

## A Conditional Probability Model for Groundwater Risk Assessment

Grey S. Nearing<sup>\*</sup>, D. M. Tartakovsky<sup>†</sup> and C. L. Winter<sup>\*\*</sup>

<sup>\*</sup> Department of Hydrology and Water Resources  
University of Arizona, Tucson, AZ USA

<sup>†</sup> Department of Mechanical and Aerospace Engineering  
University of California San Diego, La Jolla, CA USA

<sup>\*\*</sup> Department of Hydrology and Water Resources  
University of Arizona, Tucson, AZ USA  
winter@hwr.arizona.edu

**Key words:** Groundwater contamination, risk, reduced complexity model

**Summary.** We present a model of reduced complexity for assessing the probabilistic risk of groundwater pollution from a contaminant spill at a waste storage site. The model is based on a logical decomposition of sources, flow, geochemistry, biology, aquifer contamination and so on into basic events. The influence of basic events on each other is reduced to subsets of events that are conditionally independent. The conditional probabilities can be obtained subjectively from experts or from physical data. We compute the subjective probability of aquifer contamination from various combinations of events for a simple example.

### 1 INTRODUCTION

Consider the threat that a chemical spilled at a waste storage site may pose to a nearby drinking water aquifer. Predicting the fate and transport of such a contaminant has a strong physical foundation (e.g., Darcy's Law, Fick's Law of Diffusion, reactive geochemistry) that in principle would allow exact prediction if complete information about the given hydrogeologic system were available. Since complete information is never available, uncertainty about flow and transport parameters always exists, and its effects must be quantified, usually probabilistically. The physical basis for contaminant transport suggests representing system performance through stochastic partial differential equations formally corresponding to the relevant physical laws, but with parameters (including initial and boundary conditions) that are random variables or random fields. In that case, the "solution" is either the moments of system state variables related to those of system parameters or approximate probability densities for system states. An extensive literature drawing on the theory of stochastic partial differential equations exists for such applications<sup>1,2</sup>. In this paper we discuss an alternative approach based on models of reduced complexity arising from the field of probabilistic risk assessment<sup>3</sup>, and we use it to predict the likelihood of aquifer contamination from a single point failure at a waste site.

The probabilistic risk assessment (PRA) approach to predicting contaminant transport is motivated by the relative insensitivity to model complexity observed in some predictions of subsurface flow and transport. Although it seems intuitively reasonable that detailed stochastic models based on extensive data about distributions of system parameters should increase confidence in predictions, in some cases detailed models may not improve over predictions obtained from simple models<sup>4</sup>. In many cases conclusions drawn from simple models are about as reliable as those based on detailed physical models<sup>5</sup>. The relative insensitivity of predictions to model complexity in some cases has led to interest in simple probabilistic models based on a few measurements which are relatively easy to make and/or can be based on the subjective opinions of experts<sup>6,7,8</sup>.

Rather than rely on explicit system dynamics, a model of reduced complexity,  $\mu$ , represents failure of a hydrogeologic containment system as the result of interactions among coarsened system states corresponding to statements like “engineered remediation fails,” “natural attenuation is ineffective,” and so on. As such,  $\mu$  is usually expressed in terms of system failure, usually aquifer contamination, represented as a (Boolean) logical combination of the basic events. The joint probability of the logical combination of events is expressed as independent constituent probabilities via probabilistic conditioning based on physical properties, engineering principles, and expert judgement. Industrial applications of PRA generally employ extensive databases derived from engineering reliability studies of the performance of mechanical, electrical, and chemical components<sup>3</sup>. No such databases are usually available for evaluating the performance of environmental containment systems that are primarily natural and whose engineered components, e.g., remediation barriers, are unique from application to application. Hence, the individual constituent probabilities of environmental PRA must usually be based on natural theory, observed data, experiment and/or expert judgement. Bolster et al.<sup>8</sup> gives an especially clear derivation of probabilities of basic events from hydrogeologic data and theory.

The application of PRA to hydrogeologic systems is at an early stage of model development where different approaches to system representation and prediction are still being explored. Verification and validation studies await the development of a mature approach to the entire cycle of model definition from identification of basic events and model structure to the definition and evaluation of constituent probabilities based on observations, experiments, and expertise. In this paper we develop a model of aquifer contamination using the Boolean approach pioneered by Tartakovsky<sup>6</sup> and taken by Bolster et al.<sup>8</sup>. The example is important in its own right: a site where the risk of aquifer contamination from a waste spill is combated with a combination of engineered remediation and natural attenuation. The method is based on constructing a fault tree corresponding to system failure<sup>3</sup>. Time does not enter explicitly into the analysis as it does in the finite-state machine approach taken by Winter and Tartakovsky<sup>6</sup>. Despite this limitation, fault tree decompositions provide a simple expression of system structure and are an obvious starting point for the development of joint probabilities. There seem to be no obstacles to combining fault tree analyses with models that explicitly include time, but we are unaware of anyone doing that. Instead we define a containment system in Section 2 that is the same as the one examined by Tartakovsky<sup>6</sup>, propose a new logic for it in Section 3 that we contrast with Tartakovsky<sup>6</sup>, and develop a

corresponding conditional probability model in Section 4. In Section 5 we apply the probability model using the same set of component probabilities as Tartakovsky<sup>6</sup> and contrast our results with his. We offer some concluding remarks in Section 6.

## 2 EXAMPLE SYSTEM

Suppose contamination of a drinking water aquifer (reservoir) depends on (i) a contaminant spill occurring at an associated waste storage site, (ii) the failure of efforts to remediate contamination through engineering, and (iii) the inability of natural attenuation to reduce concentration before the contaminant reaches the drinking water reservoir. The system fails at given time  $T$  if contaminant concentration exceeds a level in the reservoir that is mandated by a regulatory authority like the US Environmental Protection Agency. It is implicit that the volume encompassed by engineered remediation is outside the reservoir, that engineered remediation succeeds if peak concentration is lower than the regulatory threshold when the plume reaches the reservoir, and that natural attenuation fails if the concentration in the reservoir exceeds the threshold before  $T$ . Beyond that, estimates of of system failure ( $C = \text{true}$ ) depend on the specific logical form of its hypothesized relations to the occurrence of a spill ( $S = \text{true}$ ), remediation failure ( $R = \text{true}$ ), and ineffective natural attenuation ( $N = \text{true}$ ), and also on the potential existence of preferential flow paths ( $F = \text{true}$ ) that may carry high concentrations of a contaminant past engineered and natural barriers. From here on we simply write, e.g.,  $C$  for  $C = \text{true}$  and  $C'$  for  $C = \text{false}$ , the prime indicating negation (NOT) for logical events. Following Tartakovsky<sup>6</sup>, we address the effect of preferential paths after first examining the logical representation of the contaminant system.

## 3 CONTAMINANT SYSTEM LOGIC

PRA frequently uses Boolean logic to structure relations among elements of a system<sup>3</sup>. The logical structure proposed in Tartakovsky<sup>6</sup> for the example system is,

$$\mu 1: C = S(R + N) = SR + SN$$

where addition indicates logical OR, juxtaposition corresponds to AND, and  $S$ ,  $R$ , and  $N$  are basic events<sup>3</sup>. According to  $\mu 1$  the example aquifer is contaminated ( $C$ ) if a spill occurs ( $S$ ) and either engineered remediation fails ( $R$ ) or natural attenuation is ineffective ( $N$ ), but not necessarily both. This is equivalent to contamination occurring if either a spill occurs and remediation fails ( $SR$ ) or a spill occurs and natural attenuation is ineffective ( $SN$ ). Note that  $\mu 1$  can produce seeming contradictions since (i) a spill ( $S$ ) and ineffective natural attenuation ( $N$ ) can result in aquifer contamination despite a successful remediation effort ( $R$ ). Similarly, (ii) a spill and failed remediation ( $R$ ) can result in a contaminated aquifer even if natural attenuation is effective ( $N'$ ).

Bolster et al.<sup>8</sup> offers a different perspective on the logic of this general problem by examining a case in which remediation depends specifically on a reactive barrier and also on natural attenuation. They point out that such a barrier may be improperly placed so that contaminant transport bypasses the barrier completely and any chance of remediation rests

solely on the shoulders of natural attenuation. This might, at first glance, suggest that failure of natural attenuation (N) suffices for a spill to contaminate an aquifer (C) without regard to engineered barriers. In fact, however, failure of the barrier (R) is implied in this case because the required contaminant reduction cannot be achieved through engineering efforts (due to a misplaced barrier), hence engineered remediation fails (R) with probability 1 in the Bolster et al.<sup>8</sup> case.

Further, if this arguably intuitive definition of a successful mitigation effort is accepted, as it is in Bolster et al.<sup>8</sup>, aquifer contamination is only certain if all three basic events occur together, implying the alternative logic that we propose,

$$\mu_2: C = SRN.$$

In this model contamination is certain to occur if a spill occurs, engineered remediation fails, and natural attenuation is also ineffective. We will contrast  $\mu_1$  and  $\mu_2$  throughout the remainder of this paper, but it is clear at this point that  $\mu_2$  is not subject to the contradictions identified above.

Properties of the models can be further clarified by considering sets of basic events that guarantee the aquifer will not be contaminated ( $C'$ ).  $\mu_1$  asserts that the aquifer will not be contaminated,

$$C' = S' + R' N' \quad (1)$$

if a spill does not occur ( $S'$ ) or both remediation and natural attenuation succeed ( $R' N'$ ). While it is true that the system will not fail if both remediation and natural attenuation work, this seems overly restrictive since the system will also not fail if either engineered remediation or natural attenuation is effective by itself ( $R'$  OR  $N'$ ). The logic of  $\mu_2$  reflects this situation exactly, since

$$C' = S' + R' + N'. \quad (2)$$

In this case the aquifer is not contaminated if a spill does not occur ( $S'$ ), or engineered remediation succeeds ( $R'$ ), or natural attenuation is effective ( $N'$ ). Again, success is assumed to mean that the given process is effective in reducing contaminate levels to below the level which defines a contaminated reservoir.

#### 4 PROBABILITY MODEL

Models like  $\mu_1$  and  $\mu_2$  are used in PRA to structure the probabilities of events like aquifer contamination. To simplify the probability analysis, Tartakovsky<sup>6</sup> assumes that a spill has occurred, so  $P[S] = 1$ , in which case the probability of contamination corresponding to  $\mu_1$  amounts to

$$P_1[C] = P[R + N] = P[R] + P[N] - P[RN] \quad (3)$$

and for  $\mu_2$

$$P_2[C] = P[RN]. \quad (4)$$

Tartakovsky<sup>6</sup> also allows for the possibility that preferential flow paths may exist (F) or not (F'). Using the definition of conditional probability

$$P[\text{RN}] = P[\text{RNF}] + P[\text{RNF}'] = P[\text{RN}|F] P[F] + P[\text{RN}|F'] P[F'], \quad (5)$$

since  $P[F] + P[F'] = 1$ .

A couple of simplifying assumptions can be justified at this point based on the hydrogeology:

When  $F = \text{True}$ . First, the existence of a preferential flow path couples the joint failures of remediation and natural attenuation,

$$P[\text{RN}|F] = P[\text{R}|NF] P[\text{N}|F] = P[\text{N}|F], \quad (6)$$

that is, the occurrence of NF suffices through conditioning to assure remediation will fail to occur, so  $P[\text{R}|NF] = 1$ . The reasoning is the same as discussed earlier with regard to Bolster et al.<sup>8</sup>

When  $F = \text{False}$ . When a preferential path does not exist, it can be assumed that R and N are independent,

$$P[\text{RN}|F'] = P[\text{R}|F'] P[\text{N}|F']. \quad (7)$$

Based on (6) and (7), equation (4) becomes

$$P[\text{RN}] = P[\text{N}|F] P[F] + P[\text{R}|F'] P[\text{N}|F'] P[F']. \quad (8)$$

Note that this is the form of  $\mu_2$  (Eq 4) in this setting, while (Eq 3) for  $\mu_1$  is

$$P_1[C] = P[\text{R}] + P[\text{N}] - (P[\text{N}|F] P[F] + P[\text{R}|F'] P[\text{N}|F'] P[F']). \quad (9)$$

## 5 EXAMPLE

With (3) and (8) in hand Tartakovsky<sup>6</sup> considers a case in which  $P[F] \approx 0$ . When  $P[F]$  is small, it is assumed  $P[\text{R}] \approx P[\text{R}|F']$  and  $P[\text{N}] \approx P[\text{N}|F']$ . To make the example concrete, the paper supposes  $P[F] = 0.01$ ,  $P[\text{R}] = 0.1$ ,  $P[\text{N}] = 0.5$ , and  $P[\text{N}|F] = P[\text{R}|F] = 1$ . In that case  $P_1[C] \approx 0.54$  according to  $\mu_1$ , which it should be noted, is larger than the probability of either basic event. For instance, the model concludes that the combined mitigation effects of R and N are worse than just applying R (or N) alone. Such a degree of negative synergy seems unlikely in a system where preferential paths effectively do not occur and a contaminant must navigate both the engineered system and the natural system to reach the aquifer. The corresponding result using  $\mu_2$  is  $P_2[C] \approx 0.06$ , which is reasonable in such a system.

Next, consider the opposite case where  $P[F] \approx 1$ . Based on the reasoning of Tartakovsky<sup>6</sup>, this is a case in which aquifer contamination should be very likely to occur due to the almost certain presence of a preferential path. This is a strong test of multiplicative models like  $\mu_2$  where a small value in any constituent can lead to an estimate of probability of contamination  $\ll 1$ . Yet in this case  $\mu_2$  leads to  $P_2[C] = P[\text{RN}] \approx 1$  because i) it can still be assumed that  $P[\text{N}|F] = 1$ , i.e., it is certain that a contaminant will escape natural attenuation if it migrates on a preferential path (similarly,  $P[\text{R}|F] = 1$ ) and ii)  $P[F] \approx 1$ . Hence, under model  $\mu_2$ ,  $P_2[C] \approx 1$ ,

as it should. In this case  $\mu_1$  also leads to the appropriate conclusion because  $P[N] = P[N|F]P[F] + P[N|F']P[F'] \approx 1$  (similarly  $P[R] \approx 1$ ) so  $P_1[C] = P[N] + P[R] - P[RN] \approx 1 + 1 - 1 \approx 1$ .

## 6 CONCLUSIONS

We provide a straightforward model ( $\mu_2$ ) for calculating the risk that a contaminant spill from a waste storage site will foul an associated drinking water aquifer. As a practical alternative to the specific model ( $\mu_1$ ) proposed by Tartakovsky<sup>6</sup>, it strengthens the general approach to hydrogeologic PRA taken in that paper by allowing contamination (C) only if a waste spill occurs, engineered remediation fails (R), and natural attenuation proves ineffective (N). This conforms to the logic of the storage system analyzed and to intuition. It is born out by example calculations of the probability of aquifer contamination. In particular,  $\mu_1$  yields  $P_1[C] \gg P[R]$  (similarly,  $P_1[C] > P[N]$ ) for systems without preferential flow paths, i.e., the probability of system failure is amplified in  $\mu_1$  so that it is greater than the probability that either of its constituent elements fails. Our alternative yields  $P_2[C] \ll 1$  in this case, which is consistent with the constituent probabilities and the underlying logic of the example.

## REFERENCES

- [1] Zhang, D., *Stochastic methods for flow in porous media : coping with uncertainties*. Academic Press, London, UK (2001).
- [2] Rubin, Y., *Applied stochastic hydrogeology*. Oxford University Press, Oxford, UK (2003).
- [3] Bedford, T., and R. Cooke, *Probabilistic Risk Analysis: Foundations and Methods*. Cambridge Univ. Press, New York, NY (2003).
- [4] Scheibe T.D. and Y.J. Chien, "An evaluation of conditioning data for solute transport prediction," *Ground Water* 41 (2) (2003).
- [5] Hill, M.C., "The practical use of simplicity in developing ground water models," *Ground Water* 44 (6) (2006).
- [6] Tartakovsky, D. M., "Probabilistic risk analysis in subsurface hydrology," *Geophys. Res. Lett.*, 34, L05404, doi:10.1029/2007GL029245 (2007).
- [7] Winter, C.L., D.M. Tartkovsky, "A reduced complexity model for probabilistic risk assessment of groundwater contamination," *Water Resour. Res.* 44 (6) (2008).
- [8] Bolster D., M. Barahona, M. Dentz, D. Fernandez-Garcia, X. Sanchez-Vila, P. Trinchero, C. Valhondo, D. M. Tartakovsky, "Probabilistic risk analysis of groundwater remediation strategies," *Water Resour. Res.*, 45, W06413, doi:10.1029/2008WR007551 (2009).