A Divide and Conquer Approach to Cope with Uncertainty, Human Health Risk and Decision Making in Contaminant Hydrology

Felipe P. J. de Barros[1], Diogo Bolster[2], Xavier Sanchez-Vila[3] and Wolfgang Nowak[4]

F.P.J. de Barros, Institute of Applied Analysis and Numerical Simulation/SimTech, Pfaffenwaldring 57, University of Stuttgart, 70569 Stuttgart, Germany. (felipe.debarros@simtech.uni-stuttgart.de)

[1]Institute of Applied Analysis and

Abstract. Assessing health risk in hydrological systems is an interdisciplinary field. It relies on the expertise in fields of hydrology and public health and needs powerful translation concepts to provide decision support and policy making. Reliable health risk estimates need to account for the uncertainties and variabilities present in hydrological, physiological and human behavioral parameters. Despite significant theoretical advancements in stochastic hydrology, there is still a dire need to further propagate these concepts to practical problems and to society in general. Following a recent line of work,

Numerical Simulation/SimTech,
Pfaffenwaldring 57, University of Stuttgart,
70569 Stuttgart, Germany

[2]Environmental Fluid Dynamics
Laboratories, Dept. of Civil Engineering
and Geological Sciences, University of Notre
Dame, IN 46556, USA.

[3]Dept. of Geotechnical Engineering and
Geosciences, Technical University of
Catalonia, E-08034 Barcelona, Spain.

[4]Institute of Hydraulic Engineering,
LH^2/SimTech, University of Stuttgart,
Pfaffenwaldring 61, 70569 Stuttgart,
Germany.

we use of fault trees to address the task of probabilistic risk analysis (PRA) and to support related decision and management problems. Fault trees allow to decompose the assessment of health risk into individual manageable modules, thus tackling a complex system by a structural divide and conquer approach. The complexity within inside each module can be chosen individually according to data availability, parsimony, relative importance and stage of analysis. Three differences are highlighted in the current paper when compared to previous works: (1) The fault tree proposed here accounts for the uncertainty in both hydrological and health components, (2) system failure within the fault tree is defined in terms of risk being above a threshold value whereas previous studies that used fault trees used auxiliary events such as exceedance of critical concentration levels and (3) we introduce a new form of stochastic fault tree that allows to weaken the assumption of independent subsystems that is required by a classical fault tree approach. We illustrate our concept in a simple groundwater-related setting.

1. Introduction

Assessing the impact of water pollutants on human health relies on our ability to accurately assess two things: first, the transport and possible reactions between contaminants in a hydrosystem and second, evaluating the physiological response of humans to such contaminants and the resulting adverse effects on human health [e.g., *Andricevic and Cvetkovic*, 1996; *Maxwell et al.*, 1998; *Maxwell and Kastenberg*, 1999; *Maxwell et al.*, 1999; *Benekos et al.*, 2007; *de Barros and Rubin*, 2008; *Maxwell et al.*, 2008]. Notoriously, both of these fields contain uncertainty for a variety of reasons. These include the lack of characterization data, inadequate conceptual models and the occurrence of natural variability in both hydrosystems and health components [*Bogen and Spear*, 1987; *McKone and Bogen*, 1991; *Burmaster and Wilson*, 1996; *Maxwell and Kastenberg*, 1999]. Given such uncertainties, following the traditional route of making single deterministic predictions for a given scenario has little practical purpose [*USEPA*, 2001]. This fact has been recognized in recent times by many large-scale government regulatory bodies. As a consequence, they increasingly insist on the use of probabilistic approaches that include estimates in uncertainty of risk [e.g., *Rubin et al.*, 1994; *Andricevic and Cvetkovic*, 1996; *Davison et al.*, 2005; *Persson and Destouni*, 2009].

In an ideal world with extensive computational resources, one might try to tackle such water-related health impact problems in a probabilistic framework by running high-resolution Monte-Carlo simulations of the entire interacting system at full complexity. However, the multi-component (and multi-scale) nature of these problems can often render such an approach difficult (if not impossible) to implement in practice. On the hy-

drological side of the problem, heterogeneity in many physical and chemical parameters can range over multiple orders of magnitude and lead to scale-dependence of process descriptions. Depending on the specific problem at hand and the contaminants in question, the number of required parameters can be very large, far beyond parsimony, with limited spatial resolution of the hydrosystem [*Rubin*, 2003; *Tartakovsky and Winter*, 2008]. Similarly, on the health side, natural variability in human behavior, age, body type and genetic characteristics (to mention but a few) lead to large variability in physiological parameters [e.g., *Maxwell and Kastenberg*, 1999].

Apart from the unresolved issues with natural variability that occur in both parts of the system, it is not even entirely clear that the conceptual mathematical models used in each field are fully appropriate. For example, in hydrogeology, there is an ever increasing number of field, laboratory and numerical data sets, indicating that "anomalous"behavior (i.e. non-Fickian phenomena that cannot be described by the traditional advection dispersion equation approaches) may in fact not be all that anomalous but rather the rule [e.g., *Gelhar et al.*, 1992; *Sidle et al.*, 1998; *Silliman et al.*, 1997; *Levy and Berkowitz*, 2003; *Fiori et al.*, 2007]. Such anomalies, observed in conservative transport, pose even further complications for the transport of reactive solutes [*Raje and Kapoor*, 2000; *Gramling et al.*, 2002]. However, there is a continuous emergence of new models that appear capable of capturing these effects [e.g., *Neuman and Tartakovsky*, 2009; *Benson and Meerschaert*, 2008; *Donado et al.*, 2009; *Bolster et al.*, 2010; *Edery et al.*, 2009]. On the health side, many of the mathematical dose-response models rely on linear extrapolation of data from high-dose laboratory experiments on animals [*Fjeld et al.*, 2007], which do not take into account the possibility of nonlinear behavior at lower doses [*Bogen and Spear*, 1987;

McKone and Bogen, 1991; *Burmaster and Wilson*, 1996]. In response to these limitations and uncertainties on both sides of the problem, a recent series of papers has emerged that quantified the relative gain in overall information from enhanced characterization in each component in probabilistic health risk assessment [*de Barros and Rubin*, 2008; *de Barros et al.*, 2009].

As with many applied sciences and engineering disciplines, the correct implementation of assessing health-related risk in hydrosystems is an interdisciplinary field. It relies on the expertise of hydrologists and physiologists/toxicologists as well as a potentially large number of other disciplines, depending on the specific problem being considered. Additionally, in practical situations, stakeholders (e.g. managers, politicians, judges etc.), who are given the responsibility of making decisions within such complex systems, are typically only experts in one field at most. As a result, there is a strong need to communicate information across interfaces between different fields in an efficient and comprehensible manner, which is rarely an easy task [*McLucas*, 2003]. For example, despite significant theoretical advancements in stochastic hydrogeology over the last several decades, stronger efforts are still needed to transfer this knowledge to applications [see discussions in *Rubin*, 2003; *Christakos*, 2004; *Freeze*, 2004; *Rubin*, 2004; *Pappenberger and Beven*, 2006].

2. Goals, Approach and Contribution

In this work, we propose a formal probabilistic risk analysis (PRA) that relies on the use of fault trees and can address all of the above mentioned issues. Fault trees have commonly been used in risk assessment concerning engineered systems [e.g., *Bedford and Cooke*, 2003]. However, for a variety of reasons, e.g., because hydrosystems are comprised of a mixture of natural and engineered components that complicates matters, this approach

has been receiving increasing attention in the hydrological community [*Tartakovsky*, 2007; *Winter and Tartakovsky*, 2008; *Bolster et al.*, 2009]. The basic idea of this methodology is simple and can be summarized as a divide and conquer approach: It consists of taking a large and complex system that is difficult to handle and dividing it into a series of quasi-independent simpler systems (modules) that are manageable on an individual scale. Once each of the smaller problems has been addressed, they can be recombined in a systematic manner to look at the large system. According to *Bedford and Cooke* [2003], a rigorous PRA based on fault tree should consist of the following steps:

1. Define failure of the system to be examined, where system failure must be defined a priori by stakeholders.

2. Identify the key components of the system and all events that must occur for the system to fail.

3. Construct a fault tree that visually depicts the combination of these events.

4. Develop a mathematical representation of the fault tree by the use of Boolean algebra.

5. Compute the probabilities of occurrence of each event.

6. Use these to calculate the probability of failure for the entire system.

The advantages of the divide and conquer approach is that, for a well developed fault tree, each key component/event should be quasi-independent from all others (i.e., if there is a dependence it should be weak). Therefore, each event can be tackled without explicit knowledge of all others. For example, in *Bolster et al.* [2009], each of the events was studied by a different person without mutual interaction. This opens the gateway for

interdisciplinary cooperation, as each component can be addressed independently by the most appropriate expert.

Additionally, a decision maker can use the fault tree to visually understand where risk and uncertainty emerge from in this system, without having to enter into the complex details of each component. In some sense, the fault tree acts as a translator of information between experts in different fields, thus enabling better decision making.

Another benefit of such a fault tree approach is that one can work with each individual component: For instance, one can replace the method of examining each component without having to touch the others. This can be thought of as analogous to the object-oriented approach to programming, where one only updates the necessary objects as the demand arises without having to rewrite an entire code. This enables better allocation of resources and incorporation of more advanced theories and data sets as they become available. For example, as a starting point one can use simple calculations to study each component. With such an initial estimate, one can identify the events which contribute most to the final risk or those which propagate the highest degree of uncertainty through the system. This information can be used to allocate further resources to these dominant events, and more sophisticated and detailed models can be pursued for these events as new data or advanced theoretical models become available. Moreover, it can be use towards rational allocation of resources for further data acquisition [*de Barros et al.*, 2009; *Nowak et al.*, 2010] within a dynamical and adaptive framework. Thus, fault trees can structure and guide through the entire process of PRA, from initial screening over additional investigations and refinement to the final conception of risk management strategies.

The purpose of this work is to extend the fault tree framework used by *Bolster et al.* [2009] to account for both hydrology and human physiological/behavioral characteristics. We develop this idea by unifying the framework provided by *Bolster et al.* [2009] with the ideas of *Maxwell and Kastenberg* [1999], *Maxwell et al.* [1999] and *de Barros and Rubin* [2008]. *Bolster et al.* [2009] defined system failure by exceeding a regulatory threshold concentration. In contrast, we define the ultimate prediction goal (i.e., human health risk) to be the center of attention, and define system failure as exceeding a threshold risk value [as done in *Maxwell and Kastenberg*, 1999]. Such threshold risk is often given by environmental regulation bodies for the sensitive target at stake [e.g., *USEPA*, 2001]. The novelty here lies in constructing a fault tree analysis that includes the uncertainty and variability from both hydrological and human health risk parameters. One of the new key features of this choice is that it allows us to investigate the role of health-related variability and uncertainty in decision making. For instance, if the concentration value at a drinking water supply is higher than that allowed, but if the characteristics of exposed individuals are such that little of that contamination is ingested or metabolized, then individuals might be at little or no risk.

We begin by defining the problem formulation and presenting a generic methodology for developing fault trees in hydrological systems. More precisely, we propose a stochastic fault tree method. To elucidate this process and demonstrate its strengths, we present a specific illustrative example. We consider a simple groundwater contamination scenario, illustrating how system failure and related uncertainty therein changes (i) according to the physical characteristics of the flow and transport problem and (ii) for different levels of uncertainty and variability in the health component.

3. Problem Formulation

3.1. The Co-Existence of Water-Related Health Hazards

Surface or groundwater can be polluted by the presence of many different chemicals (either organic or inorganic) as well as pathogenic microorganisms (bacteria, protozoa, and viruses) [e.g, *Molin and Cvetkovic*, 2010]. Exposure of humans to polluted water through ingestion, inhalation, or skin contact may produce a number of different diseases. Whether one of these potential diseases is developed in a given individual depends on the toxicity of the pollutant, but also on the metabolism, personal habits of an individual's water-related practices and finally, consumption and exposure habits.

Diseases can be either caused by accumulation over the years or by acute exposure, i.e., over a very short period of time. Synergetic effects may cause the same pollutant to have different toxicity in different parts of the world; e.g., lung cancer may be caused by drinking water with a high concentration of trihalomethanes, but it is also likely to be developed in people living in areas with heavy atmospheric pollution.

Obviously, for a given hazardous substance when either concentration or time of exposure increases, so does the potential (risk) of developing a disease. Actual existing models are highly disputable, since most of them are extrapolations from high-dose laboratory experiments carried out on animals such as mice to low-dose effects on humans [e.g., *McKone and Bogen*, 1991]. We denote $r_i(x,t), i = 1...N$, as the risk associated to the development of a given adverse health effect for a given pollutant, N being the number of chemicals released. In general, r_i are supposed to be small values (otherwise the problem is considered pandemic). Thus, the potential development of two or more diseases at exactly the same time can be considered negligible, and total risk can be taken as the sum

of the individual risks:

$$r(x,t) = \sum_{i=1}^{N} r_i(x,t). \tag{1}$$

3.2. Evaluating Health Risk for a Particular Substance and Exposure Pathway

The starting point for this section is to formulate human health risk for a single substance i in probabilistic terms following *de Barros and Rubin* [2008]. Depending on the particular contaminant, there are a number of models to evaluate the risk [*Maxwell and Kastenberg*, 1999; *Morales et al.*, 2000; *Fjeld et al.*, 2007; *de Barros et al.*, 2009; *Molin et al.*, 2010].

In order to simplify the discussion, let us consider a carcinogenic contaminant. The increased lifetime cancer risk r due to the groundwater ingestion pathway (chronic exposure) for an individual is expressed by an assumed linear model as [*USEPA*, 1989]:

$$r(\mathbf{x},\mathbf{t}) = \beta C(\mathbf{x},\mathbf{t}), \tag{2}$$

where concentration $C(\mathbf{x},t)$ [mg/l] is an outcome of all the relevant flow, transport and transformation processess in the system at hand. β is a lumped parameter that accounts for all the behavioral and physiological parameters:

$$\beta = \frac{IR \times ED \times EF}{BW \times AT} \times Sf_o, \tag{3}$$

where IR [l/d] denotes the ingestion rate, ED [y] represents exposure duration, EF [d/y] is the exposure frequency, BW [kg] is the body weight, AT [d] is the average time, and Sf_o is the slope factor [kg-d/mg], obtained from experiments. Note that C can represent a point concentration or a flux-averaged concentration. In most health risk applications, C corresponds either to the peak concentration or to an averaged concentration over the

exposure period at a environmentally sensitive target [see *Maxwell and Kastenberg*, 1999]. All the health parameters are values corresponding to an individual from the exposed population. These parameters contain some level of uncertainty and vary from individual to individual [*Dawoud and Purucker*, 1996]. A large degree of uncertainty is present in Sf_o because of the animal to human extrapolation [*McKone and Bogen*, 1991]. Expression (2) is merely a simplification of a more general model that includes several exposure pathways, contaminant dependencies and non-linearities [*Maxwell and Kastenberg*, 1999; *Morales et al.*, 2000; *Fjeld et al.*, 2007; *de Barros et al.*, 2009]:

$$r(\mathbf{x},\mathbf{t}) = \beta_G[C(\mathbf{x},\mathbf{t}) - C_G^*]^{m_G} + \beta_H[C(\mathbf{x},\mathbf{t}) - C_H^*]^{m_H} + \beta_S[C(\mathbf{x},\mathbf{t}) - C_S^*]^{m_S}, \qquad (4)$$

where the subscripts G stands for ingestion, H for inhalation, and S for contact through skin and β_j are coefficients that relate to the toxicities of the substance for each pathway. C_j^* is the corresponding threshold value, i.e., a value below which we do not expect any adverse effects for an individual. These threshold values are pollutant dependent. The exponents m_G, m_H and m_S determine the non-linearity of each dose-response curve [*Fjeld et al.*, 2007]. In the case of carcinogenic compounds, USEPA suggests to use a zero value. This indicates that, no matter how small the concentration is in water, risk is never null [*USEPA*, 1989, 1991]. An alternative is using the detection limit given by the chemical analytical method. Equation (4) can only be used if each individual term C_j is above C_j^*; otherwise the individual term should be removed from the equation.

For the sake of discussion but without loss of generality, we will work with the model expressed in equation (2) to demonstrate the modular character of the methodology proposed. It will serve to illustrate the purpose and exchange character of the suggested

methodology. Still, at any time, more complex risk models, such as Eq. (4) can be incorporated. The only prerequisite is that sufficient data should be available to justify any more complex choice (see works by *Troldborg et al.* [2008, 2009]) .

3.3. Stochastic Representation of Human Health Risk

According to Eq. (2), risk is the product of two quantities, both of them uncertain. Uncertainty in β comes from the imperfect characterization (and lack of knowledge) of the toxicity. However, β is also variable since its value varies from individual to individual within the exposed population. Values of β also vary according to the population cohort such as age groups and gender [*Yu et al.*, 2003; *Maddalena and McKone*, 2002]. *Maxwell et al.* [1998] and *Maxwell and Kastenberg* [1999] reported that the impact of the variability in β on risk is very significant.

The remaining issue is to evaluate the contaminant concentration at any particular point within an environmentally sensitive target (Ω_p) over a period of time t_p, $C(\mathbf{x} \in \Omega_p, t_p)$ and to quantify its uncertainty. Spatial variability and uncertainty in concentration is due to the ubiquituous heterogeneity in physical and biochemical processes, boundary conditions and contaminant release patterns. The processes involved are solute- and soil-dependent, and might include advection, diffusion, dispersion, sorption, precipitation/dissolution, redox processes, cation exchange, evaporation/condensation, microbial or chemical transformation and decay. For any given substance, an appropriate model is written as a governing equation that depends on a number of parameters. In most applications, there is a need to be careful with the problem of scales, since both the relevant processes as well as the representative parameters are often scale dependent.

Accepting that $C(\mathbf{x} \in \Omega_p, t_p)$ and β are uncertain, the resulting risk r is regarded as a random function R, with a cumulative distribution function (cdf) $F_R(r)=Pr(R \leq r)$. Thus, it is convenient to formulate risk in terms of exceeding probabilities [e.g., *Andricevic and Cvetkovic*, 1996; *de Barros and Rubin*, 2008]:

$$Pr(R > r_{crit}) = 1 - F_R(r_{crit}), \tag{5}$$

with r_{crit} representing an environmentally regulated value, for instance, $r_{crit} = 10^{-6}$ or 10^{-4} [*USEPA*, 1989].

Uncertainty in the concentration can be reduced by conditioning on measurements of either the dependent variables (e.g. concentrations, groundwater heads, river discharges, etc) or the parameters themselves (through field or laboratory tests). Details concerning different mentalities on uncertainty reduction through conditioning can be found in the literature [e.g., *Rubin*, 1991; *Kitanidis*, 1995; *Bellin and Rubin*, 1996; *Freer et al.*, 1996; *Zimmerman et al.*, 1998]. Once it is decided which components to investigate in more detail, specific methods for optimal experimental design can be used, e.g., for optimal sampling layouts [e.g., *Ucinski*, 2005; *Nowak et al.*, 2010; *Nowak*, 2010].

4. Methodology: Fault-Tree Analysis

Before one can begin any fault tree analysis, one must define the system that is being investigated. The system that we consider in this work is depicted schematically in Figure 1. This figure illustrates several sources of contamination (SO_i), a general mean flow field and a region that we define as the protection region (Ω_p). The sources of contamination could be anything from natural sources (e.g. arsenic), known spill sites, industrial regions where contamination of certain pollutants may be probable, or agriculatural lands where

certain contaminants may occur, to any other imaginable source of contamination. Similarly the protection zone could be anything like a well field, a lake or a residential area. The system defined in this work is deliberately kept generic and would of course be made more specific to a particular problem under consideration as the demand arises. Based on this generic system, we will follow the six steps outlined in the introduction. We will present a more specific illustrative example in the following section.

Step 1: Defining System Failure

We define failure of this system (SF) when risk, as defined in section 3, exceeds a critical regulatory value:

$$SF: \; r > r_{crit} \tag{6}$$

with exceedance probability given by Eq (5).

Step 2: Identifying the Key Components/Events

We use this particular step to divide the problem into two components: A hydrological contamination scenario and the consequences of contamination on human health risk. This is an important distinction because concentration exceedance does not imply that the population is at risk. For example, individuals not drinking tap water (or with exceptional physiology) might be at little or no risk. For such reasons, the combination of the concentration and the health parameters is the important factor to consider (only the joint effect can culminate in adverse health effects).

The first key component follows a similar path to the works of *Tartakovsky* [2007], *Winter and Tartakovsky* [2008] and *Bolster et al.* [2009]. It focuses on the hydrological component and is meant to establish whether it is necessary at all to consider health risk. This event is called "Critical Concentration of Exposure "(CCE_i) and is defined as the event that the concentration of a contaminant i, arriving at the protection zone, exceeds some critical concentration value. If such an event occurs, decision makers must be alerted and should become concerned about the consequences on human health. The lower-level events associated with this key event are:

$\mathbf{SO}_i$ (Source Occurrence) is the event that a contaminant exists. In many possible scenarios, the existence of a contaminant source is not deterministic. For instance, a contamination source provenient from fertilizers or pesticides within an agricultural zone may (or may not) exist and the probability of its occurrence must be quantified.

$\mathbf{P_{1,i}}$ (Plume Path 1) is the event that the plume evolving from contaminant source i bypasses the protection zone.

$\mathbf{P_{2,i}}$ (Plume Path 2) refers to the event that the same plume hits the protection zone. If such a path does not exist, due to the morphology of the hydrosystem, then there is no reason for concern

$\mathbf{NA}_i$ (Natural Attenuation) represents the event that natural attenuation can decrease concentration peaks below a defined threshold value through chemical reactions, dispersion and dilution.

The second component relates to all health risk considerations. For this component, the basic events are:

CCE$_i$ (Critical Concentration of Exposure) reflects the concentration that, when combined to a value β (see the relation in Eq. 2), will result in risk exceeding its critical value established by regulations (e.g., $r_{crit} = 10^{-6}$ or 10^{-4}), i.e., system failure will occur.

BPC$_i$ (Behavioral Physiological Component) corresponds to the event that an individual (or population cohort) that is exposed has characteristic β (see Eq. 3).

The point to note here is that CCE_i is conditioned on a value of β provenient from BPC_i, which, as highlighted in section 3, is not a single value and it varies within the population based on several parameters [e.g., *Maxwell et al.*, 1998; *Maxwell and Kastenberg*, 1999; *Maxwell et al.*, 1999; *de Barros and Rubin*, 2008; *de Barros et al.*, 2009]. As expressed in Equation (4), each individual contains a specific β (e.g., j^{th} individual with characteristic β_j). The fact that CCE_i can be defined only for a given value of β will require, in a later stage of our analysis, an extension of the conventional fault tree approach to account for all possible values from the distribution of β.

Step 3: Building the Fault Tree(s)

In step 2, we divided the problem into two sections. In this step we will draw a fault tree for each of those sections. The first branch of the fault tree addresses the hydrogeological contamination scenario, leading to the key event CCE_i. The fault tree is shown in Figure 2 and is, in some sense, a version of the fault tree discussed in *Bolster et al.* [2009]. The combination with the second branch yields the main fault tree and represents the novelty

of this work. This main fault tree replicates for each contaminant species and source and is shown in Figure 3. It illustrates visually how we have linked contamination and human health risk. The system failure (risk exceedance) for contaminant i is the joint occurrence of the events CCE_i and BPC_i.

Those readers who are familiar with fault trees might notice a particular gate (Boolean operator) below the R_{crit} event they are not familiar with. This gate is novel and we define it as an "ENSEMBLE AND"gate. It reflects the fact that the R_{crit} event must be calculated based on all possible values of β and of the concentration arriving at the protection zone. The ensemble operator $\langle ... \rangle_\beta$ indicates that the averaging should be done over the ensemble of β to obtain the risk over the average individual because $Pr[R] = \langle\ Pr[R|\beta]\ \rangle_\beta$. The fault tree without the operator $\langle ... \rangle_\beta$ would be equivalent to a tree for a single exposed individual with known characteristics and with known toxicity. In other words, the fault tree shown here is generalized for every individual of the exposed population. The fault tree depicted in Figure 3 allows us to evaluate system failure for an average individual over a specified population cohort (e.g., average senior with β specified over a range of possible values), average individual over all the exposed population (averaging over all β range) or for a single exposed individual (with specified β_j). This process represents the internal loop from the nested Monte Carlo approach proposed by *Maxwell and Kastenberg* [1999]. *Maxwell et al.* [1998] and *Maxwell and Kastenberg* [1999] showed how the variabilities within health parameters have a strong impact in human health risk prediction. As with all fault tree analysis, it is meant to act as a visual aid to the user to understand where risk can come from. Accounting for $\langle ... \rangle_\beta$ within the fault tree implies that one needs to account for the variability in its description such that one can assign the

probability of occurrence for the event R_{crit}. The "ENSEMBLE AND"gate generalizes the previous fault tree by covering over all range of population behavioral characteristics.

Step 4: Translation to Boolean Logic

This part can be viewed as the final stage in the development of the risk assessment system. The subsequent steps (items 5 and 6 in the introduction) involve the actual calculations of probabilities of all basic events and the combination thereof based on the expression that emerges from the current step. First, we will write a Boolean logic expression for the probability of event CCE_i occurring. The "AND"and "OR"operators represents multiplications and additions of probabilities respectively. As discussed (and as can be seen from the fault tree in Figure 2), the appropriate Boolean expression for failure CCE_i is given by

$$CCE_i = SO_i \,\text{AND}\, P_{2,i} \,\text{AND}\, NA_i, \tag{7}$$

with probability of occurrence:

$$Pr[CCE_i] = Pr[SO_i]\, Pr[P_{2,i}]\, Pr[NA_i], \tag{8}$$

since SO_i, $P_{2,i}$ and NA_i are completely independent of each other. Similarly, for the main fault tree depicted in Figure 3, the Boolean expression for system failure associated with each source $R_{crit,i}$ can be written as

$$R_{crit,i} = CCE_i \,\text{AND}\, BPC \tag{9}$$

with probability of occurrence:

$$Pr[R_{crit,i}] = Pr[CCE_i]\, Pr[BPC]. \tag{10}$$

If more contaminants are present ($i \geq 2$), then the total system failure (SF_{all}) is given by:

$$SF_{all} = R_{crit,1}\,\mathrm{OR}\,R_{crit,2}\,\mathrm{OR}...\mathrm{OR}\,R_{crit,N} \tag{11}$$

and the probability of system failure is given by

$$Pr[SF_{all}] = Pr[SF_1] + Pr[SF_2] + ... + Pr[SF_N]. \tag{12}$$

Steps 5 and 6, see Section 2, are straightfoward and need no further explanation. In the following section, we will develop them for an illustrative example.

5. Illustrative Example

Our goal is to show how the methodology can accommodate the entire range from simple to complex problems and solution approaches. It is seldom that a large data set is available in probabilistic health risk assessment, and we cannot always solve the problem entirely. For such reasons, it is common to make conservative assumptions and assess the worst case scenario with simple models [*Troldborg et al.*, 2009; *Bolster et al.*, 2009]. The scenario under consideration assumes almost complete absence of site-specific data, leading to crude yet conservative estimates of probabilities. Other existing methods rather than the simple one we selected for the illustration here can be found in the literature, (e.g. see *Rubin* [2003] for an extensive review). The level of complexity in the analysis of each component and event can vary according to the available information and the importance within the fault tree, and can easily be adapted interactively during

the analysis. If hydrological field data is available and more complex physical-chemical processes are involved, one may opt for numerical Monte Carlo simulations to allow more flexibility in relaxing simplifying assumptions as done in *Maxwell and Kastenberg* [1999]; *Maxwell et al.* [1999] and *de Barros et al.* [2009]. Without loss of generality, our illustrative example will focus in a groundwater contamination problem. The method shown here can also be applied to surface water bodies or to coupled catchment-scale problems [e.g., *Baresel and Destouni*, 2007; *Persson and Destouni*, 2009].

5.1. Physical Scenario and Assumptions

We consider a regional aquifer, confined, 2D depth-averaged with mean flow velocity U taken along the x-direction. A degrading contaminant is continuously released with inlet concentration C_o within a rectangular source with transverse dimension $w = y_{SR}$ - y_{SL} (see Figure 4). Once contamination has occurred, the contaminant plume might hit the environmentally sensitive target represented by a control plane (CP) situated at a distance $x = L_b$ - L_s from the source zone (see Figure 4). The schematic representation of the physical problem is given in Figure 4.

At this stage, we will evaluate the concentration field under the worst case scenario. This is a common approach in human health risk assessment since decision makers must account for safety factors when dealing with human lives [*Troldborg et al.*, 2008, 2009; *McKnight et al.*, 2010]. We assume, in accordance with the worst case scenario philosophy, that the concentration can be calculated using a 1D solution by neglecting transverse dispersion between neighboring streamlines. Further more, longitudinal dispersion is also neglected. This excludes dilution processes as described by *Kitanidis* [1994]. The only

natural attenuating factor is degradation with linear decay coefficient λ (neglecting pore-scale dispersion):

$$C(\tau) = C_o \exp[-\lambda\tau], \tag{13}$$

where $\tau = x/U$ denotes the travel time between source and control plane. In the subsequent sections, we will account for the uncertainty in τ in order to derive a simple expression for the concentration probability density function (pdf) and λ is known.

5.2. Quantifying Probabilities of Occurrence

5.2.1. Probability of Travel Paths

Here we compute the probabilities of path 1 or 2 of occurring, i.e. events P_1 and P_2 (see Section 4 for definitions). We prefer to calculate the probability of the plume bypassing the control plane ($Pr[P_1]$). Since $Pr[P_1]$ and $Pr[P_2]$ are mutually exclusive, we have:

$$Pr[P_2] = 1 - Pr[P_1]. \tag{14}$$

In order to compute the above probabilities, we must quantify the pdf of each contaminant particle within the source zone intercepting the control plane of the protection zone. Neglecting pore-scale dispersion (both longitudinal and transverse), we approximate the time of interception t_b by the mean travel time $t_b \approx \frac{L_b - L_s}{U}$ (time from the source to the control plane). In analogy to the work presented in *Bolster et al.* [2009], we assume a Gaussian model to describe the particle displacement pdf. For alternative definitions of the displacement pdf, we refer to *Dagan* [1987]; *Rubin* [2003]. Our resulting pdf is given by:

$$p_1(L_b, t_b) = \frac{1}{\sqrt{4\pi D_{\text{eff}} t_b}} e^{-\frac{(y_b - y_o)^2}{4 D_{eff} t_b}}, \tag{15}$$

where y_o is a point within the contamination zone. D_{eff} is an effective macroscopic dispersion coefficient (purely uncertainty-related) that can arise for a variety of reasons, e.g., due to heterogeneity [*Dagan*, 1989; *Rubin*, 2003] or due to temporal fluctuations in the flow field [*Dentz and Carrera*, 2005] to mention but a few. Accounting for a dispersive term in Eq. (15), but not in Eq. (13), might seem inconsistent at first sight, but it is a consistent set of worst-case assumptions.

If no particles from the source bypasses the control plane either on the left or on the right, then no interception occurs. A conservative envelope can be constructed by considering that particles from the back right point [$\mathbf{s_R} = (L_s, y_{sR})$], see Figure 4, have to pass by the outer left point of the protection zone [$\mathbf{b_L} = (L_b, y_{bL})$] and vice-versa.

$$\begin{aligned} Pr(P_1) &= Pr(P_{1,L}) + Pr(P_{1,R}) \\ &= \int_{-\infty}^{y_{bL}} \frac{1}{\sqrt{4\pi D_{eff} t_b}} e^{-\frac{(y_b - y_{sR})^2}{4 D_{eff} t_b}} dy_b \\ &\quad + \int_{y_{bR}y}^{\infty} \frac{1}{\sqrt{4\pi D_{eff} t_b}} e^{-\frac{(y_b - y_{sL})^2}{4 D_{eff} t_b}} dy_b. \end{aligned} \tag{16}$$

5.2.2. Probability of Natural Attenuation

Above, we used the back end of the source as worst case scenario for interception with the protection zone. The worst case scenario for natural attenuation is based on the front center of the source area because this yields the shortest distance (thus shortest time) for decay.

Given the uncertainty in flow parameters and scarce site characterization, we consider for illustration the travel time τ to be stochastic and lognormally distributed [e.g., *Cvetkovic et al.*, 1992]:

$$f_\tau(\tau) = \frac{e^{-\frac{(\log(\tau)-\mu_\tau)^2}{2\sigma_\tau^2}}}{\sqrt{2\pi}\sigma_\tau\tau}, \tag{17}$$

with μ_τ and σ_τ denoting the travel time mean and variance in logarithmic space and are related to the mean velocity [e.g., *Andricevic et al.*, 1994].

We can now calculate the pdf for concentration based on the travel time pdf and Eq. (13):

$$f_c(C) = \left|\frac{d\tau}{dC}\right| f_\tau(\tau), \tag{18}$$

which allows us to evaluate the probability of the concentration being above a regulatory threshold value C_{crit} at the sensitive target. Substituting Eq. (13) into Eq. (18), we obtain:

$$f_c(C) = \frac{1}{\lambda C} f_\tau\left(\frac{1}{\lambda}\ln\left[\frac{C}{C_o}\right]\right), \tag{19}$$

Eq. (18) reflects only one possible and simple choice of model for the concentration pdf under the conditions assumed in the current work for illustrative purposes. It is worth mentioning that many other models could be used in this approach under more generic conditions [e.g., *Rubin et al.*, 1994; *Bellin and Tonina*, 2007; *Cirpka et al.*, 2008]. For example, other choices for travel time distributions can be found in Ch. 10 of *Rubin* [2003] and in *Sanchez-Vila and Guadagnini* [2005]. If hydrogeological data is available, one could also follow the approach described in *Rubin and Dagan* [1992] to condition the travel time pdf.

5.2.3. Probability of Risk Exceedence

Based on Eq. (5), we can evaluate the probability that the risk will exceed a threshold value r_{crit}. Here, we present a risk distribution for the commonly used risk model given in Eq. (2). In order to evaluate the risk cdf (F_R) based on the pdf f_β of the health parameters and concentration pdf f_C we have:

$$F_R(r_{crit}) = \int_0^{r_{crit}} \int_0^{\infty} f_\beta(\beta) f_C\left(\frac{r}{\beta}\right)\frac{1}{\beta} dr d\beta \tag{20}$$

where we used statistical independence between β and C. The concentration pdf comes from Eq. (18) while f_β is determined from population studies [e.g., *Dawoud and Purucker*, 1996] or the data provided in *Maxwell et al.* [1998] and *Benekos et al.* [2007]. If a single individual with characteristics β_o is exposed, then Eq. (20) becomes:

$$\begin{aligned} F_R(r_{crit}) &= \int_0^{r_{crit}} \int_0^{\infty} \delta(\beta - \beta_o) f_C\left(\frac{r}{\beta}\right)\frac{1}{\beta} dr d\beta \\ F_R(r_{crit}) &= \frac{1}{\beta_o}\int_0^{r_{crit}} f_C\left(\frac{r}{\beta_o}\right) dr, \end{aligned} \tag{21}$$

where we used the properties of the Dirac Delta δ: $f_\beta(\beta) = \delta(\beta - \beta_o)$. This feature is incorporated in the fault tree represented in Figure 3 and illustrates how the approach can be used to cover cases for a single exposed individual and for a fully exposed population (also different population cohorts: gender and/or age dependent).

6. Results and Discussion

We illustrate the methodology by considering a simple example for cancer risk. Two species (A and B) are continuously released from their source locations and may pose a threat to human lives. The two contaminants are released in different locations, with different source dimensions and initial concentrations (to reproduce the varying range of

typical situations found in the field). Both contaminants are released from line sources with dimensions 4 m (for contaminant A) and 2 m (for contaminant B). Contaminant A is closer to the protection zone (35 m) while contaminant B is further away (60 m). These values as well as other relevant parameters are summarized in Table 1. The main sources of uncertainty under consideration here are the contaminant travel times, Eq. (17). We also account for the variability in the health-related parameter β, Eq. (3). For the current scenario, we assume that travel time standard deviation is equal to $\sigma_\tau = 0.1$ d and that $D_{\text{eff}} = 0.1$ m^2/d.

Since we have two distinct contaminants, the values for β are different. For instance, contaminant A affects a specific population cohort while contaminant B affects a different one (thus reflecting variability). In this example, we assume both values of β to be lognormally distributed with mean $\mu_{ln\beta}$ and standard deviation $\sigma_{ln\beta}$, see Table 1 (values given in logarithmic space). Figure 5 illustrates the pdf of β for contaminants A and B. Risk estimates were obtained using the linear model in Eq. (2) and their corresponding probabilities of exceeding a regulatory value are computed through the cdf provided in Eq. (20).

Given that contamination is known to exist (SO with probability 1), we need to evaluate the probabilities associated with each branch of the fault tree using the steps described in Section 4. The events and their corresponding probabilities are summarized in Table 2 for both contaminants.

With the data given in Table 1 and using Eq. (16), the probability of the plume hitting the sensitive target is $Pr[P_2] = 0.38$ for contaminant A and $Pr[P_2]= 0.26$ for contaminant B. From the results given in Figure 6, we can also extract the probabilities

of the concentration being above a regulatory threshold value C_{crit}. The probabilities of $C \geq C_{crit}$ for contaminant A is 0.18 where for contaminant B we have 0.015. This is caused by the physical setup of the problem, since the source for contaminant A is closer to the environmentally sensitive target than to the release location of contaminant B. This shows how the extension of the contaminant source as well as its distance from the protection zone influences the probabilities of the plume hitting the target and of the concentration exceedance.

Figure 7 depicts the risk cdfs for both contaminants. Assuming that the critical risk value established by the regulatory agency is $r_{crit} = 10^{-4}$, we can compute the risk exceedance probabilities $Pr(R > r_{crit})$ using Eq. (5), and obtain 0.69 and 0.54 for species A and B respectively. With Eq. (10), the probability of system failure can be obtained (values given in Table 2).

Next, we illustrate a sensitivity analysis to identify which parameters are more relevant in predicting the system failure for contaminants A and B. In addition, it serves as a first screening tool to see which parameters are dominant in each of the branches of the fault tree and may require more detailed investigation. The parameters chosen to perform the sensitivity analysis are $\boldsymbol{\theta} = \{U,\ D_{\text{eff}},\ \lambda,\ \sigma,\ \mu_{\ln\beta},\ \sigma_{\ln\beta}\ \}$. We perturb, one by one, each parameter within $\boldsymbol{\theta}$ by 10 percent and re-evaluate the probability of system failure each time. The resulting differences (between the perturbed and unperturbed case) given by $\Delta Pr[SF]$ are depicted in Figures 8 and 9 for contaminants A and B, respectively.

One striking difference between Figures 8 and 9 is the sensitivity of system failure to the health-related parameters: Contaminant A is more sensitive to the health-related parameters than contaminant B. This result aligns well with the results by *de Barros*

and Rubin [2008]. They showed that the relative significance of health-related parameters decreases with travel distance, because of the uncertainty in transverse plume position increases [*Rubin*, 1991]. Moreover, we note that both contaminants respond differently to all other parameters, with the exception of the mean velocity.

For contaminant A, the macroscopic effective dispersion parameter (D_{eff}) is less important, see Figure 8, since the source area for contaminant A is close to the environmental target. Over short travel distances, the macroscopic effective dispersion has a small probability to make the plume bypass the protection zone (event P_2). Vice-versa, D_{eff} has a larger significance in the probability of system failure for contaminant B, because its source is located farther away from the target (event P_2).

The decay coefficient, λ, is the second most important parameter relative to the others for contaminant A. Since the source for pollutant A is so close to the protection zone, decay is the only process that can significantly reduce the probability of system failure. The opposite occurs for contaminant B, since the significance of other events is higher.

Figure 10 shows how the coefficients of variation of the statistical distribution of risk changes for each perturbation in $\boldsymbol{\theta}$. This quantifies how sensitive the uncertainty is in assessing health risk to each individual parameter. In the current simple example, λ and U have stronger effects on the uncertainty of risk for both species A and B than all other parameters. We also observe that the mean and standard deviation of the health-related component ($\mu_{\ln\beta}$ and $\sigma_{\ln\beta}$) has a significant contribution in the final risk pdf. These health parameters have a stronger contribution to the spread of the risk pdf for contaminant A (closer to the source) than for B. For predictions closer to the source, characterization of the health parameters becomes important since concentrations are still high. As the

distance between the contaminant source and receptor increases, the contaminant plume's peak concentration decreases due to the physical processes involved (in our case, decay). Source dimensions and distance to the protection zone have a definite role in defining the significance of the health parameters in the final risk. Again, this agrees with the results from *de Barros and Rubin* [2008].

Although we have used a simple linear dose-response curve to evaluate cancer risk for the illustration, many other alternatives exist with varying levels of uncertainty. For instance, the work of *Yu et al.* [2003] provides detailed epidemiological dose-response curves and parameter uncertainties for arsenic that are age- and gender-dependent. Such dose-response curves are less subject to uncertainty than cancer risk models, because the latter rely on extrapolated animal-to-human data. This implies that, if the a contaminant site has several contaminants, different types of risk models could be used. This would lead to different relative contributions to uncertainty propagation in assessing system failure as discussed in *de Barros et al.* [2009].

An important and attractive feature of the methodology shown is that it allows one to observe, in a most graphical manner, the sensitivity of the probabilities in system failure for each branch of the tree. This is a crucial basis for supporting managing decisions. For example, it indicates how to allocate resources towards further site characterization via prioritization according to highest risk contributions and highest sensitivity.

7. Summary and Conclusions

In this work, we used the fault tree methodology to evaluate human health risk in a probabilistic manner. The approach breaks complex problems into individual events that can be tackled individually. The main differences between the ideas proposed here and the

previous works [*Tartakovsky*, 2007; *Winter and Tartakovsky*, 2008; *Bolster et al.*, 2009] are:

1. The fault tree proposed here accounts for the uncertainty in both hydrogeological and health component;

2. System failure is defined in terms of risk being above a threshold value;

3. We introduced of a new form of stochastic fault-tree that weakens the assumption of independent events which is necessary in conventional fault tree analysis.

Although we used only a crude and simple setting to illustrate the methodology, the approach can be used with arbitrarily more complex models. However, such simple approaches can be useful for performing a preliminary screening in PRA, see works by *Troldborg et al.* [2008, 2009]. For instance, with an initial estimate based on simple models, one can identify the events which contribute most to the final risk estimate or those that propagate the highest degree of uncertainty throughout the system. This information can then be used to invest further resources to these specific events, and more elaborate models can be used if additional data becomes available. The divide & conquer and modularity features of the proposed framework easily allow the methods or tools used in each component to be easily exchanged (and refined) in later analysis without being intrusive in other components.

Moreover, assessing health-related risk in hydrosystems is an interdisciplinary field and it relies on the expertise from a large number of disciplines (for example, hydrologists, engineers, public health, etc). As a result, communicating the information across interfaces between different fields in an efficient and comprehensible manner is needed such that reliable water management decisions are made. The divide and conquer approach

inherent to fault trees allows individual experts to work on the individual problems with clear communication interfaces given by the fault tree structure. The approach allows decision makers to better visualize the components culminating in system failure (e.g., population at risk) as well as the uncertainty emerging from each subsystem. This is appealing from the decision maker's perspective, since it does not require entering into the complex details of each component of the PRA and helps communicate probabilistic concepts to practioners. Furthermore, it acts as a translator to experts from different fields, thus aiding public authorities in policy making and water management.

Despite the fact that our work focused on a groundwater contamination application, it can be also used in other problems such as soil contamination, well vulnerability and surface waters and catchment-scale coupled problems [e.g., *Frind et al.*, 2006; *Baresel and Destouni*, 2007; *Troldborg et al.*, 2008, 2009; *Persson and Destouni*, 2009]. Furthermore, an emerging challenge consists in using the ideas discussed in this paper to tackle a fully integrated hydrosystem (groundwater, soil, surface water, etc.) where the need for dividing a complicated problem into smaller ones as well as interdisciplinary communication are even more evident [*Persson and Destouni*, 2009; *McKnight et al.*, 2010]. For instance, *Bertuzzo et al.* [2008] studied how river networks (acting as environmental corridors) affect the spreading of cholera epidemics. These authors clearly showed how hydrological, health and demographical data needs to be considered in order to capture an accurate description of the main controlling factors dictating the spread of cholera epidemics.

As pointed out in the literature, practitioners are still reluctant to embrace the concepts of uncertainty [*Pappenberger and Beven*, 2006]. Such resistance has also been a matter of discussion in a 2004 Forum published in *Stochastic Environmental Research and Risk*

Assessment [*Christakos*, 2004; *Freeze*, 2004; *Rubin*, 2004]. A common conclusion is that the dialogue between the interdisciplinary groups is of utmost importance. Thus, having a tool that allows to illustrate, in a rather simplistic manner, these concepts (uncertainties) and its impact on society (for example, through risk) provides a step towards strengthening the bridge between the scientific developments in stochastic hydrogeology and the state-of-practice.

Acknowledgments. The first and fourth authors would like to thank the German Research Foundation (DFG) for financial support of the project within the Cluster of Excellence in Simulation Technology (EXC310/1) at the University of Stuttgart. This work has been partially supported by the Spanish Ministry of Science and Innovation through projects RARA AVIS (reference CGL2009-11114) and Consolider-Ingenio 2010 (ref. CSD2009-00065). We also would like acknowledge the comments made by our reviewers.

References

Andricevic, R., and V. Cvetkovic (1996), Evaluation of Risk from Contaminants Migrating by Groundwater, *Water Resources Research*, *32*(3), 611–621.

Andricevic, R., J. Daniels, and R. Jacobson (1994), Radionuclide migration using travel time transport approach and its application in risk analysis, *Journal of Hydrology*, *163*, 125–145.

Baresel, C., and G. Destouni (2007), Uncertainty-accounting environmental policy and management of water systems, *Environ. Sci. Technol*, *41*(10), 3653–3659.

600 Bedford, T., and R. Cooke (2003), *Probalistic Risk Analysis: Foundations and Methods*,
601 Cambridge University Press.

602 Bellin, A., and Y. Rubin (1996), HYDRO_GEN: A spatially distributed random field
603 generator for correlated properties, *Stochastic Hydrology and Hydraulics*, *10*(4), 253–
604 278.

605 Bellin, A., and D. Tonina (2007), Probability density function of non-reactive solute con-
606 centration in heterogeneous porous formations, *Journal of contaminant hydrology*, *94*(1-
607 2), 109–125.

608 Benekos, I., C. A. Shoemaker, and J. R. Stedinger (2007), Probabilistic risk and uncer-
609 tainty analysis for bioremediation of four chlorinated ethenes in groundwater, *Stochastic*
610 *Environmental Research Risk Assessment*, *21*, 375–390.

611 Benson, D. A., and M. M. Meerschaert (2008), Simulation of chemical reaction via particle
612 tracking: Diffusion-limited versus thermodynamic rate-limited regimes, *Water Resour.*
613 *Res.*, *44*, W12,201, doi:10.1029/2008WR007,111.

614 Bertuzzo, E., S. Azaele, A. Maritan, M. Gatto, I. Rodriguez-Iturbe, and A. Rinaldo
615 (2008), On the space-time evolution of a cholera epidemic, *Water Resources Research*,
616 *44*(1), 1–W01,424.

617 Bogen, K. T., and R. C. Spear (1987), Integrating Uncertainty and Interindividual Vari-
618 ability in Environmental Risk Assessment, *Risk Analysis*, *7*(4), 427–436.

619 Bolster, D., M. Barahona, M. Dentz, D. Fernandez-Garcia, X. Sanchez-Vila, P. Trinchero,
620 C. Valhondo, and D. Tartakovsky (2009), Probabilistic risk analysis of groundwater
621 remediation strategies, *Water Resources Research*, *45*(6).

Bolster, D., D. A. Benson, T. LeBorgne, and M. Dentz (2010), Anomalous mixing and reaction induced by super-diffusive non-local transport, *Physical Review E*, *Submitted*.

Burmaster, D., and A. Wilson (1996), An introduction to second-order random variables in human health risk assessments, *Human and Ecological Risk Assessment: An International Journal*, *2*(4), 892–919.

Christakos, G. (2004), A sociological approach to the state of stochastic hydrogeology, *Stochastic Environmental Research and Risk Assessment*, *18*(4), 274–277.

Cirpka, O., R. Schwede, J. Luo, and M. Dentz (2008), Concentration statistics for mixing-controlled reactive transport in random heterogeneous media, *Journal of contaminant hydrology*, *98*(1-2), 61–74.

Cvetkovic, V., A. Shapiro, and G. Dagan (1992), A solute flux approach to transport in heterogenous formations 2: Uncertainty Analysis, *Water Resources Research*, *28*(5), 1377–1388.

Dagan, G. (1987), Theory of Solute Transport by Groundwater, *Annual Review of Fluid Mechanics*, *19*, 183–215.

Dagan, G. (1989), *Flow and Transport in Porous Formations*, Springer Verlag, Berlin.

Davison, A., G. Howard, M. Stevens, P. Callan, L. Fewtrell, D. Deere, and J. Bartram (2005), Water Safety Plans Managing drinking-water quality from catchment to consumer. Geneva: World Health Organisation, *Tech. rep.*, WHO/SDE/WSH/05.06.

Dawoud, E., and S. Purucker (1996), Quantitative Uncertainty Analysis of Superfund Residential Risk Pathway Models for Soil and Groundwater: White Paper, *Tech. rep.*

de Barros, F. P. J., and Y. Rubin (2008), A Risk-Driven Approach for Subsurface Site Characterization, *Water Resources Research*, *44*(W01414),

645 doi:10.1029/2007WR006,081.

646 de Barros, F. P. J., Y. Rubin, and R. Maxwell (2009), The concept of comparative in-
647 formation yield curves and its application to risk-based site characterization, *Water*
648 *Resources Research*, *45*(W06401), doi:10.1029/2008WR007,324.

649 Dentz, M., and J. Carrera (2005), Effective solute transport in temporally fluctuating flow
650 through heterogeneous media, *Water Resources Research*, *41*(8), W08,414.

651 Donado, L. D., X. Sánchez-Vila, M. Dentz, J. Carrera, and D. Bolster (2009), Multicom-
652 ponent reactive transport in multicontinuum media, *Water Resour. Res.*, *45*, W11,402,
653 doi:10.1029/2008WR006,823.

654 Edery, Y., H. Scher, and B. Berkowitz (2009), Modeling bimolecular reactions and trans-
655 port in porous media, *Geophys. Res. Lett.*, *36*, L02,407 doi:10.1029/2008GL036,381.

656 Fiori, A., I. Jankovic, G. Dagan, and V. Cvetkovic (2007), Ergodic transport through
657 aquifers of non-Gaussian log conductivity distribution and occurrence of anomalous
658 behavior, *Water Resources Research*, *43*(9), W09,407.

659 Fjeld, R., N. Eisenberg, and K. Compton (2007), *Quantitative Environmental Risk Anal-*
660 *ysis for Human Health*, first ed., Wiley.

661 Freer, J., K. Beven, and B. Ambroise (1996), Bayesian estimation of uncertainty in runoff
662 prediction and the value of data: An application of the GLUE approach, *Water Re-*
663 *sources Research*, *32*(7), 2161–2173.

664 Freeze, R. (2004), The role of stochastic hydrogeological modeling in real-world engi-
665 neering applications, *Stochastic Environmental Research and Risk Assessment*, *18*(4),
666 286–289.

667 Frind, E. O., J. W. Molson, and D. L. Rudolph (2006), Well vulnerability: a quantitative
668 approach for source water protection, *Ground water*, *44*(5), 732–742.

669 Gelhar, L. W., C. Welty, and K. Rehfeldt (1992), A critical review of data on field scale
670 disperion in aquifers, *Water Resour. Res*, *28*, 1955–1974.

671 Gramling, C., C. Harvey, and L. Meigs (2002), Reactive transport in porous media: A
672 comparison of model prediction with laboratory visualization, *Environ. Sci. Technol.*,
673 *36*, 2508–2514.

674 Kitanidis, P. (1994), The concept of the dilution index, *Water Resources Research*, *30*(7),
675 2011–2026.

676 Kitanidis, P. K. (1995), Quasi-linear geostatistical theory for inversing, *Water Resour.*
677 *Res.*, *31*(10), 2411–2419.

678 Levy, M., and B. Berkowitz (2003), Measurement and analysis of non-fickian dispersion
679 in heterogeneous porous media, *Journal of Contaminant Hydrology*, *64*, 203–226.

680 Maddalena, R. L., and T. E. McKone (2002), Developing and Evaluating Distributions
681 for Probabilistic Human Exposure Assessments, *Tech. Rep. LBNL-51492*.

682 Maxwell, R., and W. Kastenberg (1999), Stochastic Environmental Risk Analysis: An
683 Integrated Methodology for Predicting Cancer Risk from Contaminated Groundwater,
684 *Stochastic Environmental Research Risk Assessment*, *13*, 27–47.

685 Maxwell, R., S. Pelmulder, F. Tompson, and W. Kastenberg (1998), On the development
686 of a new methodology for groundwater driven health risk assessment, *Water Resources*
687 *Research*, *34*(4), 833–847.

688 Maxwell, R., W. Kastenberg, and Y. Rubin (1999), A methodology to integrate site char-
689 acterization information into groundwater-driven health risk assessment, *Water Re-*

690 *sources Research*, *35*(9), 2841–2885.

691 Maxwell, R., S. Carle, and A. Tompson (2008), Contamination, Risk, and Heterogeneity:
692 On the Effectiveness of Aquifer Remediation, *Environmental Geology*, *54*, 1771–1786.

693 McKnight, U., S. Funder, J. Rasmussen, M. Finkel, P. Binning, and P. Bjerg (2010), An
694 integrated model for assessing the risk of TCE groundwater contamination to human
695 receptors and surface water ecosystems, *Ecological Engineering*, *In Press*.

696 McKone, T., and T. Bogen (1991), Predicting the uncertainties in risk assessment, *Envi-*
697 *ron. Sci. Technol.*, *25*(10), 1674–1681.

698 McLucas, A. C. (2003), *Decision making: risk management, systems thinking and situa-*
699 *tion awareness*, Argos Press, Canberra, Australia.

700 Molin, S., and V. Cvetkovic (2010), Microbial risk assessment in heterogeneous aquifers:
701 1. Pathogen transport, *Water Resources Research*, *46*(5 (W05518)).

702 Molin, S., V. Cvetkovic, and T. Stenstrom (2010), Microbial risk assessment in heteroge-
703 neous aquifers: 2. Infection risk sensitivity, *Water Resources Research*, *46*(5(W05519)).

704 Morales, K. H., L. Ryan, T. L. Kuo, M. Wu, and C. J. Chen (2000), Risk of Internal
705 Cancers from Arsenic in Drinking Water, *Environmental Health Perspectives*, *108*(7),
706 655–661.

707 Neuman, S. P., and D. M. Tartakovsky (2009), Perspective on theories of anoma-
708 lous transport in heterogeneous media, *Adv. Water Resour.*, *32*(5), 670–680, doi:
709 10.1016/j.advwatres.2008.08.005.

710 Nowak, W. (2010), Measures of parameter uncertainty in geostatistical estimation and
711 design, *Mathematical Geosciences*, *42*(2), 199–221.

712 Nowak, W., F. P. J. de Barros, and Y. Rubin (2010), Bayesian geostatistical design:
713 Optimal site investigation when the geostatistial model is uncertain, *Water Resour.*
714 *Res.*, *46*(W03535), doi:10.1029/2009WR008,312.

715 Pappenberger, F., and K. Beven (2006), Ignorance is bliss: Or seven reasons not to use
716 uncertainty analysis, *Water Resources Research*, *42*(5), W05,302.

717 Persson, K., and G. Destouni (2009), Propagation of water pollution uncertainty and risk
718 from the subsurface to the surface water system of a catchment, *Journal of Hydrology*,
719 *377*(3-4), 434–444.

720 Raje, D., and V. Kapoor (2000), Experimental study of bimolecular reaction kinetics in
721 porous media, *Environ. Sci. Technol.*, *24*, 1234–1239.

722 Rubin, Y. (1991), Prediction of tracer plume migration in heterogeneous porous media by
723 the method of conditional probabilities, *Water Resour. Res.*, *27*(6), 1291–1308.

724 Rubin, Y. (2003), *Applied Stochastic Hydrogeology*, first ed., Oxford Press.

725 Rubin, Y. (2004), Stochastic hydrogeology - challenges and misconceptions, *Stochastic*
726 *Environmental Research Risk Assessment*, *18*, 280–281.

727 Rubin, Y., and G. Dagan (1992), Conditional estimates of solute travel time in heteroge-
728 nous formations: impact of transmissivity measurements, *Water Resources Research*,
729 *28*(4), 1033–1040.

730 Rubin, Y., M. A. Cushey, and A. Bellin (1994), Modeling of transport in groundwater for
731 environmental risk assessment, *Stochastic Hydrol. Hydraul.*, *8*(1), 57–77.

732 Sanchez-Vila, X., and A. Guadagnini (2005), Travel time and trajectory moments of con-
733 servative solutes in three dimensional heterogeneous porous media under mean uniform
734 flow, *Advances in Water Resources*, *28*, 429–439.

735 Sidle, C. R., B. Nilson, M. Hansen, and J. Fredericia (1998), Spatially varying hydraulic
736 and solute transpiort characteristics of a fractured till determined by field tracer test,
737 funen, denmark, *Water Resour. Res*, *34*, 2515–2527.

738 Silliman, S. E., L. Konikow, and C. Voss (1997), Laboratory experiment of longitudinal
739 dispersion in anisotropic porous media, *Water Resour. Res*, *23*, 2145–2154.

740 Tartakovsky, D. (2007), Probabilistic risk analysis in subsurface hydrology, *Geophysical*
741 *Research Letters*, *34*(5), 5404.

742 Tartakovsky, D. M., and C. L. Winter (2008), Uncertain future of hydrogeology, *ASCE*
743 *J. Hydrologic Engrg.*, *13*(1), 37–39.

744 Troldborg, M., G. Lemming, P. Binning, N. Tuxen, and P. Bjerg (2008), Risk assessment
745 and prioritisation of contaminated sites on the catchment scale, *Journal of contaminant*
746 *hydrology*, *101*(1-4), 14–28.

747 Troldborg, M., P. Binning, S. Nielsen, P. Kjeldsen, and A. Christensen (2009), Unsat-
748 urated zone leaching models for assessing risk to groundwater of contaminated sites,
749 *Journal of contaminant hydrology*, *105*(1-2), 28–37.

750 Ucinski, D. (2005), *Optimal measurement methods for distributed parameter system iden-*
751 *tification*, CRC.

752 USEPA (1989), Risk Assessment Guidance for Superfund Volume 1: Human Health Man-
753 ual (Part A), *Tech. Rep. Rep.EPA/540/1-89/002*.

754 USEPA (1991), Risk Assessment Guidance for Superfund Volume 1: Human Health Eval-
755 uation (Part B), *Tech. Rep. Rep.EPA/540/R-92/003*.

756 USEPA (2001), Risk Assessment Guidance for Superfund: Volume III - Part A, Process
757 for Conducting Probabilistic Risk Assessment, *Tech. Rep. Rep.EPA 540/R-02/002*.

758 Winter, C., and D. Tartakovsky (2008), A reduced complexity model for probabilistic risk
759 assessment of groundwater contamination, *Water Resources Research*, *30*(6), 2799–
760 2816.

761 Yu, W., C. Harvey, and C. Harvey (2003), Arsenic in groundwater in Bangladesh: a
762 geostatistical and epidemiological framework for evaluating health effects and potential
763 remedies, *Water Resources Research*, *39*(6), 1146.

764 Zimmerman, D. A., et al. (1998), A comparison of seven geostatistically based inverse ap-
765 proaches to estimate transmissivities for modeling advective transport by groundwater
766 flow, *Water Resour. Res.*, *34*(6), 1373–1413.

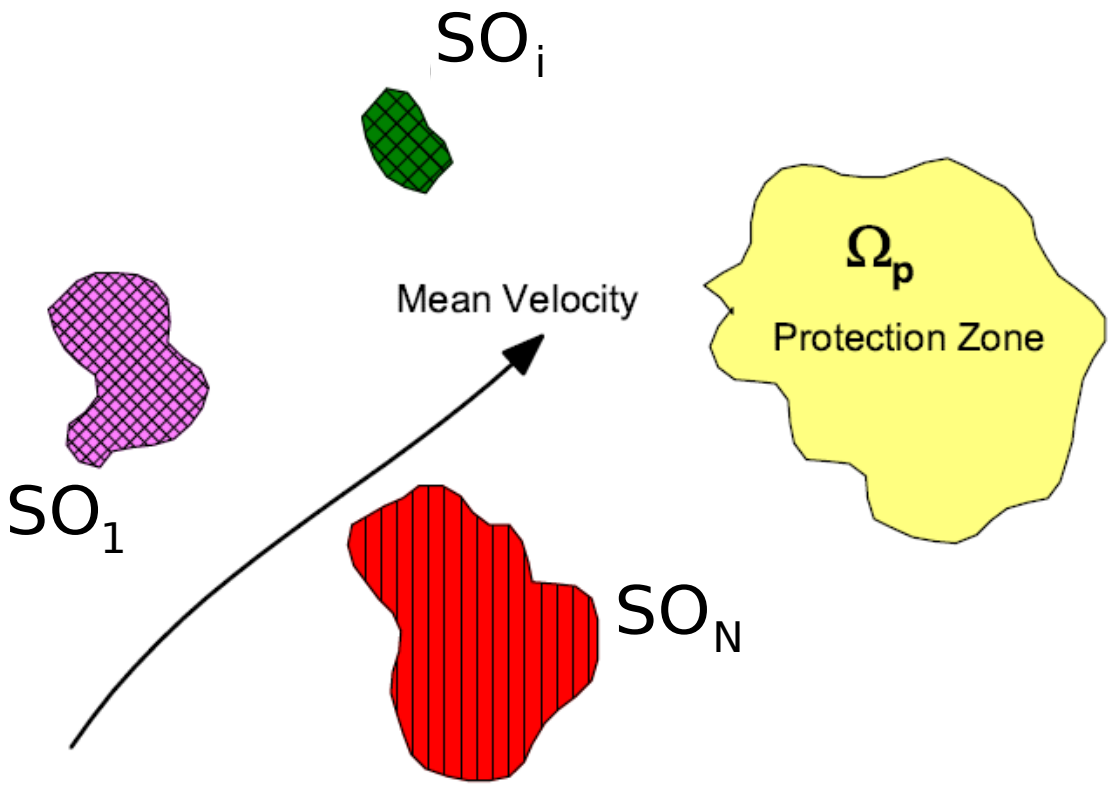

Figure 1. Schematic depiction of the contamination scenario considered in this work. Several potential sources SO_i, $i = 1,...,N$ are considered. Each source implies the combination of a potentially hazard solute located in a given (sometimes unknown) location.

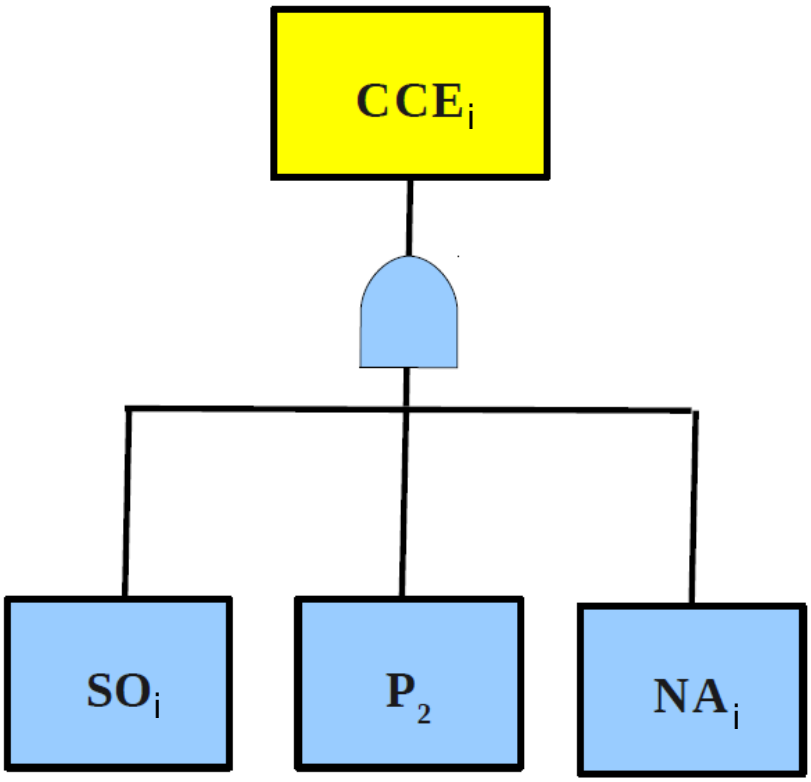

Figure 2. Fault tree for CCE_i.

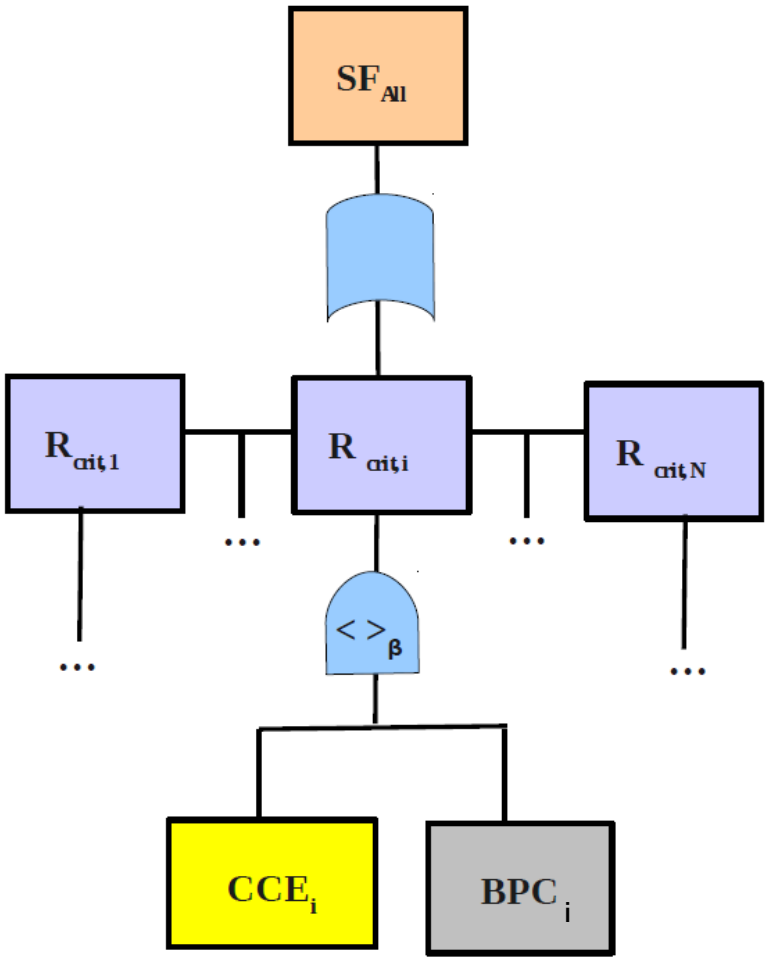

Figure 3. Fault tree for the total system failure

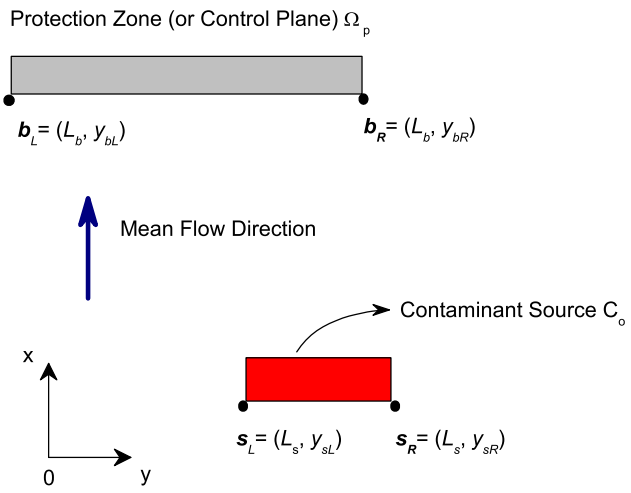

Figure 4. Schematic representation of the physical problem. A contaminant with initial concentration C_o is released. U is the mean velocity.

Data		
Parameter	A	B
C_o	1 mg/l	1.5 mg/l
λ	0.004 d^{-1}	0.002 d^{-1}
$L_b - L_s$	35 m	60 m
y_{sR}	12 m	4 m
y_{sL}	8 m	2 m
y_{bR}	1 m	1 m
y_{bL}	10 m	10 m
C_{cit}	0.1 mg/l	0.4 mg/l
$\mu_{ln\beta}$	-5.54	-6.9
$\sigma_{ln\beta}$	0.59	0.4

Table 1. Data for contaminant A and B.

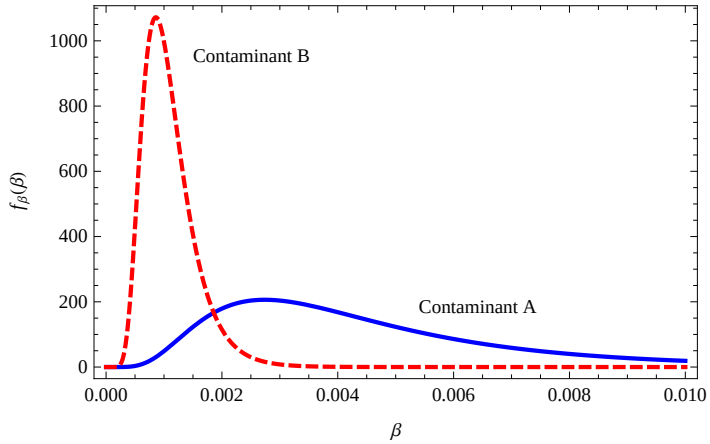

Figure 5. Distributions for the health-related parameters for contaminants A (continuous line) and B (dashed line).

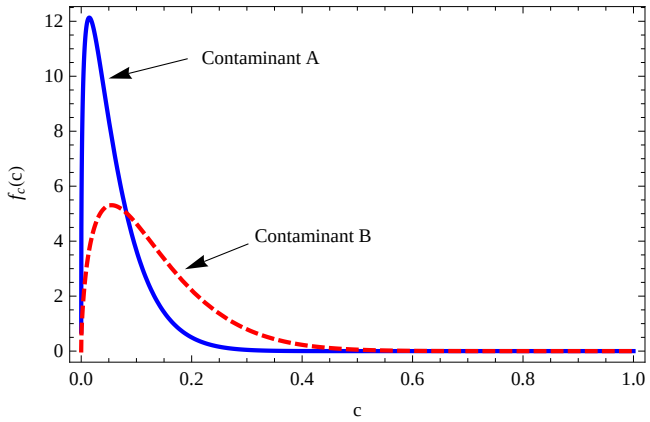

Figure 6. Concentration pdfs for contaminants A (continuous line) and B (dashed line) according to Eq. (18)

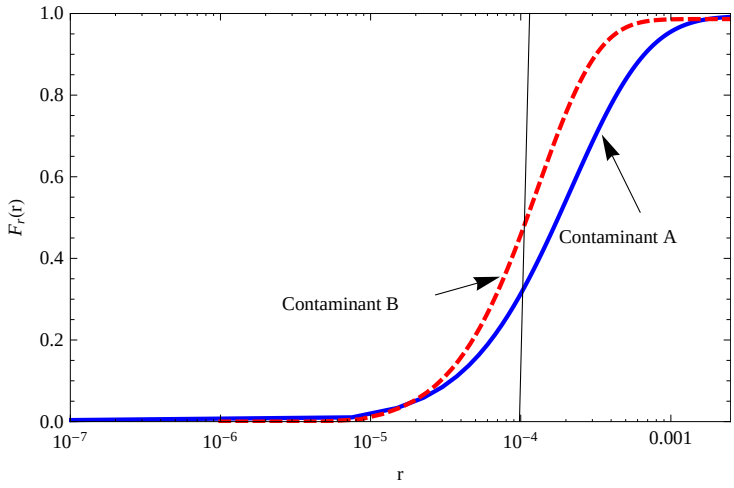

Figure 7. Risk cdf $F_r(r)$ for contaminants A (continuous line) and B (dashed line). The regulatory threshold is defined to be $r_{crit} = 10^{-4}$.

Probabilities			
Event	Parameter	A	B
SO	$Pr[SO]$	1	1
P_2	$Pr[P_2]$	0.38	0.26
NA	$Pr(C \geq C_{crit})$	0.18	0.015
R_{crit}	$Pr(r \geq R_{crit})$	0.69	0.54
SF	$Pr(SF)$	0.047	0.0022

Table 2. Computed probabilities for the hypothetical illustrative case

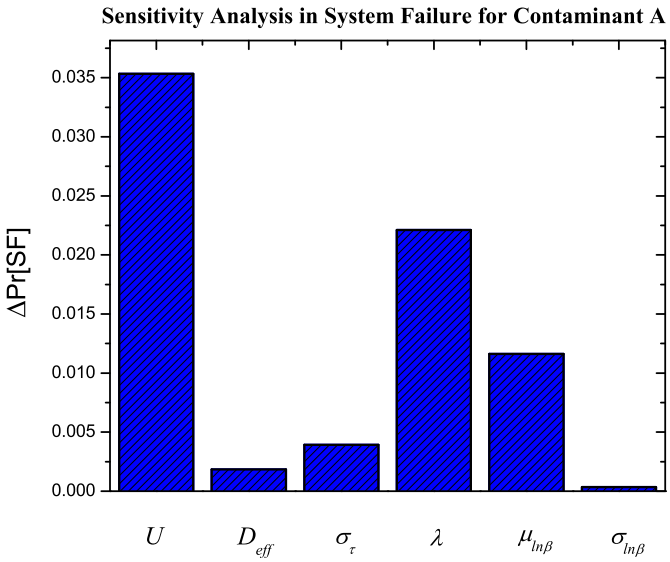

Figure 8. Sensitivity analysis for contaminant A. Change in probability of system failure $\Delta Pr[SF]$ if each parameter in $\boldsymbol{\theta}$ is perturbed by 10 percent.

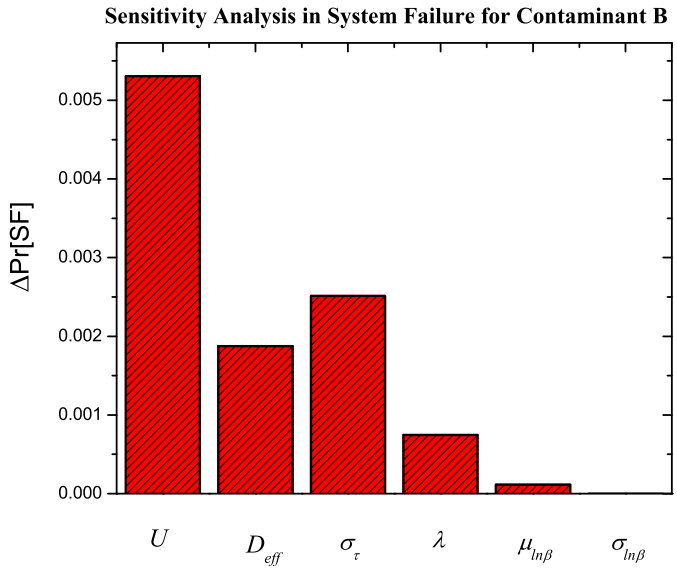

Figure 9. Sensitivity analysis for contaminant B. Change in probability of system failure $\Delta Pr[SF]$ if each parameter in $\boldsymbol{\theta}$ is perturbed by 10 percent.

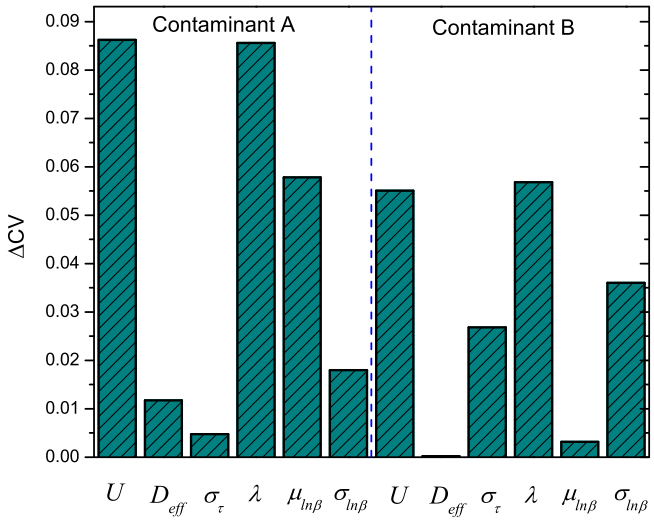

Figure 10. The dependency of the risk coefficient of variation for contaminants A and B on the perturbed parameter. Each parameter in $\boldsymbol{\theta}$ was perturbed by 10 percent. The coefficient of variation is equal to the risk standard deviation divided by its mean. Results evaluated using Eq. 20. ΔCV corresponds to the change in the coefficient of variation.