A reduced complexity model for probabilistic risk assessment of groundwater contamination

C. L. Winter[1] and Daniel M. Tartakovsky[2]

[1] We present a model of reduced complexity for assessing the risk of groundwater pollution from a point source. The progress of contamination is represented as a sequence of transitions among coarsely resolved states corresponding to simple statements like "a spill has occurred." Transitions between states are modeled as a Markov jump process, and a general expression for the probability of aquifer contamination is obtained from two basic assumptions: that the sequence of transitions leading to contamination is Markovian and that the time when a given transition occurs is independent of its end state. Additionally, we derive an asymptotic value for the probability of contamination that is equivalent to the so-called rare event approximation. First we develop the model for sites in statistically homogeneous natural porous media, and then we extend it to highly heterogeneous media composed of multiple materials. Finally, we apply the model to a simple example to illustrate the method and its potential.

1. Introduction

[2] Risk has been defined as the probability that something bad will occur [*Brillinger*, 2003]. Contaminant spills at waste sites are among the bad things that can happen to groundwater systems. To evaluate the risk of a spill at a site we must have a means of calculating $P_{UC}(t_0, t_{crit})$, the probability that the site goes from an initially (at time t_0) uncontaminated state U to a contaminated state C before some critical time t_{crit}. In general, probability of contamination can be broken into two parts: $P_{US}(t_0, t_S)$, the probability that engineering protections at the site will fail and a spill from storage (S) will occur by time t_S; and $P_{SC}(t_S, t_{crit})$, the probability that the site's hydrogeology, including active measures like barrier wells, will fail and a spill will contaminate the site's groundwater. The first problem is the domain of engineers who must assess the likelihood that storage systems at a given site will fail. Since it is not intrinsically hydrologic, we do not treat $P_{US}(t_0, t_S)$ at all in this paper except to assume it can be solved. We concentrate on the hydrogeologic problem of estimating $P_{SC}(t_S, t_{crit})$, the risk that a known spill will contaminate a local aquifer, although we give a general expression for $P_{UC}(t_0, t_{crit})$ in equations (11)–(12) below.

[3] Attempts to estimate risk by deterministic approaches face challenges and contradictions that are both linguistic (risk, after all, does imply uncertainty and probability) and operational. Deterministic approaches attempt to describe contaminant process by solving appropriate transport equations, e.g., an advection-diffusion equation (ADE) or its

non-Fickian counterparts. Such efforts assume that a model provides a valid description of contaminant transport at a site, and its parameters (coefficients) and relevant initial and boundary conditions (IBC) are known. In practice, the parameters and IBC are never known well enough to directly solve transport equations for a specific site. One of two approaches is generally taken to address this lack of data. In the first, a calibration technique is used to estimate the parameters and IBC, and the ADE is then solved as if the system were deterministic. This approach to groundwater modeling has been documented in many references [e.g., *Spitz and Moreno*, 1966]. Calibration is generally ill posed because the resulting parameter estimates are rarely unique [*Moore and Doherty*, 2006]; many quite different parameterizations will yield about the same quality of calibration, yet lead to significantly different results on data that is independent of the values used to calibrate the model. Equally important, the calibration procedure by itself does not yield probabilistic estimates of risk despite the highly uncertain system properties.

[4] A comprehensive probabilistic risk analysis (PRA) [*Tartakovsky*, 2007] provides a rigorous and scientifically defensible means for estimating hydrogeologic risk. Specifically, a comprehensive PRA should account for structural (model) and parametric uncertainties associated with subsurface processes. Structural uncertainty arises from imperfect knowledge of the geologic makeup of the subsurface and from incomplete understanding of physical and biogeochemical processes affecting the fate and transport of contaminants at any given site. Geologic uncertainty and its effects on contaminant transport can be quantified by means of the random domain decomposition [*Winter and Tartakovsky*, 2002; *Guadagnini et al.*, 2004]. Model uncertainty often manifests itself via the existence of several competing conceptual and mathematical descriptions whose ability to accurately model naturally occurring transport

---

[1]National Center for Atmospheric Research, Boulder, Colorado, USA.
[2]Department of Mechanical and Aerospace Engineering, University of California, San Diego, La Jolla, California, USA.
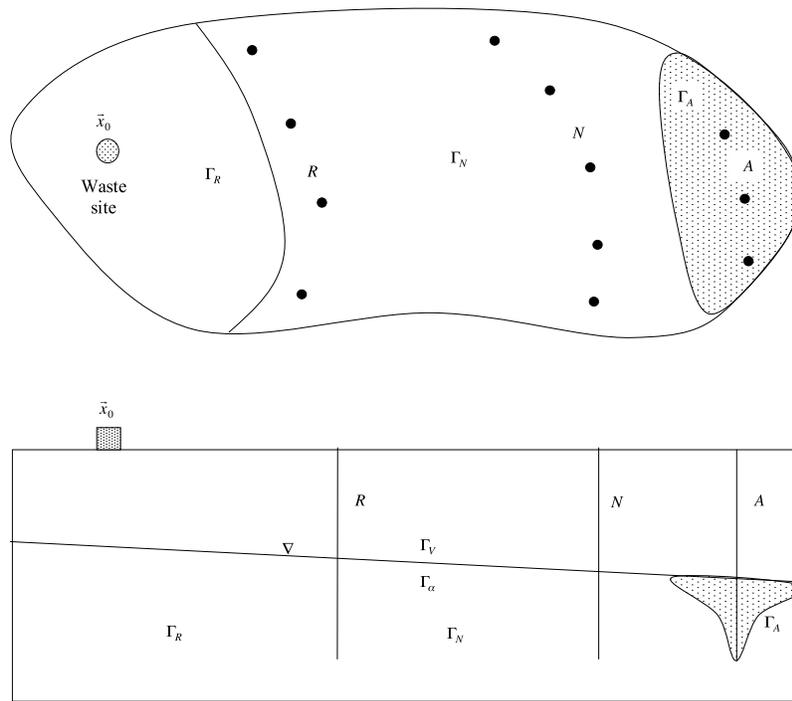
Figure 1. (top) Horizontal and (bottom) vertical cross sections of a typical remediation site.

phenomena cannot be validated by data with a required degree of fidelity. To quantify this source of uncertainty, one can use either a Bayesian maximum entropy approach [*Christakos*, 1990] or maximum likelihood Bayesian averaging [*Neuman*, 2003]. Parametric uncertainty arises from spatial heterogeneity coupled with limited and often noisy measurements of hydraulic and biogeochemical parameters, such as hydraulic conductivity and retardation coefficient. It is commonly quantified by solving stochastic ADEs or other stochastic models, including Lagrangian averaging and continuous time random walk [see *Cushman*, 1997; *Dentz and Tartakovsky*, 2006, and references therein].

[5] No matter which approach, deterministic or PRA, is taken, the upshot is a complicated model whose solution is computationally intensive, and often prohibitively so. This suggests there is room for simpler models based on a higher level of description than the detailed level of physical description captured in deterministic and stochastic models. We describe such a model in this paper. Specifically, we focus on the formalism by developing a model of reduced complexity of the risk that a given spill will eventually contaminate a set of water supply wells. A model of reduced complexity preserves the essential features of a problem, but does not represent the dynamics exactly. It is based on parameters that can be either measured in the field or laboratory, or estimated probabilistically by an expert through an "educated guess." In our case, we represent groundwater flow and transport as a sequence of transitions between discrete states of a waste site–aquifer system. These states are easy to understand and generally form an adequate description of such events for decision makers. The dynamics of contamination are represented by a Markov jump process in which random state transitions occur at random times. Such a model is consonant with the usual limitations of data available for characterizing a

site. The model requires fewer than a dozen parameters to specify the contamination process.

[6] First we define the elements of waste site dynamics at a level appropriate to a model of reduced complexity. These include the basic geometry of the model (section 2), and the states of the reduced system and allowable transitions between them (section 3). In section 4, we develop a model of risk for the reduced system and derive a basic formula for quantifying risk of contamination (equations (11)–(12) below). The setting and framework are similar to those of *Tartakovsky* [2007], and one of our contributions is to derive his operational formulas from the dynamics of our reduced complexity model. We also extend the analysis to formations composed of multiple materials (section 5) and illustrate the method by applying it to a couple of typical contamination problems (section 6). We close by discussing the model's parametric requirements and complexity (section 7).

2. Conceptualization of Remediation Site

[7] We begin by formalizing a typical point source groundwater remediation problem shown in Figure 1. A remediation system occupies a subsurface volume Γ that contains a waste site and a set $A = \{w_1, \ldots, w_{n_A}\}$ of n_A wells used to produce water from an aquifer $\Gamma_\alpha \in \Gamma$. Each well $w_a = (x_a, y_a, l_a)$, for $a = 1, \ldots, n_A$, is characterized by $\mathbf{x}_a = (x_a, y_a)$, its location on the ground surface, and by l_a, its pumping depth below the ground surface. In addition to the aquifer Γ_α, the subsurface Γ contains the vadose (unsaturated) zone Γ_V, such that $\Gamma = \Gamma_\alpha \cup \Gamma_V$ and $\Gamma_\alpha \cap \Gamma_V = 0$. Since the volume of the waste site is much smaller than Γ, we represent its spatial extent by a point $\mathbf{x}_0 = (x_0, y_0)$ on the ground surface. Other commonly encountered contamination problems, wherein a contamination source is located within

Table 1. Possible States $\Sigma(t)$ of the Contaminated Site Shown in Figure 1

State, $\Sigma(t)$	State Description
U	The site is uncontaminated
S	A spill has occurred
R	The site is undergoing remediation
N	The site is undergoing natural attenuation
C	The aquifer is contaminated

Table 2. Possible Transitions Between the System States Shown in Table 1

Transition	Outcome of the Transition
US	The site is contaminated by a spill
SR	The contaminant escapes the waste site
SU	The spill is contained on site
RN	Remediation fails
NC	Natural attenuation fails
RU	Remediation succeeds
NU	Natural attenuation succeeds

the vadose zone, can be handled in the identical fashion. The production zone of A is the volume $\Gamma_A \in \Gamma_\alpha$ affected by the pumping in A.

[8] The waste site $\mathbf{x}_0$ is contained within a remediation volume $\Gamma_R \in \Gamma$, which is upstream of Γ_A and can include both unsaturated and saturated regions. The unsaturated region can be represented by a separate volume if necessary. In general, Γ_R is bounded by some kind of barrier to contaminants migrating from $\mathbf{x}_0$ into the production zone Γ_A. For example, the barrier may be defined by a collection of pumping wells designed to stop flow between the waste site and A. A monitoring network $R \in \Gamma_\alpha$ is deployed to observe the remediation effort. In most cases, $R = \{w_1, \ldots, w_{n_R}\}$ consists of n_R wells and is located downstream of Γ_R and upstream of Γ_A. Finally, there is another region $\Gamma_N \in \Gamma$, downstream of Γ_R and upstream of Γ_A, where contamination that escapes remediation efforts may still be mitigated by natural attenuation. The effectiveness of natural attenuation is monitored by $N = \{w_1, \ldots, w_{n_N}\}$, another set of n_N wells. In many cases, these are the production wells A, although we show them as separate in Figure 1 for completeness.

3. Model's States and Transitions

[9] The essential elements of this problem can be described as sequences of transitions among system states that are shown in Table 1. We represent transitions between states as pairs, for instance SR, where the first element, S, is the initial state and the second element, R, is the result of the transition. Only a few of the 20 possible transitions can actually occur in our model system. These are shown in Table 2.

[10] The relative sequence of transitions is not arbitrary, and allowable sequences are illustrated in Figure 2. A sequence (for instance, $US \rightarrow SR$) is only allowed if the result of the first transition is the initial state of the second, i.e., if the second element of the first pair is the first element of the second. Allowable sequences can be identified by means of fault tree diagrams [*Tartakovsky*, 2007]. The sequence

$$US \rightarrow SR \rightarrow RN \rightarrow NC \qquad (1a)$$

leads to aquifer contamination and forms the so-called minimal cut [*Tartakovsky*, 2007]. As noted, we will mostly be interested in sequences in which a spill has already occurred and escaped the waste site, i.e., with the sequence

$$SR \rightarrow RN \rightarrow NC \qquad (1b)$$

but the extension of the technique to (1a) is obvious.

[11] A transition $\sigma \rightarrow \sigma'$ from a state $\sigma \in \{U, S, R, N, C\}$ to another allowable state $\sigma' \in \{U, S, R, N, C\}$ is said to

occur when a criterion based on contaminant concentration $c(\mathbf{x}, t)$ is met. Typically, the criterion is that concentration at the state σ exceeds a critical value $c_\sigma = \text{const}$ at any monitoring well w_i, i.e., that $c(w_i, t) > c_\sigma$. If the transition does occur, it occurs over a random interval of time $\tau_{\sigma\sigma'}$, which is defined as the random time taken by the system to move from state σ to state σ'. For instance, τ_{RN} is the random interval of time required for the contaminant plume to pass from detection by one of the wells of R to detection by one of the wells of N, and $\overline{\tau}_{RN}$ is its expected value. Transition durations associated with sequence (1b) are shown in Figure 2, and their precise, operational definitions are given in section 4.

[12] Besides the duration of a transition, we are also concerned with the random time τ_σ, when the system first enters a given state σ. For instance, $\tau_R = \tau_{US} + \tau_{SR}$ is the time when a contaminant first enters Γ_R. If a spill is known to have occurred already, we use t_0 to indicate the known time of the spill.

[13] The critical time is set by environmental regulations. Although in many cases $t_{crit} = \infty$, that is not always so. For instance the regulatory time at the proposed U.S. nuclear waste repository at Yucca Mountain was once set at $t_{crit} = 10,000$ years.

4. Probability Model

[14] Our goal is to estimate the probability that a spill occurring at t_0 contaminates the aquifer by a given critical
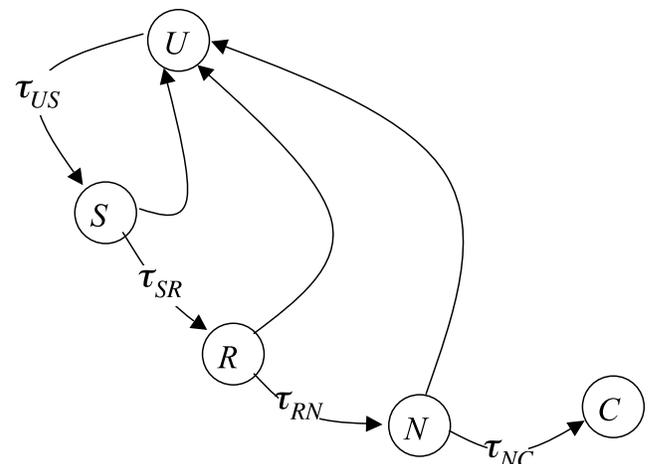
Figure 2. Allowable transitions and transition times of the model.

time t_{crit}. In general we represent the event of being in a given state σ by time t with the pair of events $\Omega(\tau_\sigma) = \sigma$ and $\tau_\sigma < t$,

$$\sigma(t) \equiv \{\Omega(\tau_\sigma) = \sigma, \tau_\sigma < t\}. \qquad (2)$$

Similarly, the transition from a given current state $\sigma(t)$ to another, say σ' at time $t' > t$, is

$$\begin{aligned}\sigma(t) \to \sigma'(t') &\equiv \{\sigma(t); \sigma'(t')\} \\ &= \{\Omega(\tau_\sigma) = \sigma, \tau_\sigma < t; \Omega(\tau_{\sigma'}) = \sigma', \tau_{\sigma'} < t'\}. \end{aligned} \qquad (3)$$

Depending on the setting, we use several alternative notations for the transition probability, i.e., probability P of the transition $\sigma(t) \to \sigma'(t')$ in (3),

$$\begin{aligned}P[\sigma(t); \sigma'(t')] &\equiv P[\Omega(\tau_\sigma < t) = \sigma; \Omega(\tau_{\sigma'} < t') = \sigma'] \\ &\equiv P_{\sigma\sigma'}[t < \tau < t']. \end{aligned} \qquad (4)$$

Finally, $P[\sigma(t); \sigma'(t'); \ldots; \sigma''(t'')]$ denotes the joint probability of multiple transitions from states $\sigma, \sigma', \ldots, \sigma''$ for times $t < t' \ldots < t''$.

[15] The probability of a sequence of transitions leading to aquifer contamination is $P[S(t_0); R(t_R); N(t_N); C(t_{crit})]$. Hence, the probability of aquifer contamination before t_{crit}, given a spill at t_0, is

$$P_{SC}(t_0, t_{crit}) = \iint P[S(t_0); \rho; \nu; C(t_{crit})] \mathrm{d}\rho \mathrm{d}\nu, \qquad (5)$$

where ρ and ν range over all possible $R(t_R)$ and $N(t_N)$, respectively. We spend the rest of this section making this statement precise. We do so in two stages for clarity.

[16] At the first stage, we treat the transition times as known, i.e., $\tau_\sigma = t_\sigma$, and decompose the joint probabilities of the state transitions into constituent conditional probabilities,

$$\begin{aligned}P[S(t_0); R(t_R); N(t_N); C(t_{crit})] &= P[C(t_{crit})|S(t_0), R(t_R), N(t_N)] \\ &\times P[N(t_N)|S(t_0), R(t_R)] \\ &\times P[R(t_R)|S(t_0)] \times P[S(t_0)]. \end{aligned} \qquad (6)$$

Of course, $P[S(t_0)] = 1$ if it is known that a spill occurred at t_0.

[17] It is reasonable to assume that the state transitions form a Markov process since for example, the probability that the system is contaminated by time t_{crit} given that $\Omega(t_N) = N$ does not depend on time t_R when $\Omega(t_R) = R$. All that matters is that natural attenuation fails at t_N. How the system got to that point is irrelevant. (State transitions are distinct from transition times that are sometimes used to model contaminant transport [see *Cushman*, 1997; *Dentz*

and Tartakovsky, 2006, and references therein]. The latter might or might not be Markovian depending on the conceptual model used to describe transport phenomena.) The Markovian assumption allows us to replace (6) with

$$\begin{aligned}P[S(t_0); R(t_R); N(t_N); C(t_{crit})] &= P[C(t_{crit})|N(t_N)] \times P[N(t_N)|R(t_R)] \\ &\times P[R(t_R)|S(t_0)] \times P[S(t_0)]. \end{aligned} \qquad (7)$$

From now on we will let $t_0 = 0$ when convenient without any loss of generality.

[18] At the second stage, we allow transition times τ_σ to be random. Consider $P_{\sigma'\sigma''}[t' < \tau < t'']$, the general transition from state σ' to state σ'' in random time τ. Now we make our second basic assumption: the probability of a transition can be represented as the product

$$P_{\sigma'\sigma''}[t' < \tau < t''] = F_{\sigma''}(t'')Q_{\sigma'\sigma''}, \qquad (8)$$

where $Q_{\sigma'\sigma''}$ is the probability of the state transition ever occurring, and $F_{\sigma'}(t'')$ is the probability that this transition will occur by t''. The Markovian nature of transition times requires [*Doob*, 1953] that $F_{\sigma''}$ be an exponential distribution defined below. If these conditions hold, (8) describes a Markov jump process from the current state σ' to a possible next state σ''. The probability that the system stays in state σ until time t_σ, i.e., that the jump has not occurred by t_σ, is given by

$$F_\sigma(t_\sigma) = \int_0^{t_\sigma} q_\sigma e^{-q_\sigma \tau_\sigma} \mathrm{d}\tau_\sigma. \qquad (9)$$

The rate parameter $q_\sigma > 0$ has units of $1/t$, and $1/q_\sigma = E[\tau_\sigma] \equiv \overline{\tau}_\sigma$ is the expected time before the transition from σ occurs. Note that $\overline{\tau}_C > \overline{\tau}_N > \overline{\tau}_R > \overline{\tau}_S$, i.e., $q_S > q_R > q_N > q_C$.

[19] Substituting (8) into (7), we obtain

$$\begin{aligned}P[S(t_0); R(t_R); N(t_N); C(t_{crit})] &= Q_{US}F_S(t_0)Q_{SR}F_R(t_R)Q_{RN}F_N(t_N) \\ &\cdot Q_{NC}F_C(t_{crit}). \end{aligned} \qquad (10)$$

Now we can complete the definition of $P_{UC}(t_0, t_{crit}) \equiv P[U(t_0) \to C(t_{crit})]$. Letting $t_0 = 0$,

$$P_{UC}(0, t_{crit}) = Q_{US}Q_{SR}Q_{RN}Q_{NC}\mathcal{I}, \qquad (11)$$

where

$$\begin{aligned}\mathcal{I} = &\int_0^{t_{crit}} \mathrm{d}\tau_{CN} q_C e^{-q_C\tau_{CN}} \int_0^{t_{crit}-\tau_{CN}} \mathrm{d}\tau_{NR} q_N e^{-q_N\tau_{NR}} \\ &\times \int_0^{t_{crit}-\tau_{CN}-\tau_{NR}} \mathrm{d}\tau_{RS} q_R e^{-q_R\tau_{RS}} \\ &\times \int_0^{t_{crit}-\tau_{CN}-\tau_{NR}-\tau_{RS}} \mathrm{d}\tau_{SU} q_S e^{-q_S\tau_{SU}}. \end{aligned} \qquad (12a)$$
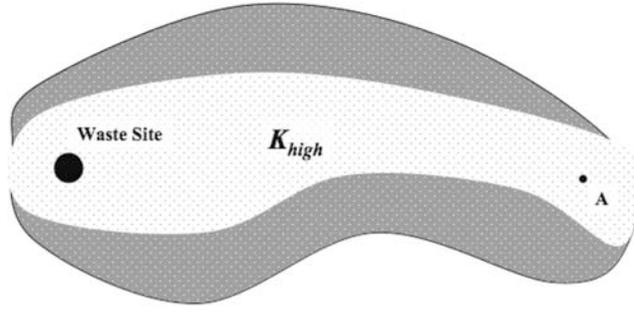
Figure 3. A schematic representation of preferential flow paths.

Evaluating the quadratures in (12a) yields

$$
\begin{aligned}
\mathcal{I} = {} & 1 - \mathrm{e}^{-q_C t_{crit}} + q_C \frac{\mathrm{e}^{-q_C t_{crit}} - \mathrm{e}^{-q_N t_{crit}}}{q_C - q_N} \\
& + q_C q_N \frac{\mathrm{e}^{-q_N t_{crit}} - \mathrm{e}^{-q_C t_{crit}}}{(q_C - q_N)(q_N - q_R)} \\
& + q_C q_N \frac{\mathrm{e}^{-q_C t_{crit}} - \mathrm{e}^{-q_R t_{crit}}}{(q_C - q_R)(q_N - q_R)} \\
& + q_C q_N q_R \frac{\mathrm{e}^{-q_C t_{crit}} - \mathrm{e}^{-q_N t_{crit}}}{(q_C - q_N)(q_N - q_R)(q_R - q_S)} \\
& + q_C q_N q_R \frac{\mathrm{e}^{-q_R t_{crit}} - \mathrm{e}^{-q_C t_{crit}}}{(q_C - q_R)(q_N - q_R)(q_R - q_S)} \\
& + q_C q_N q_R \frac{\mathrm{e}^{-q_N t_{crit}} - \mathrm{e}^{-q_C t_{crit}}}{(q_C - q_N)(q_N - q_S)(q_R - q_S)} \\
& + q_C q_N q_R \frac{\mathrm{e}^{-q_S t_{crit}} - \mathrm{e}^{-q_C t_{crit}}}{(q_N - q_S)(q_R - q_S)(q_S - q_C)}.
\end{aligned}
\tag{12b}
$$

When $t_{crit} \to \infty$, $\mathcal{I} = 1$ and

$$
P_{UC}(0, t_{crit}) = Q_{US} Q_{SR} Q_{RN} Q_{NC},
\tag{13}
$$

which is essentially the rare event approximation discussed by *Tartakovsky* [2007]. It is clear from (13) that $Q_{\sigma_1 \sigma_2}$ is the probability that the given transition (Q_{US}, Q_{SR}, Q_{RN}, or Q_{NC}) ever occurs.

5. Preferential Flow Paths

[20] Most natural porous media are heterogeneous rather than uniform, and many exhibit separable macroscopic and microscopic scales of heterogeneity [*Winter and Tartakovsky*, 2002]. On the macroscopic scale, such media consist of n_{vol} large subvolumes, or blocks, each of which is composed of a single specific material. In Figure 3, two low-conductivity blocks are separated by a high-conductivity path. The microscopic scale of heterogeneity corresponds to point-point variability within a block. Typically, within block variability contributes much less to the total variance of hydrogeologic parameters than across block variability [*Winter and Tartakovsky*, 2000]. Since we are not concerned with the detailed geometry of flow, we can ignore microscopic variability.

[21] In such systems, preferential flow paths exist at macroscopic scales if a continuous volume of highly conductive material is embedded in Γ. When preferential paths exist, we can break a given transition, say $\sigma' \to \sigma''$, into the

flow occurring in high-conductivity zones and the flow that does not. Hence, we can treat a state transition within a zone as a Markov jump process as before but split the transition into a set of transitions,

$$
P_{\sigma_1 \sigma_2}^{(i)}(0, t_2) = F_2^{(i)}(t_2) Q_{\sigma_1 \sigma_2}^{(i)}, \qquad i = 1, \dots, n_{vol}, \tag{14}
$$

each one of which occurs within a given block subvolume V_i. The hydrology of the setting limits the amount of flow between blocks, so the probability of transport between blocks is small and we ignore it. Several modeling options are available when exchanges between blocks must be accommodated, and we return to them at the end of the section.

[22] Let $\tau^{(i)}$ be the total time required to pass from S to C in V_i. We want to evaluate the probability

$$
\begin{aligned}
P[\tau_C < t_{crit}] &= P\left[\min_i \tau_C^{(i)} < t_{crit} \right] \\
&= 1 - \prod_i P\left[\tau_C^{(i)} > t_{crit} \right].
\end{aligned}
\tag{15}
$$

Often one highly conductive flow volume is surrounded by less conductive volumes, in which case

$$
P\left[\min_i \tau_C^{(i)} < t_{crit} \right] \approx \max_i P\left[\tau_C^{(i)} < t_{crit} \right].
\tag{16a}
$$

To put this in the notation of (11)–(12),

$$
\begin{aligned}
P_{SC}(0, t_{crit}) &= 1 - \prod_i \left[1 - P_{SC}^{(i)}(0, t_{crit}) \right] \\
&\approx \max_i P_{SC}^{(i)}(0, t_{crit})
\end{aligned}
\tag{16b}
$$

with $P_{SC}^{(i)}(0, t_{crit})$ denoting the transition probability in V_i.

[23] When exchanges between blocks are significant, the blocks are not really separate and may be treated as one superblock with its own time parameters based on measurements of the superblocks performance. Then the methods of section 4, especially equations (11)–(12), can be applied. Alternatively, additional states pertaining to the blocks (e.g., "The contaminant is in the ith block") and transitions among them can be defined if the data and available expertise can support the needed parameterizations.

6. Examples

[24] A couple of examples will help understand the method. We begin with the uniform porous medium sketched in Figure 1 and set the time at which a spill is known to have occurred at $t_0 = 0$. In this case, (11)–(12) reduce to

$$
P_{SC}(0, t_{crit}) = Q_{SR} Q_{RN} Q_{NC} \mathcal{I}
\tag{17}
$$

with

$$
\begin{aligned}
\mathcal{I} = {} & \frac{1 - q_N q_R (q_N - q_R) \mathrm{e}^{-q_C t_{crit}}}{(q_C - q_R)(q_C - q_N)(q_N - q_R)} \\
& + \frac{q_C q_R (q_C - q_R) \mathrm{e}^{-q_N t_{crit}} + q_C q_N (q_C - q_N) \mathrm{e}^{-q_R t_{crit}}}{(q_C - q_R)(q_C - q_N)(q_N - q_R)}.
\end{aligned}
\tag{18}
$$
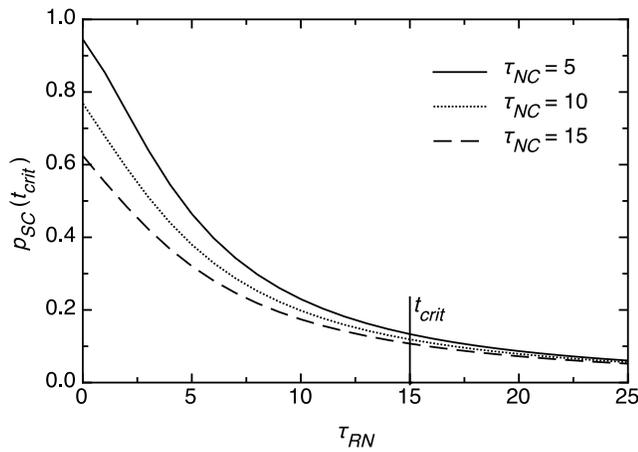
Figure 4. Probability of contamination $P_{SC}(t_{crit})$, scaled with $Q_{SR}Q_{RN}Q_{NC} = 0.01$ as a function of τ_{RN}, the time lag introduced by remediation.

The factor $Q_{SR}Q_{RN}Q_{NC}$ is the probability, irrespective of time, that a sequence of jumps will eventually lead to aquifer contamination. Setting all of these probabilities to 1 results in the most conservative estimate of the risk of contamination.

[25] A few general remarks are in order. Most waste sites are small in comparison to Γ and do not have provisions for containing potential spills, so we let $Q_{SR} = 1.0$ in this example; that is, it is certain that a spill will escape the waste site. On the other hand, remediation measures can be instituted after a spill, and effective measures will lead to small Q_{RN}, so we set $Q_{RN} = 0.01$. Unless the parent material has special geochemical properties, natural attenuation will often ultimately fail, so in our examples $Q_{NC} = 1.0$. The upshot is a value,

$$Q_{SR}Q_{RN}Q_{NC} = 0.01. \qquad (19)$$

Since modifications needed to adapt $Q_{SR}Q_{RN}Q_{NC}$ to other cases are clear, we turn our attention to $\mathcal{I}$, the remaining factor in (17).

[26] Of the three rates, only the remediation rate $q_N = 1/\overline{\tau}_N$ ($\overline{\tau}_N = \overline{\tau}_{SR} + \overline{\tau}_{RN}$) can be controlled by engineering. The other two, q_R and q_C, depend on natural properties of the hydrogeologic system, so we keep them basically the same in the three examples. As noted, in most cases the volume of the waste site and the contaminant's path through the unsaturated zone will be much smaller than the volume available for natural attenuation, hence $q_R \gg q_C$ (equivalently, $\tau_R \ll \tau_C$). We let $q_R = 0.1$ in our examples. The rate q_C, of course, is affected by q_N since it corresponds to the expected elapsed time $\overline{\tau}_C = \overline{\tau}_{SR} + \overline{\tau}_{RN} + \overline{\tau}_{NC}$. Hence, we maintain a constant difference $\overline{\tau}_{NC} = 5, 10, 15$ in the example shown in Figure 4, which plots $\mathcal{I}$ as a function of τ_{RN}, the time lag introduced by remediation. The longer remediation remains effective, the lower is the probability that the site will be contaminated before t_{crit} (= 15, in the example).

[27] Finally, we illustrate a case of preferential flow paths in the setting of Figure 3. In this example we have two subvolumes, the first of which is highly conductive, while the second is not. The corresponding parameters are shown

in Table 3. They result in $P[\min_i \tau_i < t_{crit}] = 0.652$ and $\max_i P[\tau_i < t_{crit}] = 0.646$, so the approximation $\max_i P[\tau_i < t_{crit}] \approx P[\min_i \tau_i < t_{crit}]$ is pretty good.

7. Discussion

[28] We have presented a simplified probabilistic model that is suitable for evaluating the risk to an aquifer of a contaminant spill at a waste site. The model is based on decomposing the process of contaminant transport into sequences of transitions among discrete states. The model represents transitions between states as the outcomes of a Markov jump process, which is a standard model for processes composed of discrete states. Model definition consists of two steps: (1) state specification and ordering and (2) parameterization of the jump process. Selection of both is a matter of an interplay between diverse sets of data and expert opinion.

[29] The actual state-wise decomposition depends on details of a given site, but the basic methodology of decomposition is the same for any site. State selection and ordering are largely the job of an expert relying on whatever sources of hydrogeologic information are available for characterizing features of the site. Decomposition requires about the same level of expertise and information as specifying the distribution of hydrogeologic parameters and IBC in a typical deterministic groundwater model.

[30] The reduced model is based on two assumptions: (1) state transitions are Markovian and (2) the time when a transition occurs does not depend on the transition's end state. The first assumption seems reasonable, since the underlying process of contamination is Markovian, regardless of whether it is described by an advection-dispersion equation, or particle tracking, or stochastic Lagrangian models. Given the concentration of contaminant at two times $t' < t''$ in the past, predictions of future states of contaminant concentration only depend on the more recent state t''. The second assumption is stronger. It can be understood in light of regulatory requirements. Consider transitions from state N, "The system is undergoing natural attenuation." The system can only go in one of two directions from N: either to U, "The system is uncontaminated," or to C, "The system is contaminated." We say the system is in C if contaminant concentration exceeds a threshold, c_N, by time t_{crit}, otherwise it goes to U. In either event, the direction of the transition does not depend on the time.

[31] The reduced model has the great advantage of relying on a small number of fairly intuitive parameters. The first are simple transition probabilities $Q_{\sigma'\sigma''}$ of ever going from the current state σ' to a possible next state σ''. They can be based on field experiments, simulations, or expert opinion. Time is introduced through another set of parameters, the expected time $\overline{\tau}_{\sigma'\sigma''}$ needed to go from

Table 3. Probabilities for Two Subvolumes Shown in Figure 3 With $t_{crit} = 2.0$

Subvolume	q_R	q_N	q_C
Conductive	3.0	2.0	1.0
Nonconductive	0.3	0.2	0.1

σ' to σ''. These can be measured by simple tracer tests, or estimated through simulation or by expert opinion. In all, only a dozen or so parameter values must be set, even in a complex groundwater setting with preferential flow paths. As noted, the intuitive meaning of the parameters allows values to be set, discussed, and defended by experts. Of course, the parameters are only estimates and hence are uncertain, so Bayesian methods should be used to account for parametric uncertainty.

[32] The complexity of the model is reduced in several senses when it is compared to models based on deterministic or stochastic differential equations. In the first place, the reduced model requires many fewer degrees of freedom than a deterministic transport model in which parameters, e.g., hydraulic conductivity and dispersivity, must be set at every point of a computational domain. The reduced model is also computationally simpler than solving transport equations numerically, which requires a number of iterations through time over at least a two-dimensional grid. The computational demand of applying Monte Carlo techniques to calculate the statistics of stochastic equations is even greater. By contrast, (11)−(12) can be solved with a calculator and a pencil. The data requirements of (11)−(12) are also much lower than those needed to parameterize transport equations, be they stochastic or not. Furthermore, the reduced model needs simpler kinds of data. A large number of observations must be collected to characterize heterogeneous flow and transport parameters deterministically or statistically. Then the spatial statistics of the parameters must be calculated to generate realizations or moment approximations. A high degree of expert judgment must go into choosing the number of samples, their locations, and then the form of the parameter's structure functions. Compare that with the reduced model which, in principle, can be parameterized by tracer tests or expert opinion. To sum up, the reduced complexity risk assessment is to the comprehensive risk quantification, what a simple mass balance equation (i.e., a lumped-parameter model) is to a spatially distributed transport model.

[33] Being a model of reduced complexity, the proposed approach to risk assessment provides a coarse-grained analysis. When a more detailed assessment is required, one can employ the comprehensive framework for risk analysis [*Tartakovsky*, 2007], which allows one to deal with structural (model) and parametric uncertainties. Not surprisingly, this approach is significantly more computationally demanding.

References

Brillinger, D. R. (2003), Three environmental probabilistic risk problems, *Stat. Sci.*, *18*(4), 355–383.

Christakos, G. (1990), A Bayesian maximum entropy view to the spatial estimation problem, *Math. Geol.*, *22*(7), 763–777.

Cushman, J. H. (1997), *The Physics of Fluids in Hierarchical Porous Media: Angstroms to Miles*, Kluwer Acad., New York.

Dentz, M., and D. M. Tartakovsky (2006), Delay mechanisms of non-Fickian transport in heterogeneous media, *Geophys. Res. Lett.*, *33*, L16406, doi:10.1029/2006GL027054.

Doob, J. L. (1953), *Stochastic Processes*, John Wiley, New York.

Guadagnini, L., A. Guadagnini, and D. M. Tartakovsky (2004), Probabilistic reconstruction of geologic facies, *J. Hydrol.*, *294*(1–3), 57–67.

Moore, C., and J. Doherty (2006), The cost of uniqueness in groundwater model calibration, *Adv. Water Resour.*, *29*(4), 55–77.

Neuman, S. P. (2003), Maximum likelihood Bayesian averaging of uncertain model predictions, *Stochastic Environ. Res. Risk Assess.*, *17*(5), 291–305.

Spitz, K., and J. Moreno (1966), *A Practical Guide to Groundwater and Solute Transport Modeling*, Wiley-Interscience, Hoboken, N. J.

Tartakovsky, D. M. (2007), Probabilistic risk analysis in subsurface hydrology, *Geophys. Res. Lett.*, *34*, L05404, doi:10.1029/2007GL029245.

Winter, C. L., and D. M. Tartakovsky (2000), Mean flow in composite porous media, *Geophys. Res. Lett.*, *27*(12), 1759–1762.

Winter, C. L., and D. M. Tartakovsky (2002), Groundwater flow in heterogeneous composite aquifers, *Water Resour. Res.*, *38*(8), 1148, doi:10.1029/2001WR000450.

———————————

D. M. Tartakovsky, Department of Mechanical and Aerospace Engineering, University of California, San Diego, La Jolla, CA 92093, USA. (dmt@ucsd.edu)

C. L. Winter, National Center for Atmospheric Research, Boulder, CO 80307-3000, USA. (lwinter@ucar.edu)