The concept of comparative information yield curves and its application to risk-based site characterization

Felipe P. J. de Barros,[1] Yoram Rubin,[1] and Reed M. Maxwell[2]

Received 30 July 2008; revised 11 March 2009; accepted 19 March 2009; published 2 June 2009.

[1] Defining rational and effective hydrogeological data acquisition strategies is of crucial importance as such efforts are always resource limited. Usually, strategies are developed with the goal of reducing uncertainty, but less often they are developed in the context of their impacts on uncertainty. This paper presents an approach for determining site characterization needs on the basis of human health risk. The main challenge is in striking a balance between reduction in uncertainty in hydrogeological, behavioral, and physiological parameters. Striking this balance can provide clear guidance on setting priorities for data acquisition and for better estimating adverse health effects in humans. This paper addresses this challenge through theoretical developments and numerical simulation. A wide range of factors that affect site characterization needs are investigated, including the dimensions of the contaminant plume and additional length scales that characterize the transport problem, as well as the model of human health risk. The concept of comparative information yield curves is used for investigating the relative impact of hydrogeological and physiological parameters in risk. Results show that characterization needs are dependent on the ratios between flow and transport scales within a risk-driven approach. Additionally, the results indicate that human health risk becomes less sensitive to hydrogeological measurements for large plumes. This indicates that under near-ergodic conditions, uncertainty reduction in human health risk may benefit from better understanding of the physiological component as opposed to a more detailed hydrogeological characterization.

Citation: de Barros, F. P. J., Y. Rubin, and R. M. Maxwell (2009), The concept of comparative information yield curves and its application to risk-based site characterization, *Water Resour. Res.*, *45*, W06401, doi:10.1029/2008WR007324.

1. Introduction

[2] Obtaining accurate predictions of potential human health risks from groundwater contamination is a challenge. The main difficulty relies on the fact that many of the factors that constitute risk are uncertain. Among these, we highlight two classes of parameters: (1) hydrogeological and (2) physiological.

[3] Hydrogeological parameters are necessary to estimate fate and transport of pollutants in the subsurface as well as the level of contamination to which humans potentially will be exposed. Because of aquifer heterogeneity [*Dagan*, 1984; *Rubin and Dagan*, 1992; *Rubin*, 2003], the input values for hydrogeological parameters between measurement locations can influence the flow field and, consequently, the concentration values calculated by the model. Since we lack the full knowledge of the subsurface structure, we must account its uncertainty to fill the spatial gap not covered by measurements [*Beckie*, 1996; *Rubin*, 2003].

[4] Physiological parameters are needed in order to link contaminant concentration to human health risk. Uncertainty within this component comes from dose response studies [*McKone and Bogen*, 1991; *Chiu et al.*, 2007]. The dose response relationship is often obtained by performing laboratory experiments on animals at high doses and later extrapolating the results to humans. In addition, dose response relationships are extrapolated to the lower doses relevant to regulatory concerns. Thus, because of these extrapolations, dose response models are uncertain. Besides the physiological component, human behavioral characteristics, such as ingestion rate of tap water, also add uncertainty and variability in the risk related parameters [*Burmaster and Wilson*, 1996; *Maxwell et al.*, 1998; *Daniels et al.*, 2000].

[5] Understanding the impact from each of these factors in human health risk provides a rational guidance toward answering questions such as: What is the expected risk uncertainty reduction if additional measurements of hydraulic conductivity are sampled? Given the uncertainty present in physiology, when is a detailed site characterization campaign necessary?

[6] Several studies have investigated risk due to groundwater contamination in a probabilistic framework. For example, risk–cost benefit analysis can be found in work by *Massmann and Freeze* [1987], *Freeze et al.* [1990] and *James and Gorelick* [1994]. They studied the trade-offs between financial costs and risk. Uncertainty is accounted within the hydrogeological parameters using a Bayesian

[1]Department of Civil and Environmental Engineering, University of California, Berkeley, California, USA.
[2]Department of Geology and Geologic Engineering, Colorado School of Mines, Golden, Colorado, USA.

0043-1397/09/2008WR007324$09.00

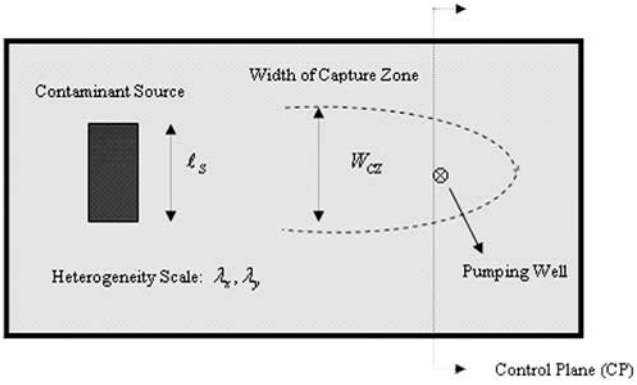

Figure 1. Illustration of the various length scales that define the flow and transport and consequently risk. Shown are width of the contaminant source (ℓ_S), capture zone width (W_{cz}), and the representative geostatistical correlation lengths (λ_x, λ_y).

framework. In these previous studies, costs are associated with probability of failure (contamination above some regulatory threshold value occurring) and the worth of data was addressed. Implication of aquifer heterogeneity in risk is addressed by *Rubin et al.* [1994]. A few other articles investigated the dependence of risk to hydrogeological and physiological parameters under different types of contaminants (i.e., radionuclide or organic) [*Bogen and Spear*, 1987; *Andricevic et al.*, 1994; *Andricevic and Cvetkovic*, 1996; *Maxwell et al.*, 1998; *Maxwell and Kastenberg*, 1999]. *Maxwell et al.* [1999] addressed how increased sampling of hydraulic conductivities affects reduction of uncertainty in human health risk. *Benekos et al.* [2007] extended the studies performed by *Maxwell and Kastenberg* [1999] for multispecies transport.

[7] To investigate the relative impact of uncertainty reduction in the hydrogeological component and in physiological component on the final risk estimate, *de Barros and Rubin* [2008] developed a metric based on the concept of information entropy that allows one to quantify the relative impact of information gathered on human health risk. This metric is used within a graphical tool that compares alternative strategies for risk uncertainty reduction.

[8] However, the role of flow and transport scales in determining characterization needs in a risk-driven approach has not received much attention. There is still need for further investigation when counterbalancing the effects of hydrogeological site characterization with physiological uncertainty as a function of flow and transport scales. Hydraulic properties can vary on different scales and the value of hydrogeological information is dependent on these physical scales. Physical scales include the characteristic lengths that characterize subsurface heterogeneity, flow and transport. Such scales, as shown in Figure 1, are source size relative to the correlation length of aquifer heterogeneity, size and configuration of the exposure endpoint (screened well or control plane), pore-scale and capture zones induced by the action of pumping. Furthermore, little attention has been given to the implication of different risk models in defining characterization needs and this issue will also be addressed.

[9] In this paper, we employ the concepts presented by *de Barros and Rubin* [2008] to investigate the significance of various length scales that define the risk problem and their impact on hydrogeological site characterization. We extend the ideas from *de Barros and Rubin* [2008] to introduce the concept of comparative information yield curves in order to quantify the relative impact of uncertainty reduction of flow and health parameters in risk. The theoretical aspects of this concept are presented along with its implications on site characterization applications. This paper addresses the following fundamental question: Are there physical flow and transport characteristics in which uncertainty reduction in human health risk will benefit more from uncertainty reduction from human physiology or hydrogeology? We wish to investigate the role of these physical scales in determining characterization needs.

[10] The above question is relevant since assessing the value of data acquisition is an issue of concern in real life applications. Questions concerning the expansion of existing, and sometimes even substantial, measurement networks, or issues regarding selecting between alternative targets for additional investment are of primary concern. Such efforts may not always be justified, because they can potentially yield only marginal improvement in the predictive capability. Because of an ever-increasing demand on site characterization, many sampling techniques are available that vary in resolutions and offer direct or indirect information on the parameters relevant for modeling [*Hubbard and Rubin*, 2000; *Kowalsky et al.*, 2005]. Thus having rational guide to manage all these alternatives becomes relevant since we live in a resource constrained world.

[11] The present work is structured as follows: section 2 gives the general mathematical statement of the problem; section 3 shows how to use the entropy concept in human health risk; section 4 presents an illustration case; results, together with discussion, are given in section 5. Finally conclusions and future considerations are given in section 6. A notation section containing the majority of the symbols used throughout this work is also included.

2. Mathematical Statement of the Problem

[12] Given the uncertainty present in all components of human health risk assessment, it is rational to use a probabilistic framework to quantify risk due to groundwater contamination. Our objective is to obtain the ensemble distributions of human health risk for the exposed population. For our work, we will consider r to represent the increased lifetime cancer risk. This is not a limitation and noncancer risks can also be used within the framework. Here, $F_R(r)$ denotes the corresponding risk cumulative distribution function (cdf). $F_R(r)$ is evaluated for a given vector of hydrogeological parameters, $\boldsymbol{\theta}_H$, field site measurements, $\{m\}$, and for a given matrix containing the population's health-related parameters $\boldsymbol{\Theta}_P$.

[13] The vector $\boldsymbol{\theta}_H$ contains the parameters that characterize the space random function (SRF) of the hydrogeological variables [*Dagan*, 1984, 1987; *Rubin and Dagan*, 1992; *Rubin*, 2003] such as mean value and variance of the log conductivity, integral scales as well as other flow and transport related parameters such as porosity, source concentration, pumping rates and dispersion coefficients. These parameters have a physical and chemical nature and can be deterministic or stochastic.

[14] The exposed population with I individuals is characterized by the matrix $\Theta_P = \{\boldsymbol{\theta}_{P,1}, \boldsymbol{\theta}_{P,2}, \boldsymbol{\theta}_{P,3}, \ldots \boldsymbol{\theta}_{P,i}, \ldots \boldsymbol{\theta}_{P,I}\}$ where $\boldsymbol{\theta}_{P,i}$ is the vector of behavioral and physiological characteristics of the ith individual. Each $\boldsymbol{\theta}_{P,i}$, where $i = 1, 2, \ldots, I$, varies from individual to individual. The typical parameters present in $\boldsymbol{\theta}_{P,i}$ are, for example, the ingestion rate per body weight, exposure duration and cancer potency factors as well as their statistical moments if uncertainty exists. Statistical distributions for these parameters for different pathways are found in the literature [see *Maxwell et al.*, 1998; *Binkowitz and Wartenberg*, 2001; *Portier et al.*, 2007]. The conditional risk cdf for the ith individual of the exposed population is given as follows:

$$F_R\left(r|\boldsymbol{\theta}_H, \boldsymbol{\theta}_{P,i}, \{m\}\right) = \Pr[R < r]. \tag{1}$$

With equation (1), risk estimates can be obtained given an appropriate risk model. Most important, for a given regulatory acceptable risk value, for example $r = 10^{-6}$, equation (1) provides the probability of risk reliability or exceedance. This may be accomplished by calculating the complementary cumulative distribution function, $\Pr[R > r] = 1 - F_R$.

3. Use of Entropy to Quantify the Impact of Information on Risk

[15] In the recent work of *de Barros and Rubin* [2008], information entropy was used to develop a metric that relates the amount of information in hydrogeology to the amount of information in physiology. This metric, denoted in the present work by α, was used to investigate uncertainty trade-offs between hydrogeological parameters (such as hydraulic conductivity K) and physiological parameters (cancer potency factor). Before explaining the form of α and how it functions, some definitions are required. We first introduce the concept of information yield curves. Afterward, we extend the theoretical aspects toward applications to site characterization.

3.1. Concept of Information Yield Curves

[16] Following the work of *de Barros and Rubin* [2008], let us define E_H as the information entropy for hydrogeological parameters (including transport variables such as chemical parameters) and E_P as the entropy for physiological and behavioral parameters. The entropies are defined as [*Christakos*, 1992]

$$\begin{aligned} E_H &= -\int f_H(\boldsymbol{\theta}_H|I_H, \{m_a\}) \ln[f_H(\boldsymbol{\theta}_H|I_H, \{m_a\})] d\boldsymbol{\theta}_H \\ E_P &= -\int f_P(\boldsymbol{\theta}_P|I_P, \{s_a\}) \ln[f_P(\boldsymbol{\theta}_P|I_P, \{s_a\})] d\boldsymbol{\theta}_P, \end{aligned} \tag{2}$$

where f_H and f_P are the continuous probability density functions (pdf's) for the vector of hydrogeological parameters $\boldsymbol{\theta}_H$ and for the health-related parameters $\boldsymbol{\theta}_P$, respectively. The integration in equation (2) is performed over the entire parameter space. For the sake of notation, we have omitted the subscript i from $\boldsymbol{\theta}_{P,i}$ as defined in section 2. Equation (2) represents the total amount of information from each component at an initial stage of knowledge. These entropies can be evaluated with hydrogeological prior knowledge I_H, with a small amount of available hydraulic data $\{m_a\}$, physiological prior information I_P and finally, available health-related sample data $\{s_a\}$. From the distributions necessary to estimate E_H and E_P we can evaluate a corresponding $F_R(r)$, defined by equation (1), and consequently its statistical moments. As more information becomes available, either from flow or health physics, E_H and E_P would decrease since the uncertainty in both f_H and f_P is reducible with additional data collection.

[17] Being able to estimate the values of E_H and E_P with no a priori information allows one to investigate relative value of information in human health risk. This is necessary since decision makers need to decide where to invest resources toward risk uncertainty reduction. At this early stage of the risk analysis, only a small amount of information is available through prior knowledge or initial data. In order to decide whether or not more data is needed, one must evaluate its impact in the human health risk distribution. We now denote the unknown (to be sampled) hydrogeological measurements by $\{m_{na}\}$ and the unknown health-related by $\{s_{na}\}$. The following equations are the entropies averaged over all possible measurement values:

$$\begin{aligned} E_{H,O} &= \left\langle -\int \hat{f}_H(\boldsymbol{\theta}_H|I_H, \{m_a\}, \{m_{na}\}) \right. \\ &\qquad \left. \cdot \ln\left[\hat{f}_H(\boldsymbol{\theta}_H|I_H, \{m_a\}, \{m_{na}\})\right] d\boldsymbol{\theta}_H \right\rangle \\ E_{P,O} &= \left\langle -\int \hat{f}_P(\boldsymbol{\theta}_P|I_P, \{s_a\}, \{s_{na}\}) \right. \\ &\qquad \left. \cdot \ln\left[\hat{f}_P(\boldsymbol{\theta}_P|I_P, \{s_a\}, \{s_{na}\})\right] d\boldsymbol{\theta}_P \right\rangle \\ &\text{with } E_{P,O} \le E_P \text{ and } E_{H,O} \le E_H, \end{aligned} \tag{3}$$

with $E_{H,O}$ and $E_{P,O}$ are the expected entropy values over all possible measurements values that $\{m_{na}\}$ and $\{s_{na}\}$ can take. They are evaluated with the inferred pdf's $\hat{f}_H$ and $\hat{f}_P$ such that $E_{P,O} \le E_P$ and $E_{H,O} \le E_H$, see equation (2). A general numerical procedure that can be used to obtain the entropies in equation (3) is as follows.

[18] 1. Generate a possible realization of N_o measurements for $\{m_{na}\}$ and S_o measurements for $\{s_{na}\}$ from prior knowledge. This requires the assumption that the models from which the measurements are generated are known. For example, a Gaussian or exponential geostatistical model and a dose response model.

[19] 2. Using the data drawn from this realization, the parameter's pdf's, $\hat{f}_H$ and $\hat{f}_P$, are inferred. These parameters can be the mean or variance of the log conductivity data and integral scales or cancer potency factor. As the number of measurements increases, these pdf's become more informative. Details concerning the pdf estimation procedure is given in Appendix A.

[20] 3. With $\hat{f}_H$ and $\hat{f}_P$, the conditional entropies $E_{H,1}$ and $E_{P,1}$ can be calculated. Here, the second subscript corresponds to the first realization of the data sets $\{m_{na}\}$ and $\{s_{na}\}$. These entropies are conditional on the generated data and a known model.

[21] 4. Repeat steps 1–3 for several realizations of the data $\{m_{na}\}$ and $\{s_{na}\}$ such that two vectors with elements $E_{H,j}$ and $E_{P,j}$ are obtained. Here the subscript $j = 1, \ldots, J_{MAX}$ corresponds to the realizations. J_{MAX} is the maximum number to realizations.

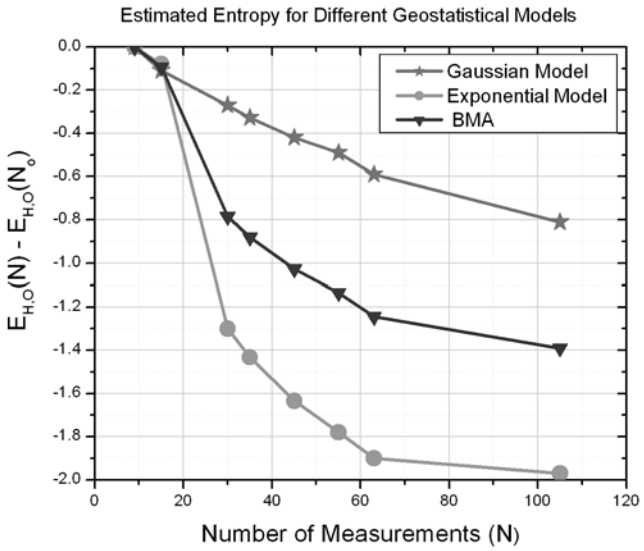

Figure 2. Entropy averaged over all possible measurement values generated by a geostatistical model. $E_{H,O}(N)$ and $E_{H,O}(N_o)$ are the entropies evaluated with N and N_o measurements, respectively, with $N \geq N_o$. If the model of the underlying formation is unknown, the methodology can account for the Bayesian model averaging (BMA). Here M_1 and M_2 correspond to the Gaussian and exponential models, respectively, with $\omega_1 = \omega_2 = 0.5$.

[22] 5. With the entropies $E_{H,j}$ and $E_{P,j}$, with $j = 1, \ldots, J_{MAX}$, the values for the ensembles averages, $E_{H,O}$ and $E_{P,O}$, are obtained.

[23] 6. Repeat steps 1–5 to evaluate the impact of an additional amount of data ($N > N_o$) in $\{m_{na}\}$ and $\{s_{na}\}$ ($S > S_o$).

[24] The assumption in this outlined procedure is that some information about the site needs to be known. This includes the prior parameter pdf and the use of expert opinions or information borrowed from geologically similar formations (see Appendix A). If model uncertainty exists (for example: the geostatistical model of the underlying geological formation or the shape of dose response model), the current framework can incorporate Bayesian model averaging [*Hoeting et al.*, 1999; *Neuman*, 2003]. This is done by assigning different weights to each entropy ensemble evaluated for a given model and then averaging them. Mathematically this is equivalent to $E_{H,O} = \sum \omega_i \times (E_{H,O}|M_i)$, where ω_i is the ith weight for the corresponding ith model denoted by M_i. The term $(E_{H,O}|M_i)$ is the ensemble averaged entropy given a geostatistical model. In many situations, the conceptual model for flow and transport could change as more data is collected. For instance, extra hydrogeological data may give evidence to the presence of a leaking aquitard or strong vertical pressure gradients, which would cause revisions of the initial conceptual model. Even if there is still uncertainty within the conceptual model, this additional data helps in updating the weights, ω_i, in the Bayesian model averaging procedure. The current framework allows for this model updating process as more data is collected and obtain new predictions.

[25] For increasing number, N, of measurements in both $\{m_{na}\}$ and $\{s_{na}\}$, the average entropy estimates decrease as shown in Figure 2. The vertical axis represents the difference between the entropy evaluated with increasing N measurements and the initial entropy calculated with N_o measurements (with $N \geq N_o$). This plot was obtained for the hydrogeological parameters by making use of steps given above together with Appendix A. A similar plot can be done with $E_{P,O}$ by averaging over all possible, nonavailable, physiological and behavioral data. Figure 2 also shows how different geostatistical models, exponential and Gaussian, can lead to different entropy estimates. It also illustrates the Bayesian model averaging result if the geostatistical model is uncertain. Equal weights were assigned to each model for this demonstration.

[26] Now that we have presented the necessary definitions, we can write the following entropy differences for both hydrogeological and physiological parameters:

$$\Delta E_H = E_H - E_{H,O}$$
$$\Delta E_P = E_P - E_{P,O}. \quad (4)$$

Equation (4) defines the differences between the expected entropies, $E_{H,O}$ and $E_{P,O}$, given in equation (3) and the current entropy stages denoted by E_H and E_P. By reducing uncertainty from both physiology and hydrogeology, ΔE_H and ΔE_P tend to values closer to zero. The metric that relates uncertainties from each risk component is given below:

$$\alpha = \frac{10^{\Delta E_H}}{10^{\Delta E_P}}, \quad (5)$$

where ΔE_H and ΔE_P are defined in equation (4). As explained by *de Barros and Rubin* [2008], loss of information in $\boldsymbol{\theta}_H$ means α increasing to values greater than one. This is obtained by increasing ΔE_H while keeping ΔE_P equal to zero. If uncertainty increases in $\boldsymbol{\theta}_P$, then α values are bounded between zero and one (keeping ΔE_H fixed and equal to zero). When α equals to one, we have $\Delta E_P = \Delta E_H = 0$. The point $\alpha = 1$ is considered the *base case* from which the relative contribution of information will be quantified. Figure 3 illustrates the α concept.

[27] With equation (5) we can obtain a series of values to the right and left of $\alpha = 1$ and evaluate, through simulations,

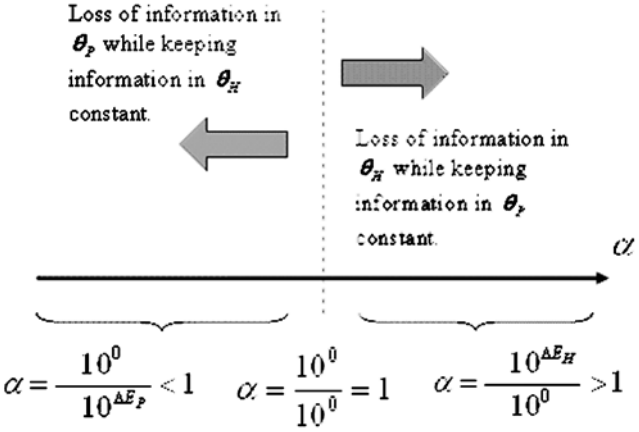

Figure 3. Graphical explanation of the α concept. At $\alpha = 1$, we have reached entropies $E_{P,O}$ and $E_{H,O}$. For each value of α, a corresponding risk variance or coefficient of variation is obtained. The plot of α versus the risk coefficient of variation is denoted here as the comparative information yield curves.

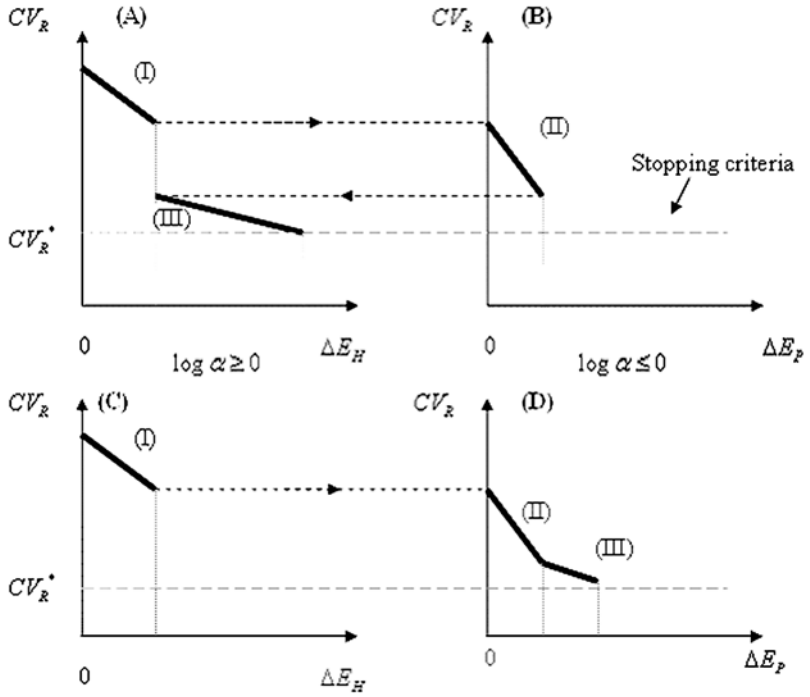

Figure 4. Illustration of the second alternative for the comparative information yield curves to investigate risk uncertainty reduction strategies between the hydrogeological and physiological component. Shown are (a and b) one sampling strategy and (c and d) a different sampling strategy. CV_R is the coefficient of variation of risk, while CV^*_R is a stopping criterion associated with an environmental regulation.

their corresponding uncertainty levels in human health risk. These corresponding uncertainty levels can be represented by risk variance, 95th confidence intervals, or the risk coefficient of variation ($CV_R = \sigma_R/\mu_R$). For the current work, we will adopt the change of the coefficient of variation in the following way:

$$\Delta CV_R = \frac{CV_R - CV_R^o}{CV_R}, \tag{6}$$

where CV_R^o corresponds to risk evaluated with the entropies $E_{H,O}$ and $E_{P,O}$. We will obtain a series of these α versus ΔCV_R curves for several different physical scenarios to investigate uncertainty trade-offs. These graphs are denoted here as the comparative information yield curves. Summarizing, the value of α denotes a change in entropy values. It is a metric for comparing two stages of information. A financial cost value can be obtained by relating α to a given sampling strategy. For this particular α value, a corresponding uncertainty reduction will occur. We represented this uncertainty reduction by ΔCV_R; however, other representative measures besides ΔCV_R can also be evaluated from the Monte Carlo simulations (such as 95th confidence intervals).

3.2. Application

[28] Since $E_{H,O}$ and $E_{P,O}$ are speculative projections, in the sense that it needs to be defined in equation (4) and (5), one may want additional formulations of the approach described previously. An alternative application of the entropy concept for investigating uncertainty trade-offs in human health risk is obtained by changing the definition of $E_{H,O}$ and $E_{P,O}$, given in equation (3) and (4), such that we have $E_H \leq E_{H,O}$ and $E_P \leq E_{P,O}$. This means that the values for $E_{H,O}$ and $E_{P,O}$ correspond to the current state of information and are denoted as base case entropies. Note that this new inequality differs from the definition given in equation (4). This would bypass the need to calculate the entropy ensemble averages, $E_{H,O}$ and $E_{P,O}$, as given in equation (3) and (4). On the basis of a set of initial data or prior information, an estimate of $E_{H,O}$ and $E_{P,O}$ is obtained such that $\Delta E_P = \Delta E_H = 0$ corresponds to the initial uncertain case together with a corresponding coefficient of variation for risk. Now, with this alternative approach, $E_{H,O}$ and $E_{P,O}$ represents the available amount of information at the early stage of characterization. As more data are collected, new estimates of E_H and E_P are obtained and their values will be lower than the corresponding $E_{H,O}$ and $E_{P,O}$. As shown in the following paragraph and in Figure 4, a graphical approach could be used not only to investigate the value of information in uncertainty reduction in human health risk as more data are acquired but also to compare different sampling strategies by estimating their respective value of information using the six-step procedure described earlier in section 3.1.

[29] Figure 4 shows the application of this alternative definition ($E_H \leq E_{H,O}$ and $E_P \leq E_{P,O}$) and how one can reduce the uncertainty in risk by using different sampling strategies. Figures 4a and 4b illustrate one sampling strategy: Reducing uncertainty from hydrogeology (segment I), then physiology (segment II) followed by hydrogeology again

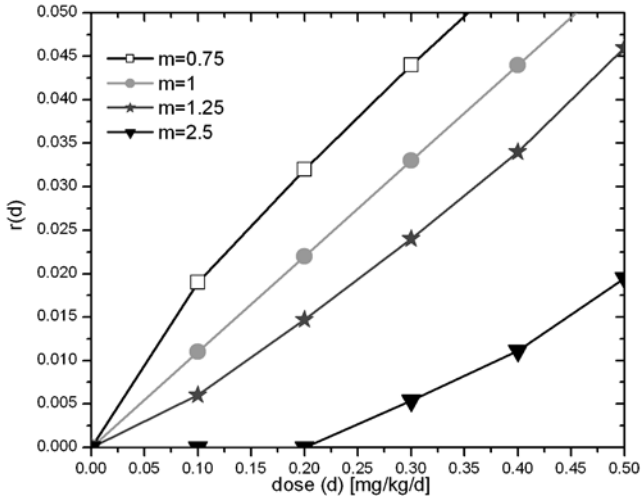

Figure 5. Example of dose response relationship at a function of m as shown in equation (7). The gray curve with circles represents the linear model used by the *EPA* [1989] with $m = 1$.

(segment III). A different strategy is shown in Figures 4c and 4d, by reducing uncertainty from hydrogeology (segment I), then physiology (segment II) and then physiology again (segment III). The sampling stops when a regulatory target is reached. Figure 4 gives a stepwise approach allowing one to direct efforts to obtain the best information yield.

[30] As mentioned before, the advantage of this alternative is that it avoids the need to prespecify $E_{H,O}$ and $E_{P,O}$ values as defined in equations (3) and (4) thus allowing one to construct the plots such as the one given in Figure 4. Also, the information yield curves based on this alternative definition offers a stepwise approach illustrated in Figure 4 that allows one to revise the conceptual model as more information becomes available. At each step, efforts can be allocated where the sampling strategy offers the best yield. The usefulness of this second approach will be illustrated in the end of this paper. Both alternatives for the use of entropy in risk will be discussed. However, it is important to state that the second alternative allows one to select where to invest resources in a more practical manner.

4. Illustration Case

[31] Consider a bounded 2-D flow in an aquifer with spatially variable and isotropic hydraulic conductivity $K(\boldsymbol{x})$ and $Y = \ln K$. Because of incomplete information of the system, K is characterized by its SRF and is considered here as statistically stationary. Its covariance structure model is assumed to be exponential and isotropic with σ_Y^2 being the variance of Y and λ the correlation length of heterogeneity. A contaminant plume, considered here as a collection of particles, is released within a rectangular source domain with transversal length ℓ_S. Each particle represents a mass of contaminant and travels along a streamline of the flow field and are used to determine spatial contaminant distributions that may cause adverse health effects. We simulate the case of a hypothetical PCE contamination problem. The prescribed pressure head along the longitudinal direction are used as boundary conditions. Zero flux boundary conditions are assumed along the transversal direction. A drinking water well with pumping rate Q represents the environmentally sensitive location. The governing flow and transport equations are given in Appendix B.

[32] Flow and transport are solved numerically using a Monte Carlo procedure. At each realization, the flow and transport problem is solved for a specific image of the aquifer's properties, generated using the turning bands method [*Tompson et al.*, 1989; *Rubin*, 2003]. Specific information concerning the numerical codes used in this work can be found in work by *Ashby and Falgout* [1996], *Maxwell and Kastenberg* [1999], *Jones and Woodward* [2001] and *Kollet and Maxwell* [2006]. Technical details concerning the numerical implementation are summarized in Appendix B. The framework presented can be used with analytical and numerical methods. Our choice for numerical implementation of flow and transport is for illustration and not to depend on simplifying assumptions. There exists a large amount of work published in the literature with analytical solutions that could be used to build the comparative information yield curves [*Rubin et al.*, 1994; *Cushey and Rubin*, 1997; *Sanchez-Vila and Rubin*, 2003] and many are summarized by *Rubin* [2003]. These same analytical solutions were also used to investigate human health risk [see *Andricevic et al.*, 1994; *Andricevic and Cvetkovic*, 1996] and served as the basis of the work of *de Barros and Rubin* [2008] in which the comparative information yield curves were used to evaluate uncertainty trade-offs. In the following we will describe the exposure pathways considered in this work as well as the input data used in the simulations.

4.1. Exposure Pathways and Risk Formulation

[33] We consider risk due to groundwater ingestion and inhalation for illustration of the methodology. These two pathways were shown to have a stronger impact in human health risk [*Maxwell et al.*, 1998]. Because of the nature and complexity of cancer mechanisms, cancer risk models are generally derived from dose response curves. These curves are based on toxicological studies and are often determined experimentally by observing adverse effects in animals for increasing applied doses (or concentrations) [*Fjeld et al.*, 2007]. The dose response curve results are then extrapolated to humans. A common challenge is determining what shape the dose response relationship in the extrapolated zone (where uncertainty is highest), also known as the low-dose zone, should take. For low doses, risk models can assume both linear and nonlinear forms [*Environmental Protection Agency* (*EPA*), 2005; *Fjeld et al.*, 2007]. Figure 5 shows linear and nonlinear models for the dose response curves.

[34] In the present formulation we will treat the risk model in the following functional form:

$$r = CPF_G \times [LADD_G]^m + CPF_H \times [LADD_H]^m, \qquad (7)$$

where CPF_G and CPF_H are the cancer potency factors for the ingestion and inhalation pathway, respectively. These are also known as the cancer slope factors. The parameter m determines the nonlinearity of the model. The value for m comes from fitting the model to toxicological data available from dose response experiments. $LADD_G$ and $LADD_H$ are the average daily doses for tap water ingestion and inhalation

Table 1. Data Used in Flow, Transport, and Health Risk Models[a]

	Value
Hydrogeological Parameters	
Q	5 and 50 m^3 d^{-1}
$Pe = \lambda/\alpha_L$	100 and ∞
λ	100 m
n	0.3
R_f	1
L	3000 m
W	2500 m
$\boldsymbol{x_w}$	(2500 m, 1000 m)
Behavioral Parameters	
IR/BW	0.033 L (d kg)$^{-1}$
AT	70 years
HR/BW	0.39 m^3 (d kg)$^{-1}$
W_S	480 h^{-1}
EF	350 d a^{-1}
ED	30 years
TE_S	0.5
VR_S	12 mg m^{-3}
ET_S	0.13 h d^{-1}

[a]Behavioral parameters are representative of the 50th fractile of variability. See the notation section.

during shower. The average daily dose is a function of the concentration (C) and behavioral and exposure parameters:

$$LADD_G = C \times \left(\frac{IR}{BW}\right) \times \frac{EF \times ED}{AT};$$
$$LADD_H = AC_S \times ET_S \times \left(\frac{HR}{BW}\right) \times \frac{EF \times ED}{AT}. \quad (8)$$

These behavioral and exposure parameters are the well-known EPA risk variables such as ingestion rate per body weight (IR/BW), exposure duration (ED), average lifetime (AT) and exposure frequency (EF), inhalation rate per body weight (HR/BW) and shower exposure time (ET_S) [*EPA*, 1989, 2001]. The indoor air concentration is denoted by $AC_S = C(W_S \times TE_S)/VR_S$ with W_S being the tap water use rate, TE_S the transfer efficiency from tap water to air and VR_S is the air exchange rate [*EPA*, 1989, 2001; *Maxwell et al.*, 1998]. For our work, we will use the peak concentrations to evaluate risk. Other works studied the effects of averaging concentration over the exposure duration [*Maxwell and Kastenberg*, 1999; *Hassan et al.*, 2001; *Maxwell et al.*, 2008] and the implications of using average versus peak concentration in hydrogeological site characterization [*de Barros and Rubin*, 2008]. To obtain the classic EPA linear low-dose model, we set $m = 1$ [*EPA*, 1989, 2001].

4.2. Input Data Used in the Case Study

[35] Table 1 summarizes the deterministic data used for input in the simulations. The domain with longitudinal dimension L and width W (size: $50\lambda \times 32\lambda$) is discretized into a regular rectangular grid. Each grid block has dimensions $\Delta x_1 = \Delta x_2 = \lambda/5$ [*Rubin et al.*, 1999; *Lawrence and Rubin*, 2007]. As mentioned previously, flow and transport are solved within the Monte Carlo approach and 300 realizations were performed.

[36] To answer the research questions posed in the introduction, we select an aquifer with parameters summarized in Table 2. This aquifer, denoted as the baseline, was selected from several realizations simulated with geometric mean $K_G = 1$ m d^{-1} and $\sigma_Y^2 = 1$ (see Table 2). From this baseline aquifer, we sampled values of hydraulic conductivity in fixed intervals of 8λ, 4λ and 2λ in a subdomain ($18\lambda \times 16\lambda$) horizontally centered with the contaminant source and the environmentally sensitive target. We denote by $\{m_1\}$ the measurement density associated with the sampling interval 8λ, $\{m_2\}$ with 4λ and finally $\{m_3\}$ with 2λ.

Table 2. Hydrogeological Data Used in the Conditional Simulations[a]

Sampling Strategy	K_G (m d^{-1})	σ_Y^2	N^*	$Var[\sigma_Y^2]$	$Var[K_G]$
"Base" aquifer	1	1	NA	NA	NA
$\{m_1\}$: 8λ	1.5	1.2	9	0.295	1.06
$\{m_2\}$: 4λ	0.9	0.60	25	0.035	0.16
$\{m_3\}$: 2λ	0.76	0.71	81	0.012	0.02

[a]Here N^* denotes the number of measurements sampled. NA means not applicable.

[37] For the present investigations, we assume that σ_Y^2 and K_G are uncertain parameters and its statistical distributions can be inferred as shown in Appendix A. Both σ_Y^2 and K_G vary between conditional simulation according to the three mentioned sampling densities shown in Table 2. Hence, f_H in equation (2) corresponds to the pdf for σ_Y^2 and K_G. If a distribution is assumed, say, lognormal, the statistical moments of σ_Y^2 and K_G can be estimated by using the maximum likelihood function [*Rubin*, 2003] (see Appendix C). Table 2 summarizes the estimated parameters from the sampled data set used in flow simulation.

[38] From the physiological side, we assume that cancer potency factors are the uncertain parameter and uniformly distributed [*McKone and Bogen*, 1991]. Thus, f_P in equation (2), represents the uniform pdf for CPF_G and CPF_H. Table 3 summarizes the upper and lower bounds used in the following simulations. The coefficient of variation (CV) is also included in Table 3. We evaluate risk for different levels of parametric uncertainty in CPF_G and CPF_H. Because of the lack of data, we assume, without loss of generality, that the hydrogeological and physiological parameters are independent. This assumption is not a limitation in our work. If correlations between both components are known (for example, concentration data and the cancer potency factors), then joint entropies can be evaluated with the corresponding joint pdf between hydrogeological and physiological parameters [see *Christakos*, 1992]. For the current work, $E_{H,O}$ is evaluated with estimated uncertain parameters from denser measurement grid $\{m_3\}$

Table 3. Uniform Distribution Parameters for CPF_G and CPF_H[a]

	CPF_G			CPF_H		
Case	Minimum	Maximum	CV	Minimum	Maximum	CV
1	0.001	0.025	0.53	0.0012	0.002	0.14
2	0.005	0.025	0.38	0.0015	0.002	0.08
3	0.01	0.025	0.24	0.0017	0.002	0.05
4	0.015	0.025	0.14	0.0017	0.0019	0.03

[a]Units of $[(\text{kg-d})\ (\text{mg})^{-1}]^m$; see equation (7).

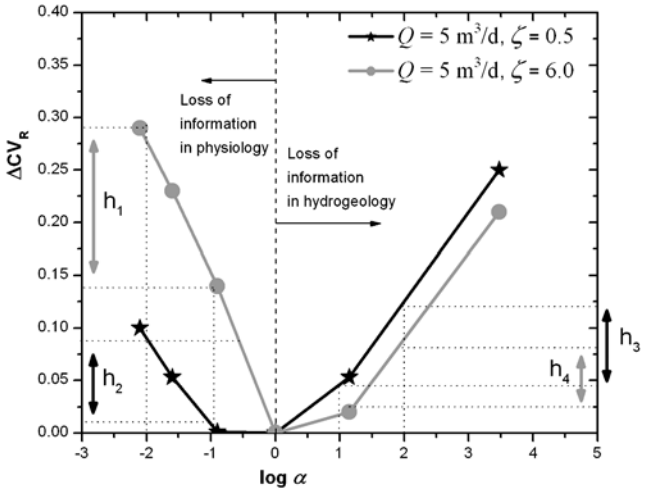

Figure 6. Illustration of the comparative information yield curves concept and the relative contribution of information for $Q = 5$ m^3 d^{-1} and $Pe \rightarrow \infty$ given source sizes $\zeta = 0.5$ and $\zeta = 6$. Here $h_1 > h_2$ and $h_3 > h_4$ for a fixed change in $\log\alpha$. Risk is evaluated with a linear model provided in equation (7) with $m = 1$.

given in Table 2 and $E_{P,O}$ is evaluated with the statistical distributions in case 4 from Table 3.

5. Results and Discussion

[39] In this section, results are presented on the basis of the data set given previously. We first address the interplay between the plume scale, capture zones, and pore-scale dispersion. The differences between using a screened well versus a control plane to evaluate the concentration in defining characterization needs within a risk driven approach is also addressed. Finally, we illustrate the how the value of information depends on the risk model used (i.e., linear versus nonlinear model). Our discussion and analysis are based on the concept of comparative information yield curves described in section 3.

5.1. On Plume Scale, Capture Zones, and Pore Scale

[40] Here, we investigate the dependence of risk uncertainty reduction on plume size by making use of the entropy concept defined in section 3. Our goal is to evaluate the risk cdf conditioned on the contaminant source size and measurements, $F_R(r|\zeta, \{m_i\})$ where ζ is the ratio between the source width (ℓ_S) and the heterogeneity correlation length (λ) (see Figure 1). Results are shown for a small ($\zeta = 0.5$) and large ($\zeta = 6$) contaminant source given measurements densities $\{m_1\}$, $\{m_2\}$ and $\{m_3\}$.

[41] The comparative information yield curves in Figure 6 shows that the effect of conditioning in reduction of risk uncertainty is much more beneficial for small source when conditioning on hydrogeological data. However, gaining information on the physiology side has much more effect in risk uncertainty reduction when the source is large ($\zeta = 6$).

[42] For a given change in α at points $\alpha > 1$, we observe that the corresponding change in ΔCV_R is greater for the smaller plume case ($\zeta = 0.5$) than for the larger plume ($\zeta = 6$). This means that hydrogeological data acquisition has a stronger impact on risk uncertainty for smaller plumes by comparing h_3 and h_4 shown in Figure 6 ($h_3 > h_4$). One can also fix a change in ΔCV_R and compare the slopes of the curves and the corresponding changes in $\log\alpha$ to the left and right of $\alpha = 1$. A similar effect was observed by *Maxwell et al.* [1999] by comparing conditional risk cdf's for a 3-D flow and transport test case. We will explore in more detail the physical mechanisms behind this result.

[43] This effect can be explained as follows: As the scale of the solute body increases, the plume approaches the ergodic state. This means that the plume's centroid becomes less affected by small-scale fluctuations captured by hydraulic conductivity measurements [*Rubin et al.*, 1999; *Rubin*, 2003]. On the contrary, for small contaminant sources ($\zeta < 1$), additional data contributes to reducing uncertainty about the location of the contaminant plume as well as the small-scale fluctuations of the streamlines. For example, a set of additional measurements may inform whether or not the contaminant plume will bypass the drinking water well.

[44] The opposite effect is noted for $\alpha < 1$. Here we observe that for a given change in α, the larger ΔCV_R corresponds to the larger plume ($h_1 > h_2$). For larger plumes, uncertainty reduction from the physiological side causes a larger uncertainty reduction in the human health risk cdf when compared to the smaller plume case. This is quite intuitive since there is no or little uncertainty whether a larger plume will reach the environmental target. The only uncertainty is how severe the impact would be on the exposed population. This depends more on the population's physiological characteristics than on flow and transport processes.

[45] Now, we wish to extend this result to illustrate its dependence on the scale of the capture zone induced by aquifer pumping (see Figure 7). Juxtaposition of Figures 6 and 7 shows that by increasing the pumping rate Q, the benefit of additional K sampling vanishes, regardless of source dimensions. The probability that the plume will reach the drinking well increases for larger Q, thus the additional data used to increase the accuracy of the plume's

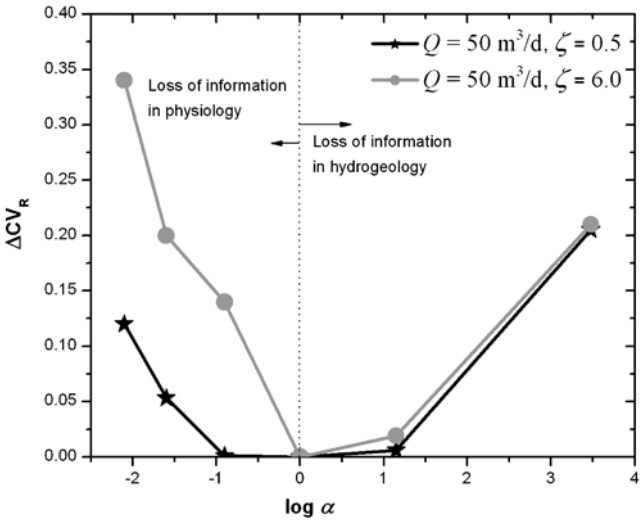

Figure 7. Illustration of the comparative information yield curves concept and the relative contribution of information for $Q = 50$ m^3 d^{-1} and $Pe \rightarrow \infty$ given source sizes $\zeta = 0.5$ and $\zeta = 6$. Risk is evaluated with a linear model provided in equation (7) with $m = 1$.

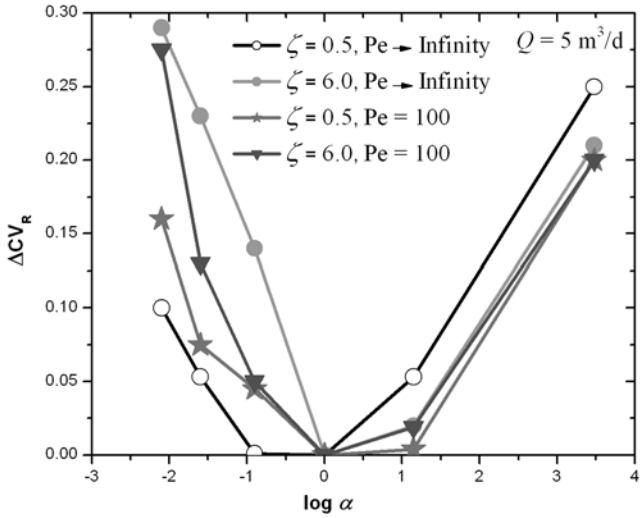

Figure 8. The influence of pore-scale dispersion in the analysis with $Pe = \lambda/\alpha_L$. Results were obtained for $Pe = 100$ and $Pe \rightarrow \infty$ given a fixed pumping rate $Q = 5\ m^3\ d^{-1}$. The longitudinal dispersivity is $\alpha_L = 1\ m^2$, and the transversal dispersivity is $\alpha_T = 0.1\ m^2$. Risk is evaluated with a linear model provided in equation (7) with $m = 1$.

location has a smaller impact and becomes less relevant. On the other hand, improved physiological characterization is more beneficial for the bigger plume than the smaller one (similar to the conclusions drawn from Figure 6) because in the absence of uncertainty on whether the plume will be captured by the well, the only impact on risk uncertainty reduction is from the physiological side.

[46] The effect of pore-scale dispersion on characterization needs is demonstrated in Figure 8. The Péclet number is defined as $Pe = \lambda/\alpha_L$ with α_L being the longitudinal dispersivity. The Péclet number is varied in this case through different α_L values. Figure 8 shows that at small Péclet numbers, the benefits of K sampling diminish independently of the plume's dimension. Two cases for comparison are shown: An infinite Péclet scenario and a finite Péclet scenario ($Pe = 100$). Larger pore-scale dispersion smooths out details captured by hydrogeological site characterization for both large and small plumes. The role of a finite Péclet number in heterogeneous flows is to increase the rate of concentration variance destruction thus smoothing out the concentration field [*Fiorotto and Caroni*, 2002; *Rubin*, 2003; *Caroni and Fiorotto*, 2005]. This is observed in Figure 8 for $\alpha > 1$. By removing pore-scale dispersion, the effect of plume size starts to play a role in defining characterization efforts as shown in Figure 8 for points $\alpha > 1$. For large Pe, the plume centroid is influenced more by heterogeneity and hydraulic data contributes to risk uncertainty reduction. Furthermore, if the plume is small and transport is dominated by advective processes, pore-scale effects as well as macrodispersion plays less of a role, thus increasing the importance of hydrogeological data acquisition. The information yield curve for this case ($\zeta = 0.5$ and $Pe \rightarrow \infty$) is represented in Figure 8 for points $\alpha > 1$.

[47] For $\alpha < 1$, we have the same results as shown in Figures 6 and 7. Note that, for finite Péclet, the curves corresponding to $\zeta = 0.5$ and $\zeta = 6$ are grouped closer compared to Figures 6 and 7. This is because dispersion tends to dilute the concentration field. On the contrary, by observing the slopes of the curves depicted in Figure 7, larger Péclet numbers imply larger variance in the concentration leading to higher probability for having larger concentration values. In the case of high Péclet, plume scale makes a large difference in determining whether or not physiological uncertainty is important. For instance, Figure 8 shows that the physiological side becomes more important to characterize for the $\zeta = 0.5$ and $Pe \rightarrow \infty$ information yield curve. Summarizing, elements that reduce hydrogeological uncertainty about the environmental target being hit (larger plume, larger dispersivity, etc.) increase the value of physiological characterization.

5.2. On the Significance of Concentration Averaging

[48] Evaluation of human health risk may yield different results depending how the concentration is calculated. Some analysis makes use of the concentration at one or more fixed points in space represented by a drinking well [*Maxwell et al.*, 1998; *Maxwell and Kastenberg*, 1999; *Maxwell et al.*, 1999; *Benekos et al.*, 2007], whereas other studies have used the total solute mass flux (Q_s) over a control plane [*Andricevic et al.*, 1994; *Andricevic and Cvetkovic*, 1996, 1998]. Dividing the total solute discharge (Q_s) by the fluid volumetric discharge over the control plane (Q_f) yields the flux-averaged concentration $C_f = Q_s/Q_f$ [*Kreft and Zuber*, 1978] (see discussion by *Rubin* [2003, p.163]).

[49] The differences in the definition of the concentration are elucidated in the information yield curves present in Figure 9. We obtained Figure 9 by making use of the flux-averaged concentration over the entire control plane (see Figure 1). When comparing with Figures 6 and 7, Figure 9 shows how the control plane approach dampens the effect of the differences in plume size and pumping. Note that for each pumping scenario, the curves for large and small plumes are closer together when compared to Figures 6

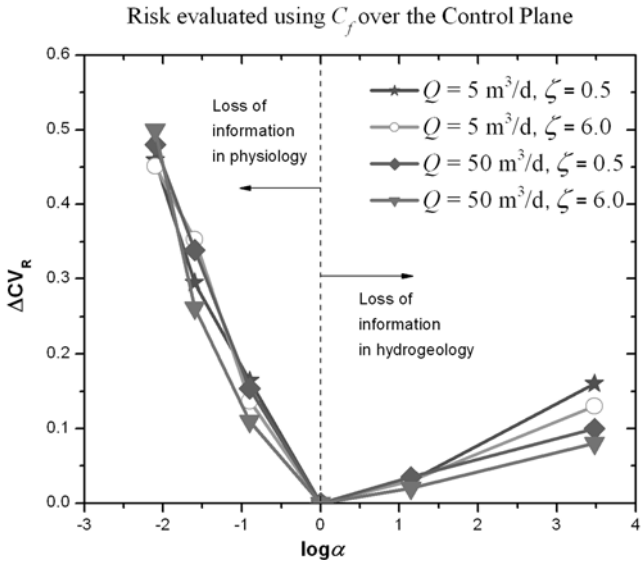

Figure 9. Measuring the relative contribution of information for pumping rates $Q = 5$ and $Q = 50\ m^3\ d^{-1}$ and source sizes $\zeta = 0.5$ and $\zeta = 6$. Results were evaluated using the flux-averaged concentration over the compliance plane. Risk is evaluated with a linear model provided in equation (7) with $m = 1$.

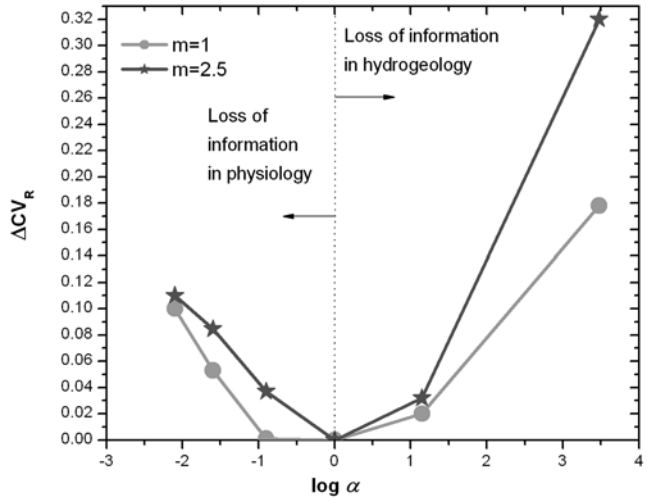

Figure 10. Sensitivity of human risk toward different dose response models. Results were evaluated with $\zeta = 0.5$ and $Q = 5\ m^3\ d^{-1}$.

and 7. This is more evident to the curves to the left of $\alpha = 1$ where the relative gain of information in the physiological component of risk is quantified. What this result illustrates is that parametric uncertainty reduction from the physiological component is less dependent on both the plume's dimension and the scale of the capture zone when human health risk is evaluated with the solute discharge over a control plane. One possible explanation is that when evaluating C_f over the control plane, the total mass of the solute present in that particular slice of the domain is being captured independently of its spatial distribution. Even if the peak of Q_s (say, above a certain regulatory threshold value) occurs along a streamline that bypasses the well, the presence of the chemical (and it peak value) will still be lumped into C_f since the averaging process is over the entire control plane. This averaging procedure over the control plane also leads to smaller differences observed in the curves to the left of $\alpha = 1$ when compared to Figures 6–8 since the breakthrough curves for C_f are smoother. From a regulator's point of view, an erroneous interpretation of the concentration term in risk may lead to unnecessary cleanup costs. For example, the control plane approach may account for the contaminant mass along a streamline that bypasses the drinking water well leading to remediation costs. Still, from the information yield curves present in Figure 9, it is possible to observe that human health risk uncertainty reduction benefits more from physiological characterization.

[50] As for the effects of additional measurements of K, the extra dilution added by averaging Q_s by Q_f smoothes out local details captured by characterization. In other words, the control plane approach adds an enhanced diffusive mechanism that removes some of the conditional effect gathered through site characterization and may mislead decision makers. However, the control plane approach can be very helpful if regulations are based on travel times as shown by *Andricevic et al.* [1994] and *Andricevic and Cvetkovic* [1996].

5.3. Effect of Alternative Risk Models

[51] In the present subsection we illustrate the sensitivity of human health risk toward different dose response models (see equation (7) and Figure 5). As explained previously, the main uncertainty in risk models is at the low dose (or low concentrations) [*EPA*, 2005; *Chiu et al.*, 2007; *Fjeld et al.*, 2007]. For instance, PCE is known to cause cell leukemia and kidney tumors in rodents; however, the shape of its dose response in humans is uncertain [*EPA*, 1998]. Here, we wish to point out how different dose response models can lead to different characterization needs.

[52] For illustration purposes, in the next results we will use a linear model ($m = 1$ in equation (4)) and a nonlinear model ($m = 2.5$ in equation (7)). The linear model assumes zero risk only at zero concentration and is normally considered conservative [*EPA*, 1989, 2001, 2005]. However, in recent years, the scientific community as well as environmental regulations acknowledges that the use of a nonlinear model maybe more adequate depending on the amount of available data used to construct the dose response model [*EPA*, 2005; *Chiu et al.*, 2007; *Fjeld et al.*, 2007]. The applicability of these nonlinear models may be expanded to both cancer and noncancer risks [*EPA*, 2005]. Now we illustrate how different risk models would possibly manifest in information yield curves.

[53] In Figure 10, we compare different risk models, their sensitivity to hydrogeological data acquisition and consequently parametric uncertainty reduction. Figure 10 shows how hydrogeological sampling has a stronger implication in risk uncertainty reduction for the nonlinear model than for the linear model. This result indicates that when using a nonlinear model to evaluate risk, the data worth of sampling hydraulic conductivity increases toward risk uncertainty reduction. Thus characterizing the behavior of the flow field becomes more important for this class of models. Although this result is shown only for carcinogenic risk, it may also have implications for some noncarcinogenic compounds with threshold doses where an adverse effect is observed. In such cases, the worth of hydrogeological information might increase given that better understanding of the flow patterns lead to better estimation of the concentration (or dose) values since the knowledge of being above or below such threshold values becomes very important.

5.4. On the Definition of $E_{H,O}$ and $E_{P,O}$

[54] As presented previously in section 3, an alternative way to investigate uncertainty trade-offs is to change the definition of $E_{H,O}$ and $E_{P,O}$ such that we have $E_H \leq E_{H,O}$ and $E_P \leq E_{P,O}$. This implies that $E_{H,O}$ and $E_{P,O}$ corresponds to the most uncertain case (see Figure 4). These entropies are now evaluated with the most uncertain distributions, corresponding to the small amount of information available a priori. In this new definition, $E_{H,O}$ and $E_{P,O}$ are now the starting points of uncertainty reduction. This avoids the need to prespecify $E_{H,O}$ and $E_{P,O}$ values as defined in section 3 thus allowing more flexibility.

[55] Figure 11 depicts how decision makers could investigate risk uncertainty reduction strategies by plotting both the coefficient of variation of risk (CV_R) versus ΔE_H and ΔE_P (equivalent to $\log\alpha$). We have used the data given in Tables 1–3 to obtain plots similar to the diagram in Figure 4. As a starting point, risk is evaluated with $E_{H,O}$ and $E_{P,O}$ (initial information available, most uncertain case) as well as its corresponding $CV_R = CV_R^O$. By collecting additional data, one may perform conditional simulations from both the hydrogeological and physiological side (see Figures 11a

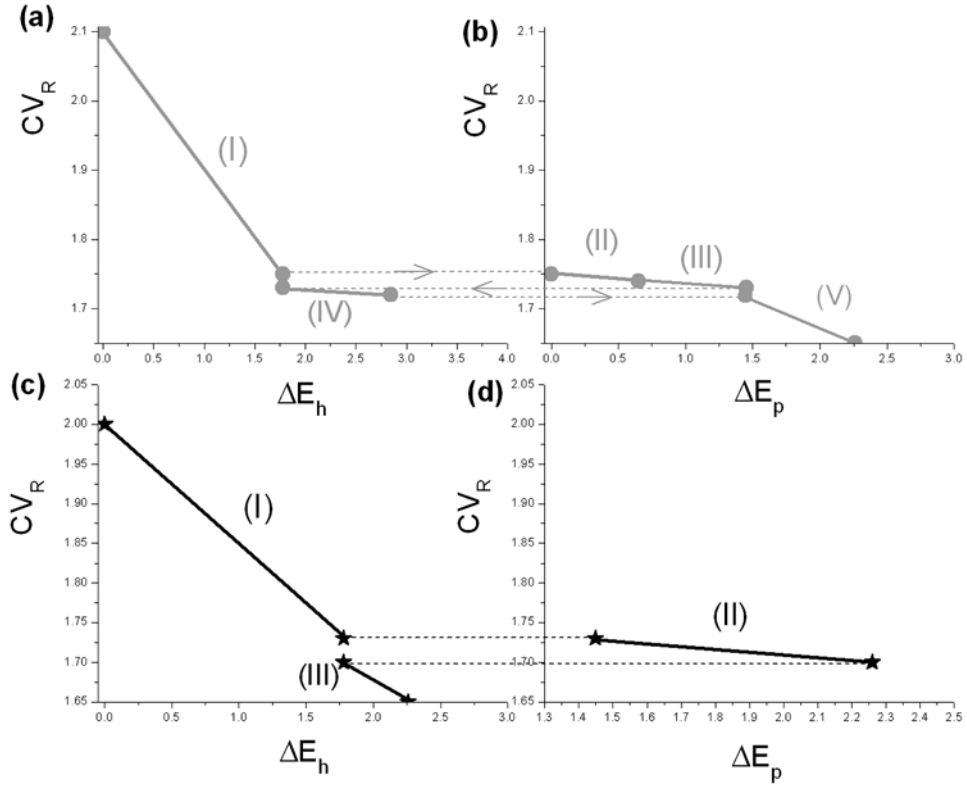

Figure 11. Illustration of the yield information curves within the stepwise approach given in section 3 comparing two sampling strategies using a linear risk model (m = 1). (a and b) Strategy using five uncertainty reduction segments. (c and d) Strategy using three uncertainty reduction segments.

and 11b) and evaluate a new CV_R. For instance, from the starting point CV_R^O, uncertainty reduction in risk can be done in a five-step procedure by collecting K data (Figure 11a, segment I), then through physiology or behavioral parameters (Figure 11b, segments II and III), then more K measurements (Figure 11a segments IV) and finally health variables again (Figure 11b, segment V). The stopping criterion is when CV_R meets the regulatory standards. Thus the necessity of additional sampling is risk driven and can be decided upon on the basis of an acceptable risk value (say, the 95th percentile confidence level or a coefficient of variation) that is in agreement with probabilistic risk assessment guidelines [*EPA*, 2001].

[56] However, this graphical approach can be useful to compare different characterization strategies by making use of the estimation procedure given in section 3 to evaluate entropies for a priori unknown data. Figures 11a and 11b showed a five-step procedure described in the previous paragraph. By summing up all the ΔE_H and ΔE_P needed to reduce CV_R from 2.1 to 1.65, one may come with an estimate of the sampling efforts. Yet, a different strategy, three-step procedure, is given in Figures 11c and 11d which can yield a different summed entropy value when compared to the one given by the five-step procedure (see Figures 11c and 11d, segments I, II and III). By associating the risk uncertainty reduction with the corresponding total change in entropy (ΔE_H and ΔE_P) one may opt for the cheapest strategy to reach a compliance goal set up by environmental agencies. For example, the costs in hydrogeology could be associated with slug tests and sampling (laboratory experiments) while in physiology and health-related parameters acquisition costs can be associated with number of animals used in toxicological studies or a more detailed survey of the behavioral characteristics of the exposed population.

[57] Next, we show how sampling efforts can differ when using the same data acquisition strategy but different risk models. Figures 12a and 12b were evaluated using a linear risk model while Figures 12c and 12d uses a nonlinear model. One can see that the slopes of the curves in Figures 12a and 12b are different than Figures 12c and 12d for each corresponding segment. Also, Figures 12a and 12c depicts how the total change in ΔE_H, represented by summing δh_1 and δh_2 corresponds to two distinct changes in CV_R represented by Δ_H. The total change in ΔE_P ($\delta p_1 + \delta p_2 + \delta p_3$) and the associated total change in CV_R, denoted by Δ_P, is highlighted in Figures 12b and 12d.

[58] In summary, the usefulness of the stepwise approach given in Figures 11 and 12 is that it allows one to see how different uncertainty reduction strategies could lead to different costs (associated with data) by avoiding the necessity of evaluating speculative values for both $E_{H,O}$ and $E_{P,O}$ required for equations (3) and (4) in section 4.

6. Summary and Conclusions

[59] In this work, we discussed the theoretical and practical aspects of the information yield curves within a risk-driven context. The relevance of transport and flow scales in defining characterization needs in human health risk is addressed. Through numerical experimentation, conditions are identified where hydrogeological site characterization,

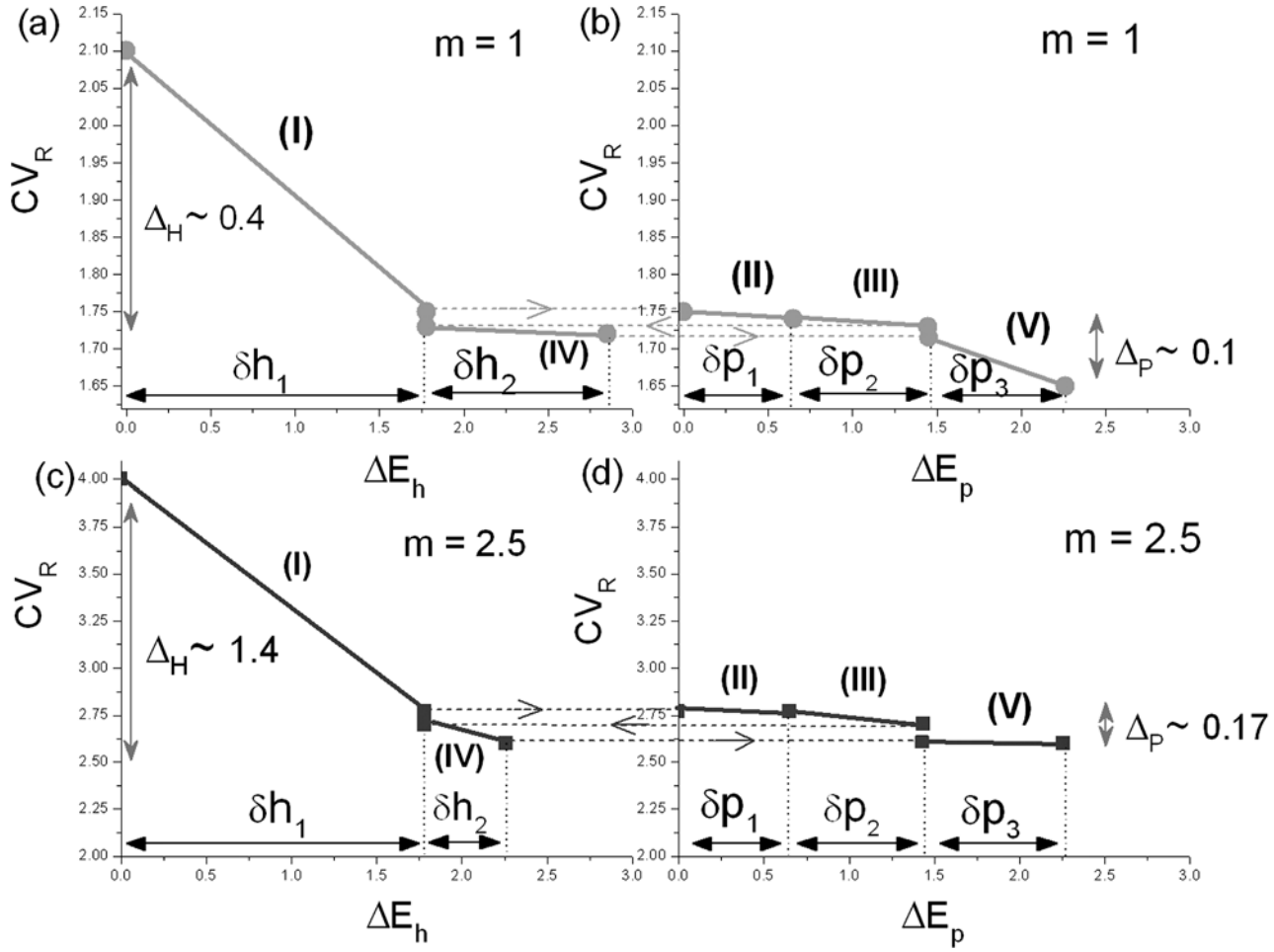

Figure 12. Illustration of the yield information curves within the stepwise approach given in section 3 comparing the same sampling strategy given two risk models (m = 1 and m = 2.5). (a and b) Strategy using m = 1 (linear risk model). (c and d) The same strategy but using m = 2.5 (nonlinear model).

through measurements of hydraulic conductivity, has a stronger impact in uncertainty reduction in risk as well as conditions in which physiological uncertainty reduction has a significant impact. In order to achieve this, we investigated the interplay between plume dimension, capture zones induced by pumping action, Péclet number and sampling scales for different conditional simulations. We have quantified the relative gain of information through uncertainty reduction from both physiology and flow physics for a fixed, although not limited to, fractile of human variability. Results were analyzed for the low-dose risk curves. On the basis of our simulations, physical configuration (2-D groundwater flow and transport) and risk pathways, we highlight the following points.

[60] 1. The role of the plume's dimension proved important in defining characterization needs in the risk-driven context. Results show that uncertainty reduction in human health risk benefits more from hydrogeological site characterization if the contaminant source is small relative to the heterogeneity correlation length. The human health risk cdf is less sensitive to measurements of hydraulic conductivity if the contaminant source is large.

[61] 2. The value of information not only depends on plume's dimension but also on its interplay with the pumping rate related to the scale of capture zone. For high pumping rates, thus larger capture zones, the value of information from the hydrogeological component becomes less dependent on the plume's dimension. The opposite occurs as the pumping rate decreases and the plume's dimension begins to gain a role in defining hydrogeological sampling needs.

[62] 3. Results indicate that uncertainty reduction in risk may benefit more from parametric uncertainty reduction from the physiological component as opposed to hydrogeological if the plume's dimension approaches ergodicity.

[63] 4. The significance of plume dimension in defining hydrogeological characterization needs is also dependent on the phenomenon occurring at pore scale. For high Péclet conditions, plume size relative to the heterogeneity scale has a role in defining characterization efforts. When pore-scale dispersion effects are increased (lower Péclet), the knowledge of whether the plume is large or small becomes less relevant in defining hydrogeological characterization strategies.

[64] 5. Similar conclusion was obtained when comparing concentration measured in a well versus the flux-averaged concentration at a control plane. The differences between concentration pumped by a well and concentration at a control plane is highlighted and can also lead to significant different characterization needs from both physiological and hydrogeological perspective.

[65] 6. We also showed how different risk models have different effects in risk uncertainty reduction and defining characterization needs.

[66] For this work, we have made extensive use of information entropy to investigate uncertainty trade-offs in a graphical manner. We denote these entropy plots as information yield curves. This is a useful concept since it allows one to easily view the relative contribution of information in risk from the physiological and the hydrogeological component. An important difference regarding the use of these entropy plots as opposed to cdf is that one can assign changes in entropy for both physiology and hydrogeology for a fixed uncertainty reduction in human health risk. The challenge in this approach is to assign estimated financial values to these α values. This way, decision makers can verify which uncertainty reduction campaign is cheaper for a given uncertainty reduction in risk estimates. Translating α values into financial terms allows one to cast the analysis in a cost-benefit framework as studied by *Massmann and Freeze* [1987] and *Freeze et al.* [1990]. Section 3 provided a discussion of Information Yield Curves and how to make this concept practical in real site characterization problems. For our simulations, the mean value of risk did not vary significantly from each conditioning case and all were found to be within the same order of magnitude. However, if the mean value of risk varies significantly, other measures of uncertainty besides ΔCV_R might be more informative (for instance: relative entropy or 95th confidential interval). In this study, the uncertainty in the hydrogeological component was within SRF parameters. In this study we did not account for uncertainty in the chemical reactions, although this is not a limitation in the framework or the numerical implementation. These parameters can be accounted in $\boldsymbol{\theta}_H$. Another option would be to generalize the metric α to other dimensions to account for specific subgroups of parameters. For example: a subgroup for chemical parameters, another for the SRF parameters and finally for source characteristics.

[67] It is important to note that the framework and results presented here can be extended to different types of data used for conditioning. In addition, other sources of uncertainty can be incorporated into the framework. We have used a two-dimensional model to answer the research questions addressed in the introduction. Reproducing these numerical simulations and problem configuration in a three-dimensional physical model may enhance the results obtained. For an instance, *Maxwell et al.* [1999] reported that finer sampling over the vertical direction is necessary and relevant to predict the expected plume path. Our approach can be extended to account for variability within the exposed population. Here we calculated the comparative information yield curves for a single fractile in population variability (see section 2 for discussions on $\boldsymbol{\Theta}_P$ and $\boldsymbol{\theta}_{P,i}$); however, one may obtain information yield curves for different fractiles. Given this, 3-D surfaces of information yield could be evaluated. As for the pore-scale dispersion analysis, the slopes of the information yield curves are affected by both longitudinal and transversal Péclet numbers. For instance, increasing the transversal dispersion coefficient, more mass will be transferred between streamlines, thus smoothing the concentration field. This effect is reflected in the information yield curves. Nevertheless, one of the novelties of the present work is the illustration of the importance of considering flow and transport scales when defining characterization needs toward better resource allocation within a risk-driven approach. Most importantly, the current paper introduced the theoretical and practical aspects of the information yield curves in human health risk assessment. This approach allows one to investigate uncertainty trade-offs from the health-related parameters and physical parameters.

Appendix A

[68] Given a set $\boldsymbol{Y}$ of $Y_i = \ln K_i$ measurements from a random field generator we are able to estimate a pdf for the uncertain parameter. From this sample, we obtain the SRF parameters, for example, $\boldsymbol{\theta}_H = \{m_Y, \sigma_Y^2, \lambda\}$ where m_Y and σ_Y^2 are the mean and variance of Y and λ is its correlation length. We need to infer the distribution of $\boldsymbol{\theta}_H$ given the measurements in $\boldsymbol{Y}$, $\hat{f}_H(\boldsymbol{\theta}_H|\boldsymbol{Y})$. The procedure is based on Bayes' theorem:

$$\hat{f}_H(\boldsymbol{\theta}_H|\boldsymbol{Y}) = \frac{f_{prior}(\boldsymbol{\theta}_H) f_Y(\boldsymbol{Y}|\boldsymbol{\theta}_H)}{f_Y(\boldsymbol{Y})}, \tag{A1}$$

where the assumption of a prior pdf, $f_{prior}(\boldsymbol{\theta}_H)$, is needed. Assuming that the pdf $f_Y(\boldsymbol{Y}|\boldsymbol{\theta})$ is multivariate Gaussian, we have

$$f_Y(\boldsymbol{Y}|\boldsymbol{\theta}_H) = \frac{1}{(2\pi)^{N/2}\|\boldsymbol{C}_{YY}\|} \mathrm{Exp}\left[-\frac{1}{2}[\boldsymbol{Y}-\boldsymbol{m}_Y]\boldsymbol{C}_{YY}^{-1}[\boldsymbol{Y}-\boldsymbol{m}_Y]\right], \tag{A2}$$

with $\boldsymbol{C}_{YY}$ is the geostatistical correlation model that depends on $\boldsymbol{\theta}_H = \{m_Y, \sigma_Y^2, \lambda\}$ and $\|\boldsymbol{C}_{YY}\| \equiv \mathbf{Det}(\boldsymbol{C}_{YY})$. With equation (A2), the estimated pdf in equation (A1) can be obtained.

Appendix B

[69] A two-dimensional depth-averaged, saturated, steady state flow is considered. The flow domain is considered bounded and defined by the aquifer's longitudinal length L and width W. The equation that governs flow is given as follows:

$$\nabla \cdot [bK(\boldsymbol{x})\nabla\varphi] = \sum_w Q_w \delta(\boldsymbol{x}-\boldsymbol{x}_w), \tag{B1}$$

where b is the average depth of the aquifer, φ the pressure head, Q_w is the pumping rate of the wth pump well at location $\boldsymbol{x}_w$. We consider no-flow boundary conditions on the transversal direction (x_2) and prescribed pressure head in the longitudinal direction (x_1). Flow occurs from the left to right. Assuming instantaneous linear chemical interactions with the soil particles and that the chemical, with initial

concentration C_o, is instantaneously released within the aquifer along a line source, we write

$$R_f \frac{\partial C}{\partial t} + \boldsymbol{v}\nabla C - \nabla \cdot [\boldsymbol{D}(\boldsymbol{x})\nabla C] = \sum_w \frac{C_w Q_w}{n}\delta(\boldsymbol{x} - \boldsymbol{x}_w), \quad \text{(B2)}$$

with n being the effective porosity and $\boldsymbol{v}$ is the Eulerian velocity vector obtained through Darcy's law and equation (2), R_f is the retardation factor, $\boldsymbol{D}$ is the dispersion coefficient tensor, C is the concentration and finally C_w is the concentration at the pumping well.

[70] ParFlow was used to solve the flow field in the aquifer [*Ashby and Falgout*, 1996; *Jones and Woodward*, 2001; *Kollet and Maxwell*, 2006]. ParFlow is a watershed flow code that uses a multigrid preconditioned conjugate gradient algorithm to efficiently solve the linear system resulting from the discretization of the flow equation.

[71] The contaminant transport is solved using a Lagrangian particle tracking algorithm with very minimal numerical dispersion and conservation of mass [*Maxwell and Kastenberg*, 1999; *Maxwell and Tompson*, 2006; *Maxwell et al.*, 2007]. This code, SLIM-FAST, simulates migration of dissolved, neutrally buoyant and reactive chemical in saturated porous media. To represent concentration and the spatial/temporal distribution of the contaminant, an explicit Lagrangian random walk particle method is implemented in the code. SLIM-FAST also benefits from the quasianalytical formulation presented by *Schafer-Perini and Wilson* [1991].

Appendix C

[72] In the case where N^* measurements of $Y = \ln K$ are available, the negative log likelihood function for a multivariate normal pdf becomes [*Rubin*, 2003]

$$-\ln L(\boldsymbol{\theta}_H | Y_i) = \frac{N^*}{2}\ln(2\pi) + \frac{1}{2}\ln\|\boldsymbol{C}_{YY}\| + \frac{1}{2}\sum_{i=1}^{N^*}\sum_{j=1}^{N^*}\frac{(Y_i - \langle Y_i\rangle)(Y_j - \langle Y_j\rangle)}{C_Y(\boldsymbol{x}_i, \boldsymbol{x}_j)}, \quad \text{(C1)}$$

where $\|\boldsymbol{C}_{YY}\|$ is the determinant of the variance-covariance matrix of order N^* by N^*. $C_Y(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is the spatial covariance model. For our results, we used the case of an exponential isotropic $C_Y(\boldsymbol{x}_i, \boldsymbol{x}_j)$ such that $\boldsymbol{\theta}_H = (m_Y, \sigma_Y^2, \lambda)$. The maximum likelihood estimators are those that minimize equation (C1).

Notation

b	depth of the aquifer [L].
AT	average lifetime.
AC_S	indoor air concentration.
C, C_w	resident concentration and concentration at the well [$M\,L^{-3}$].
C_Y	covariance model.
CPF_H, CPF_G	cancer potency factor [(kg d) $(\text{mg})^{-1}]^m$, where m is the value that determines the nonlinearity in the risk model (see equation (7)).
CV_R^o	coefficient of variation for risk evaluated with base entropies [−].
CV_R	coefficient of variation for risk [−].
$LADD_G$, $LADD_H$	intake dose [mg (kg d)$^{-1}$].
$\boldsymbol{D}$	mechanical dispersion tensor [$L^2\,T^{-1}$].
E_H, E_P, $E_{P,O}$ and $E_{H,O}$	entropies and base entropies [−].
ED	exposure duration.
EF	exposure frequency.
ET_S	shower exposure time.
f_H, f_P	probability density functions for the hydrogeological and health parameters.
$\hat{f}$	estimate of a pdf.
h_1, h_2, h_3, h_4	markings on Figure 6.
HR/BW	inhalation rate per body weight.
K	hydraulic conductivity [$L\,T^{-1}$].
IR/BW	ingestion rate per body weight.
ℓ_S	dimension of the contaminant cloud in the x_2 direction [L].
$\{m_i\}$	sampling grid.
Q	pumping well [$L^3\,T^{-1}$].
Q_f	volumetric fluid discharge [$L^3\,T^{-1}$].
Q_s	total solute discharge [$M\,T^{-1}$].
n	porosity [−].
N^*	number of measurements [−].
r	increased lifetime cancer risk [−].
R_f	retardation factor [−].
TE_S	transfer efficiency from tap water to air.
VR_S	air exchange rate.
$\boldsymbol{v}$	Eulerian velocity field [$L\,T^{-1}$].
$Var[\sigma_Y^2]$	variance of σ_Y^2.
$\boldsymbol{x}$	Euclidean space vector [L].
$\boldsymbol{x_w}$	location of the wth pumping well [L].
Y	logarithm of K [−].
W_S	tap water use rate.
ΔE_H, ΔE_P, ΔCV_R	entropy differences; change in the risk coefficient of variation [−].
δh_i, δp_i	markings in Figure 11.
α; α_L, α_T	metric used to relate entropies; dispersivities in the x_1 and x_2 direction [L^2].
ξ	lag distance.
φ	pressure head [L].
λ	correlation length [L].
μ_R	mean value of risk.
σ_R	standard deviation for risk.
σ_Y^2	variance of Y [−].
$\boldsymbol{\theta}_H$, $\boldsymbol{\theta}_P$	vector for hydrogeological and physiological parameters.
ζ	ratio between the size of the contaminant source and the characteristic length of the subsurface [−].

[73] **Acknowledgments.** The first author gratefully acknowledges the financial provided by the CAPES Institution (Coordenação de Aperfeiçoamento de Pessoal e Nível Superior) from Brazil and for a helpful technical discussion with Ram Rajagopal and Wolfgang Nowak. We also wish to express our acknowledgments to the three anonymous reviewers who contributed to the clarity of the work. This research was supported by the U.S. DOE Office of Biological and Environmental Research, Environ-

mental Remediation Science Program (ERSP), through DOE-ERSP grant DE-FG02-06ER06-16 as part of the Hanford 300 Area Integrated Field Research Challenge Project.

References

Andricevic, R., and V. Cvetkovic (1996), Evaluation of risk from contaminants migrating by groundwater, *Water Resour. Res.*, *32*(3), 611–621, doi:10.1029/95WR03530.

Andricevic, R., and V. Cvetkovic (1998), Relative dispersion for solute flux in aquifers, *J. Fluid Mech.*, *361*, 145–174, doi:10.1017/S0022112098008751.

Andricevic, R., J. Daniels, and R. Jacobson (1994), Radionuclide migration using travel time transport approach and its application in risk analysis, *J. Hydrol.*, *163*, 125–145, doi:10.1016/0022-1694(94)90026-4.

Ashby, S. F., and R. D. Falgout (1996), A parallel multigrid preconditioned conjugate gradient algorithm for groundwater flow simulations, *Nucl. Sci. Eng.*, *124*, 145–159.

Beckie, R. (1996), Measurement scale, network sampling scale, and groundwater model parameters, *Water Resour. Res.*, *32*(1), 65–76, doi:10.1029/95WR02921.

Benekos, I. D., C. A. Shoemaker, and J. R. Stedinger (2007), Probabilistic risk and uncertainty analysis for bioremediation of four chlorinated ethenes in groundwater, *Stochastic Environ. Res. Risk Assess.*, *21*, 375–390, doi:10.1007/s00477-006-0071-4.

Binkowitz, B. S., and D. Wartenberg (2001), Disparity in quantitative risk assessment: A review of input distributions, *Risk Anal.*, *21*(1), 75–90, doi:10.1111/0272-4332.211091.

Bogen, K. T., and R. C. Spear (1987), Integrating uncertainty and interindividual variability in environmental risk assessment, *Risk Anal.*, *7*(4), 427–436, doi:10.1111/j.1539-6924.1987.tb00480.x.

Burmaster, D. E., and A. M. Wilson (1996), An introduction to second-order random variables in human health risk assessments, *Hum. Ecol. Risk Assess.*, *2*(4), 892–919.

Caroni, E., and V. Fiorotto (2005), Analysis of concentration as sampled in natural aquifers, *Transp. Porous Media*, *59*, 19–45, doi:10.1007/s11242-004-1119-x.

Chiu, W. A., C. Chen, K. Hogan, J. C. Lipscomb, C. S. Scott, and R. Subramaniam (2007), High-to-low dose extrapolation: Issues and approaches, *Hum. Ecol. Risk Assess.*, *13*, 46–51, doi:10.1080/10807030601107544.

Christakos, G. (1992), *Random Field Models in Earth Sciences*, Academic, San Diego, Calif.

Cushey, M., and Y. Rubin (1997), Field scale transport of nonpolar organic solutes in 3-D heterogeneous aquifers, *Environ. Sci. Technol.*, *31*(5), 1259–1268, doi:10.1021/es9601999.

Dagan, G. (1984), Solute transport in heterogeneous porous formations, *J. Fluid Mech.*, *145*, 151–177, doi:10.1017/S0022112084002858.

Dagan, G. (1987), Theory of solute transport by groundwater, *Annu. Rev. Fluid Mech.*, *19*, 183–215, doi:10.1146/annurev.fl.19.010187.001151.

Daniels, J. I., K. T. Bogen, and L. C. Hall (2000), Analysis of uncertainty and variability in exposure to characterize risk: Case study involving trichloroethylene groundwater contamination at Beale Air Force Base in California, *Water Air Soil Pollut.*, *123*, 273–298, doi:10.1023/A:1005231732471.

de Barros, F. P. J., and Y. Rubin (2008), A risk-driven approach for subsurface site characterization, *Water Resour. Res.*, *44*, W01414, doi:10.1029/2007WR006081.

Environmental Protection Agency (EPA) (1989), Risk assessment guidance for Superfund volume 1: Human health manual (Part A), *Rep. EPA/540/1-89/002*, Washington, D. C.

Environmental Protection Agency (EPA) (1998), Appendix D: Dose response assessment, in *Cleaner Technology Substitutes Assessment: Professional Fabricare Processes*, *Rep. EPA 744-B-98-001*, pp. D1–D2, Washington, D. C.

Environmental Protection Agency (EPA) (2001), Risk assessment guidance for Superfund volume III—Part A, process for conducting probabilistic risk assessment, *Rep. EPA/540-R-02-002*, Off. of Emerg. and Remedial Response, Washington, D. C.

Environmental Protection Agency (EPA) (2005), Guidelines for carcinogen risk assessment, *Rep. EPA/630/P-03/001F*, Washington, D. C.

Fiorotto, V., and E. Caroni (2002), Solute concentration statistics in heterogeneous aquifers for finite Péclet values, *Transp. Porous Media*, *48*, 331–351, doi:10.1023/A:1015744421033.

Fjeld, R. A., N. A. Eisenberg, and K. L. Compton (2007), *Quantitative Environmental Risk Analysis for Human Health*, 1st ed., John Wiley, Hoboken, N. J.

Freeze, A. R., J. Massmann, L. Smith, T. Sperling, and B. James (1990), Hydrogeological decision analysis: 1. A framework, *Ground Water*, *28*(5), 738–766, doi:10.1111/j.1745-6584.1990.tb01989.x.

Hassan, A. E., R. Andricevic, and V. Cvetkovic (2001), Computational issues in the determination of solute discharge moments and implications for comparison to analytical solutions, *Adv. Water Resour.*, *24*, 607–619, doi:10.1016/S0309-1708(00)00071-3.

Hoeting, J., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999), Bayesian model averaging: A tutorial, *Stat. Sci.*, *14*(4), 382–417, doi:10.1214/ss/1009212519.

Hubbard, S., and Y. Rubin (2000), Hydrogeological parameter estimation using geophysical data: A review of selected techniques, *J. Contam. Hydrol.*, *45*(1–2), 3–34, doi:10.1016/S0169-7722(00)00117-0.

James, B. R., and S. M. Gorelick (1994), When enough is enough: The worth of monitoring data in aquifer remediation design, *Water Resour. Res.*, *30*(12), 3499–3513, doi:10.1029/94WR01972.

Jones, J. E., and C. S. Woodward (2001), Newton-Krylov-multigrid solvers for large-scale, highly heterogeneous, variably saturated flow problems, *Adv. Water Resour.*, *24*, 763–774, doi:10.1016/S0309-1708(00)00075-0.

Kollet, S. J., and R. M. Maxwell (2006), Integrated surface-groundwater flow modeling: A free-surface overland flow boundary condition in a parallel groundwater flow model, *Adv. Water Resour.*, *29*(7), 945–958, doi:10.1016/j.advwatres.2005.08.006.

Kowalsky, M. B., S. Finsterle, J. Peterson, S. Hubbard, Y. Rubin, E. Majer, A. Ward, and G. Gee (2005), Estimation of field-scale soil hydraulic and dielectric parameters through joint inversion of GPR and hydrological data, *Water Resour. Res.*, *41*, W11425, doi:10.1029/2005WR004237.

Kreft, A., and A. Zuber (1978), On the physical meaning of the dispersion equation and its solution for different and initial boundary conditions, *Chem. Eng. Sci.*, *33*, 1471–1480, doi:10.1016/0009-2509(78)85196-3.

Lawrence, A., and Y. Rubin (2007), Block-effective macrodispersion for numerical simulations of sorbing solute transport in heterogeneous porous formations, *Adv. Water Resour.*, *30*, 1272–1285, doi:10.1016/j.advwatres.2006.11.005.

Massmann, J., and R.A. Freeze (1987), Groundwater contamination from waste management sites: The interaction between risk-based engineering design and regulatory policy 1, *Water Resour. Res.*, *23*(2), 351–367, doi:10.1029/WR023i002p00351.

Maxwell, R. M., and W. E. Kastenberg (1999), Stochastic environmental risk analysis: An integrated methodology for predicting cancer risk from contaminated groundwater, *Stochastic Environ. Res. Risk Assess.*, *13*, 27–47, doi:10.1007/s004770050030.

Maxwell, R. M., and A. F. B. Tompson (2006), SLIM-FAST: A user's manual, *Rep. UCRL-SM-225092*, Lawrence Livermore Natl. Lab., Livermore, Calif.

Maxwell, R. M., S. D. Pelmulder, F. B. Tompson, and W. E. Kastenberg (1998), On the development of a new methodology for groundwater-driven health risk assessment, *Water Resour. Res.*, *34*(4), 833–847, doi:10.1029/97WR03605.

Maxwell, R. M., W. E. Kastenberg, and Y. Rubin (1999), A methodology to integrate site characterization information into groundwater-driven health risk assessment, *Water Resour. Res.*, *35*(9), 2841–2885, doi:10.1029/1999WR900103.

Maxwell, R. M., C. Welty, and R. W. Harvey (2007), Revisiting the Cape Cod bacteria injection experiment using a stochastic modeling approach, *Environ. Sci. Technol.*, *41*(15), 5548–5558, doi:10.1021/es062693a.

Maxwell, R. M., S. F. Carle, and A. F. B. Tompson (2008), Contamination, risk, and heterogeneity: On the effectiveness of aquifer remediation, *Environ. Geol.*, *54*, 1771–1786, doi:10.1007/s00254-007-0955-8.

McKone, T. E., and T. B. Bogen (1991), Predicting the uncertainties in risk assessment, *Environ. Sci. Technol.*, *25*(10), 1674–1681, doi:10.1021/es00022a002.

Neuman, S. (2003), Maximum likelihood Bayesian averaging of uncertain model predictions, *Stochastic Environ. Res. Risk Assess.*, *17*, 291–305, doi:10.1007/s00477-003-0151-7.

Portier, L., K. Tolson, and S. M. Robert (2007), Body weight distributions for risk assessment, *Risk Anal.*, *27*(1), 11–26, doi:10.1111/j.1539-6924.2006.00856.x.

Rubin, Y. (2003), *Applied Stochastic Hydrogeology*, 1st ed., Oxford Univ. Press, Oxford, U. K.

Rubin, Y., and G. Dagan (1992), Conditional estimates of solute travel time in heterogeneous formations: Impact of transmissivity measurements, *Water Resour. Res.*, *28*(4), 1033–1040, doi:10.1029/91WR02759.

Rubin, Y., A. Bellin, and M. Cushey (1994), Modeling of transport in ground water for environmental risk assessment, *Stochastic Hydrol. Hydraul.*, *8*(1), 57–78, doi:10.1007/BF01581390.

Rubin, Y., A. Sun, R. Maxwell, and A. Bellin (1999), The concept of block-effective macrodispersivity and a unified approach for grid-scale- and plume-scale-dependent transport, *J. Fluid Mech.*, *395*, 161–180, doi:10.1017/S0022112099005868.

Sanchez-Vila, X., and Y. Rubin (2003), Travel time moments for sorbing solutes in heterogeneous domains under nonuniform flow conditions, *Water Resour. Res.*, *39*(4), 1086, doi:10.1029/2002WR001399.

Schafer-Perini, A. L., and J. L. Wilson (1991), Efficient and accurate front tracking for two-dimensional groundwater flow models, *Water Resour. Res.*, *27*(7), 1471–1485, doi:10.1029/91WR00720.

Tompson, A. F. B., R. Ababou, and L. W. Gelhar (1989), Implementation of the three-dimensional turning bands field generator, *Water Resour. Res.*, *23*(10), 2227–2243.

F. P. J. de Barros and Y. Rubin, Department of Civil and Environmental Engineering, University of California, 626 Davis Hall, Berkeley, CA 94720-1710, USA. (barros@berkeley.edu)

R. M. Maxwell, Department of Geology and Geologic Engineering, Colorado School of Mines, 1500 Illinois Street, Golden, CO 80401, USA.