $z \rightarrow 0$ and $z \rightarrow 1$. The behavior of the CDF (A.1) in proximity of these discontinuities can be studied by means of Taylor expansion. For $z=0$ we obtain:

$$
F(\varepsilon) = \varepsilon^{\omega\langle Z\rangle}\gamma\left\{\frac{1}{\omega\langle Z\rangle} + \frac{1-\omega(1-\langle Z\rangle)}{1+\omega\langle Z\rangle}\varepsilon + O\left(\frac{\varepsilon^2}{\gamma}\right)\right\}
$$
(A.2)

where $\gamma = \Gamma[\omega]/(\Gamma[\omega\langle Z\rangle]\Gamma[\omega(1-\langle Z\rangle)])$, in which $\Gamma[x] = \int_0^\infty t^{x-1}e^{-t}\mathrm{d}t$ is the Euler gamma function, and $z=\varepsilon \ll 1$. By taking the limit of the expansion (Eq. (A.2)) for $\omega \rightarrow 0$ we obtain:

$$
\lim_{\omega\to 0} F(\varepsilon) = 1 - \langle Z\rangle
$$
(A.3)

Thus, $F(\varepsilon) = 1-\langle Z\rangle$ no matter how small $\varepsilon > 0$ is, whereas it jumps to zero for $z=0$.

Expanding now the Eq. (A.1) around $z=1$ we obtain:

$$
F(1-\varepsilon) = 1 - \gamma\varepsilon^{\omega(1-\langle Z\rangle)} + \gamma(1-\langle Z\rangle)\varepsilon^{\omega(1-\langle Z\rangle)+1} + \dots
$$
(A.4)

which for $\omega \rightarrow 0$ converges to

$$
\lim_{\omega\to 0} F(1-\varepsilon) = 1 - \langle Z\rangle
$$
(A.5)

The discontinuity in $z=1$ is then similar to that observed in $z=0$, with $F(1-\varepsilon) = 1-\langle Z\rangle$ regardless how small ε is, while $F(z=1)=1$. Whereas, $F(z)$ is constant and equal to $1-\langle Z\rangle$ between $z=0$ and $z=1$. Thus, for $\omega \rightarrow 0$ ($\beta \rightarrow \infty$) the concentration CDF assumes the following form:

$$
\lim_{\omega\to 0} F(z) = [1 - \langle Z\rangle]H(z) + \langle Z\rangle H(z-1)
$$
(A.6)

where H is the Heaviside step function (Abramowitz and Stegun, 1972), and the concentration pdf (Eq. (7)) reduces to the expression (9) obtained by Dagan (1982) with a different reasoning.

Appendix B. Moments of the RF Z for small β values

The normalized statistical moments of order $v>2$, given by the expression (11), can be rewritten by means of the definition of hypergeometric function (Gradshteyn and Ryzhik, 1980, p. 1045):

$$
\begin{aligned}
\mu_v' = (-1)^v\langle Z\rangle^v&\left[\langle Z\rangle(1-\langle Z\rangle)\frac{\beta}{1+\beta}\right]^{-v/2} \\
&\times \sum_{k=0}^{v}\frac{(-v)_k(\langle Z\rangle/\beta)_k}{(1/\beta)_k}\frac{1}{k!\langle Z\rangle^k}
\end{aligned}
$$
(B.1)

where $(a)_k = \Gamma[a+k]/\Gamma[a]$ is the rising factorial and $\Gamma[a] = \int_0^\infty t^{a-1}e^{-t}\mathrm{d}t$ is the Euler gamma function.

By expanding the rising factorials in Eq. (B.1) in Taylor series around $\beta=0$, we obtain:

$$
\begin{aligned}
\mu_v' = (-1)^v\langle Z\rangle^v&\left[\langle Z\rangle(1-\langle Z\rangle)\frac{\beta}{1+\beta}\right]^{-v/2} \\
&\times\left[1 + \sum_{k=1}^{v}(-1)^k\frac{v(v-1)\dots(v-k+1)}{k!}\frac{\Psi_k(\langle Z\rangle, \beta)}{\langle Z\rangle^k}\right]
\end{aligned}
$$
(B.2)

with

$$\Psi_k(\langle Z \rangle, \beta) = \frac{[\langle Z \rangle + (k-1)\beta][\langle Z \rangle + (k-2)\beta]\ldots[\langle Z \rangle + \beta]\langle Z \rangle}{[1 + (k-1)\beta][1 + (k-2)\beta]\ldots[1 + \beta]}$$
(B.3)

In order to study the structure of Eq. (B.2) it is convenient to treat separately the odd and even moments. When ν is odd, let say $\nu = 2j+1$, Eq. (B.2) assumes the following form:

$$\mu'_{2j+1} = \beta^{1/2} f_{2j+1,0}(\langle Z \rangle, \beta) + \beta^{3/2} f_{2j+1,1}(\langle Z \rangle, \beta) + \ldots$$
$$+ \beta^{(2j+1)/2} f_{2j+1,j}(\langle Z \rangle, \beta), \quad j = 1, 2, \ldots. \quad (B.4)$$

and the functions $f_{2j+1,k}$, $k = 0, 1, \ldots, j$, which depend on both β and $\langle Z \rangle$, share the important property that for $\beta \to 0$ they are function of $\langle Z \rangle$ only. A direct consequence of this property is that $\mu'_{2j+1} \to 0$ for $\beta \to 0$.

When ν is even, let say $\nu = 2j$, Eq. (B.2) assumes the following form:

$$\mu'_{2j} = \frac{(2j-1)!!}{[1+2\beta][1+3\beta]\ldots[1+(2j-1)\beta]}$$
$$+ \beta g_{2j,1}(\langle Z \rangle, \beta) + \beta^2 g_{2j,2}(\langle Z \rangle, \beta) + \ldots \quad (B.5)$$
$$+ \beta^{2j-2} g_{2j,2j-2}(\langle Z \rangle, \beta), \quad j = 1, 2, \ldots.$$

Similarly to the previous case, the functions $g_{2j,k}$, $k = 1, 2, \ldots, 2j-2$ depend on both $\langle Z \rangle$ and β, but reduce to a polynomial in $\langle Z \rangle$ for $\beta \to 0$. This leads us to conclude that $\lim_{\beta \to 0} \mu'_{2j} = (2j-1)!!$.

The Eqs. (B.4) and (B.5) show that for $\beta \ll 1$ the difference between the moments of Z and those of a Normal distribution with the same first two moments reduces to zero as $\beta \to 0$ with $\sqrt{\epsilon}$ and ϵ as leading term for the odd and even moments, respectively.

References

Abramowitz, M., Stegun, I.A., 1972. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th Printing. Dover, New York.

Andricevic, R., 1998. Effects of local dispersion and sampling volume on the evolution of concentration fluctuations in aquifers. Water Resour. Res. 34 (5), 1115–1129.

Arfken, G., 1985. Mathematical Methods for Physicists, 3rd edition. Academic Press, Orlando, Florida.

Bellin, A., Rubin, Y., 1996. HYDRO_GEN: a spatially distributed random field generator for correlated properties. Stoch. Hydrol. Hydraul. 10 (4), 253–278.

Bellin, A., Salandin, P., Rinaldo, A., 1992. Simulation of dispersion in heterogeneous porous formations: statistics, first-order theories, convergence of computations. Water Resour. Res. 28 (9), 2211–2227.

Bellin, A., Rubin, Y., Rinaldo, A., 1994. Eulerian–Lagrangian approach for modeling of flow and transport in heterogeneous geological formations. Water Resour. Res. 30 (11), 2913–2924.

Caroni, E., Fiorotto, V., 2005. Analysis of concentration as sampled in natural aquifers. Transp. Porous Media 59, 19–45.

Cassiani, M., Franzese, P., Giostra, U., 2005. A PDF micromixing model of dispersion for atmospheric flow: 1. Development of the model, application to homogeneous turbulence and to neutral boundary layer. Atmos. Environ. 39, 1457–1469.

Chatwin, P.C., Sullivan, P.J., 1990. A simple and unifying physical interpretation of scalar fluctuation measurements from many turbulent shear flows. J. Fluid Mech. 212, 533–556.

Chatwin, P.C., Lewis, D.M., Sullivan, P.J., 1995. Turbulent dispersion and the beta distribution. Environmetrics 6, 395–402.

Cobb, L., 1981. Stochastic differential equations for the social sciences. In: Cobb, Loren, Thrall, Robert M. (Eds.), Mathematical Frontiers of the Social and Policy Sciences. Westview Press, Inc., Boulder, Colorado, pp. 37–68. Ch. 2.

Conover, W.J., 1999. Practical Nonparametric Statistics, Third edition. John Wiley and Sons, New York.

Dagan, G., 1982. Stochastic modeling of groundwater flow by unconditional and conditional probabilities, 2. The solute transport. Water Resour. Res. 18 (4), 835–848.

Dagan, G., 1991. Dispersion of a passive solute in non-ergodic transport by steady velocity fields in heterogeneous formations. J. Fluid Mech. 233, 197–210.

Dagan, G., Fiori, A., 1997. The influence of pore-scale dispersion on concentration statistical moments in transport through heterogeneous aquifers. Water Resour. Res. 33 (7), 1595–1605.

Dopazo, C., O'Brian, E.E., 1974. An approach to the autoignition of a turbulent mixture. Acta Astronaut. 1, 1239–1266.

Fiori, A., 2001. On the influence of local dispersion in solute transport through formations with evolving scales of heterogeneity. Water Resour. Res. 37 (2), 235–242.

Fiori, A., 2002. An asymptotic analysis for determining concentration uncertainty in aquifer transport. J. Hydrol. 284, 1–12.

Fiori, A., Dagan, G., 1999. Concentration fluctuation in transport by groundwater: comparison between theory and experiments. Water Resour. Res. 35 (1), 105–112.

Fiori, A., Dagan, G., 2000. Concentration fluctuations in aquifer transport: a rigorous first-order solution and applications. J. Contam. Hydrol. 45 (1–2), 139–163.

Fiori, A., Berglund, S., Cvetkovic, V., Dagan, G., 2002. A first-order analysis of solute flux statistics in aquifers: the combined effect of pore-scale dispersion, sampling, and linear sorption kinetics. Water Resour. Res. 38 (8). doi:10.1029/2001WR000678.

Fiorotto, V., Caroni, E., 2002. Solute concentration statistics in heterogeneous aquifers for finite Pèclet values. Transp. Porous Media 48, 331–351.

Fisher, H.B., List, E.J., Koh, R.C.Y., Imberger, J., Brooks, N.H., 1979. Mixing in Inland and Coastal Waters. Academic, San Diego, Calif.

Fitts, C.R., 1996. Uncertainty in deterministic groundwater transport models due to the assumption of macrodispersive mixing: evidence from Cape Cod (Massachusetts, U.S.A. and Borden (Ontario, Canada) tracer tests. J. Contam. Hydrol. 23, 69–84.

Gardiner, C.W., 1985. Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences. Springer, Berlin.

Gelhar, L.W., 1993. Stochastic Subsurface Hydrology. Prentice Hall, Englewood Cliffs, New Jersey.

Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press, New York.

Goovaerts, P., AvRuskin, G., Meliker, J., Slotnick, M., Jacquez, G., Nriagu, J., 2005. Geostatistical modeling of the spatial variability of arsenic in groundwater of southeast Michigan. Water Resour. Res. 341 (7).

Gradshteyn, I.S., Ryzhik, I.M., 1980. Table of Integrals, Series, and Product. Academic Press, Inc., Orlando, Florida.

Hess, K.M., Wolf, S.H., Celia, M.A., 1992. Large-scale natural gradient tracer test in sand and gravel, Cape Cod, Massachusetts, 3. Hydraulic conductivity variability and calculated macrodispersivities. Water Resour. Res. 28 (8), 2011–2017.

Kapoor, V., Gelhar, L.W., 1994. Transport in three-dimensionally heterogeneous aquifers, 1. Dynamics of concentration fluctuations. Water Resour. Res. 30 (6), 1775–1788.

Kapoor, V., Kitanidis, P.K., 1998. Concentration fluctuations and dilution in aquifers. Water Resour. Res. 34 (5), 1181–1193.

Kitanidis, P.K., 1988. Prediction by the method of moments of transport in a heterogeneous formation. J. Hydrol. 102, 453–473.

Kitanidis, P.K., 1994. The concept of the dilution index. Water Resour. Res. 30 (7), 2011–2026.

Leblanc, D.R., Garabedian, S.P., Hess, K.M., Gelhar, L.W., Quadri, R.D., Stollenwerk, K.G., Wood, W.W., 1991. Large-scale natural gradient tracer test in sand and gravel, Cape Cod, Massachusetts: 1. Experimental design and observed tracer movement. Water Resour. Res. 27 (5), 895–910.

Pannone, M., Kitanidis, P.K., 1999. Large-time behavior of concentration variance and dilution in heterogeneous media. Water Resour. Res. 35, 623–634.

Pope, S.B., 1965. Turbulent Flow. Cambridge University Press.

Quinodoz, H.A.M., Valocchi, A.J., 1990. Macrodispersion in heterogeneous aquifers: numerical experiments. In: Moltyaner, G. (Ed.), Transport and Mass Exchange Processes in Sand and Gravel Aquifers: Field and Modeling Studies. Atomic Energy of Canada Limited, Chalk River Ont., Ottawa. October.

Rubin, Y., Bellin, A., 1998. Conditional simulation of geologic media with evolving scales of heterogeneity. In: Sposito, G. (Ed.), Scale Dependence and Scale Invariance in Hydrology. Cambridge University Press, New York, pp. 398–420. Ch. 14.

Thierrin, J., Kitanidis, P.K., 1994. Solute diluition at the Borden and Cape Cod groundwater tracer tests. Water Resour. Res. 30 (11), 2883–2890.

Vanderborght, J., 2001. Concentration variance and spatial covariance in second-order stationary heterogeneous conductivity fields. Water Resour. Res. 37 (7), 1893–1912.

Villermaux, J., Devillon, J.C., 1972. Représentation de la coalescence et de la redispersion des domaines de ségrégation dans un fluide par un modéle d'interaction phénoménologique. In: Elsevier (Ed.), Second International Symposium On Chemical Reaction Engineering, New York.

Wandel, A.P., Smith, N.S.A., Klimenko, A.Y., 2003. Verification of Markov hypothesis for conserved scalar process: validation of conditional moment closure for turbulent combustion. ANZIAM J. 44, C802–C819.

Zhang, D., Neuman, S.P., 1996. Effect of local dispersion on solute transport in randomly heterogeneous media. Water Resour. Res. 32 (9), 2715–2723.