Advances in Water Resources 36 (2012) 108–120

Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres

A Bayesian approach to integrate temporal data into probabilistic risk analysis of monitored NAPL remediation

Daniel Fernàndez-Garcia [a,*], Diogo Bolster [b], Xavier Sanchez-Vila [a], Daniel M. Tartakovsky [c]

[a] Technical University of Catalonia, Barcelona, Spain
[b] Department of Civil Engineering and Geological Sciences, University of Notre Dame, Indiana, USA
[c] Department of Mechanical and Aerospace Engineering, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA

ARTICLE INFO

Article history:
Available online 2 August 2011

Keywords:
Risk assessment
Fault tree analysis
Monitoring systems
Remediation
DNAPL

ABSTRACT

Upon their release into the subsurface, non-aqueous phase liquids (NAPLs) dissolve slowly in groundwater and/or volatilize in the vadose zone threatening the environment and public health over extended periods of time. The failure of a treatment technology at any given site is often due to the unnoticed presence of dissolved NAPL trapped in low permeability areas and/or the remaining presence of substantial amounts of pure phase NAPL after remediation efforts. The design of remediation strategies and the determination of remediation endpoints are traditionally carried out without quantifying risks associated with the failure of such efforts. We conduct a probabilistic risk analysis (PRA) to estimate the likelihood of failure of an on-site NAPL treatment technology. The PRA integrates all aspects of the problem (causes, pathways, and receptors) without resorting to extensive modeling. It accounts for a combination of multiple mechanisms of failure of a monitoring system, such as bypassing, insufficient sampling frequency and malfunctioning of the observation wells. We use a Bayesian framework to update the likelihood of failure of the treatment technology with observed measurements of concentrations at nearby monitoring wells.

1. Introduction

The ubiquity of subsurface NAPL contamination has led to a rapid growth in the number of alternative remediation strategies. While most of them perform well in controlled laboratory conditions, many often fail in the field. The main reasons of such failures are the uncertainty introduced by subsurface heterogeneity and the presence of unidentified rate-limited mass transfer processes. The impact of uncertainty on both the characterization of NAPL plumes and the relative performance of alternative remediation strategies is well documented [36,41]. However, quantification of this uncertainty and its propagation through a modeling effort are still in their infancy. Understanding this is key to identifying an optimal remediation strategy and to other aspects of decision making [2,24].

Nonlinearities in the equations used to predict the subsurface migration and fate of NAPLs along with the large number of uncertain parameters entering these equations, complicate both uncertainty quantification and decision-making. A standard approach for quantifying uncertainty in such complex phenomena is to treat relevant flow and transport parameters as random fields, which renders corresponding governing equations stochastic [5,8,43]. Solving these equations, either by stochastic averaging or with Monte Carlo simulations (MCS), typically yields an ensemble mean and variance of a NAPLs concentration in the subsurface. The former provides predictions of "likely" NAPL behavior (migration and fate), and the latter serves as a measure of predictive uncertainty.

Such solutions are often insufficient, since many environmental regulations are formulated in terms of probabilities of a contaminant concentration exceeding a certain mandated value. The first two statistical moments of a NAPLs concentration (its mean and variance) do not provide such information, because the concentration statistics are typically highly non-Gaussian [33,40,10]. Instead, full probability density functions (PDFs) of NAPL concentrations have to be computed. While MCS can in principle be used to compute the spatio-temporal evolution of the NAPLs concentration histograms (and, hence, corresponding PDFs), a large number of uncertain parameters, and uncertainty about the locations and strength of multiple sources, potential receptors and pathways of exposure, can render MCS computationally unfeasible.

Probabilistic risk analyses (PRAs) [44,47,1] alleviate the high-dimensionality problem by invoking a system approach to decontamination efforts. We develop a PRA framework to evaluate the risk of failure of a generic remediation effort. The framework links the various system components (exposure, sources and pathways)

* Corresponding author.
E-mail address: daniel.fernandez.g@upc.edu (D. Fernàndez-Garcia).

0309-1708/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.advwatres.2011.07.001

with their fate and transport phenomena. The method allows one to update the probability of failure given the observed real-time measurements of concentrations. This is achieved by integrating a Bayesian interpretation of the system into the PRA.

In Section 2 we construct a representative fault tree (which lies at the foundation of our PRA) for a generic NAPL decontamination effort. Section 3 contains a probabilistic representation of this fault tree in terms of Boolean algebra, which enables one to express the probability of system's failure through the combination of the probabilities of failure of its constitutive parts (basic events). The latter probabilities are computed in Section 4. In Section 5 we outline a Bayesian methodology to refine these probabilities by incorporating available data and in Section 6, we apply our general framework to a representative synthetic NAPL decontamination scenario. The main results of our analysis are summarized in Section 7.

2. Fault tree analysis

2.1. A general pollution problem

A plethora of methods to conduct a probabilistic risk assessment (PRA) of a system include failure mode and effect analysis (FMEA), multi-barrier analysis (MBA), event tree analysis (ETA), and fault tree analysis (FTA). While the goal of such approaches is the same, their philosophy is slightly different. For example, ETA and MBA methods are built on the concept of *forward logic* where for a system to fail an initiating event must occur and then be followed by a series of consequent events. In order for a system to fail under an ETA, an *accident sequence* must occur. On the other hand FTA is built on *backward logic*. In this case one identifies a particular failure of the system and defines it as the top event. Then one seeks the combination of all possible events that may contribute to this failure. Using Boolean operations 'AND', 'OR' and 'NOT', one can write down which events or combination of events lead to the top event and thus failure.

We focus on the FTA approach [44]. In particular, in this section we develop a general methodology to construct fault trees for contamination problems. In order to do so we must first define the meaning of system failure and the events that lead to it. In general, a contamination problem involves a mixture of contaminants $\{C_1, \ldots, C_n\}$ coming from several sources and moving towards several receptors $\{\Omega_1, \ldots, \Omega_m\}$. The system failure is defined as the event that the concentration of any of the contaminants surpasses some critical value in any given receptor within a legally mandated time interval $t \leqslant T$. Defining the event CC_{ij} as surpassing the critical contaminant concentration C^*_{ij} for pollutant i and receptor j

$$CC_{ij} = \left\{ C_i(\mathbf{x} \in \Omega_j, t \leqslant T) > C^*_{ij} \right\}, \tag{1}$$

we can formally write system failure as the occurrence of any of these events, i.e.,

$$SF = \{CC_{11}\} \cup \{CC_{12}\} \cup \cdots \cup \{CC_{nm}\}. \tag{2}$$

For event CC_{ij} to take place, the following basic events (or subevents) must occur:

- **Contaminant sources** – a given contaminant must be present at a particular source. We denote this combination contaminant/source with a subscript i. In many cases we cannot be certain that a contaminant source exists, but rather suspect its potential presence (e.g., in a highly agricultural zone there is a significant probability that a contaminant source related to particular pesticides commonly used in the area may exist) and we must deal with a probabilistic approach. The occurrence of this event is denoted by CS_i.
- **Potential receptors** – a receptor j must be susceptible to adverse impact by any of the contaminants. Receptors can include individual people, wildlife, water reservoirs or environmentally sensitive zones. We abbreviate this event as PR_j.
- **Pathways** – a path p connecting the contaminant source i with the receptor j must potentially exist. This includes natural flow fields (homogeneous or heterogeneous), preferential flow paths, diffusive paths, capture zones, and so on. We abbreviate these events as PW_{ijp}, where p is the pathway subscript.
- **Fate and transport** – mechanisms of natural attenuation or remediation have not reduced the contaminant concentration sufficiently along the p pathway. We abbreviate these events as FAT_{ijp}.

A generic fault tree for any number of contaminant sources, receptors, pathways, and processes provides a graphic representation of all the events that must occur for the system to fail. Fig. 1 depicts such a general tree. This fault tree reveals that for the system to fail two important mechanisms must occur: a pathway must connect the contamination source with a given receptor, and the transport processes occurring along these pathways cannot preclude the adverse effects of contamination. All processes taking place along a pathway are combined within a single FAT_{ijp} event. Although not strictly necessary, this is done to favor a compact mathematical description of all processes occurring along one pathway event.

A generic fault tree consists of different potential events, whose inter-connections can be represented with Boolean operators. Once a fault tree is constructed, the probability of each individual event must be evaluated. This in turn enables the evaluation of the overall probability of system failure. Calculating the probability of individual events is often no easy task and can rely on a variety of methods from simple numerical or analytical models [1], to PDF equations [45,46,39], to surveys [25], to expert opinion [27], to reliability data bases [7]. In all cases when estimating the probability of an event it may be advisable to err on the side of caution and be conservative in choosing the worst-case scenarios.

Such conservative approaches should typically be made a priori, aiming to identify those events that lead to the highest occurrence of risk. More resources can then be dedicated to analyze these specific events and to refine their probability of occurrence. In this work we illustrate how Bayesian methods can be integrated into a probabilistic risk analysis to refine an initial estimate of the probability of failure into its optimal value. In most cases, if the initial estimate is truly conservative, the probability of failure would decrease with increasing knowledge (Bayesian updating). On the contrary, it is also possible to envision scenarios where risk increases with new information, indicating that the prior information was far from what was expected. We explore this point in Section 6.8.

All of the events potentially leading to failure must be integrated into a well-posed fault tree, which entails defining the system failure and the events that can lead to it. Since a fault tree is essentially a visualization of a Boolean logic expression it is important to recognize that all events must be binary – i.e., true or false. This means that situations where an event can take on multiple states (e.g., high, medium or low levels of concentration) must be tackled with a different approach (see, for example, the multi-state reliability theory [28]).

2.2. Remediation actions

Observation wells are typically put in place to monitor and control a chosen remediation action. This is an important step for monitored natural attenuation projects, and a necessary component for many others. Let us consider a monitoring system consisting of n_w observation wells located at $\{\mathbf{x}_1, \ldots, \mathbf{x}_{n_w}\}$. For the

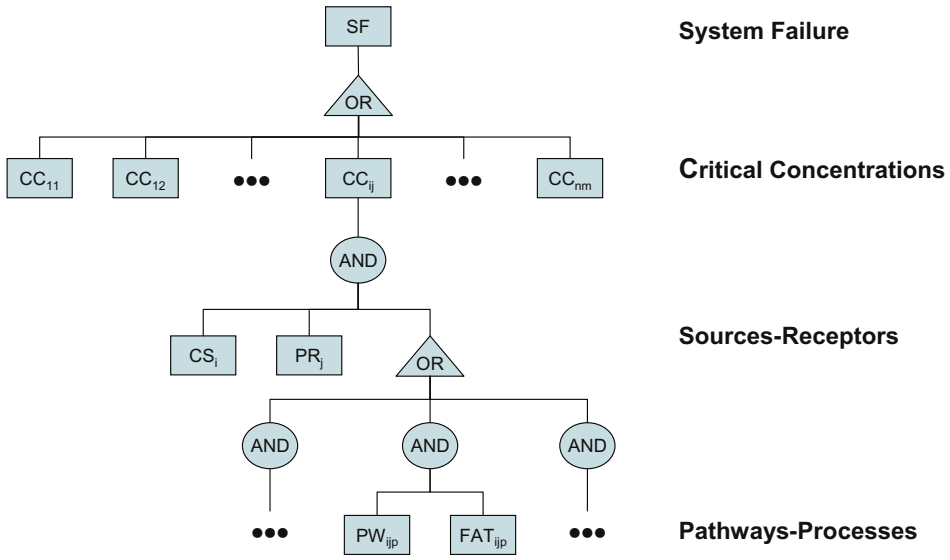

Fig. 1. Generic fault tree of a contamination problem.

monitoring system to fail, the contamination needs to go undetected by all the observation wells; otherwise the contamination will be detected and further corrective actions should be taken (e.g., shut-down of a producing well). To incorporate the monitoring system, for each observation well k, we distinguish two complementary *pathway* events that can cause an observation well's failure:

- **Bypassing** – the contaminant i can bypass the observation well k, reaching the receptor j within the time interval T. This includes the existence of preferential channels, unrepresented flow and transport mechanisms, and so on. We abbreviate this event as BP_{ijk}.
- **Not bypassing** – the observation well k can potentially detect the contamination within the time interval T. We abbreviate this event as $\overline{BP}_{ijk}$.

Although the occurrence of bypassing BP_{ijk} leads directly to the failure of the monitoring system, its complementary event $\overline{BP}_{ijk}$ leads to failure if the following two events occur:

- **Observation failure** – the observation well k fails due to malfunction, allowing the contaminant to pass undetected through the monitoring system. This can include electromechanical problems, measurement errors, human activities and others. We abbreviate this event as OBS_k.
- **Small sampling frequency** – the observation well k can potentially detect the arrival of contamination but the time lag between samples is too large to detect the plume before it reaches the potential receptor within the time interval T. We denote this event as SA_{ijk}.

Fig. 2 shows a generic fault tree for a monitoring system. In this fault tree, one still identifies the main components of a contamination problem (sources, receptors, pathways and processes), which are now combined with the features of a monitoring system. In particular, the observation wells represent intermediate events between the two main levels of the hierarchy in the fault tree, i.e., sources/receptors and pathways/processes. These intermediate events are linked through an AND gate to express that the monitoring system will only fail if all observation wells fail simultaneously. That is, if any of the observation wells detects the contamination the system will not fail as further corrective measures can be taken. Each observation well has two potential pathway events: *bypassing* and *not bypassing*. When no bypassing occurs, the failure of the system can still happen due to the occurrence of two potential events, i.e., the observation well fails either due to malfunctioning or the selection of a small sampling frequency.

3. Development of minimal cuts

Fault tree analyses (FTA) evaluate the system failure by expressing it in terms of a series of basic events using Boolean logic. Expressing the Boolean operators applied to two basic events A and B as

$$\begin{aligned} &A\text{'AND'}B \equiv A \cdot B \equiv A \cap B, \\ &A\text{'OR'}B \equiv A + B \equiv A \cup B, \end{aligned} \tag{3}$$

the fault tree in Fig. 1 represents the system failure as

$$\text{'System fails (SF)'} = \sum_{ij} CC_{ij} = \sum_{ijp} CS_i \cdot PR_j \cdot PW_{ijp} \cdot FAT_{ijp}. \tag{4}$$

The next step is to identify the *failure modes* of our system. These failure modes are known as minimal cut sets. Essentially, minimal cut sets represent the smallest combinations of basic events that are sufficient for the system to fail. It follows from (4) that the smallest collection of basic events that leads to the system failure is given by the following minimal cut sets: $\{CS_i \cdot PR_j \cdot PW_{ijp} \cdot FAT_{ijp}\}$. Expressing these cut sets as $SF = M_1 + M_2 + \cdots$, we quantify the probability of system failure by

$$P[SF] \approx \sum_a P[M_a] - \sum_{a<b} P[M_a \cdot M_b] + \sum_{a<b<c} P[M_a \cdot M_b \cdot M_c] - \cdots, \tag{5}$$

which is the *inclusion–exclusion* law of probability.

Let us now concentrate on the failure of a monitoring system. In this case the fault tree shown in Fig. 2 represents the system failure as

$$\text{'System fails (SF)'} = \sum_{ij} CC_{ij} = \sum_{ij} CS_i \cdot PR_j \cdot \left[\prod_k (B_{ijk} + S_{ijk} + O_{ijk}) \right], \tag{6}$$

where B_{ijk}, S_{ijk} and O_{ijk} represent the following groups of basic events

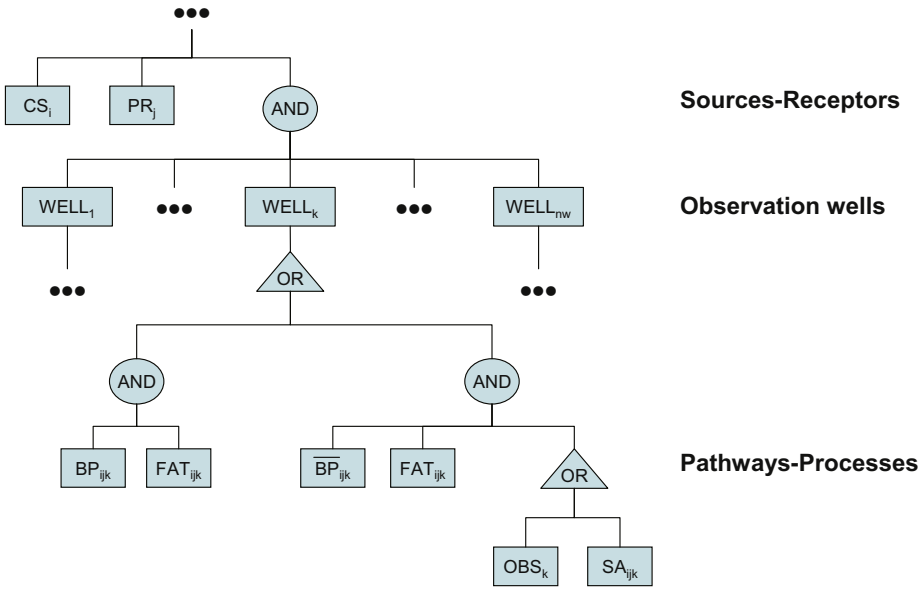

Fig. 2. Adapted fault tree for a monitoring system.

$$\begin{aligned} B_{ijk} &= BP_{ijk} \cdot FAT_{ijk}, \\ S_{ijk} &= \overline{BP}_{ijk} \cdot FAT_{ijk} \cdot SA_{ijk}, \\ O_{ijk} &= \overline{BP}_{ijk} \cdot FAT_{ijk} \cdot OBS_k. \end{aligned} \tag{7}$$

To identify the minimal cut sets, we need to reduce the Boolean expression (6) to a sum of combinations of basic components. Let us introduce the events BSO_{ijq} as

$$\sum_{q=1}^{n_q} BSO_{ijq} \equiv \prod_{k=1}^{n_w} (B_{ijk} + S_{ijk} + O_{ijk}), \quad n_q = 3^{n_w}. \tag{8}$$

Then (6) can be rewritten as

$$\text{'System fails (SF)'} = \sum_{ijq} CS_i \cdot PR_j \cdot BSO_{ijq}. \tag{9}$$

The definition of BSO_{ijq} in (8) represents the product $\prod_{k=1}^{n_w}(B_{ijk} + S_{ijk} + O_{ijk})$ as a full ternary tree diagram in which a given node, for example S_{ijk-1}, has three children $\{B_{ijk}, S_{ijk}, O_{ijk}\}$. Each path of the tree diagram depicts one BSO_{ijq} event. Thus, for a given contaminant source i and receptor j, the event BSO_{ijq} represents the simultaneous occurrence of n_w distinct basics events. Each of these events is selected sequentially out of the n_w groups $\{S_{ij1}, B_{ij1}, O_{ij1}\}, \{S_{ij2}, B_{ij2}, O_{ij2}\}, \ldots, \{S_{ijn_w}, B_{ijn_w}, O_{ijn_w}\}$. Each BSO_{ijq} event represents a different sequence of multiple failures associated with the monitoring system. For instance, the subset $\{B_{ij1}, B_{ij2}, \ldots, B_{ijn_w}\}$ accounts for the case of total bypass of the solute plume from the monitoring system. The subset $\{O_{ij1}, O_{ij2}, \ldots, O_{ijn_w}\}$ represents the simultaneous failure of all observation wells due to malfunction or human activities. This mechanism of multiple failures is a salient feature of a monitoring system and has not traditionally been included in previous risk-based transport models [21,19,30,23,22].

Expression (9) illustrates that the smallest collection of basic events that leads to failure of the monitoring system is given by the minimal cut sets $\{CS_i \cdot PR_j \cdot BSO_{ijq}\}$. In this case, quantifying the probability of system failure can be significantly simplified. The different BSO_{ijq} events cannot occur at the same time for a given contaminant source i and receptor j. Indeed, if the plume bypasses the observation well k, the events O_{ijk} and S_{ijk} cannot occur; likewise, if the well fails then sampling cannot occur. Formally, the OR gate that connects these basic events can be thought of as an exclusive OR gate, where one and only one event can occur. Thus, they must be considered mutually exclusive and

$$P[BSO_{ijq} \cdot BSO_{ijr}] = 0, \quad q \neq r. \tag{10}$$

Let us assume, for the sake of simplicity, that the occurrence of CS_i is independent of the occurrence of PR_j, and that all their characteristics are known (e.g., from expert opinion). Then

$$P[SF] = \sum_{ijq} P[CS_i]P[PR_j]P[BSO_{ijq}|CS_i, PR_j]. \tag{11}$$

It is reasonable to assume both that malfunction of a given observation well is independent of the occurrence of the other basic events and that the probability of its occurrence is the same for all observation wells, $P[OBS_k] = P[OBS]$. These two assumptions facilitate the computation of probabilities of the basic events. Consider, for example, a monitoring system that consists of two observation wells. Then the BSO_{ijq} events involving the observation failure can be estimated as

$$\begin{aligned} P[S_{ij1} \cdot O_{ij2}|CS_i, PR_j] &= P[S_{ij1} \cdot \overline{B}_{ij2}|CS_i, PR_j]P[OBS], \\ P[B_{ij1} \cdot O_{ij2}|CS_i, PR_j] &= P[B_{ij1} \cdot \overline{B}_{ij2}|CS_i, PR_j]P[OBS], \\ P[O_{ij1} \cdot O_{ij2}|CS_i, PR_j] &= P[\overline{B}_{ij1} \cdot \overline{B}_{ij2}|CS_i, PR_j](P[OBS])^2. \end{aligned}$$

In summary, expression (11) replaces the difficult task of computing $P[SF]$ with more tractable problems of computing probabilities of basic events that can be estimated by solving an appropriate stochastic transport model. As discussed in the following section, many different approaches can be used to solve these stochastic models. Our Bayesian probabilistic risk assessment approach can be used in conjunction with any of these methods.

4. Computation of probabilities

In order to compute these probabilities, for each set of basic components ijk (source, receptor, pathways), one needs to solve a stochastic transport model or a set of alternative models. These models provide a mathematical description of all transport phenomena included in FAT_{ijk}. These models are typically expressed in terms of stochastic partial differential equations (PDEs) with boundary and initial conditions applied to the variable C_i and

conditioned to the contaminant source and receptor in question. This is formally written as

$$F_{ijk}[C_i(\mathbf{x},t);\boldsymbol{\theta}]=0, \tag{12}$$

where $\boldsymbol{\theta}$ is a vector of system parameters. We assume that different contaminants do not interact with each other. For more complex geochemical problems, a solution of this system can be formally written as a system of n coupled equations with n unknowns,

$$F_{ijk}[C_1(\mathbf{x},t),\ldots,C_n(\mathbf{x},t);\boldsymbol{\theta}]=0,\quad i=1,\ldots,n. \tag{13}$$

The randomness of the PDEs stems from uncertainties about subsurface processes. These uncertainties can be either *structural* (arising from errors in a conceptual model) or *parametric* (arising from the imperfect knowledge of $\boldsymbol{\theta}$).

Structural uncertainty can be implemented in PRA studies by evaluating the relative performance of several competing models instead of a single one. By weighing predictions of each model, one can ascertain the different degrees of reliability among models (e.g., [3,18,29]).

Stochastic contaminant hydrology presents many different approaches and theories for solving stochastic PDEs, ranging from small-perturbation approaches [4,14], to moment [26] and PDF [39,45] equations, to Monte Carlo simulations [38,34]. While analyzing a system as a whole is often computationally prohibitive due to a large number of uncertain parameters (i.e., high dimensionality of the parameter space), evaluating probabilities of basic events in the fault trees is more feasible since the latter depend on fewer parameters.

Solutions of a stochastic PDE of subsurface transport are given in terms of the PDFs of C_i or related quantities (e.g., travel times). To transform these into the basic probabilities in (11), we need to express the occurrence of the basic events B_{ijk}, S_{ijk} and O_{ijk} in terms of concentrations. Given a connection between the contaminant source i and the receptor j, we define the occurrence of events $\{B_{ijk}|CS_i, PR_j\}$ as

$$\{B_{ijk}|CS_i,PR_j\}=\left\{C_i(\mathbf{x}_j,t\leqslant T)>C^*_{ij},C_i(\mathbf{x}_k,t\leqslant T)<C^*_{ij}\right\}, \tag{14}$$

where $\mathbf{x}_j$ and $\mathbf{x}_k$ are the locations of the receptor j and observation well k, respectively. Its complementary events are written as

$$\{\overline{B}_{ijk}|CS_i,PR_j\}=\left\{C_i(\mathbf{x}_j,t\leqslant T)>C^*_{ij},C_i(\mathbf{x}_k,t\leqslant T)>C^*_{ij}\right\}. \tag{15}$$

Let us consider a monitoring system of n_w observation wells, and let $\tau_i(\mathbf{x}_k, \mathbf{x}_j)$ denote the travel time of contaminant i moving from an observation well located at $\mathbf{x}_k$ to a receptor located at $\mathbf{x}_j$. We define the occurrence of $\{S_{ijk}|CS_i, PR_j\}$ as

$$\{S_{ijk}|CS_i,PR_j\}=\left\{C_i(\mathbf{x}_j,t\leqslant T)>C^*_{ij},C_i(\mathbf{x}_k,t\leqslant T)>C^*_{ij},\tau_i(\mathbf{x}_k,\mathbf{x}_j)<\Delta t_f\right\}, \tag{16}$$

where Δt_f is the time lag between measurements. These expressions do not depend on the chosen stochastic transport model. Thus, our PRA methodology to estimate the probability of system failure of a contamination problem can be considered general and not restricted to the chosen stochastic transport model.

5. Integration with Bayesian methods

5.1. The Bayes formalism

Bayesian methods employ the Bayes theorem to update a belief about a hypothesis. In subsurface hydrology, the Bayesian formalism has been typically used in geostatistical inverse methods to account for uncertainty in the chosen stochastic model of heterogeneity (e.g., the mean and the covariance function of the natural log of transmissivity) [20,48,12]. More recently, a Bayesian approach was used in inverse modeling, data assimilation, and conditional simulation of spatial random fields [35]. Any of these approaches can be integrated into PRA to update the probability of system failure given some real-time observations of concentrations.

Let $f(\boldsymbol{\theta})$ denote a prior (not informed by concentration measurement at the site) multivariate distribution of the parameters involved in the transport model (12). As concentration measurements of contaminant i, $\mathbf{c}_i=(c_{i1},c_{i2},\ldots)$, become available, $f(\boldsymbol{\theta})$ is updated into a posterior multivariate distribution $f(\boldsymbol{\theta}|\mathbf{c}_i)$ according to the Bayes theorem,

$$f(\boldsymbol{\theta}|\mathbf{c}_i)=\frac{f(\mathbf{c}_i|\boldsymbol{\theta})f(\boldsymbol{\theta})}{\int f(\mathbf{c}_i|\boldsymbol{\theta})f(\boldsymbol{\theta})d\boldsymbol{\theta}}, \tag{17}$$

where $f(\mathbf{c}_i|\boldsymbol{\theta})$ is the likelihood function, $L(\boldsymbol{\theta})=f(\mathbf{c}_i|\boldsymbol{\theta})$.

The prior distribution is often subjectively estimated based on the assessment of an expert. A few measurements result in small changes in the prior, while a large number of concentration measurements would give rise to a posterior distribution that can significantly depart from the prior. After enough measurements are taken, the posterior distribution becomes independent of the prior. For these reasons, when uncertainties are significant, it is advisable to use priors displaying a sufficiently large variance.

The likelihood function $L(\boldsymbol{\theta})$ can be estimated using either *parametric* or *non-parametric* approaches. Parametric approaches describe the unknown likelihood function with a finite set of parameters. The multi-normal distribution is often adopted for this purpose. In non-parametric approaches, the likelihood function can take any form. Although this can be computationally demanding, it enables one to identify the true underlying structure.

5.2. Stochastic generation of posterior parameters

After updating the distribution function of the parameters $f(\boldsymbol{\theta}|\mathbf{c}_i)$, we randomly generate multiple realizations of $\boldsymbol{\theta}$ and simulate the concentration values $C_i(\mathbf{x},t;\boldsymbol{\theta})$. The main difficulty arises from the fact that, although in some cases it is possible to initially consider the parameters as independent, this no longer holds once measurements are used to update the posterior distributions. In other words, $f(\boldsymbol{\theta}|\mathbf{c}_i)$ is no longer composed of statistically independent variables. To overcome this difficulty, we use the sequential simulation algorithm of a multivariate distribution [16,15]. This method decomposes $f(\boldsymbol{\theta}|\mathbf{c}_i)$ into a product of conditional univariate distributions such that

$$f(\theta_1,\theta_2,\ldots,\theta_N|\mathbf{c}_i)=f(\theta_N|\theta_{N-1},\ldots,\theta_1,\mathbf{c}_i)f(\theta_{N-1}|\theta_{N-2},\ldots,\theta_1,\mathbf{c}_i)\ldots f(\theta_1|\mathbf{c}_i), \tag{18}$$

where N is the number of random variables. This decomposition allows one to draw realizations of $\boldsymbol{\theta}=(\theta_1,\theta_2,\ldots,\theta_N)$, conditioned on $\mathbf{c}_i$, by recursively generating the N univariate conditional PDFs of the decomposition.

6. An example: a DNAPL-polluted site

6.1. Problem description

We use a synthetic example to illustrate the advantages and limitations of the proposed methodology. In particular, we conduct an integrated Bayesian probabilistic risk assessment study of a DNAPL-contaminated site to determine how the likelihood of failure of a monitoring system changes with time. For simplicity and illustrative purposes, we consider a system that consists of one contaminant source and one receptor, but remains complex in

terms of pathways and observations. This is done to stress the subtleties involved in the implementation of a monitoring system.

We consider the case where a chlorinated solvent contaminant (TCE), in the form of a DNAPL, has been accidentally released into the subsurface. This type of contamination has been widely recognized as one of the most serious groundwater problems worldwide [42,36]. Driven by gravity and helped by their low viscosity and relatively low surface tension, these liquids can penetrate beneath the water table. Once in the aquifer, DNAPLs are essentially immiscible with water but can slowly dissolve into groundwater with concentrations that are sufficiently high to threaten the environment and public health.

We begin by assuming that a free-phase TCE source exists within an area of $45 \times 90\,m^2$ (Fig. 3). The exact distribution of free-phase TCE in this area is unknown. One potential receptor is located 122 m from this area. The monitoring system consists of two observation wells OBS1 and OBS2. The former is placed down-gradient close to the source with the intent of providing information about the source size and the amount of pollutant released. The latter is located closer to the receptor, intended to alert the population in case an undetected episode of contamination occurs.

To reflect real-like conditions we consider a scenario in which information is scarce. The principle source of information is the concentration values measured at the observation wells. These measured concentrations were generated numerically with a transport code that considers that the TCE source is immobile but slowly dissolving with time in a highly heterogeneous aquifer.

6.2. Reference transport simulation

We consider a heterogeneous two-dimensional aquifer of dimensions $L_x = L_y = 200$. Groundwater is driven by a mean uniform gradient oriented along the x coordinate. The aquifer is confined, with constant heads at $x = 0$ and $x = 200$, and no-flow at the remaining model boundaries. All aquifer properties are constant except for transmissivity, which varies randomly in space.

We generated one realization of the transmissivity field $T(\mathbf{x})$ using the stochastic field generator GCOSIM3d [16]. The computational domain is discretized into blocks of size $\Delta x = \Delta y = 1.0$. The natural log of transmissivity $Y = \ln T$ follows a multi-normal distribution with mean $\langle Y \rangle = 4.8$ and variance $\sigma_Y^2 = 3.2$. The correlation structure is described by an anisotropic spherical variogram,

$$\gamma(\mathbf{h}) = \sigma_Y^2 \mathrm{sph}(\|\mathbf{r}(\mathbf{h})\|), \tag{19}$$

$$\mathrm{sph}(r) = \begin{cases} 1.5r - 0.5r^3 & \text{if } r \leqslant 1 \\ 1 & \text{otherwise,} \end{cases} \tag{20}$$

where $\mathbf{h}$ is the distance vector between two points, and $\mathbf{r}(\mathbf{h})$ is the transformed distance vector associated with a unitary equivalent isotropic variogram,

$$\mathbf{r}(\mathbf{h}) = \begin{bmatrix} \cos\chi & -\sin\chi \\ \sin\chi & \cos\chi \end{bmatrix} \cdot \begin{bmatrix} 1/a_x & 0 \\ 0 & 1/a_y \end{bmatrix} \cdot \begin{bmatrix} h_x \\ h_y \end{bmatrix}, \tag{21}$$

and $r = \|\mathbf{r}(\mathbf{h})\|$. This is obtained after a counterclockwise axes rotation and scaling of the ranges. Stratification is not oriented in the mean direction of flow so that χ is the angle with respect to the x axis. The parameter values used in the subsequent simulations are summarized in Table 1. The resultant transmissivity map is displayed in the background of Fig. 3.

The source of dissolved TCE was simulated by imposing a time-varying flux concentration $C_s(t)$ at an arbitrary injection line within the source (Fig. 3). Several mass-depletion models have been proposed in the literature to describe the time-dependent release of dissolved DNAPL mass into an aquifer. We use the model [32] that relates the flux concentrations of dissolved DNAPL leaving a source zone to mass removal by

$$\frac{C_s(t)}{C_0} = \left[\frac{M(t)}{M_0}\right]^{\beta}, \tag{22}$$

where C_s is the flux concentration of dissolved DNAPL leaving the source zone, C_0 is the initial flux concentration, $M(t)$ is the mass of DNAPL at time t, and M_0 is the initial DNAPL mass. The parameter β accounts for changes in interfacial surface area as the source mass diminishes, and thus reflects the effects of source-zone architecture, flow-field dynamics, and mass transfer processes. For finger-dominated residual DNAPL $\beta > 1$, whereas for DNAPL in the form of pools and lenses $\beta < 1$ is small [31]. This is due to the fact that finger-dominated source zones generally exhibit higher initial mass transfer coefficients per unit mass than lens-dominated domains. The mass depletion of DNAPL in the source zone is given by [31]

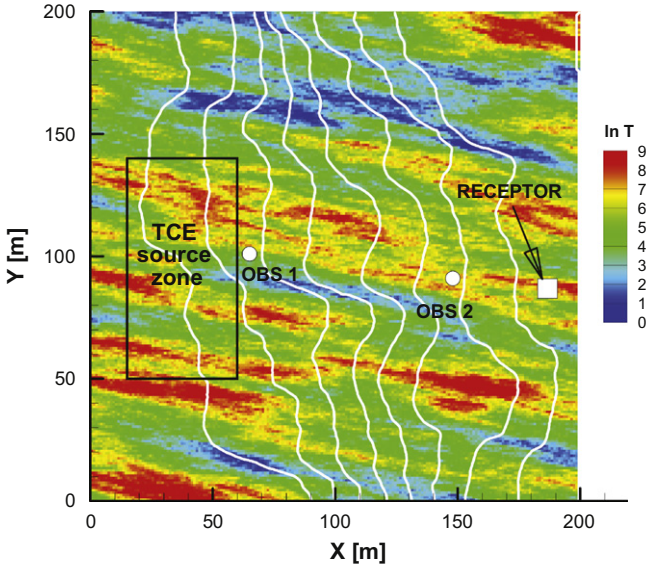

Fig. 3. Setup of the synthetic example. The map of the natural log of transmissivity is displayed in the background (color pixels) with corresponding equipotential lines (white lines), and pathways corresponding to 50 particles evenly located at the source (black lines). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Flow problem parameters adopted during the numerical simulations performed in a heterogeneous medium.

Parameter	Value
System	
Observation well 1	(65, 101)
Observation well 2	(148, 91)
Receptor	(186, 86)
Flow problem	
Number of cells in x-direction, n_x	200
Number of cells in y-direction, n_y	200
Cell size in x-direction, Δx [L]	1.0
Cell size in y-direction, Δy [L]	1.0
Mean hydraulic gradient in x-direction, J_x [–]	0.01
Mean hydraulic gradient in y-direction, J_y [–]	0.0
Heterogeneous field	
Mean of $\ln T$ [L^2/T]	4.8
Variance of $\ln T$ [–]	3.2
Maximum variogram range, a_{max} [L]	150.0
Minimum variogram range, a_{min} [L]	15.0
Clockwise angle of rotation, χ [degrees]	10.0

$$\frac{M(t)}{M_0} = \begin{cases} \exp(-Bt) & \beta = 1, \\ [(\beta-1)Bt + M_0^{1-\beta}]^{1/(1-\beta)} & \beta \neq 1. \end{cases} \quad (23)$$

The mass depletion rate constant B incorporates the effects of mass transfer processes and flow field dynamics occurring in the source zone. An analysis of the field data [6] shows the range of the values of β for different remediation technologies (Table 2), with values ranging from 0.15 to 5.2. In our example, we chose $\beta = 1$, somewhat an average value of those reported in Table 2.

Forward transport simulations of dissolved TCE were performed using the random walk particle tracking code RW3D-MT [9,37]. The code uses velocity fields obtained from the finite difference groundwater flow model MODFLOW2000 [17]. Breakthrough curves were estimated by reconstructing concentrations from particle distributions using the optimal Kernel Density estimation method [11]. Table 3 summarizes the transport parameters used in the forward simulation.

In this synthetic example, the solute travels through the heterogeneous aquifer towards the receptor. The path taken by the solute is shown in Fig. 3, which depicts the trajectory of 50 particles evenly selected out of the 10^5 injected particles representing the entire plume. The location of the source zone and the initial concentration released are not known with certainty and are the subject of the stochastic analysis.

The numerical results of the concentration breakthrough curves obtained at the observation wells and at the receptor are shown in Fig. 4. We consider a sampling campaign consisting of five measurements, which are also depicted in Fig. 4.

Table 2
Review of depletion exponent values (after [6]).

Remediation action	J/J_0	M/M_0	β
Natural gradient water flush	0.92	0.77	0.32
In situ chemical oxidation	0.0	0.05	–
Natural gradient water flush	0.91	0.68	0.24
Surfactant enhanced aquifer remediation	0.17	0.06	0.63
Cosolvent	0.16	0.1	0.80
Surfactant enhanced aquifer remediation	0.09	0.25	1.74
Cyclodextrin flushing	0.25	0.02	0.35
Ethanol flush	0.48	0.36	0.72
n-Propanol flush	0.19	0.2	1.03
Surfactant enhanced aquifer remediation	0.09	0.36	2.36
In situ chemical oxidation	0.11	0.18	1.29
Emulsified zero-valent iron	0.44	0.28	0.64
Six phase heating	0.05	0.05	1.00
Cosolvent	0.53	0.36	0.62
Pump-and-treat	0.09	0.66	5.80
Six phase heating	0.3	0.02	0.31
Surfactant enhanced aquifer remediation	0.46	0.28	0.61
In situ chemical oxidation	0.68	0.08	0.15
In situ chemical oxidation	0.01	0.06	1.64
Rotary steam stripping	0.24	0.3	1.19

Table 3
Transport parameters to generate the synthetic example.

Parameter	Value
Line of injection [L]	(53, 90)–(53, 112)
Injected concentration at $t = 0$, C_0 [M/L^3]	406.7
Coefficient B [T^{-1}]	10^{-3}
Mass depletion exponent, β [–]	1.0
Porosity, ϕ [–]	0.3
Longitudinal dispersivity, α_L [L]	0.3
Transverse dispersivity, α_T [L]	0.1
Retardation coefficient, R [–]	1.0
Biodegradation decay constant, λ [T^{-1}]	0.0
Courant number [–]	0.05
Number of particles [–]	1.0×10^5

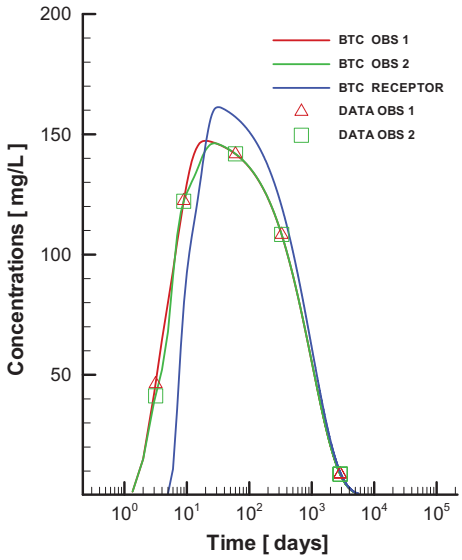

Fig. 4. Synthetic breakthrough curves obtained at the observation wells. Symbols represent the actual observations of concentrations.

6.3. Fault tree analysis

Next we adapt the general fault tree of a contamination problem, developed in Section 2, to our specific problem. In particular, the failure of the system is defined as the event that concentrations surpass EPA groundwater drinking standard levels for TCE ($C^* = 0.005$ ppm) at the receptor within a time interval T,

$$SF = \{C(\mathbf{x}_R, t \leqslant T) > C^*\}, \quad (24)$$

where $\mathbf{x}_R$ is the location of the receptor. The fault tree events are shown in Table 4 and follow the definition of system components discussed in Section 2. The receptor is assumed to exist with certainty, $P[PR] = 1$, but the contaminant source (presence of DNAPL) is assumed uncertain and quantified by $P[NAPL] = 0.5$. The event SA represents the failure of the system due to a small sampling frequency, which is fixed to $\Delta t_f = 2$ days. The events OBS_1 and OBS_2 represent potential failure of the corresponding observation wells due to mechanical or human activities, their probabilities are set to $P[OBS] = 0.1$. In accordance with (11), the fault tree analysis for this problem results in the probability of system failure,

$$\begin{aligned} \frac{P[SF]}{P[NAPL]} = {} & P[B_1 \cdot B_2|NAPL] + P[B_1 \cdot S_2|NAPL] + P[B_1 \cdot \overline{B}_2|NAPL]P[OBS] \\ & + P[S_1 \cdot B_2|NAPL] + P[S_1 \cdot S_2|NAPL] + P[S_1 \cdot \overline{B}_2|NAPL]P[OBS] \\ & + P[\overline{B}_1 \cdot B_2|NAPL]P[OBS] + P[\overline{B}_1 \cdot S_2|NAPL]P[OBS] \\ & + P[\overline{B}_1 \cdot \overline{B}_2|NAPL](P[OBS])^2. \end{aligned} \quad (25)$$

Table 4
Glossary of event abbreviations.

Event	Abbreviation
System failure	SF
Presence of free-NAPL phase (CS)	$NAPL$
Presence of a receptor	PR
Bypass of the observation wells 1 and 2	BP_1, BP_2
No bypass of the observation wells 1 and 2	$\overline{BP}_1$, $\overline{BP}_2$
Failure of observation wells	OBS
Failure of sampling frequency at wells 1 and 2	SA_1, SA_2
Natural attenuation fails (FAT)	NA
Event $\overline{BP}_1 \cdot NA \cdot SA_1$	S_1
Event $\overline{BP}_2 \cdot NA \cdot SA_2$	S_2
Event $BP_1 \cdot NA$	B_1
Event $BP_2 \cdot NA$	B_2

The details of the calculation of these probabilities are given in Sections 6.4, 6.5, 6.6, 6.7.

6.4. Stochastic model

We consider a stochastic transport model based on an advection–dispersion equation (ADE) with equivalent homogeneous coefficients that are random. The reasoning behind this is that we are not interested in completely reproducing the actual profile of concentrations at a given point and time. Instead, our goal is determine whether a simplified model is capable of providing valuable information about the contamination (location, extension and depletion rate of the DNAPL source). We assume that the same stochastic transport model is applicable to all *FAT* events. Fig. 5 shows a sketch of the simplified transport model. Prompted by this, for any given realization of the parameters, a solution of the assumed ADE at a given downstream location is a time convolution

$$C(x',y',t)=\int_0^t C_s(t-\tau)F_y(y'-y'_0,\tau)G_x(x'-x'_0,\tau)e^{-\lambda\tau}\,d\tau, \tag{26}$$

where

$$G_x(x,t)=\frac{1}{\sqrt{4\pi D'_\ell t}}\exp\left[-\frac{(x-v't)^2}{4D'_\ell t}\right] \tag{27}$$

and

$$F_y(y,t)=\frac{1}{2}\mathrm{erfc}\left(\frac{y-L_y/2}{\sqrt{4D'_t t}}\right)-\frac{1}{2}\mathrm{erfc}\left(\frac{y+L_y/2}{\sqrt{4D'_t t}}\right). \tag{28}$$

Here D_ℓ and D_t are the longitudinal and transverse dispersion coefficients, $D'_\ell=D_\ell/R$, $D'_t=D_t/R$, $v'=v/R$, R is the retardation factor, λ is the first-order biodegradation constant, and (x_0, y_0) denotes the location where the dissolved DNAPL exits the source zone. The plume travels in an arbitrary and unknown direction x', which is rotated about the origin by an acute angle γ. Therefore, $x'=x\cos\gamma+y\sin\gamma$ and $y'=-x\sin\gamma+y\cos\gamma$. This allows us to account for the observation that in most scenarios the mean plume direction is not necessarily given by the direction of the mean hydraulic gradient.

The stochasticity of the system comes directly from the uncertainty in the equivalent homogeneous coefficients. Despite a large number of existing remediation strategies, they typically fail to remove enough DNAPL mass to reduce aquifer concentrations to levels that meet groundwater standards [36,41]. The complex nature of the DNAPL architecture, the hard task of properly characterizing soil attributes, and the difficulty in describing field-scale DNAPL mass transfer processes render predictions of the actual performance of an individual remediation project highly uncertain. In our analysis, this uncertainty is represented by five uncertain parameters: the initial concentration $\{C_0\}$ at the release area, the size and location of this area $\{y_0, L_y\}$, the depletion rate $\{B\}$, and the apparent velocity $\{v\}$ observed in the observation wells. All five parameters are treated as random variables. In particular, the randomness of $\{v, y_0, L_y\}$ is crucial in emulating the occurrence of preferential channels associated with large velocities and smaller plume sizes.

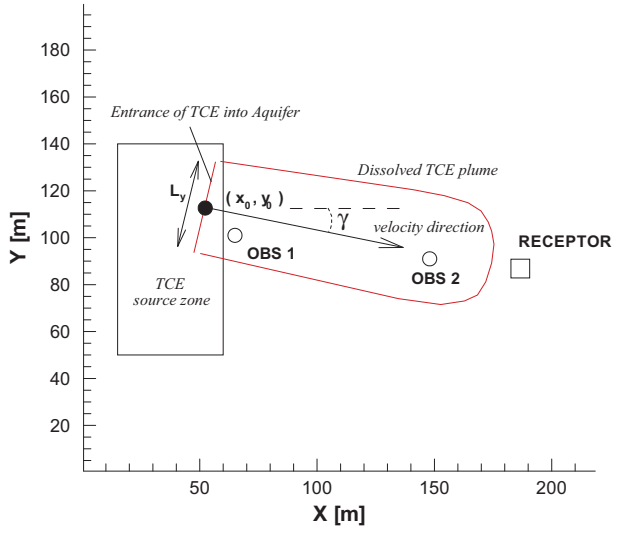

Fig. 5. Conceptual representation of the simplified transport model.

6.5. Computation of probabilities

The computation of $P[SF]$ is made up of the following sequential steps: For any given probability distribution of the parameters $f(\boldsymbol{\theta}|\mathbf{c})$:

- We generate 100 realizations of the random variables $\boldsymbol{\theta}$ using the sequential simulation algorithm described in Section 5.2;
- In each realization, we simulate the concentration breakthrough curves at the observation wells and receptor through the transport model given in Section 6.4;
- From the breakthrough curves, we estimate the travel time of the contaminant moving from the observation wells to the receptor;
- For each time interval T, the resulting outcome from the previous step is translated into probabilities through (14)–(16).

6.6. Priors

Prior distributions of these random variables are reported in Table 5. Following the principle of indifference, we assign a uniform distribution to most priors. The effective velocity is an exception, since its mean $\bar{v}$ and variance σ_v^2 are computable from the stochastic theories [4,14],

$$\bar{v}=\frac{K_g J}{\phi},\quad \sigma_v^2\approx\sigma_x^2\frac{\bar{v}^2}{d_{ij}^2}, \tag{29}$$

where $\sigma_x^2=2A_\ell^{cm}d_{ij}$ is the variance of the plume's center of mass in the x direction, d_{ij} is the travel distance from the source-zone i to the $\mathbf{x}_j$ location, and A_ℓ^{cm} refers to the variability of the center of mass

Table 5
Model parameters and prior distribution functions.

Parameter	Priors
Porosity, ϕ	N.R. (0.3)
Solute velocity, v [L/T]	*Norm*(4.21, 400, 0, 50)
Longitudinal dispersivity, a_ℓ [L]	N.R.(10.0)
Transverse dispersivity, a_t [L]	N.R.(3.0)
Rotation angle, γ [degrees]	N.R.(−10.0)
Retardation coefficient, R [–]	N.R.(1.0)
Biodegradation decay constant, λ [T^{-1}]	N.R.(0.0)
Source size in y-direction, L_y [L]	*Unif*(1, 67)
Source location in x-direction, x_0 [L]	N.R.(53)
Source location in y-direction, y_0 [L]	*Unif*(60, 130)
Initial flux concentration, C_0 [M/L^3]	*Unif*(300, 1000)
Depletion rate, B [T^{-1}]	*Unif*(10^{-1}, 10^{-4})
Depletion exponent, β [–]	N.R.(1.0)

N.R.(v) means that the variable is not assumed random and fixed to v. *Unif*(a,b) is the uniform distribution that lies within the interval (a,b). *Norm*(m, σ^2, a, b) is the truncated normal distribution with mean m, variance σ^2, and bounded below and above by a and b.

location among all possible aquifer realizations. We approximate A_ℓ^{cm} by [13]

$$A_\ell^{cm} \approx \sigma_Y^2 I_x, \tag{30}$$

where I_x is the integral scale of $Y = \ln T$ in the x direction. The approximation (30) provides a conservative estimate of risk, since it is an upper bound value and real dispersion is always smaller than this value.

6.7. Incorporation of data

A likelihood functions $L(\boldsymbol{\theta})$ quantifies the probability of observing the data given $\boldsymbol{\theta}$. If $\mathbf{c} = (c_1, c_2, \ldots)$ is an independent sample of concentrations, then

$$L(\boldsymbol{\theta}) = f(\mathbf{c}|\boldsymbol{\theta}) = \prod_m f(c_m|\boldsymbol{\theta}). \tag{31}$$

When model predictions yield concentration values that are closer to measurements for a given $\boldsymbol{\theta}$, the corresponding $L(\boldsymbol{\theta})$ value should increase. To incorporate this feature, we take $f(c_m|\boldsymbol{\theta})$ to follow a truncated power law density function

$$f(c_m|\boldsymbol{\theta}) \propto |c_m - C_m(\boldsymbol{\theta})|^{-n}, \tag{32}$$

where n is a hyper-parameter, c_m is the mth measurement of concentration, and $C_m(\boldsymbol{\theta})$ is its corresponding model prediction given $\boldsymbol{\theta}$. The larger the n, the sharper the likelihood function. This indicates that a larger relative weight is given to the vector of parameters that actually lead to matching observed concentrations, which provides peaked posterior distributions. Note that the likelihood function (31) and (32) considers measurement errors to be negligible. If this is not the case, one can incorporate this source of uncertainty into $L(\boldsymbol{\theta})$.

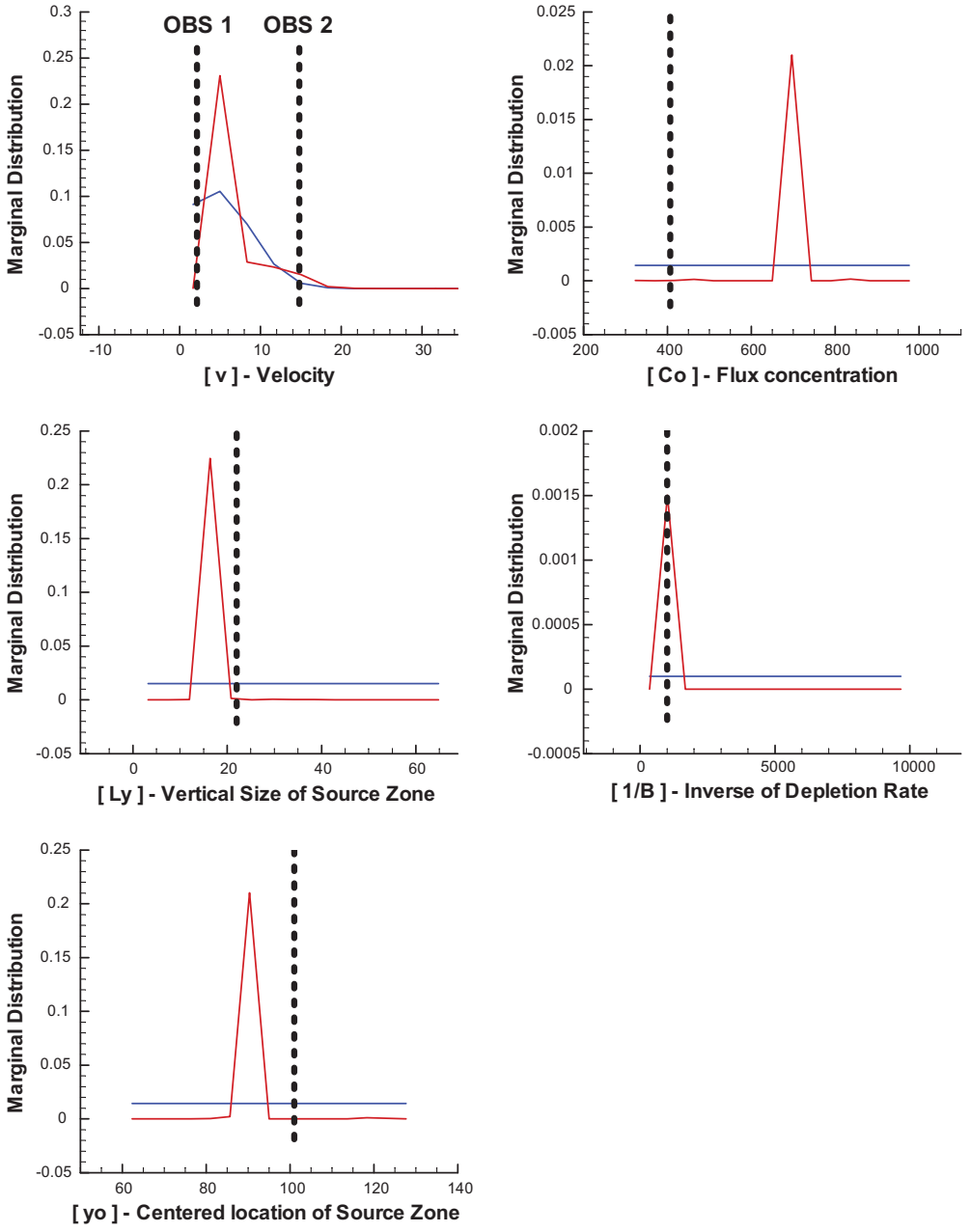

Fig. 6. Prior (blue lines) and posterior (red lines) marginal distributions of the transport model parameters. Vertical dashed lines denote the original true values of the parameters. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

6.8. Results

Fig. 6 shows the update of the prior marginal distributions of the parameters,

$$f(\theta_i) = \int f(\theta_1, \ldots, \theta_N) d\theta_1 \cdot \ldots \cdot d\theta_{i-1} \cdot d\theta_{i+1} \cdot \ldots \cdot d\theta_N. \quad (33)$$

These integrals were computed numerically. As previously discussed, the transport parameters in our model are essentially effective parameters. As such, their posterior distribution functions must reflect the effects of heterogeneity. This is clearly seen for the effective velocity and the effective initial flux concentration (the first and zeroth moments of the breakthrough curves). The effective velocity exhibits a bimodal posterior distribution, with modes close to the apparent velocity of the solute plume (i.e., the ratio of the travel distance to the mean arrival time) observed at wells OBS1 and OBS2. The vertical dashed lines denote the apparent velocity values observed at the wells.

The solute mass that passes through the observation wells depends mostly on the intrinsic peculiarities of the transmissivity field and the hydraulic connection between a source and a receptor. As a result, its effective mean value exhibits large deviations from its original mean estimate. While this may seem to be a weak point of our model, we contend the opposite. In the presented example, the mean flow velocity is not really informative of the system's behavior, which is controlled by the fast pathway in the vicinity of the source. This implies that the actual effective velocity is larger than its mean value, while the effective source size is smaller.

Note that the travel distance is small compared to the number of integral scales that are needed in order to invoke ergodicity. Fig. 2b illustrates this point by accounting for the possibility that some of the particles can avoid the observation wells, while a large number of particles concentrate very close to the source and near OBS1. An ensemble mean description of this problem is clearly not enough to capture these phenomenon.

Conversely, our model is capable of predicting the depletion rate of the DNAPL mass. Although the aquifer is very heterogeneous ($\sigma_Y^2 = 3.2$), the slow dissolution of DNAPL at the source mainly controls the late-time behavior of the breakthrough curves. As a result, the model is capable of accurately predicting the depletion rate. The size and location of the source are slightly underestimated, due to the convergence of particle pathways and random effects.

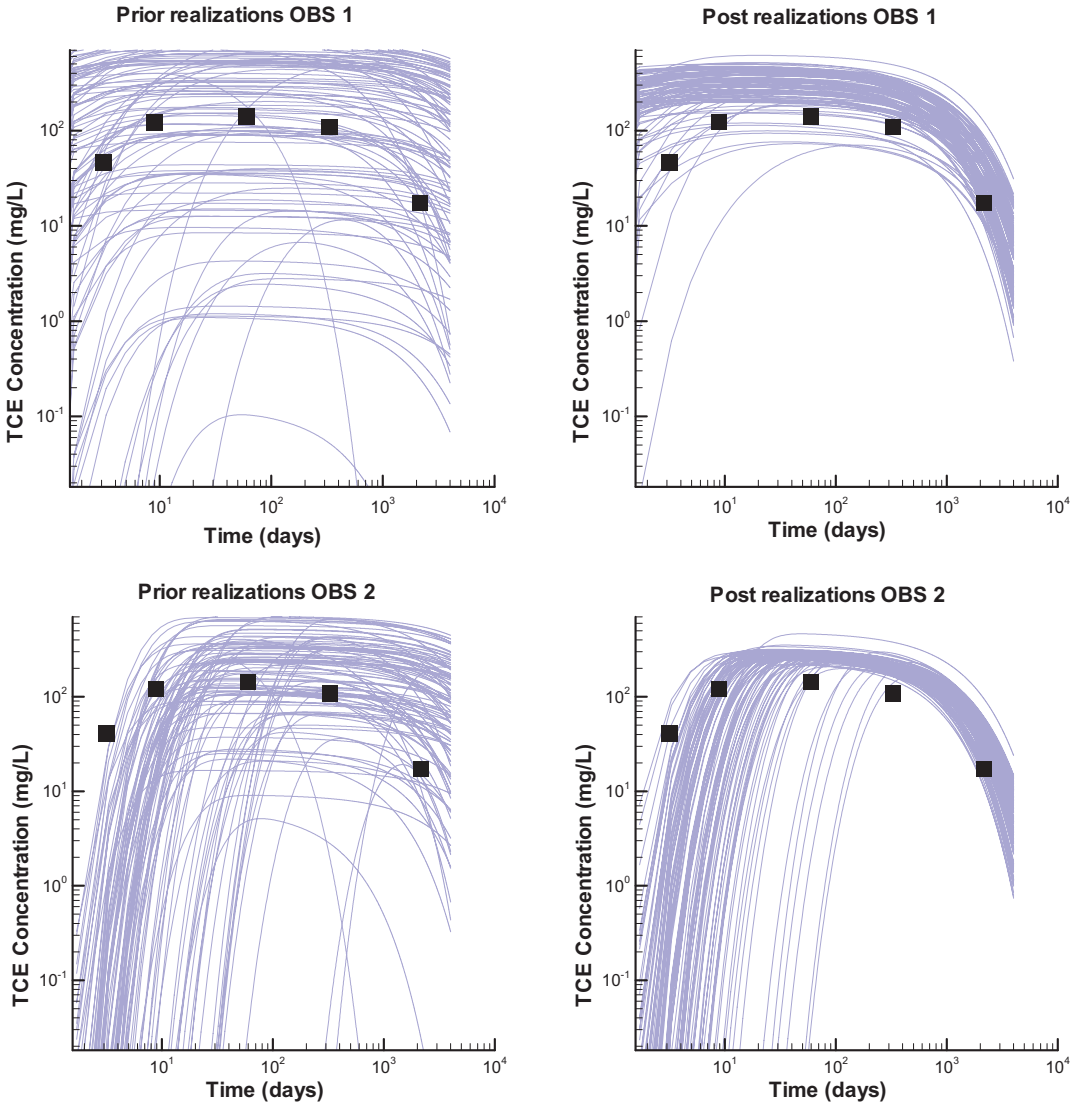

Fig. 7. Stochastic generation of breakthrough curves using realizations of the different parameters drawn from the prior and the posterior marginal distributions. Square symbols denote the observed measurements of concentrations.

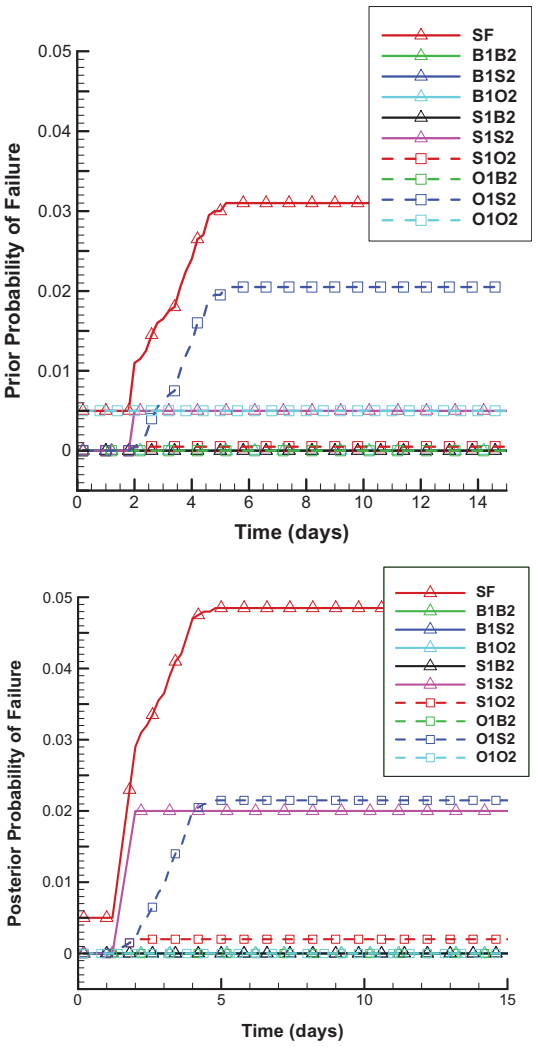

Fig. 8. Time evolution of the probability of system failure.

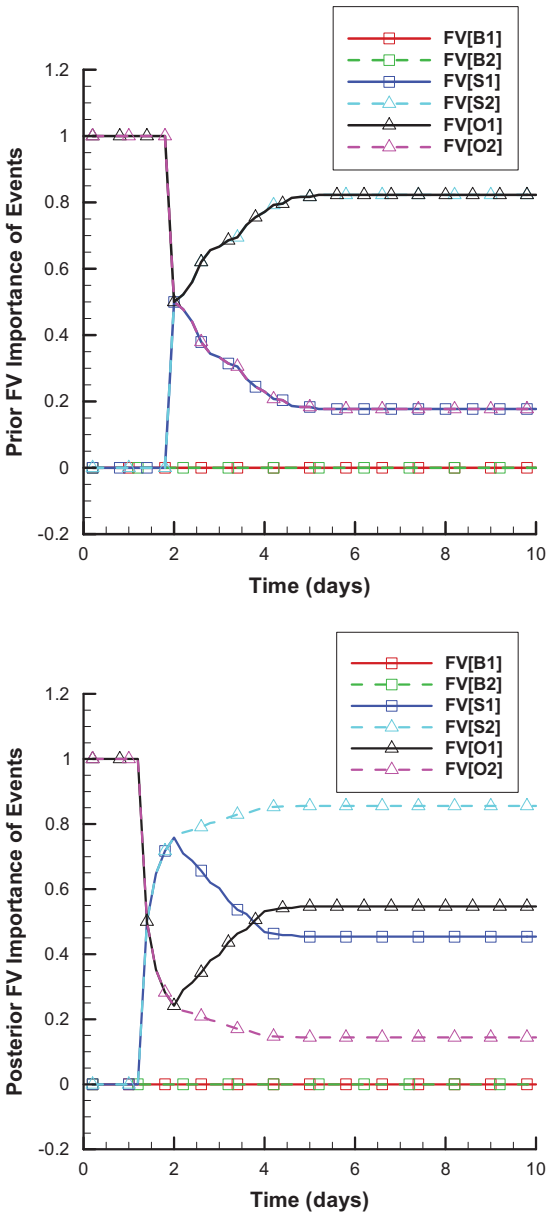

Fig. 9. Time evolution of the FV importance of events.

The posterior multivariate distributions of the system parameters were used to generate their equally likely realizations conditioned on the data. Fig. 7 shows 100 realizations of the concentration breakthrough curves at the observation wells. These realizations correspond to parameter values drawn from either the prior (blue[1] lines) or the posterior (red lines) marginal distributions with the sequential simulation algorithm. The Bayesian update automatically recalibrates the system parameters to improve predictions of our conceptual model. The bias in parameter estimates is an indication that the parameter values converge to their equivalent homogeneous counterparts.

Most importantly, our approach updates the probability of system failure over time, as depicted in Fig. 8. The update suggests that the risk of failure is higher than originally expected. This important result demonstrates that the addition of information does not necessarily imply a reduction of risk. The sampling established that the the observation wells are capable of detecting the leakage of dissolved TCE from the source zone. Hence, the likelihood of bypassing the monitoring system (event B_1B_2) becomes negligible. At the same time, the plume's apparent velocity is larger than its original estimate, identifying the existence of a preferential pathway. As a result, for the constant sampling frequency, the posterior likelihood of system failure increases to account for the simultaneous occurrence of a low sampling frequency at the two observation wells (event S_1S_2).

Estimation of the probabilities of the basic events in the fault tree enables us to identify the relative importance of these events for system failure. The latter can be used to either redesign or prioritize the remediation and future dynamics of the basic events. This feature has not been previously explored in probabilistic risk analyses (PRAs) in subsurface hydrology.

We demonstrate how a PRA can be used as a key tool for decision making. Several alternative definitions have been used to quantify the importance of basic events. The Fussel–Vesely (FV) importance of a basic event X is defined as the conditional probability of X given that the system fails, i.e., $P[X|SF]$. It can be estimated as

$$FV[X] = P[X|SF] \approx \frac{\sum_{M:X\in M} P[M]}{\sum_{M} P[M]}, \tag{34}$$

[1] For interpretation of color in Fig. 7, the reader is referred to the web version of this article.

where M denotes a cut set, and M: $X \in M$ refers to the cut sets involving X. For instance, the FV importance of S_2 (low sampling frequency in observation well 2) is given by

$$FV[S_2] \approx \frac{P[S_1 \cdot S_2|NAPL] + P[\overline{B}_1 \cdot S_2|NAPL]P[OBS] + P[B_1 \cdot S_2|NAPL]}{P[SF]/P[NAPL]}. \quad (35)$$

Fig. 9 shows the temporal evolution of the FV importance of the basic events involved in the monitoring system. Prior to sampling and for large times, the joint occurrence of well malfunctioning close to the source and insufficient sampling frequency near the receptor is the most critical mechanism by which the system fails (events O_1 and S_2). After sampling, the data suggests that the probability of having fast paths in the system is higher than initially postulated. As a result, the conditional probability of event S_1 increases.

7. Conclusions

Managing risk during remediation efforts is cumbersome due to the difficulty of combining a large number of uncertainties associated with subsurface heterogeneities, rate-limited mass transfer processes, and the existence of multiple potential sources, receptors, and pathways of exposure. We developed a general PRA framework to evaluate the risk of failure of a typical site remediation project that links multiple system components (exposure, sources and pathways) with their fate and transport phenomena as well as with a monitoring system. The proposed PRA method is used to determine the likelihood of a remediation effort's failure to prevent the contaminant plume from reaching a receptor, without being detected by a monitoring system.

Our PRA approach accounts for uncertainty about the size and location of the contamination, the sampling frequency at observation points and the probability that the contaminant bypasses the observation wells. It can be readily combined with any existing transport model of a polluted site to quantify the probabilities of basic events. A Bayesian interpretation of these probabilities within the PRA framework allows for the use of measurements from observation wells to update the probability of system failure over time. A fault tree analysis, an integral part of the PRA, attributes the failure of a monitoring system not to the failure of a single observation well but rather to the combination of multiple joint effects that can lead the monitoring system to fail.

We used a synthetic example of the management of a DNAPL-contaminated site to illustrate the importance of this feature. The example considers the release of dissolved TCE into a heterogeneous aquifer. The monitoring system consists of two observation wells. Our results show that the method can identify the sequence of events by which the monitoring system is most likely to fail. Prior to sampling, the design of the system dictates that the most critical sequence of events is the failure of the observation well close to the source due to malfunctioning, followed by the failure of the observation well closest to the receptor due to a small sampling frequency. As more information is added, the Bayesian interpretation of the problem allows us to automatically recalibrate the probability of system failure. Thus, after sampling, the PRA identifies the joint failure of the observation wells due to small sampling frequency is an important issue.

The example also illustrates that our PRA framework can be used as a tool for decision making. In order to identify the most critical events in the system, we introduced a quantitative measure of the Fussel–Vesely importance of an event. This information can be used to prioritize future developments or redesign the system.

Acknowledgments

This work was supported by the Water Catalan Agency (Agència Catalana de l'Aigua), by the Spanish Ministry of Science and Innovation with Projects CONSOLIDER-Ingenio 2010 (CSD2009-00065), RARAAVIS (CGL2009-11114), and HEROS (CGL 2007-66748). The work of D.M.T. was supported by Applied Mathematics program of the US DOE Office of Advanced Scientific Computing Research.

References

[1] Bolster D, Barahona M, Dentz M, Fernàndez-Garcia D, Sanchez-Vila X, Trinchero P, et al. Probabilistic risk analysis of groundwater remediation strategies. Water Resour Res 2009;45:W06413. doi:10.1029/2008WR007551.
[2] Cardiff M, Liu X, Kitanidis PK, Parker J, Kim U. Cost optimization of DNAPL source and plume remediation under uncertainty using a semi-analytic model. J Contam Hydrol 2010;113(1–4):25–43.
[3] Christakos G. A Bayesian maximum entropy view to the spatial estimation problem. Math Geol 1990;22(7):763–77.
[4] Dagan G. Flow and transport in porous formations. Berlin: Springer-Verlag; 1989.
[5] Dean DW, Illangasekare TH, Russell TF. A stochastic differential equation approach for modeling of NAPL flow in heterogeneous porous media. Tech. rep., University of Colorado at Denver and Health Sciences Center, Denver, CO; 2005.
[6] DiFilippo EL, Brusseau ML. Relationship between mass-flux reduction and source-zone mass removal: analysis of field data. J Contam Hydrol 2008;98:22–35.
[7] EIREDA. European industry reliability data handbook. Tech. rep., C.E.C J.R.C, ISEI 21020 ISPRA, Varese, Italy; 1991.
[8] Ewing RP, Berkowitz B. Stochastic pore-scale growth models of DNAPL migration in porous media. Adv Water Resour 2001;24(3–4):309–23.
[9] Fernàndez-Garcia D, Illangasekare TH, Rajaram H. Differences in the scale dependence of dispersivity estimated from temporal and spatial moments in chemically and physically heterogeneous porous media. Adv Water Res 2005;28:745–59.
[10] Fernàndez-Garcia D, Rajaram H, Illangasekare TH. Assessment of the predictive capabilities of stochastic theories in a three-dimensional laboratory test aquifer: effective hydraulic conductivity and temporal moments of breakthrough curves. Water Resour Res 2005;41:W04002. doi:10.1029/2004WR003523.
[11] Fernàndez-Garcia D, Sanchez-Vila X. Optimal reconstruction of concentrations, gradients and reaction rates from particle distributions. J Contam Hydrol 2011;120–121:99–114.
[12] Feyen L, Ribeiro PJ, Smedt FD, Diggle PJ. Bayesian methodology to stochastic capture zone determination: conditioning on transmissivity measurements. Water Resour Res 2002;38:1164. doi:10.1029/2001WR000950.
[13] Gelhar L, Axness C. Three-dimensional stochastic analysis of macrodispersion in aquifers. Water Resour Res 1983;19(1):161–80.
[14] Gelhar LW. Stochastic subsurface hydrology. Upper Saddle River, NJ: Prentice-Hall; 1993.
[15] Gómez-Hernández JJ, Cassiraga FE. Theory and practice of sequential simulation. In: Armstrong M, Dowd P, editors. Geostatistical simulation. Kluwer; 1993.
[16] Gómez-Hernández JJ, Journel AG. Joint simulation of multi-Gaussian random variables. In: Soares A, editor. Geostatistics tróia'92, vol. 1. Kluwer; 1993.
[17] Harbaugh AW, Banta ER, Hill MC, McDonald MG. The US Geological Survey modular ground-water model-user guide to modularization concepts and the ground-water flow process, open-file report 00-92. Tech. rep., US Geological Survey, Denver, CO; 2000.
[18] Hoeting JA, MADigan D, Raftery AE, Volinsky CT. Bayesian model averaging: a tutorial. Stat Sci 1999;14(4):382–417.
[19] James B, Gorelick S. When enough is enough: the worth of monitoring data in aquifer remediation design. Water Resour Res 1994;30(12):3499–513.
[20] Kitanidis PK. Parameter uncertainty in estimation of spatial functions: Bayesian analysis. Water Resour Res 1986;22(4):499–507. doi:10.1029/WR022i004p00499.
[21] Massmann J, Freeze RA. Groundwater contamination from waste management sites: the interaction between risk-based engineering design and regulatory policy, 1. Methodology. Water Resour Res 1987;23(2):351–67.
[22] Maxwell RM, Kastenberg WE, Rubin Y. A methodology to integrate site characterization information into groundwater-driven health risk assessment. Water Resour Res 1999;35(9):2841–55.
[23] Maxwell RM, Pelmulder SD, Tompson AFB, Kastenberg WE. On the development of a new methodology for groundwater-driven health risk assessment. Water Resour Res 1998;34(4):833–47.
[24] Mayer A, Endres KL. Simultaneous optimization of dense non-aqueous phase liquid (DNAPL) source and contaminant plume remediation. J Contam Hydrol 2007;91(3–4):288–311.
[25] McClelland GH, Schulze W, Hurd B. The effect of risk beliefs on property values: a case study of a hazardous waste site. Risk Anal 1990;10(4):467–85. doi:10.1111/j.1539-6924.1990.tb00534.x.

[26] Morales-Casique E, Neuman SP, Guadagnini A. Non-local and localized analyses of non-reactive solute transport in bounded randomly heterogeneous porous media: theoretical framework. Adv Water Res 2006;29(9):1399–418.
[27] Mosleh A, Apostolakis G. The development of a generic database for failure rates. In: ANS/ENS international topical meeting on probabilistic safety methods and applications, San Francisco; 1985.
[28] Natvig B. Recent development in multistate reliability theory. Probabilistic models in mechanics of solids and structures. Berlin: Springer; 1985.
[29] Neuman SP. Maximum likelihood Bayesian averaging of uncertain model predictions. Stoch Environ Res Risk Assess 2003;17(15):291–305.
[30] Nowak W, de Barros F, Rubin Y. Bayesian geostatistical design: task-driven optimal site investigation when the geostatistical model is uncertain. Water Resour Res 2010;46(W03535). doi:10.1029/2009WR008312.
[31] Parker JC, Park E. Evaluation of an upscaled model for DNAPL dissolution kinetics in heterogeneous aquifers. Adv Water Resour 2005;28:1280–91.
[32] Parker JC, Park E. Effects of mass reduction, flow reduction and enhanced biodecay of DNAPL source zones. Transport Porous Med 2008;73:95–108. doi:10.1007/s11242-007-9164-x.
[33] Ptak T, Piepenbrink M, Martac E. Tracer tests for the investigation of heterogeneous porous media and stochastic modelling of flow and transport – a review of some recent developments. J Hydrol 2004;294(1–3):122–63.
[34] Riva M, Guadagnini A, Fernàndez-Garcia D, Sanchez-Vila X, Ptak T. Relative importance of geostatistical and transport models in describing heavily tailed breakthrough curves at the Lauswiesen site. J Contam Hydrol 2008;101:1–13.
[35] Rubin Y, Chen X, Murakami H, Hahn M. A Bayesian approach for inverse modeling, data assimilation, and conditional simulation of spatial random fields. Water Resour Res 2010;46:W10523. doi:10.1029/2009WR008799.
[36] Saenton S, Illangasekare TH, Soga K, Saba TA. Effects of source zone heterogeneity on surfactant-enhanced NAPL dissolution and resulting remediation end-points. J Contam Hydrol 2002;59(1–2):27–44.
[37] Salamon P, Fernàndez-Garcia D, Gómez-Hernández JJ. Modeling mass transfer processes using random walk particle tracking. Water Resour Res 2006;42:W11417. doi:10.1029/ 2006WR004927.
[38] Salamon P, Fernàndez-Garcia D, Gómez-Hernández JJ. Modeling tracer transport at the MADE site: the importance of heterogeneity. Water Resour Res 2007;43:W08404. doi:10.1029/ 2006WR005522.
[39] Sanchez-Vila X, Guadagnini A, Fernàndez-Garcia D. Conditional probability density functions of concentrations for mixing-controlled reactive transport in heterogeneous aquifers. Math Geosci 2009;41:323–51. doi:10.1007/s11004-008-9204-2.
[40] Schubert M, na PP, Balcázar M, Meissner R, Lopez A, Flores JH. Determination of radon distribution patterns in the upper soil as a tool for the localization of subsurface NAPL contamination. Radiat Meas 2005;40(2–6):633–7.
[41] Soga K, Page J, Illangasekare TH. A review of NAPL zone remediation efficiency and the mass flux approach. J Hazard Mater 2004;110:13–27.
[42] Soga K, Page JWE, Illangasekare TH. A review of NAPL source zone remediation efficiency and the mass flux approach. J Hazard Mater 2004;110(1–3):13–27.
[43] Tartakovsky AM, Meakin P, Huang H. Stochastic analysis of immiscible displacement of the fluids with arbitrary viscosities and its dependence on support scale of hydrological data. Adv Water Resour 2004;27(12):1151–66.
[44] Tartakovsky DM. Probabilistic risk analysis in subsurface hydrology. Geophys Res Lett 2007;34:L05404. doi:10.1029/2007GL029245.
[45] Tartakovsky DM, Broyda S. PDF equations for advective–reactive transport in heterogeneous porous media with uncertain properties. J Contam Hydrol 2010. doi:10.1016/j.jconhyd.2010.08.009.
[46] Tartakovsky DM, Dentz M, Lichtner PC. Probability density functions for advective–reactive transport in porous media with uncertain reaction rates. Water Resour Res 2009;45:W07414. doi:10.1029/2008WR007383.
[47] Winter CL, Tartakovsky DM. A reduced complexity model for probabilistic risk assessment of groundwater contamination. Water Resour Res 2008;44:W06501. doi:10.1029/2007WR006599.
[48] Woodbury AD, Rubin Y. A full-Bayesian approach to parameter inference from tracer travel time moments and investigation of scale effects at the Cape Cod experimental site. Water Resour Res 2000;36(1):159–71. doi:10.1029/1999WR900273.