

Probabilistic risk analysis in subsurface hydrology

Daniel M. Tartakovsky^{1,2}

Received 7 January 2007; revised 5 February 2007; accepted 14 February 2007; published 15 March 2007.

[1] We present a general framework for probabilistic risk assessment (PRA) of subsurface contamination. PRA provides a natural venue for the rigorous quantification of structural (model) and parametric uncertainties inherent in predictions of subsurface flow and transport. A typical PRA starts by identifying relevant components of a subsurface system (e.g., a buried solid-waste tank, an aquitard, a remediation effort) and proceeds by using uncertainty quantification techniques to estimate the probabilities of their failure. These probabilities are then combined by means of fault-tree analyses to yield probabilistic estimates of the risk of system failure (e.g., aquifer contamination). Since PRA relies on subjective probabilities, it is ideally suited for assimilation of expert judgment and causal relationships. **Citation:** Tartakovsky, D. M. (2007), Probabilistic risk analysis in subsurface hydrology, *Geophys. Res. Lett.*, *34*, L05404, doi:10.1029/2007GL029245.

1. Introduction

[2] Accurate and verifiable predictions of subsurface flow and transport are notoriously illusive due to the heterogeneity of a typical subsurface environment, the lack of sufficient site characterization, and, occasionally, the inadequacy of conceptualizations and mathematical descriptions of relevant physical and bio-geochemical processes. These factors introduce fundamental uncertainty about subsurface systems and cast doubts on the feasibility and, in fact, desirability of obtaining a single deterministic prediction of system's behavior.

[3] With a possible exception of hydrogeology, both scientific and engineering communities have fully embraced the importance of dealing with, and quantifying, predictive uncertainty. In earth sciences, for example, the NRC report by the *Senior Seismic Hazard Analysis Committee (SSHAC)* [1997] states unambiguously that any seismic hazard analysis (SHA) must be probabilistic and that the main focus of any probabilistic SHA (PSHA) should be on uncertainty quantification.

[4] Many probabilistic analyses distinguish two types of uncertainty: epistemic and aleatory. The former is defined as “the uncertainty attributable to incomplete knowledge about a phenomenon that affects our ability to model it” and the latter as “the uncertainty inherent in a nondeterministic (stochastic, random) phenomenon” [SSHAC, 1997]. However, the review of the NRC report by the *National Research Council (NRC)* [1997] concluded that “unless

one accepts that all uncertainty is fundamentally epistemic, the classification of PSHA uncertainty as aleatory or epistemic is ambiguous.” The review further concludes that the epistemic/aleatory separation is somewhat artificial and not needed in practical applications. Fortunately, hydrogeologic applications are virtually free of this controversy, since most subsurface processes and parameters (e.g., hydraulic conductivity) are inherently deterministic, so that one is primarily concerned with epistemic uncertainty.

[5] Epistemic uncertainty in subsurface phenomena has been the subject of stochastic hydrogeology, whose progress over the last several decades can be discerned from the earliest monograph by *Shvidler* [1964] to the latest by *Rubin* [2003]. (Mathematical representations of deterministic parameters, e.g., hydraulic conductivity, as random fields explain a seeming contradiction of using stochastic tools to deal with epistemic uncertainty.) Despite significant theoretical advancements, including an explanation of anomalous behavior of contaminant transport in heterogeneous subsurface environments [e.g., *Cushman*, 1997], stochastic hydrogeology remains a largely academic pursuit.

[6] In this letter, we introduce a general framework for probabilistic risk assessment (PRA) that bridges this gap by presenting information (e.g., the risk of aquifer contamination in 100 years is X and the best remediation strategy is A) in a way that is more accessible and useful to practitioners and decision makers than that routinely used by scientists (e.g., the ensemble mean and standard deviation of a contaminant concentration at the water table in 100 years will be \bar{C} and σ_C , respectively). While this framework is applicable to a variety of subsurface contamination problems, we demonstrate its basic ideas and concepts by considering a simple example described in the following section.

2. Probabilistic Risk Analysis (PRA)

[7] The concentration of a contaminant in groundwater, C , is fundamental in determining environmental health risks. For example, if the contaminant in question is carcinogenic, its impact on human health can be quantified by the Excess Lifetime Cancer Risk (ELCR) factor [U.S. Environmental Protection Agency (U.S. EPA), 1992]

$$\text{ELCR} = \alpha C, \quad \alpha = \frac{\text{IR} \times \text{EF}}{365 \text{ days} \times \text{BW}}, \quad (1)$$

where IR is human ingestion rate, EF is exposure frequency, and BW is average body weight. While any or all parameters entering the expression for α can be uncertain, it is common to take IR = 2 liters/day, EF = 350 days/year, and BW = 70 kg. According to U.S. Environmental Protection Agency [2004], the levels of a carcinogen (e.g., benzene) in groundwater are considered safe (i.e., health risks are believed to be acceptable) if ELCR is within the

¹Department of Mechanical and Aerospace Engineering, University of California, San Diego, La Jolla, California, USA.

²Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico, USA.

range $[10^{-4}, 10^{-6}]$. A key goal of probabilistic risk assessment is to determine the probability that the carcinogen's concentration C in groundwater exceeds, at time $t = T$, the EPA mandated value of $C^* = \alpha^{-1}$ ELCR.

[8] Most present-day analyses of groundwater contamination lack probabilistic risk assessment and uncertainty quantification of the kind routinely required in other engineering and environmental disciplines [Covello and Mumpower, 1985]. Specifically, a comprehensive risk analysis should provide answers to the following three questions: "What can happen? How likely is it to happen? Given that it occurs, what are the consequences?" [Bedford and Cooke, 2003]. Several recent studies [e.g., Batchelor et al., 1998; Chen et al., 2003; Wang and McTernan, 2002; Li et al., 2006] addressed some of these questions, but not all.

[9] A typical subsurface system consists of the vadose zone and a saturated zone (aquifer), both of which are heterogeneous and often composed of multiple geologic facies. Consider a contaminant that migrates through the vadose zone from a (localized or distributed) source (e.g., a soil surface spill or a compromised subsurface waste storage facility) to the water table. We say that the subsurface system "fails" at time $t = T$, if the contaminant concentration at the water table exceeds the EPA mandated levels of $C^* = \alpha^{-1}$ ELCR. Our objective is to assess both the likelihood of the system failure (aquifer contamination) and the efficiency of alternative remediation strategies.

[10] Following Bedford and Cooke [2003], we start by constructing a *fault tree* (Figure 1), which relates the occurrence of the system failure, i.e., aquifer contamination, to the failures of its constitutive parts (basic events), i.e., the occurrence of a spill, the failure of natural attenuation, and the failure of a remediation effort. The term "natural attenuation" is used here in a broad sense to include not only "the combination of natural biological, chemical, and physical processes that act without human intervention to reduce the mass, toxicity, mobility, or concentration of the contaminants" [Alvarez and Illman, 2006], but also usual transport mechanisms (advection, diffusion, and dispersion) that enable the contaminant to migrate through the vadose zone. The Boolean operators AND and OR indicate a collection of basic events that would lead to aquifer contamination.

[11] The second step is to identify *minimal cut sets* of the system, i.e., the smallest collections of events that lead to aquifer contamination. The fault tree in Figure 1 reveals two such minimal cuts: {Spill occurs (SO), Natural attenuation fails (NA)} and {Spill occurs (SO), Remediation effort fails (RE)}.

[12] The third step is to represent the fault tree in Figure 1 by a Boolean expression. Recalling that the Boolean operators AND and OR applied to two events X and Y can be written as $X \text{ AND } Y \equiv X \cdot Y \equiv X \cap Y$ and $X \text{ OR } Y \equiv X + Y \equiv X \cup Y$, the Boolean expression corresponding to the fault tree in Figure 1 is

$$\begin{aligned} \text{"Aquifer contamination(AC)"} &= \text{SO} \cdot (\text{NA} + \text{RE}) \\ &= \text{SO} \cdot \text{NA} + \text{SO} \cdot \text{RE}. \end{aligned} \tag{2}$$

The latter expression is known as a *cut set representation* of the fault tree in Figure 1.

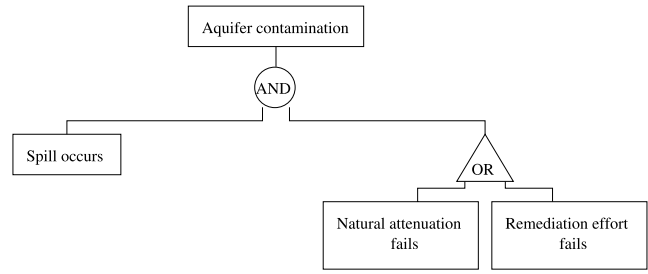


Figure 1. Fault tree for a possible aquifer contamination.

[13] The final step is to use (2) to compute $P[AC]$, the probability of aquifer contamination at time $t = T$, as

$$\begin{aligned} P[AC] &= P[(\text{SO} \cap \text{NA}) \cup (\text{SO} \cap \text{RE})] \\ &= P[\text{SO} \cap \text{NA}] + P[\text{SO} \cap \text{RE}] - P[\text{SO} \cap \text{NA} \cap \text{RE}]. \end{aligned} \tag{3}$$

The uncertainty about the basic event "Spill occurs (SO)" relates not only to the occurrence of a contaminant release per se, but also to its precise location, strength, toxicity, duration, etc. Let $P[\text{SO}]$ denote the probability that "Spill occurs (SO)", and $P[X|Y]$ denote the probability of the occurrence of X conditioned on the occurrence of Y . Then (3) can be rewritten as $P[AC] = \{P[\text{NA}|\text{SO}] + P[\text{RE}|\text{SO}] - P[\text{NA} \cap \text{RE}|\text{SO}]\} P[\text{SO}]$, which leads to

$$P[AC|\text{SO}] = P[\text{NA}|\text{SO}] + P[\text{RE}|\text{SO}] - P[\text{NA} \cap \text{RE}|\text{SO}]. \tag{4}$$

Suppose, for the sake of simplicity, that the spill has already occurred and that all of its characteristics are known with certainty. Then $P[\text{SO}] = 1$, and the probability of aquifer contamination is given by

$$P[AC] = P[\text{NA}] + P[\text{RE}] - P[\text{NA} \cap \text{RE}]. \tag{5}$$

[14] In engineering applications, the probability of a component failure, e.g., $P[\text{NA}]$ or $P[\text{RE}]$, is typically small and the probability of a simultaneous failure of more than one components, e.g., $P[\text{NA} \cap \text{RE}]$, is often an order of magnitude smaller than that. While the latter observation is not universal, it allows one to further simplify (5) by employing a *rare event approximation* [Bedford and Cooke, 2003],

$$P[AC] \approx P[\text{NA}] + P[\text{RE}], \tag{6}$$

in which the probability of a system failure depends exclusively on the probability of failure of its constitutive parts. The validity of this approximation in the hydrogeologic context is discussed in the following section.

3. PRA in Hydrogeology

[15] The procedure described above is used extensively in probabilistic risk assessments for complex artificial systems, such as nuclear power plants and space shuttles [Bedford and Cooke, 2003]. Its use in the hydrogeologic context is much more challenging due to the following reasons.

3.1. Definition of Probabilities

[16] All probabilities in modern PRAs are *subjective*, i.e., probability is defined as “degree of belief, of one individual” [Bedford and Cooke, 2003]. Subjective probabilities in hydrogeology must be used to deal with both parametric and structural (model) uncertainties, since many fundamental issues in subsurface flow and transport are yet to be resolved [Christakos, 1992; Rubin, 2003]. This is witnessed, for example, by the ongoing debates on whether hydraulic conductivity can be accurately represented by multi-Gaussian random fields and whether subsurface transport can be adequately modeled with advection-dispersion equations. The subjective beliefs of this author result in negative answers to both of these questions, but other opinions, and mathematical models based on these opinions, should be incorporated into a rigorous PRA to avoid a systematic predictive bias.

[17] In other words, the ability to incorporate expert knowledge into quantitative subsurface modeling becomes paramount. Fortunately, several formal approaches to dealing with diverging expert opinions originally developed in economics [e.g., Otway and von Winterfeldt, 1992] can be adapted in hydrogeology, just as it has been done in seismology [SSHAC, 1997; NRC, 1997]. Their practical implementation is a separate topic that lies outside the scope of the present letter.

3.2. Computation of Probabilities

[18] PRAs used in most engineering applications rely on reliability databases and manufacturing specifications to estimate the probabilities of basic events. In hydrogeology, one must solve (stochastic) differential equations to compute probabilities, which reflect structural (model) and parametric uncertainties associated with subsurface processes. Recent advances in stochastic hydrogeology and other scientific and engineering fields allow one to quantify these uncertainties and, thus, to compute the probabilities of the basic events in (3). Some of the relevant approaches are mentioned below.

[19] Structural uncertainty arises from imperfect knowledge of the geologic makeup of the subsurface and from incomplete understanding of physical and bio-geochemical 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 describe naturally occurring transport 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].

[20] Parametric uncertainty arises from spatial heterogeneity coupled with limited and often noisy measurements of hydraulic and bio-geochemical parameters, such as hydraulic conductivity and retardation coefficient. Its quantification is the traditional *raison d’être* of stochastic hydrogeology [Shvidler, 1964], which routinely employs Monte Carlo simulations and moment analyses [Rubin, 2003] and, more

recently, polynomial chaos expansions [e.g., Ghanem, 1998]. These approaches are currently used to compute the mean and variance of contaminant concentration and require further assumptions to compute the probabilities in (3). Alternatively, one can attempt to derive deterministic equations for the probability density function of contaminant concentration [Shvidler and Karasaki, 2003; Tartakovsky et al., 2003].

3.3. Dependent Probabilities

[21] The rare event approximation (6) provides a conservative estimate of the probability of aquifer contamination, which might prove to be overly pessimistic in many situations. When the rare event approximation becomes invalid (e.g., if the probabilities of failure of both natural attenuation and the remediation effort are larger than 0.5), (5) must be used instead.

[22] In the example considered above, the failure of both natural attenuation and a remediation effort often stems from a *common cause*, such as “the presence of a preferential flow path” (PF) between the source of contamination and the water table. In principle, the methods for uncertainty quantification described above are capable of computing not only the probabilities of basic events, such as the failures of natural attenuation $P[NA]$ and a remediation effort $P[RE]$, but also the probability of their joint failure $P[NA \cap RE]$. However, their practical implementation might prove to be computationally prohibitively expensive.

[23] The computational burden can be reduced if one may assume that the presence of a preferential flow path is the only common cause of the failure of both natural attenuation and the remediation effort. This assumption results in a conservative estimate of contamination risks and implies that PF completely couples the occurrence of NA and RE but does not necessarily cause them

$$P[NA \cap RE|PF] \approx P[NA|PF]. \quad (7)$$

Let PF' denote the absence of a preferential flow path, whose probability is $P[PF'] \equiv 1 - P[PF]$. Since

$$P[NA \cap RE] = P[NA \cap RE|PF]P[PF] + P[NA \cap RE|PF']P[PF'], \quad (8)$$

the approximation (7) yields

$$P[NA \cap RE] \approx P[NA|PF]P[PF] + P[NA \cap RE|PF']P[PF']. \quad (9)$$

Assume next that, in the absence of a preferential flow path, NA and RE are independent, $P[NA \cap RE|PF'] \approx P[NA|PF']P[RE|PF']$. Then (9) becomes

$$P[NA \cap RE] \approx P[NA|PF]P[PF] + P[NA|PF']P[RE|PF']P[PF']. \quad (10)$$

Finally, if $P[PF] \ll 1$ it is reasonable to assume that $P[NA|PF'] \approx P[NA]$ and $P[RE|PF'] \approx P[RE]$, so that

$$P[NA \cap RE] \approx P[NA|PF]P[PF] + P[NA]P[RE]P[PF']. \quad (11)$$

An in-depth discussion of conditional independence in Bayesian systems, which encompasses the analysis above, is given by Pearl [2000]. Expression (11) is analogous to the *common cause approximation* used in reliability analysis [Bedford and Cooke, 2003]. The probability of aquifer contamination can now be computed by combining (5) and (11).

[24] The proposed approach can be used to estimate the probability of success of alternative remediation strategies by minimizing the effect of dependent probabilities. For example, the existence of a preferential flow path might have more severe implications for the success of an in situ remediation effort than for the success of an attempt to construct a hydrologic barrier.

4. Problem-Specific PRAs

[25] An additional significant reduction in the computational burden associated with a comprehensive uncertainty quantification required by modern PRAs can be achieved by making them problem specific. For example, a wide range of contamination problems, including the one described in section 2, do not require point-wise probabilistic predictions of contaminant behavior. Instead, mass-balance calculations are often sufficient to assess probabilities of the occurrence of basic events by quantifying epistemic uncertainty in appropriate lumped-parameter models.

[26] These models result in simple, closed-form expressions for the bulk behavior of a contaminant and its response to various remediation strategies. A typical example of such analyses is provided by Rabideau *et al.* [1999], who modeled the remediation of TCE-contaminated soils by air sparging. Their analytical solutions depend on a number of parameters, some of which are measured (e.g., flow rates) and some are fitted to data (e.g., sorption coefficient). In any field application, the values of these parameters are uncertain and should be modeled probabilistically.

[27] Suppose that a probabilistic analysis of the lumped-parameter model and site-characterization data resulted in the following (subjective) probabilities: The probability that TCE reaches the water table at time $t = T$ through a preferential flow path (PF) is $P[\text{PF}] = 0.01$; The probabilities of failure of natural attenuation and the remediation effort at time $t = T$ are $P[\text{NA}] = 0.5$ and $P[\text{RE}] = 0.1$, respectively; If TCE were to migrate through the preferential flow path, the probabilities of failure of both natural attenuation and the remediation effort at time $t = T$ are $P[\text{NA}|\text{PF}] = P[\text{RE}|\text{PF}] = 1$.

[28] The probability of aquifer contamination at time $t = T$ computed with the rare event approximation (6) is $P[\text{AC}] = 0.6$, while its counterpart computed with the common cause approximation (5) and (11) is $P[\text{AC}] \approx 0.54$. Note that the contribution of the low-probability common cause PF to both the probability of the joint failure of natural attenuation and the remediation effort and the probability of aquifer contamination is quite significant. Another interesting observation is that the rare event approximation (6) gives a reasonably accurate risk estimate, even though the probability of failure of natural attenuation is 50%.

[29] This simple analysis can be used to obtain rough estimates of the risks posed by subsurface contamination. A

more detailed and rigorous PRA would require the use of the probabilistic and stochastic tools described in section 2.

5. Summary

[30] The general framework for probabilistic risk assessment (PRA) we presented in this letter is capable of handling subsurface phenomena that range from contamination of municipal water supplies to subsurface carbon-dioxide sequestration to oil recovery. This framework can be used to make decisions under uncertainty, including (1) determination of the viability of natural attenuation and other alternative remediation strategies, (2) optimization of data collection and monitoring campaigns, (3) selection of the optimal use of a contaminated site, and (4) assessment of subsurface water resource vulnerability. Key features of this approach are (1) the comprehensive treatment of structural (model) and parametric uncertainties inherent in subsurface flow and transport, and (2) the use of subjective probabilities, i.e., the reliance on expert knowledge.

[31] Any PRA of subsurface processes must be flexible and extensible enough to make optimal use of existing site characterization data and to accommodate new information, including new data, and conceptual models. The extensibility is critical for both the long-term relevance of the framework and its impact on the development and deployment of effective tools for monitoring and/or remediation of contaminated sites. It can be achieved by using some of the appropriate probabilistic tools for quantification of various types and levels of uncertainty that contribute to the overall predictive uncertainty. The flexibility comes from the modular use of some or all of these techniques and the possibility of incorporating other approaches.

[32] Applications of the PRA approach to complex subsurface problems might necessitate a computerized construction of fault trees. Commercially available PRA software often produces fault trees that are non-coherent, i.e., contain, in addition to the operators AND and OR, inverse operators of Boolean algebra. A probabilistic analysis of non-coherent trees is often preceded by the construction of *binary decision diagrams* [Bedford and Cooke, 2003].

[33] **Acknowledgments.** The prescient comments of the two reviewers have helped to improve the original draft of the manuscript and are greatly appreciated. This research was supported by the Office of Science (BER), U.S. Department of Energy, Cooperative Agreement DE-FC02-07ER64324.

References

- Alvarez, P. J. J., and W. A. Illman (2006), *Bioremediation and Natural Attenuation*, John Wiley, Hoboken, N. J.
- Batchelor, B., J. Valdes, and V. Araganth (1998), Stochastic risk assessment of sites contaminated by hazardous wastes, *J. Environ. Eng.*, 124, 380–388.
- Bedford, T., and R. Cooke (2003), *Probabilistic Risk Analysis: Foundations and Methods*, Cambridge Univ. Press, New York.
- Chen, Z., G. H. Huang, and A. Chakma (2003), Hybrid fuzzy-stochastic modeling approach for assessing environmental risks at contaminated groundwater systems, *J. Environ. Eng.*, 129, 79–88.
- Christakos, G. (1990), A Bayesian maximum entropy view to the spatial estimation problem, *Math. Geol.*, 22(7), 763–777.
- Christakos, G. (1992), *Random Field Models in Earth Sciences*, Elsevier, New York.
- Covello, V. T., and J. Mumpower (1985), Risk analysis and risk management: An historical perspective, *Risk Anal.*, 5(2), 103–120.
- Cushman, J. H. (1997), *The Physics of Fluids in Hierarchical Porous Media: Angstroms to Miles*, Springer, New York.

- Ghanem, R. (1998), Scales of fluctuation and the propagation of uncertainty in random porous media, *Water Resour. Res.*, 34(9), 2123–2136.
- Guadagnini, L., A. Guadagnini, and D. M. Tartakovsky (2004), Probabilistic reconstruction of geologic facies, *J. Hydrol.*, 294, 57–67.
- Li, J., L. Liu, G. Huang, and G. Zeng (2006), A fuzzy-set approach for addressing uncertainties in risk assessment of hydrocarbon-contaminated site, *Water Air Soil Pollut.*, 171, 5–18.
- National Research Council (NRC) (1997), *Review of Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on Uncertainty and Use of Experts*, Natl. Acad. Press, Washington, D. C.
- Neuman, S. P. (2003), Maximum likelihood Bayesian averaging of uncertain model predictions, *Stochastic Environ. Res. Risk Assess.*, 17(5), 291–305.
- Otway, H., and D. von Winterfeldt (1992), Expert judgment in risk analysis and management: Process, context, and pitfalls, *Risk Anal.*, 12(1), 83–93.
- Pearl, J. (2000), *Causality: Models, Reasoning, and Inference*, Cambridge Univ. Press, New York.
- Rabideau, A. J., J. M. Blayden, and C. Ganguly (1999), Field performance of air-sparging system for removing TCE from groundwater, *Environ. Sci. Technol.*, 33(1), 157–162.
- Rubin, Y. (2003), *Applied Stochastic Hydrogeology*, Oxford Univ. Press, New York.
- Senior Seismic Hazard Analysis Committee (SSHAC) (1997), Recommendations for probabilistic seismic hazard analysis: Guidance on uncertainty and use of experts, *Tech. Rep. NUREG/CR-6372*, U.S. Nucl. Regul. Comm., Washington, D. C.
- Shvidler, M. I. (1964), *Filtration Flows in Heterogeneous Media; A Statistical Approach*, Springer, New York.
- Shvidler, M., and K. Karasaki (2003), Probability density functions for solute transport in random field, *Transp. Porous Media*, 50(3), 243–266.
- Tartakovsky, D. M., P. C. Lichtner, and R. J. Pawar (2003), PDF methods for reactive transport in porous media, *IAHS Publ.*, 277, 162–167.
- U.S. Environmental Protection Agency (U.S. EPA) (1992), Guidelines for exposure assessment, *Rep. EPA/600/Z-92/001*, Washington, D. C.
- U.S. Environmental Protection Agency (2004), Integrated risk information system: Benzene, *Rep. CASRN 71-43-2*, Washington, D. C.
- Wang, T. A., and W. F. McTernan (2002), The development and application of a multilevel decision analysis model for the remediation of contaminated groundwater under uncertainty, *J. Environ. Manage.*, 64, 221–235.
- 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)