Title

Hydrogeological Modeling and Water Resources Management: Improving the Link Between Site Characterization, Prediction, and Decision Making

Author

Harken, Bradley John

Publication Date

2017

Peer reviewed|Thesis/dissertation

Hydrogeological Modeling and Water Resources Management: Improving the Link Between Site Characterization, Prediction, and Decision Making

by

Bradley Harken

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Engineering - Civil and Environmental Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Yoram Rubin, Chair
Associate Professor Sally Thompson
Professor David Brillinger

Summer 2017

Hydrogeological Modeling and Water Resources Management: Improving the Link Between Site Characterization, Prediction, and Decision Making

Abstract

Hydrogeological Modeling and Water Resources Management: Improving the Link Between Site Characterization, Prediction, and Decision Making

by

Bradley Harken

Doctor of Philosophy in Engineering - Civil and Environmental Engineering

University of California, Berkeley

Professor Yoram Rubin, Chair

Groundwater resources are an important part of any water resources management strategy, but must be managed with care to ensure the availability of water in sufficient quantity and quality to meet the ever growing demand from industrial, agricultural, and municipal uses. Hydrogeological models have a great potential to support the successful management of groundwater resources by making predictions regarding water availability and contaminant migration. However, complications arise due to the uncertainty stemming from incomplete characterization and natural variability in subsurface hydraulic properties. The uncertainty relating to these properties and model selection propagates through the modeling process, producing uncertain hydrogeologic predictions. Many formal methods exist for coping with this uncertainty in a theoretically sound manner which could improve decision making, but widespread adoption of these methods has been slow or nonexistent. This dissertation explores potential reasons for this hindered adoption of these methods in hydrogeological practice, despite success of similar methods in other industries. Issues examined include practicability of stochastic methods, the role of regulations, and the translation of uncertainty to risk at the knowledge-decision interface.

This dissertation presents a framework which addresses these challenges. The framework utilizes statistical hypothesis testing and an integrated approach to the planning of hydrogeological site characterization, modeling prediction, and water resources decision making. Benefits of this framework include aggregated uncertainty quantification and risk evaluation, simplified communication of risk between various stakeholders, and improved defensibility of decisions. The framework acknowledges that 100% certainty in decision making is impossible to obtain–rather, the focus is on providing a systematic way to make decisions in light of this inevitable uncertainty and on determining the amount of field information needed to make a decision under conditions meeting predefined criteria. In this manner, quantitative evaluation of any field campaign design is possible *before* data is collected. This can be done beginning from any knowledge state, and updating as more information becomes available is also possible.

The framework is presented in general and demonstrated in two synthetic case studies predicting 1) contaminant arrival time and 2) enhanced cancer risk due to groundwater contamination. Results from the arrival time study indicate that the effectiveness of field campaigns depends not only on the environmental performance metric being predicted but also its threshold value in decision making. Results also demonstrate that improved parameter estimation does not necessarily lead to better decision making, emphasizing the need for goal-oriented characterization design. The case study in predicting enhanced cancer risk involves hydrogeological characterization as well as population characterization. Population characterization can involve physiological and behavioral parameters. This case study explores the relationship between hydrogeological characterization, population characterization, cancer risk, and water resources decision making.

To my parents, James David and Keely Ann Harken

who taught me to work hard, dream big, and care for others.

Contents

Contents ii

List of Figures iv

List of Tables viii

1 Introduction and Background 1
- 1.1 Groundwater Resources . . . 1
- 1.2 Conceptual Models and Environmental Performance Metrics . . . 1
- 1.3 Spatial Heterogeneity and Geostatistics . . . 3
- 1.4 Site Characterization . . . 4
- 1.5 Uncertainty: Inverse Modeling and Stochastic Forward Modeling . . . 8
- 1.6 Risk: Making Decisions Under Uncertainty . . . 8

2 Challenges in Adoption of Stochastic Methods 10
- 2.1 Status Quo 2004 . . . 11
- 2.2 Status Quo 2016 . . . 13
- 2.3 Misconception: Data Requirements . . . 14
- 2.4 Regulations and Uncertainty . . . 14
- 2.5 Communication of Uncertainty . . . 15
- 2.6 A Comparative Study: Hydrogeology and Petroleum . . . 16
- 2.7 Summary and Outlook . . . 21

3 Hypothesis Testing in Hydrogeological Modeling and Water Resources Management 22
- 3.1 Objective . . . 24
- 3.2 Modeling Predictions as Hypotheses . . . 25
- 3.3 Modeling Predictions in Decision Making . . . 26
- 3.4 Field Campaign Design . . . 26
- 3.5 Communication of Uncertainty . . . 28
- 3.6 Defensibility of Decisions . . . 29
- 3.7 Hypothesis Testing in Water Resources Decision Making . . . 30

3.8 Level of Significance 31
3.9 Field Data 31
3.10 General Procedure 33
3.11 Calculation of Error Probabilities 37
3.12 Conditional Error Probabilities 38
3.13 Summary 39

4 Case Study: Contaminant Arrival Time **42**
4.1 Statistical & Physical Setup 42
4.2 Field Campaign Setup 44
4.3 Monte Carlo Methodology 44
4.4 Case Study Results and Discussion 44
4.5 Summary 52

5 Case Study: Increased Cancer Risk **57**
5.1 Introduction 57
5.2 Decision Making 58
5.3 ILCR Definition 60
5.4 Case Study Setup 61
5.5 Results and Discussion 65

6 Conclusions **70**

Bibliography **72**

List of Figures

1.1 Example of spatial heterogeneity of aquifer parameters. Visualized here is the variability of the log-transform of hydraulic conductivity in three-dimensional space. 3
1.2 Example of geostatistical estimation by the method of Simple Kriging in one dimension. The true value represented by the black line is the quantity to be estimated, but is observed only where the black circles are. The method produces estimates represented by the solid red line, as well as a variance, portrayed by the dashed red lines. As we can see, the estimation variance decreases close to measurements and increases farther away. Kriging methods exhibit the property of *exactitude*, which means that estimates always honor measured values. 5
1.3 Demonstration of geostatistical simulation in one dimension. The solid red line represents the true value of the the quantity $Z(x)$, but is observed only where the red circles are. As opposed to *estimation* (see Figure 1.2), which produces a smooth best guess, *simulation* produces a random output which obeys that same spatial pattern as the underlying process. The conditional simulations, represented by the dashed black lines, both exhibit the same pattern of spatial variability as the solid red line, and exactly honor the observed values. 6

2.1 Literature search results for university, government, and industry author affiliations in hydrogeology and petroleum. 20

3.1 Relationship between hydrogeological characterization and prediction, water resources management, and regulation. While previous work on field campaign design focuses on improving parameter estimates (arrow A), the framework presented in this paper broadens the scope to allow a more goal-oriented approach (arrow B). In addition, this framework facilitates the relationships between decision making, risk, and regulations (arrows C and D). 23
3.2 Four possible outcomes of hypothesis testing. These four possibilities stem from the binary nature of both the true state of the of the system being modeled and the conclusion which is made after making inferences and modeling. 27

3.3 Conventional process for characterization, modeling, prediction, and decision making in hydrogeology and water resources management. Prior Info refers to geological descriptions of the site along with *ex situ* data from similar sites. Data Acquisition Design refers to the specification of the type, quantity, and location of field measurements to be taken and Data refers to the information obtained from such measurements. Decision criteria refers to the threshold value of the EPM being predicted (e.g. MCL), as well as the level of significance. Inverse modeling refers to whichever process by which the parameters of the site are inferred using the data. Forward modeling is the process by which numerical models are used to predict EPM(s) using the information obtained from Inverse Modeling. Finally, a decision regarding the management of water resources is made. 34
3.4 Process for characterization, modeling, prediction, and decision making using the framework presented in this paper. First, all steps leading up to and including final decision making are simulated with an ensemble of baseline fields (synthetic realities), which enables assessment of the field campaign design before data is collected. The criteria for assessment of the data is the probability that the data collected will lead to an erroneous decision. If this probability is low enough, then the steps are executed according to plan. This framework enables the simple communication of uncertainty by allowing decisions to be communicated in terms of the probability that an incorrect decision was made. While every step of the process is open for review and scrutiny, the framework allows the aggregation of uncertainty from each step into a simple description (risk) which is more useful to water resources managers than e.g. descriptions of uncertainty in variogram parameters. Such open communication of uncertainty in decisions improves the defensibility of decisions because it moves discussion from what may or may not be correct to what level of risk can be considered acceptable, which is defined outside the realm of engineering design. 35
3.5 Flowchart of overall procedure computing error probabilities starting with prior information about geostatistical parameters, simulation of baseline fields and decisions, as well as computation of error probabilities. In many cases, simulation of an ensemble of geostatistical parameters is necessary in order to create an ensemble of baseline fields which accoutns for this variability. A more detailed description of the treatment of a single baseline field is provided by Figure 3.6. . 39
3.6 Flowchart describing the computational procedure for each baseline field. For a descripiton of the ensemble baseline fields, see Figure 3.5. 40

4.1 Area of travel paths illustrated by plotting the 50% and 90% quantiles of lateral displacement versus longitudinal displacement. The contaminant source is indicated by the red square, and the green rectangle indicates the control plane defining the environmentally sensitive target. The travel path quantiles were computed by simulating using unconditional particle trajectories. 45

4.2 The 8 campaign designs G to be tested. Left-hand-side (subscript A): measurements concentrated along travel path. Right-hand-side (subscript B): measurements spread throughout the domain. $N = 4, 8, 16, 32$ corresponding to subscripts $1, 2, 3, 4$, respectively. 46

4.3 Rejection regions with respect to τ_{crit} (non-dimensionalized). "Reject" means that H_0^G was rejected in favor of H_a^G. In other words, rejection indicates that the specified field campaign design is sufficient. The rejection region for each campaign corresponds to the values of τ_{crit} for which $\langle \phi_\alpha^G \rangle$ exceeds α, as indicated by Figure 4.4. The region for which H_0^G is accepted is indicated by absence of shading. Prior Information Scenarios 1,2, and 3, respectively. (Scenarios described in section 4.1) 48

4.4 $\langle \phi_\alpha \rangle$ resulting from all 8 campaign designs, plotted against τ_{crit}, nondimensionalized by the travel length and the average velocity. As we can see, $\langle \phi_\alpha \rangle$ approaches zero as τ_{crit} appraoches zero, and also as τ_{crit} gets large, for all measurement configurations. We also see that increasing the quantity of measurements improves performance, with diminishing marginal returns. Furthermore, for a given quantity of measurements, the configuration with measurements focused along the travel path (A) performed better than the configuration with measurements spread throughout the domain (B). Prior Information Scenarios 1,2, and 3, respectively. (Scenarios described in section 4.1) 49

4.5 P_α^G, $\langle \phi_\alpha^G \rangle$, and $\langle I \rangle$ plotted against τ_{crit}, which is nondimensionalized by the travel path length and the average velocity. $\langle I \rangle$ coincides with the cumulative distribution function of τ due to the definition of I (equation 5.1). $\langle \phi_\alpha^G \rangle$ represents the probability that H_0^I is true and H_0^I is rejected, which happens most often when τ_{crit} is near $\langle \tau \rangle$. P_α^G, on the other hand, is defined as the probability that H_0^I is rejected, conditional to H_0^I being true, and is thus equal to $\frac{\langle \phi_\alpha \rangle}{\langle I \rangle}$. Due to this, P_α approaches one as τ_{crit} approaches zero. What this indicates is that as τ_{crit} decreases, the probability of H_0^I being true approaches zero. As this event becomes more unlikely, it becomes nearly impossible to predict. Thus, the occurrence probability $\langle \phi_\alpha^G \rangle$ remains small while the conditional probability P_α^G becomes large. Prior Information Scenarios 1,2, and 3, respectively. (Scenarios described in section 4.1) 50

4.6 Standard deviation of I, ϕ_α, and P_α, plotted against non-dimenisonalized τ_{crit}. The figure shows that uncertainty in I and ϕ_α is highest when $\tau_{crit}\langle v \rangle / L$ is near one, while uncertainty in P_α is highest for lower values of τ_{crit}. Prior Information Scenarios 1,2, and 3, respectively. (Scenarios described in section 4.1) 53

4.7 Root mean square error in estimating μ_Y, σ_Y^2, and I_Y resulting from each measurement configuration, for Scenario 2. As we can see, the estimates improve with increasing number of measurements, and measurements spread throughout the domain (denoted B) performed better than the configurations with measurements concentrated along the travel path (denoted A). 54

4.8 Root mean square error in estimating μ_Y, σ_Y^2, and I_Y resulting from each measurement configuration, for Scenario 3. As we can see, the estimates improve with increasing number of measurements, and measurements spread throughout the domain (denoted B) performed better than the configurations with measurements concentrated along the travel path (denoted A). 55

5.1 Empirical pdf showing uncertainty in lumped parameter β which results from uncertainty in its components (see Table 5.4). Since β varies over only a few orders of magnitude, the uncertainty in β contributes relatively little to uncertainty in predicting enhanced cancer risk. 65
5.2 Barplot showing the Kullback-Leibler Divergence resulting from the simulated physiological information G_1^β, G_2^β, G_3^β, respectively. 66
5.3 This figure shows how the probability $Pr[R > r_{crit}]$ varies with respect to r_{crit} for three selected baseline fields. The solid black line indicates the synhetic truth associated with each baseline field, and the colored lines indicate the predicted response conditional to the four different measurement configurations. Type I error would occur for values of r_{crit} at which the black line is greater than α but the colored lines are less than α, which never occurs in this scenario. Type II error would occur if the inverse was true. 67
5.4 This figure shows the probability of the null hypothesis being true, conditional to five information states. The black line is conditional only the prior distributions $f_\beta(\beta)$ and $f_\theta(\theta)$. The colored lines are conditional the four different measurement configurations. This plot relates to Figure 5.3 by indicating whether or not the probabilities exceed α, and averaged over all baseline fields. 69
5.5 This figure shows the occurrence probability for Type II error for the four measurement configurations G_1^H, G_2^H, G_3^H, and G_4^H. 69

List of Tables

2.1 Summary of 2004 commentary on issues preventing adoption of stochastic methods in practice along with changes in 2016 commentary. 13
2.2 Terms for communicating uncertainty in event outcomes. From the Intergovernmental Panel on Climate Change Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties 16

4.1 Three prior information scenarios for the geostatistical parameters μ_Y, σ_Y^2, and I_Y used in the case study. 43

5.1 Non-hydrogeological quantities necessary to predict enhanced cancer risk to a population from a nearby groundwater contaminant plume. 61
5.2 Prior information for the SRF parameters for $Y = \ln K$ used in the case study. . 62
5.3 Values of non-hydrogeological parameters which are assumed deterministically known for the enhanced cancer risk case study. 64
5.4 Probabilistic description of non-hydrogeological parameters which were stochastically characterized for the enhanced cancer risk case study. 64

Acknowledgments

It seems surreal that my studies at UC Berkeley are nearing completion. It has been a long journey since I arrived here as a Master's student, and I have numerous people to thank for their help along the way. First and foremost, I thank my PhD advisor, Yoram Rubin. It has truly been an honor to work with such an expert in stochastic hydrogeology. From the beginning of my studies until the end, his guidance and insistence to never lose sight of the "big picture" have proven invaluable. There's no way I could have gotten this far without his help. I would also like to thank my collaborators for their assistance. I am extremely grateful to Peter Dietrich and Thomas Kalbacher at the Helmoltz–UFZ in Leipzig, Germany for hosting me as a guest scientist and also for their helpful advice throughout the research process. I am thankful to Felipe de Barros who has been a tremendous help as a source of both research advice and relentless positivity. I am also grateful for the advice of the faculty that have served on my committees–Sally Thompson, Tina Chow, Cari Kaufman, David Brillinger, and Nicholas Sitar. I also owe thanks to the labmates I've had througout my time here–Matthew Over, Michelle Newcomer, Ching-Fu Chang, Nigel Chen, Falk Heße, Changhong Wang, Andreas Geiges, Jon Sege, Thuy Nguyen, Carlos Osorio Murillo, Hu Ting, Bo Tan, Jianqin Chen, and Nura Kawa have all been helpful in some way, including advice on coding and writing as well as moral support. Karina Cucchi and Heather Savoy were especially supportive during the final stages of my dissertation and I am deeply grateful for their help.

I'm incredibly thankful to my entire family for their support throughout this process. I'm grateful to the many roommates and neighbors at the "Dwight House" that I've had throughout the years, especially Patrick Gorman, Rebecca Usoff, and Tina Ding who formed my "Berkeley family" during my time here. I'm thankful for the support of these friends, who helped me get through the stresses of graduate school, reminded me to relax every once in awhile, and were always ready to discuss anything from the philosophy of science to the latest football scores. I am thankful to the numerous cafes around Berkeley, especially Abe's Cafe and Pipeline Coffee, which have kept me caffeinated throughout the many many hours of reading, coding, and writing over the years.

Chapter 1

Introduction and Background

1.1 Groundwater Resources

Groundwater is an important resource, comprising 96 percent of all liquid freshwater in the world. It is used all over the world to meet municipal, agricultural, and industrial demands. Nearly half of all drinking water consumed worldwide comes from groundwater, as does over 40 percent of water used for irrigation (*Smith et al.*, 2016). In the United States, nearly 150 million people rely on groundwater for drinking water. In a twenty year timespan, over 20 percent of groundwater samples have had at least one contaminant present at levels potentially harmful to human health (*DeSimone et al.*, 2014). In light of threats from scarcity and contamination, providing groundwater resources in sufficient quantity and quality is becoming an ever more challenging task. In management of groundwater resources, decisions must be made regarding e.g. selection of water sources and allocation of remediation resources. To ensure safe and sustainable management of groundwater resources, these decisions should be made based on many different types of information. On one hand, information regarding site-specific geologic, hydrologic, and hydraulic properties is necessary. This information allows models to be used to make predictions which support groundwater management decisions (*Mays and Todd*, 2005). On the other hand, these decisions should also take social, economic, and political factors into consideration (*Smith et al.*, 2016).

1.2 Conceptual Models and Environmental Performance Metrics

While much of the groundwater present on Earth exists deep below the surface, the majority of the groundwater resources extracted by humans for the aforementioned uses comes from the shallow subsurface, which is replenished by precipitation (*Smith et al.*, 2016). Much of this water is stored in, and flows through, the pore spaces of natural soil material. This flow through natural porous media can be modeled using Darcy's law, which is given by

$$\mathbf{q}(\mathbf{x}) = K(\mathbf{x})\nabla H(\mathbf{x}) \tag{1.1}$$

where $\mathbf{x}$ indicates the spatial location, in one, two, or three dimensions. $\mathbf{q}$ is the spatially dependent specific discharge vector $[L/T]$, and $H(\mathbf{x})$ is the hydraulic head $[L]$. $K(\mathbf{x})$ is the hydraulic conductivity $[L/T]$ (*Fetter*, 2001). Darcy's law together with the law of conservation of mass (assuming constant fluid density) provides the groundwater flow equation:

$$\nabla \cdot (K(\mathbf{x})\nabla H(\mathbf{x})) = S\frac{\partial H}{\partial t} \tag{1.2}$$

where S is the storage coefficient $[-]$. In steady-state conditions, the right-hand-side of Equation is equal to zero. Groundwater velocity $\mathbf{v}(\mathbf{x})$ is related to specific discharge by the soil porosity $n(\mathbf{x})$, which is defined as the volumetric proportion of the porous media which is void space $[-]$(*Fetter*, 2001).

$$\mathbf{v}(\mathbf{x}) = \frac{\mathbf{q}(\mathbf{x})}{n(\mathbf{x})} \tag{1.3}$$

Given hydraulic conductivity, porosity, storage coefficients, boundary conditions, and initial conditions, Equation 1.2 can be solved to provide the hydraulic head and in turn, discharge and velocity. Provided the solution to the groundwater flow problem, the transport of solutes at a regional scale can be described by the advection-dispersion equation (ADE), which is given by e.g. *Dagan* (1989):

$$\frac{\partial C}{\partial t} = \nabla \cdot (\mathbf{D}\nabla C) - \mathbf{v} \cdot \nabla C + s \tag{1.4}$$

where C is spatially and temporally dependent solute concentration $[M/L^3]$, $\mathbf{v}$ is groundwater velocity (Equation 1.3), and $\mathbf{D}$ is the dispersion tensor $[L^3/T]$. The dispersion tensor represents the non-advective transport, and depending on the spatial scale being modeled, can be used to model the effect of dispersion at various scales (e.g. *de Barros and Rubin* (2011)). s represents sources or sinks of solutes, which may occur as a result of many natural or anthropogenic causes $[M/(L^3T)]$.

With these equations, any of a number of different numerical methods may be applied to make predictions which aid in management of groundwater resources (*Anderson et al.*, 2015). When dealing with contaminants, predictions may involve predicting some *Environmental Performance Metric* (EPM), such as contaminant arrival time or contaminant concentration at some specified location (*de Barros et al.*, 2012).

In other scenarios, such as predicting the effect of a groundwater contaminant plume on a nearby population, predicting hydrogeological processes is not enough. In these cases, other information is necessary. For example, in the case of predicting enhanced cancer risk of a population due to a population due to a nearby contaminant plume, information is needed regarding exposure pathways, physiology, and behavior of the population (e.g. *Maxwell et al.* (1998a), *de Barros and Rubin* (2008)). This concept is explored further in Chapter 5.

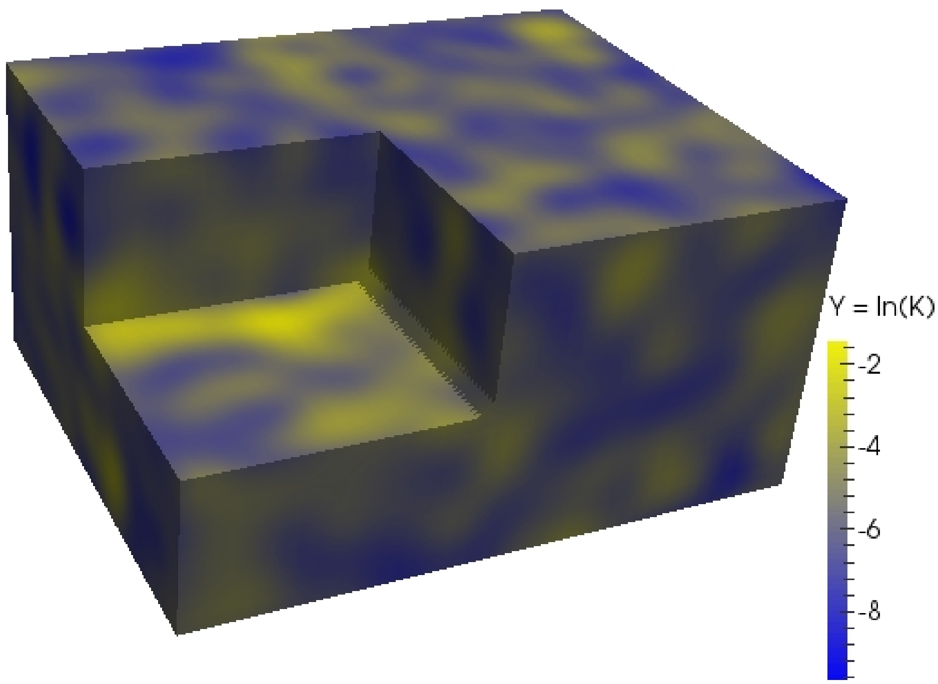

Figure 1.1: Example of spatial heterogeneity of aquifer parameters. Visualized here is the variability of the log-transform of hydraulic conductivity in three-dimensional space.

1.3 Spatial Heterogeneity and Geostatistics

One of the more challenging aspects of modeling hydrogeologic processes in natural media is the spatial variability of the media and its properties. Parameters such as hydraulic conductivity and porosity, as well as their spatial heterogeneity, have significant effect on the behavior of flow and transport in the subsurface (*Dagan* (1989), *Rubin* (2003)). The heterogeneity of parameters is demonstrated in Figure 1.1, which shows the variability of hydraulic conductivity in a synthetic aquifer.

Due to the spatial heterogeneity of parameters such as these, a complete description of their value at every location throughout a modeling domain is impossible, and some amount of uncertainty will always be present. In light of this, these parameters are described stochastically rather than deterministically. Fundamental probability theory provides tools for describing uncertain quantities referred to as Random Variables (RVs) (*Mukhopadhyay*, 2000). Any quantity which depends on a RV is in turn also an RV, which means any EPM prediction is also uncertain.

Dagan (1989), *Kitanidis* (1997), and *Rubin* (2003) provide many tools for modeling, predicting, and simulating RVs which display spatial patterns, which are referred to Space Random Functions (SRFs). SRFs are random variables which have a single distribution but can be observed at any number of points in space. Spatial patterns are often described using drift in the mean value or a spatial autocorrelation function, or both. An example of a hydrogeologic variable which displays a drift in the mean is hydraulic head in an aquifer exhibiting uniform-in-the-average flow. The average value of hydraulic head decreases in the direction of flow, though small-scale variations occur. Spatial autocorrelation functions are often described using variograms, which model the self-similarity of spatially variable quantities.

Given some set of measurments of an SRF, a variety of geostatistical methods can be used to estimate or simulate values at locations which have not been measured. Commonly used methods include Kriging (depicted in Figure 1.2 and Sequential Gaussian Simulation. Geostatistical simulation, which is depcited in Figure 1.3 often forms a component of Monte Carlo (MC) simulations, which involve simulation of ensemble of some variable, then computation via a transfer function of an ensemble of an output variable. MC methods are a powerful tool, which enable probabilistic analysis of many processes in hydrogeology (e.g. *Rubin* (2003)).

1.4 Site Characterization

Hydrogeological modeling has great potential to facilitate successful decision making in management of groundwater resources. In order to make decisions about a specific location, however, the location must be characterized. Generally speaking, the procedure has a few different components: prior information, field data, and inverse modeling.

Prior Information

First, all information available about the site is collected. This may include geological descriptions, maps, satellite data, or other forms of information. Specific information from geologically similar sites, referred to as *ex situ* data, can be used a starting point. This type of information is referred to as *a priori* information. *A priori* information can be used for preliminary analyses of the site. Very often, the Bayesian framework is used for site characterization, in which case the *a priori* information is formulated as prior distributions of parameters to be inferred from field data. Prior distributions can be informative or non-informative (e.g. *Kass and Wasserman* (1996), *Ulrych et al.* (2001), and *Tang et al.* (2016)).

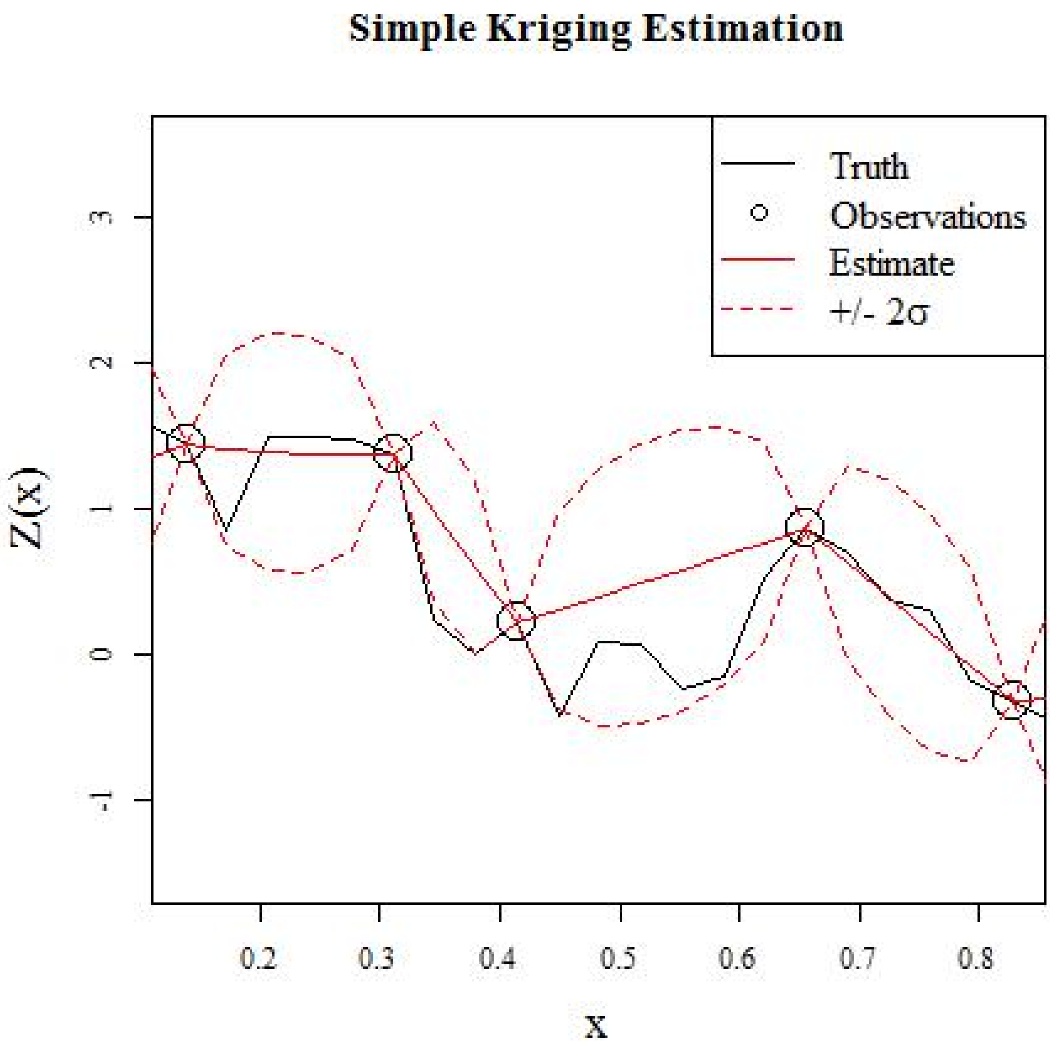

Figure 1.2: Example of geostatistical estimation by the method of Simple Kriging in one dimension. The true value represented by the black line is the quantity to be estimated, but is observed only where the black circles are. The method produces estimates represented by the solid red line, as well as a variance, portrayed by the dashed red lines. As we can see, the estimation variance decreases close to measurements and increases farther away. Kriging methods exhibit the property of *exactitude*, which means that estimates always honor measured values.

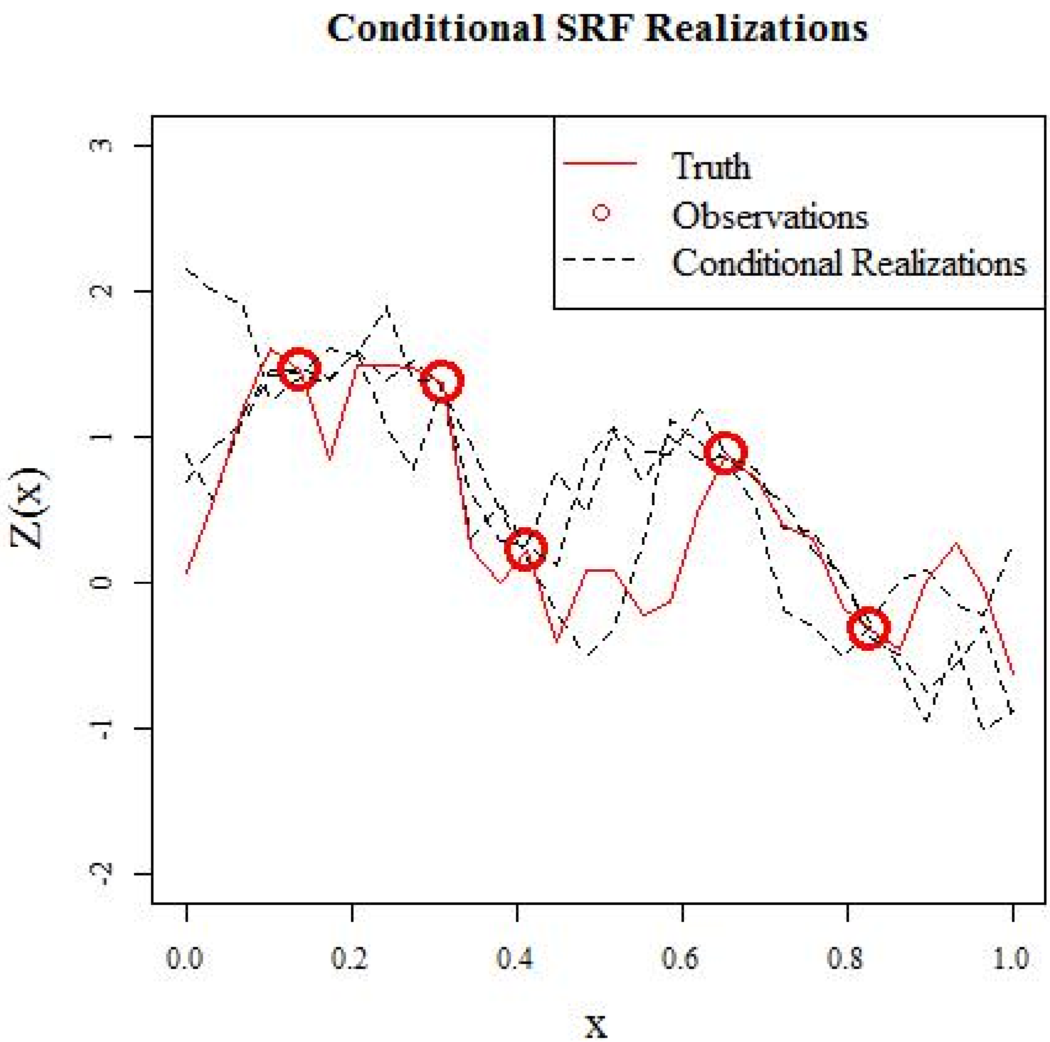

Figure 1.3: Demonstration of geostatistical simulation in one dimension. The solid red line represents the true value of the the quantity $Z(x)$, but is observed only where the red circles are. As opposed to *estimation* (see Figure 1.2), which produces a smooth best guess, *simulation* produces a random output which obeys that same spatial pattern as the underlying process. The conditional simulations, represented by the dashed black lines, both exhibit the same pattern of spatial variability as the solid red line, and exactly honor the observed values.

Field Methods

There are many methods which can be used to estimate various parameters regarding a site and its hydraulic conditions. Some methods involve collecting samples and analyzing them in a laboratory, while others involve *in situ* testing. The various methods can provide information about different parameters such as porosity, hydraulic conductivity, etc. and at different scales (*Fetter* (2001), *Rubin and Hubbard* (2006)).

Traditional Methods

Many methods for characterizing hydraulic parameters of aquifers which have been used for some time are laboratory tests and pumping tests. For laboratory tests, undisturbed soil samples are taken from the site and analyzed in a laboratory to determine parameters such as grain size distribution, porosity, and hydraulic conductivity. Grain size distribution of a soil sample can be determined using a combination of sieve analysis and hydrometer analysis. Hydraulic conductivity can be determined in a laboratory using any of a variety of permeability type testing (*Fetter*, 2001).

Pumping tests involve pumping water from a borehole in the field and using nearby monitoring wells to observe the change in hydraulic head over time. For transient analysis, the Theis solution can be used which also provide information about the transient behavior of soils, allowing estimates of storage coefficients (*Cherry and Freeze*, 1979). Steady-state cone of depression analysis can be done using the Thiem solution (*Fetter*, 2001). These tests provide information about the behavior of aquifers on a relatively large scale, and do not always provide much information about smaller scale heterogeneity.

Direct Push Methods

Direct push technology has been developed relatively recently (e.g. *Butler et al.* (2002), *Rubin and Hubbard* (2006)). Direct push methods involve probes which are driven into the soil and involve *in situ* hydraulic analysis. The use of direct push permeameters is described by *Butler et al.* (2007), which involve injecting a fluid at a known flowrate from a screened portion of the probe, and measuring the pressure response along other portions of the probe. The direct push injection logger rapidly provides information about relative conductivity with great resolution along a profile, which can be correlated to nearby permeameter tests (*Dietrich et al.*, 2008). These technologies offer the advantages of providing a great quantity of information in relatively little time, allowing for greater resolution of the spatial patterns of variability (*Köber et al.*, 2009).

Inverse Modeling

Inverse modeling is the process of utilizing observations from field or laboratory testing to infer parameters of the site to be modeled. Parameters which can be inferred are not only the hydraulic properties themselves, but also their patterns of spatial variability. Many different inverse modeling methods have been developed in the past few decades, each with their own advantages and disadvantages. Discussion of many of these methods can be found in e.g. *Yeh* (1986), *Carrera et al.* (2005), and *Zhou et al.* (2014).

1.5 Uncertainty: Inverse Modeling and Stochastic Forward Modeling

Uncertainty associated with any EPM prediction is inevitable. Uncertainty stems from the spatial heterogeneity of the aquifer material, unknown boundary and initial conditions, data scarcity, uncertainty in conceptual model definition, uncertainty in inferred physical parameters as well as uncertainty in their statistical distributions and parameters. Throughout the past several decades, the field of stochastic hydrogeology has undergone significant theoretical development and matured into numerous practical tools (e.g. *Dagan* (1989) and *Rubin* (2003)). Stochastic hydrogeology, simply put, is the science of modeling hydrogeological phenomena while also accounting for and quantifying uncertainty. With stochastic methods, not only can EPM predictions be provided, but also a description of the uncertainty surrounding such predictions. However, very often stochastic models are not used in practice, ignoring or not properly accounting for uncertainty, as discussed in a recent debate series ((*Cirpka and Valocchi*, 2016), (*Fiori et al.*, 2016), (*Fogg and Zhang*, 2016), (*Rajaram*, 2016), (*Sanchez-Vila and Fernndez-Garcia*, 2016)). This topic is explored further in Chapter 2.

1.6 Risk: Making Decisions Under Uncertainty

Water resource management decisions often depend on the predicted value of some EPM, which can never be predicted with complete certainty. However, stochastic methods can be employed to quantify uncertainty, improving the decision making process. *U.S. Environmental Protection Agency* (2014) described the importance of acknowledging such uncertainty in management decisions, saying, "If uncertainty and variability have not been well characterized or acknowledged, potential complications arise in the process of decision-making."

The issue of how to make a decision when presented with uncertainty has been dealt with in classical statistics by the method of hypothesis testing (*Navidi*, 2015). Hypothesis testing enables the translation of *uncertainty* stemming from all the sources described above into a simple description of *risk*, which simply describes the probability of making an incorrect decision. *Höllermann and Evers* (2017) pointed out that this translation is often the missing link between practicing hydrologists and water resources managers. This is discussed further in Chapters 2 and 3. When dealing with uncertain outcomes, hypothesis testing allows the discussion to shift from questioning whether a given outcome will occur to discussing precisely how probable is the occurrence of such an outcome. The question, then, is about how much risk is acceptable when making such a decision. *Farber and Findley* (2010) shed insight on this concept, stating:

> “A basic fact about risk management is that there is a difference between a ’safe’ world and a ’zero-risk’ world. In many situations, ’safety’ cannot be absolute but must entail an ’acceptable’ level of risk, however and by whomever that level may be defined. In some cases the decision may be made by an ’expert’ risk manager, while in other cases it may be a function of the political process.” (*Farber and Findley*, 2010)

A Balancing Act

Successful water resources management should take social, economic, and political factors into consideration in addition to hydrogeological information. There are often many roles and many stakeholders are involved in the making of water resources management decisions. For instance, hydrogeologists perform analysis, water resources managers make decisions based on these analyses, all while communicating results and decisions to regulators as well as the general public. As indicated by *Farber and Findley* (2010), some amount of risk must be accepted. The hypothesis testing framework allows for input from regulators, policy makers, and the general public in weighing the consequences of incorrect decisions in determination of what level of risk is acceptable. In other words, it is the role of hydrogeologists to determine the probability of incorrect decisions, not whether or not to act on such probabilities.

Field Campaign Design

The significant cost of obtaining field data of any type motivates the careful planning of field campaigns in order to best balance cost and characterization efficacy. This balancing process is complicated by the nonlinear relationship between design inputs (e.g. well location and pumping rates), performance outputs, and characterization efficacy. *Freeze and Gorelick* (1999) discussed these challenges in the context of remediation design, which faces similar challenges. *Abellan and Noetinger* (2010) provided a method for optimizing hydrogeological field campaigns with respect to improved geostatistical parameter estimates, and *Nowak et al.* (2012) related field campaign design to decision making risk. However, challenges remain in simply classifying a field campaign as adequate or not.

With sufficient site characterization, uncertainty can be reduced to acceptable levels for defensible decisions. This raises two questions, however: 1) how much uncertainty is acceptable? and 2) how much information is necessary for such a level of uncertainty? The answer to question 1 is a choice defined by social and political values, and codified by regulation. Chapter 3 presents a framework for determining the answer to question 2.

Chapter 2

Challenges in Adoption of Stochastic Methods

Stochastic hydrogeology has undergone significant development over the past few decades, beginning with fundamental concepts regarding uncertainty of subsurface properties and processes and maturing into advanced methodologies for quantification of uncertainty in characterization and modeling efforts of many varieties (e.g. *Dagan* (1989), *Rubin* (2003)). However, there appears to be a disconnect between the research community and practicing hydrogeologists when it comes to the application of stochastic hydrogeology. Previous series of articles have detailed the perspectives of stochastic hydrogeology researchers on why this disconnect exists (*Zhang and Zhang* (2004), *Sudicky* (2004), *Freeze* (2004), *Molz* (2004), *Neuman* (2004), *Christakos* (2004), *Ginn* (2004), *Rubin* (2004), *Winter* (2004), *Dagan* (2004), *Cirpka and Valocchi* (2016), *Fiori et al.* (2016), *Fogg and Zhang* (2016), *Rajaram* (2016), *Sanchez-Vila and Fernndez-Garcia* (2016))

This chapter aims review the evolution of thoughts on this disconnect, to evaluate strengths and weaknesses and to provide additional perspectives by highlighting misconceptions of stochastic methods, evaluating external factors that affect the adoption of stochastic methods in practice, and considering other impediments on the way to application.

First, we define the meaning of stochastic in this context. We consider any method regarding uncertainty using formal statistical theories. Discussion is not limited to geostatistical methods, but considers the uncertainties related to non-spatial variables and model choice in addition to the uncertainty imposed by spatial heterogeneity.

The conversation of how practitioners and scientists consider uncertainty can be found in other contexts. For example, *Höllermann and Evers* (2017) elicited perspectives from practitioners and scientists in the water management sector on how uncertainty enters their decision-making process. The authors point out how practitioners differ from scientists regarding what sources of uncertainty are of greatest concern and in which direction uncertainties flow between knowledge and decisions. Of interest is the translation from uncertainty to risk, as practitioners are prone to do implicitly and as decision makers can use directly. In order to make stochastic research more palatable for practitioners and decision-makers, stochastic theory should be complemented with the translation of uncertainty into application-specific guidelines on interpreting risk and aiding decision making.

To begin the stochastic hydrogeology-specific discussion, the comments in two series of articles addressing stochastic hydrogeology in practice are summarized. These series were in the journal Stochastic Environmental Research and Risk Assessment (SERRA) in 2004 and in the journal Water Resources Research (WRR) in 2016. The 2004 SERRA series asked nine researchers the following two specific questions: 1) "Why have there not been many real-world applications of stochastic theories and approaches, despite the significant progress in developing such rigorous theories and approaches for studying fluid flow and solute transport in heterogeneous media?", and 2) "In your opinion, what must be done in order to render stochastic theories and approaches as routine tools in hydrogeologic investigation and modeling?". The 2016 WRR series is more open-ended, asking teams of researchers to debate on stochastic subsurface hydrology from theory to practice. These two series were chosen because they address similar questions, they combine multiple perspectives from a variety of researchers, and the 12-year time span between them allows us to consider the evolution of ideas on this topic. However, due to the more open nature of the 2016 WRR series, not all of the discussion points can be directly compared between the two series. Of note is that issues related specifically to modeling of transport were covered in greater detail in 2016 than in 2004, and are excluded from consideration here in favor of the wider scope of the questions regarding stochastic methods in general.

Even though the 2016 WRR series is so recent, this discussion is written to compare the perspectives of the 2016 WRR series to the 2004 SERRA series as a baseline for comparison and to provide additional discussion on external perspectives such as from the petroleum industry and the legal system. The remainder of the chapter is organized as follows: given first is a summary and comparison the 2004 SERRA and 2016 WRR series, followed by discussion of issues pertaining to regulations, the legal systems, and incentive structure along with a comparative analysis between the hydrogeologial and petroleum disciplines.

2.1 Status Quo 2004

The 2004 SERRA series contains perspectives from nine stochastic hydrogeology researchers in response to the two questions listed above. Based on their responses to the question regarding why stochastic hydrogeology is not used in practice, we have compiled a list of five over-arching topics appearing across the nine papers (see Table 1). Those five topics are:

1. The influence of regulations and the court system: This topic appears with comments regarding how hydrogeology practitioners are motivated to use status quo, as opposed to state-of-the-art. It was mentioned how very often hydrogeological analyses are used in deliberation in the sometimes adversarial relationship between regulatory agencies and potentially responsible parties, with the deliberation often taking place in court. In

these deliberations, hydrogeological analyses are used to argue liability, or lack thereof. There is a perception that acknowledging any uncertainty in the analyses weakens the argument, thus disincentivizing the use of stochastic methods.

2. The role of higher education: This topic appeared either in terms of the lack of proper training in stochastic methods for hydrogeology practitioners or in terms of the theory being at a level not digestible by students.

3. The lack of appropriate measurement technology and/or data: This topic mainly appears as that the then-current technology did not provide measurements attributable to the parameters and scales of stochastic models and that, in practice, too much data is needed to fit those parameters.

4. The lack of user-friendly software that applies theory: This topic mainly appears as the call for user-friendly software that integrates multiple forms of information and that is computationally efficient enough for practitioners to use.

5. The lack of applicability of theory to real-world problems: Arguments here were either directed to the theoretical research as not applicable in real-world problems (e.g. oversimplification of spatial structure, or minimizing uncertainty in parameters when other uncertainties need to be minimized), or at least there is a lack of applications to showcase the applicability of theories in real-world problems.

6. Value: there was mention of a perception that relative to adhering to traditional deterministic methods, stochastic characterization and prediction are simply not worth the additional effort. This point can be considered directly related to Number 1. If risk-based predictions were mandated by regulatory agencies, there would be no question of value and incentive for adoption of uncertainty quantification would be provided by compliance.

The second question in this series requests suggestions on what needs to be done to have the field of stochastic hydrogeology adopted by practitioners. The suggestions provided by the nine authors transcend the five topics listed above, so they are presented separately below:

1. Cross-disciplinary collaborations: Have cross-disciplinary collaborations with hydrogeology practitioners for translating theory to practice, numerical modelers for embedding stochastic tools in software, and decision-makers or health-related researchers for making stochastic analyses relevant. These applications should be done and there should be literature that is application-oriented and readable for people outside the stochastic hydrogeology community.

2. Software: Develop flexible, powerful, and efficient tools for integration frameworks, embed stochastic methods into pre-existing and popular software, or make stand-alone stochastic software capable of coupling with flow/transport models.

Mentioned in 2004	Remains in 2016	Resolved by 2016	Added in 2016
Influence of Regulations	Resistance to probabilistic analyses		
Role of Higher Education	Lack of Professionals Trained in Stochastic Methods		
Measurement Technology and Availability of Data	Stochastic methods perceived to require more data and be more expensive	Measurement Technology has advanced. Stochastic analysis should use all available data via e.g. Bayesian Methods	
User-friendliness of concepts	Lack of usable software tools	Some forward models have incorporated stochastic methods, but too few	Misconception that risk assessment is not stochastic
Theory too limited or not showcased in applications	Research does not have the same goals as practice and often has too many unrealistic assumptions and simplifications		Need a goal-oriented framework for characterization and modeling.

Table 2.1: Summary of 2004 commentary on issues preventing adoption of stochastic methods in practice along with changes in 2016 commentary.

3. Geological realism: Build catalogs of properties and geological structures or improve theory for incorporating geological realism into stochastic models (*Molz*, 2004).

4. Education: Improve education and understanding of stochastic concepts, its data needs, and its relationship with respect to deterministic methods.

5. Other comments on where improvements can be made included the notion that regulations should embrace stochastic methods and that, in general, research topics take time to translate into industry and adoption will improve with time.

2.2 Status Quo 2016

The 2016 WRR series was not intended to be an update on the 2004 SERRA series and the foci dont fully overlap, but topics shared in common are still indicative of the evolution of thought over the 12 years. For comparison to a more recent series of papers addressing similar questions, we searched for the same five topics in the 2016 WRR series to gauge progress since the SERRA 2004 series. Since no new significantly different topics were discerned, we provide the summary of the 2016 WRR series as it pertains to the 2004 topics have and how they have changed. How those topics evolved per 2016 series perspective can be seen in Table 2.1. The topics most commonly mentioned in the 2016 WRR series are Topic 4 (user-friendliness of concepts) and Topic 5 (theory too limited or not showcased in applications). Concerns regarding the availability of useful software tools remain (*Cirpka and Valocchi*, 2016), however there is some recognition that few stochastic modules were introduced into forward models (*Fiori et al.*, 2016). The extensive discussion on geological realism (*Fogg and Zhang*, 2016) and transport processes (*Cirpka and Valocchi*, 2016) in the 2016 WRR series is encapsulated in the issue of simplicity-both concepts relate to the practical concerns of applicability. Of interest, however, is the mention of goal-oriented frameworks by *Fiori et al.* (2016), which is akin to the translation of uncertainty into risk at the knowledge-decision interface that *Höllermann and Evers* (2017) recommends. More discussion on goal-oriented framework is presented in the following chapters.

2.3 Misconception: Data Requirements

Many of the comments in the series refer to data requirements as one of the setbacks to adoption of stochastic methods. The underlying idea here is that the stochastic paradigm requires more data than the deterministic one. However, this idea is misleading. Returning to the definition of stochastic, which is accounting for uncertainty in a theoretically sound manner, it follows that uncertainty can be accounted for and quantified, regardless of the amount of data. Properly accounting for uncertainty can involve many steps. For example, determining statistical models and inferring parameters of those models can be precursors to forward modeling. Undoubtedly, having a greater quantity of data will improve the processes of model selection and parameter estimation. However, to think that in absence of large datasets it is better to simply ignore uncertainty can be perilous. When data is most scarce is when uncertainty is highest, and most important to be accounted for instead of ignored (*Rubin*, 2003). When field data from a specific site is scarce, other information can be used to constrain uncertainty and provide a better description of the site and the processes occurring within. As mentioned by *Cirpka and Valocchi* (2016), Bayesian methods can use a variety of forms of information including prior knowledge and indirect data to help constrain the uncertainty in these parameters.

2.4 Regulations and Uncertainty

The challenges with respect to regulations can be broadly summarized in the following points.

1. Regulations prefer simple point estimates over probabilistically defined quantities.

2. Regulations often define compliance as having some quantity above or below some threshold value. For example, contaminant concentrations must be below the specified limit (MCL).

3. It can be difficult to define and communicate the relationship between all sources of uncertainty (parameter, model, etc.). In this way, demonstrating compliance with regulations is facilitated by providing simple, deterministic quantities and not acknowledging uncertainty in such estimates.

While the first point may be true of many regulating agencies and many specific regulations, it is not always the case, nor does it have to be. Presented below are examples of regulating agencies coping with probabilistically defined quantities. A framework which addresses the last two points is presented and demonstrated in the following chapters.

This section explores this topic by examining the regulations known as Risk-Based Corrective Actions (RBCA). RBCA focuses on cancer risk, defined as the upper bound lifetime probability of an individual developing cancer as a result of exposure to a particular level of a potential carcinogen (US EPA, 1989). For example (following *Smalley et al.* (2000), a risk of 10^{-6} represents an increased probability of one in one million of developing cancer. Assessment of the exposure risk would require analysts to consider all the elements that combine to affect this risk, which would include source, contaminant transport pathways, exposure pathways, and toxicology parameters. RBCA thus specifically and explicitly recognized the stochastic nature of making predictions under uncertainty. The rationale underlying RBCA is extended to other focus areas of the EPA, such as "risk-based decision making in underground storage tank corrective action programs."

The claim that regulations dictate that hydrological predictions be expressed as single values to this day is supported by *Höllermann and Evers* (2017), who surveyed scientists and practitioners about the role of uncertainty in their professions. The survey results indicated that practitioners are not opposed to uncertainty assessment within their internal analyses before reporting to regulators. The willingness of practitioners to use stochastic concepts in their internal analyses alongside the lack of reporting those probabilistic results indicates that there is a component of the environmental regulation and policy sector that is stifling the widespread adoption of stochastic analyses. This effect could be reversed if regulations were to mandate probabilistic descriptions of quantities upon which decision making rests.

2.5 Communication of Uncertainty

Another issue in the management of uncertainty is communication. While quantified levels of uncertainty can be readily utilized by engineers for analysis, sometimes a simpler terminology is more suitable for communicating to a wider audience. Climate science is one example of a different field where uncertainty pervades models and predictions. To address

Term	Likelihood of the Outcome
Virtually certain	99-100% probability
Extremely likely	95-100% probability
Very likely	90-100% probability
Likely	66-100% probability
More likely than not	50-100% probability
About as likely as not	33-66% probability
Unlikely	0-33% probability
Very unlikely	0-10% probability
Extremely unlikely	0-5% probability
Exceptionally unlikely	0-1% probability

Table 2.2: Terms for communicating uncertainty in event outcomes. From the Intergovernmental Panel on Climate Change Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties

the challenges associated with communicating uncertainty, the Intergovernmental Panel on Climate Change provided a set of terms ranging from 'virtually certain' to 'exceptionally unlikely' to communicate in plain language the probability of uncertain events. These terms are shown in Table 2.2. Adoption of such terminology in hydrogeology could be beneficial in communicating uncertain events to the many stakeholders, which range from water resources managers, regulating bodies, governing bodies, and even the general public. The challenge however, is relating between uncertainty in e.g. model selection and parameterization to uncertainty in digestible outcomes, such as the probability of water supplies becoming contaminated. The framework presented in the next chapter provides a mechanism for exactly this: aggregating all uncertainty into a single quantity describing the likelihood of making an incorrect decision in water resources management (*Intergovernmental Panel on Climate Change*, 2010).

2.6 A Comparative Study: Hydrogeology and Petroleum

Several of the contributors to the 2004 and 2016 paper series have pointed to a lack of education in stochastic hydrogeology, along with stochastic analyses being too costly and not providing enough benefit, as main reasons for the lack of adoption of these methods in practice (*Dagan* (2004), *Molz* (2004), *Neuman* (2004), *Sudicky* (2004), *Winter* (2004), *Cirpka and Valocchi* (2016), *Fiori et al.* (2016)). In response, we aim to compare various aspects affecting the adoption of new methods in hydrogeology and petroleum engineering, where stochastic methods for characterizing, mapping, and modeling the subsurface have become commonplace in industry (*Jonkman et al.* (2000), *Floris et al.* (2001), *Liu and Oliver* (2003), *Oliver and Chen* (2010), *Rwechungura et al.* (2011)). The aspects we aim to compare include incentive structure, education, and interaction between academic and industrial organizations.

Incentive

To begin comparing the two industries, we start by acknowledging a difference in basic incentive structure, which in turn leads to a difference in operations and adoption of new practices. In the petroleum industry, any improvement in subsurface characterization and mapping capabilities would ultimately lead to a company being more profitable. Reservoir characterization and reserve estimates have a major impact on the bottom-line of a company, affect interactions with Wall Street (*Misund and Osmundsen*, 2015), and are regulated by the Securities and Exchange Commission (SEC). This relationship between petroleum engineering and the economy relies on guidelines set by both the SEC, as well as the Society for Petroleum Engineers, on how to communicate the certainty in the inherently uncertain reserve estimates (*Harrell and Gardner*, 2003). By having both professional and regulatory bodies dictating the use of probabilistic methods, the petroleum engineering field has significant incentive to stay at the forefront of stochastic methods innovation and application. In addition, the enormous capital costs associated with new and continued wellfield operations motivates using the most informative, risk-based predictions of costs versus benefits. In summary, petroleum firms are extremely motivated to create the most accurate depiction of underground resources in order to both accurately report reserve estimates and to best allocate costly drilling resources. Simply put, there is a direct link between improved characterization and improved profitability.

On the other hand, improvements in characterization and modeling do not always directly link to improvements in profitability for firms involved with groundwater investigation and contaminant remediation. As pointed out by *Ginn* (2004), there are numerous factors which affect the perceived costs and benefits of adopting stochastic methods, thus obfuscating the incentive structure. Very often, hydrogeologists involved in groundwater contamination projects must explain and justify steps taken, methods used, and conclusions reached to a wide range of people, including scientists in other fields, site owners, regulators, attorneys, and the general public. In light of these factors, it may seem more justifiable to adhere to established, proven methods because adopting new methods may seemingly invite criticism. Furthermore, as *Freeze* (2004) pointed out, it may often be the case that hydrogeologists are eager to use stochastic methods for uncertainty quantification, while their clients are reluctant to do so due to the perception that admitting any uncertainty may be construed as ignorance and thus leaving the client vulnerable to penalties from regulatory agencies.

However, these widespread perceptions do not necessarily represent the full extent of possible costs. It is worth remembering that when dealing with groundwater contamination, there is a lot at stake. For example, incomplete descriptions of subsurface characteristics, plume behavior, and remediation efficacy could lead to costly outcomes such as necessitating the procurement of new water sources, or even worse, exposing populations to harmful contaminants. In light of this, when the cost-benefit analysis is undertaken with appropriate scope, the benefits of protecting our water resources clearly outweigh the costs of the marginal effort associated with adopting stochastic methods and uncertainty quantification.

Education

Several contributions to the 2004 and 2016 series pointed to a lack of availability of university courses in stochastic methods and, in turn, a lack of hydrogeologists trained in stochastic methods as one of the setbacks to widespread adoption of stochastic methods in practice (*Neuman* (2004), *Winter* (2004)). Due to the prevalence of these methods in the petroleum industry, one explanation for the disparity could be a difference in the prevalence of education in these methods across the two disciplines. Continuing the comparative analysis, we aim to compare the prevalence of courses in these subjects in the disciplines most likely to be studied by future petroleum engineers and practicing hydrogeologists. Three academic disciplines were selected: petroleum engineering, civil/environmental engineering, and earth sciences. We surveyed the top ten schools in the US News World Report 2016 rankings for graduate programs in petroleum engineering and earth sciences. Rankings for civil engineering are done separately from environmental engineering, and the top ten of the two rankings produced thirteen unique universities, which were the universities included for this survey.

Gathering information is made complicated because the amount of information about course offerings was variable from school to school. For some schools, course titles and course descriptions were available, while for others only course titles were available. For a couple programs, no course information was publicly available online. While rankings exist for Earth Sciences as a subject, it is difficult to make direct comparisons due to the various organizational structures.

For schools where course information was found, seven out of nine petroleum engineering schools had a course where geostatistics is the main focus (i.e. in the course title), and one school had a course where geostatistics is mentioned as part of the course. On the other hand four out of twelve of the civil/environmental programs had a course where geostatistical methods were the main focus. Two programs had a course that included geostatistics or stochastic hydrogeology in the course description, but were seemingly not a main focus. Only one Earth Sciences program had a course focusing on stochastic hydrogeology.

Many civil/environmental engineering or Earth sciences program have at least one course on the topics of probability, statistics, or uncertainty analysis in some regard, but not specifically related to hydrology. While this survey does have limitations due to the variable amount of information available regarding frequency of course offerings, enrollment statistics, and neglecting the possibility of cross-department course enrollment, the results are clear: courses in geostatistics and stochastic subsurface modeling are much more prevalent in petroleum engineering than in civil/environmental engineering and Earth sciences.

Petroleum engineering has a clear connection to the petroleum industry, with masters and PhD programs serving as a pipeline for engineers for this industry. Practicing hydrogeologists, on the other hand, may come from a variety of educational backgrounds. Practitioners may have degrees in civil/environmental engineering, earth sciences, or others. However, the proportion of students in these departments who go on to become hydrogeologists is very small. Civil/environmental engineering departments, for example, provide coursework in structural engineering, transportation engineering, etc. where hydrogeology is, at best, one very small subset of the department. A similar statement could be made about earth sciences, where students take courses in tectonics, volcanology, and many other fields, where contaminant transport in the shallow subsurface is a very small subset of the topics covered in an earth sciences department.

One of the most obvious pathways for novel concepts and methods to move from research to industry is via academic training of future practitioners. The fact that petroleum engineering is considered its own discipline by many academic institutions facilitates this transition. In hydrogeology, on the other hand, the opposite is true. The underlying premise is that development of new methods occurs only in academic institutions. However, it is possible for research and development to occur within industrial institutions, either independently or in collaboration with academic institutions. The extent to which this occurs in these two disciplines is explored in the next subsection.

Research Collaboration

The vast majority of research in any scientific field comes from institutions belonging to one of three categories: academic, governmental, and industrial. While the fundamental research goals of academic institutions can vary widely, the fundamental goals of governmental and industrial research bodies can be more clearly distinguished. For governmental research bodies, the goal is to create knowledge that serves as a public good. For industrial research bodies, the goal is to advance the state-of-the-art of the industry and, in doing so, improve the profitability of the company. Thus, significant participation in research by industrial institutions can act as a catalyst to more widespread adoption of novel methods, because research is more motivated by the needs of practitioners.

Since participation by industrial institutions in research is one avenue for adoption of new methods in practice, we aim to compare this activity in the two fields. To accomplish this, we performed a bibliometric survey using Web of Science similar to the one performed by *Rajaram* (2016). The goal of the survey was to quantify the proportion of research contributions coming from academic, governmental, and industrial affiliations in each of the two disciplines. Thus, we aimed to perform a survey of the literature in the petroleum and the hydrogeological fields to compare the cross-collaboration between the different types of institutions that produce research in the two fields. Due to the differences between hydrogeology and petroleum described above, our hypothesis is that the body of research in petroleum engineering will have significantly greater participation by industrial institutions than in hydrogeology.

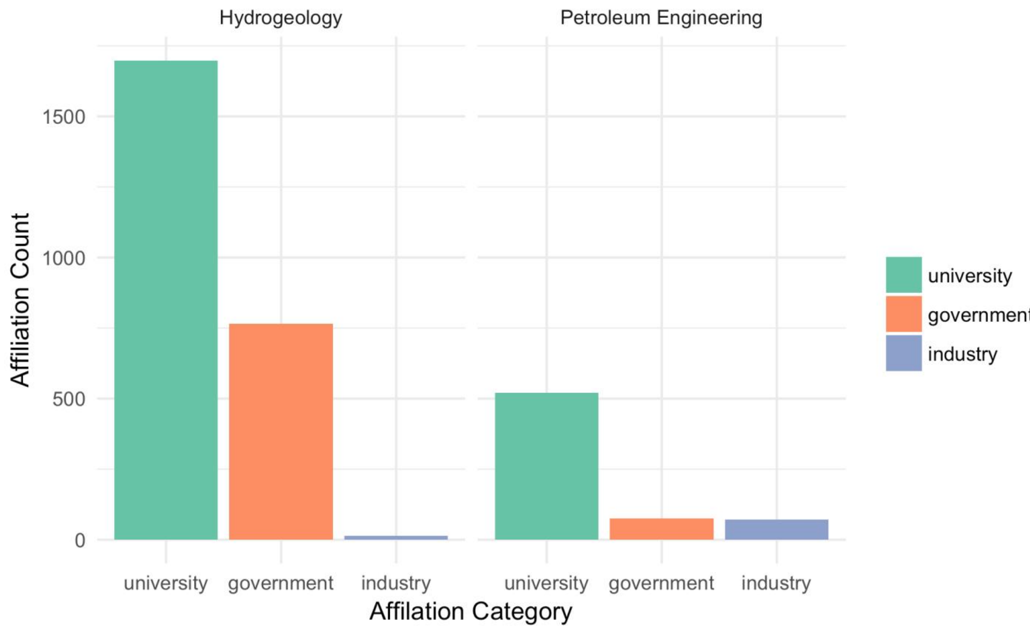

Figure 2.1: Literature search results for university, government, and industry author affiliations in hydrogeology and petroleum.

The bibliometric survey was performed in June 2017 by searching for all articles matching keywords related to stochastic methods in the two fields. The author affiliations of the resulting articles were recorded, and ordered by number of articles for each institution. All institutions which appeared on at least five articles were recorded and categorized as one of academic, governmental, or industrial. The number of appearances of each category was calculated for each of the two disciplines, with results presented in Figure 2.1.

For both hydrogeology and petroleum engineering, the majority of affiliations are academic, which may be expected for any field, but the ratio of academic to industrial affiliations is at least 106:1 for hydrogeology and at most 7.3:1 for petroleum engineering. Hydrogeology has approximately three times as many publication results as petroleum engineering, but petroleum engineering still has more industry appearances at 69-71 affiliations compared to hydrogeologys 14 industry appearances. While this affiliation search is not completely exhaustive for all publications, languages, and possible search terms, the results are conclusive: there is clearly a larger presence of industrial institutions contributing to research in the petroleum field compared to hydrogeology. These results confirm the hypothesis that there is much greater participation in research by industrial institutions in the petroleum field than in hydrogeology. In turn, it can be argued that this indicates a stronger connection between theoretical development and practice: advances in theory are both motivated by practice and smoothly adopted by practice.

This difference in cross-collaboration between academic, governmental, and industrial institutions between petroleum engineering and hydrogeology affect the adoption of new techniques between the two fields. For example, since petroleum engineering has a greater influence from industry, petroleum engineering research can benefit from both the greater financial resources and the focused direction towards characterizing petroleum reservoirs. Hydrogeology has a greater influence from government research, of which hydrogeology is a small fraction. Similar to how hydrogeology is a minority in its own academic departments, it is also a minority when it comes to the research conducted at the USGS or DOE laboratories that are in charge of exploring topics ranging from earthquake risk to renewable energy.

Despite being closely aligned in theoretical bases, hydrogeology and petroleum fields have many differences, beginning with their fundamental incentive structures. The direct nature of the relationship between methodological improvements and increased profitability provides impetus for advancement in the petroleum industry. However, this relationship is not as direct in hydrogeology, which necessitates the intervention of legislation and regulation to provide such impetus. In absence of such external forcing, the status quo in hydrogeology persists. The constant desire for improvement in the petroleum industry both motivates and is facilitated by greater connection between research and practice. This greater connection manifests itself in the differences in education and research discussed above.

2.7 Summary and Outlook

This chapter has presented a description and analysis of issues preventing the widespread adoption of stochastic methods by practicing hydrogeologists. The idea that these methods are not practical was shown to be a misconception, evidenced by the widespread adoption of very similar methods in the petroleum engineering industry. One of the most notable differences between these two industries. In environmental applications, incentive for innovation is lacking and must be substituted by regulations. In order to smoothly transition into a system where regulations mandate uncertainty quantification and risk analysis, the following questions must be addressed:

1. How to quantify uncertainty in binary predictions related to threshold values?
2. How to translate uncertainty in e.g. model selection and parameter estimation to risk as it pertains to decision making?
3. How to establish a framework where risk criteria can be defined by regulators which enables simple demonstration of compliance by practicioners?

In the following chapters, a framework which addresses these questions is presented and demonstrated.

Chapter 3

Hypothesis Testing in Hydrogeological Modeling and Water Resources Management

This chapter presents the hypothesis testing framework as applied to decision making in water resources management. As mentioned earlier, challenges to adoption of stochastic methods in practice stem from difficulty in communicating between uncertainty in model selection, parameterization, etc. to risk in decision making. Challenges in field campaign design remain due to the difficulty in quantifying the efficacy of field data, especially before data is collected. The framework presented in this chapter addresses these challenges.

Efficacy of a characterization campaign can be hard to define, due to the complex process by which field data is used to define and parameterize conceptual models, define and parameterize statistical models and models of spatial variability, and also be used as conditioning points in forward models. *Abellan and Noetinger* (2010) presented a method of optimizing field campaigns, where the field campaign with the greatest information gain was selected. Information gain was defined as the difference between the prior and posterior distributions of geostatistical parameters using the Kullback-Leibler divergence.

The place of field campaigns in water resources decision making is demonstrated in Figures 3.1 and 3.4. First, prior information about the site is used to make preliminary assessments and design the field campaign. After field data is collected, inverse modeling is performed which provides *a posteriori* estimates of model parameters, as well as deciding between alternative statistical models and conceptual models. Then, forward modeling is performed where numerical models are used to predict EPMs relevant to the decision to be made. Uncertainty in prior information, field data, inverse modeling, and numerical modeling is propagated to these final EPM predictions, complicating the decision making process. For this reason, the importance of goal-oriented characterization is emphasized. In other words, it is important to consider a justifiable decision as the ultimate objective of designing field campaigns, rather than focusing on improved estimates of geostatistical parameters. As Chapter 4 shows, improved estimates of geostatistical parameters do not necessarily indicate

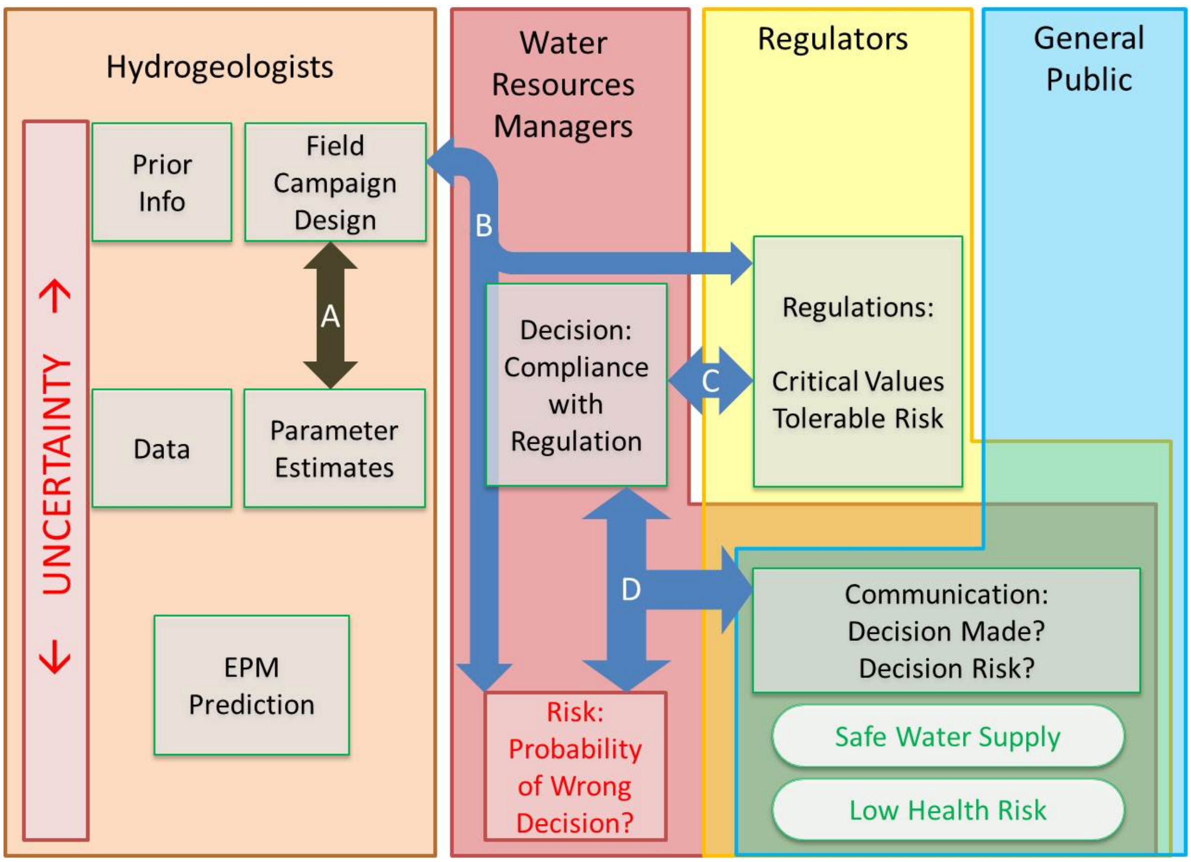

Figure 3.1: Relationship between hydrogeological characterization and prediction, water resources management, and regulation. While previous work on field campaign design focuses on improving parameter estimates (arrow A), the framework presented in this paper broadens the scope to allow a more goal-oriented approach (arrow B). In addition, this framework facilitates the relationships between decision making, risk, and regulations (arrows C and D).

improved decision making. In some application, such as assessing health risk to a potentially exposed population, hydrogeological characterization and modeling play only one part in the overall risk assessment, as shown by e.g. *Maxwell et al.* (1999) and *de Barros and Rubin* (2008).

Nowak et al. (2012) presented a framework for optimizing field campaigns, where parameterization of geostatistical models was taken as a step in computing the objective, which was water resources decision making, using a hypothesis testing framework. The present work similarly focuses on defensible decisions as the objective, but differs by allowing for any number of proposed field campaign designs to be analyzed resulting in a binary outcome: either the proposed design is adequate or not.

3.1 Objective

The objective of this chapter is to present a framework which enables integrated planning and holistic evaluation of the many processes involved in using hydrogeological models to inform water resources decision making. These processes can involve many different roles, including hydrologists, water resources managers, regulators, and even the general public (see Figure 3.1). The benefits of using such a framework are as follows.

- Simple translation of uncertainty in hydrogeological models to risk in decision making: Uncertainty is inevitable in all steps of hydrogeological characterization and prediction. This framework facilitates aggregation of these uncertainties into a single quantity describing the risk of making an incorrect decision, which can be readily used by those tasked with decision making.

- Goal-oriented characterization: While much work has been done to enable design of field campaigns which are optimal with respect to inferring parameters, this framework acknowledges that parameter estimation is merely one step on the way to the goal of effective decision making. As such, the framework enables design of field campaigns which will enable decision making which complies with regulation-based risk criteria. Simply put, this framework offers arrow B as an alternative to A in Figure 3.1.

- Simple specification of required risk criteria by regulators and demonstration of compliance by managers: Zero-risk conditions for decision making are impossible to achieve, so a balance must be struck between resources allocated towards reducing risk and some level of risk which is considered acceptable. This framework allows for regulators to define a level of risk which is deemed acceptable, and enables managers to demonstrate compliance with this requirement. Referring again to Figure 3.1, this benefit is visualized as arrow C.

- Simple communication of decisions and associated risks between managers, regulators, and the general public: The holistic manner in which this framework treats hydrogeologic uncertainty and decision risk enables managers to communicate decisions and associated risks to regulators and the general public. Instead of describing risk in terms of e.g. the uncertainty in variogram parameters, the framework allows communication in terms of decision risk, which is accessible to stakeholders who are not trained in hydrogeology. In Figure 3.1, this is visualized as arrow D.

The framework presented here can be simply summarized by the flowchart presented in Figure 3.4. The framework is not a substitute for any of the steps of hydrogeological analysis. Rather, it enables integrated analysis of the steps and aggregated evaluation of uncertainty, which in turn facilitates evaluation of risk in decision making. This is done by probabilistically simulating an ensemble of synthetic baseline fields which represent the distribution of plausible representations of the site under consideration. Then with each of these representations, each step of the process is simulated, beginning with data collection and leading to simulated decision, which can be compared to the correct decision. After repeating with the entire distribution of representations, a probabilistic description of the occurrence of correct and incorrect decisions can be obtained which in turn enables calculation of risk. If this outcome is adequate, the steps are then implemented. Otherwise, the steps are reformulated and reevaluated.

3.2 Modeling Predictions as Hypotheses

The first step in achieving these objectives is to formulate the EPM predictions on which decision making depend as hypotheses which can be tested. *Baker* (2017), *Bloschl* (2017), *McKnight* (2017), *Neuweiler and Helmig* (2017), and *Pfister and Kirchner* (2017) discussed the benefits of hypothesis testing in answering research questions in hydrology, and also the challenges which arise due to the fact that we are not able to run repeated, controlled experiments when modeling multi-scale processes within natural systems. Here we focus on using hypothesis testing to make management decisions related to water resources, and to cope with the inability to run repetitive experiments, we make use of extensive probabilistic simulation. Such management decisions, when dealing with a binary outcome, can be cast into the hypothesis testing framework following the method presented by *Nowak et al.* (2012).

The binary nature of the management questions described above lends to defining null and alternative hypotheses, which can be treated statistically. Water resources management, regulation, and policy making can be aided by hydrological models, which enable the prediction of various EPMs including contaminant concentration or arrival time, sustainable yield, or enhanced cancer risk (*Nowak et al.*, 2012). However, these predictions are inherently uncertain, an attribute which should not be neglected in making management or regulation decisions (*Oreskes et al.* (1994), *Beven* (2002)). In seeking improvements of this decision making process, we treat EPM predictions as hypotheses, subject to statistical treatment. In many cases, what is ultimately important is not simply an estimate of the EPM, but also its relation to some threshold value. For example, in some applications it is necessary to predict the concentration of a contaminant, which in turn is compared to an MCL. A course of action (i.e. to remediate or not) is then decided based on whether or not the concentration exceeds the MCL. In this context, the goal of the field and modeling campaigns should be seen as predicting whether or not the concentration exceeds the MCL, rather than solely predicting the concentration.

3.3 Modeling Predictions in Decision Making

These hypotheses can be formally tested as part of the decision making process as follows. We start by defining the EPM of concern, along with its critical, or threshold value, which represents a value or range of values of special concern. In some cases, the critical value could be a maximum allowable concentration, degradation time of a contaminant, or minimum sustainable yield. These hypotheses conform to the tradition of the null hypothesis being the undesirable, risky, or dangerous outcome. In the context of decision making in water resources management, it could, for example, represent the outcome that water supply becomes contaminated at levels greater than the MCL. The alternative hypothesis, conversely, represents the desirable outcome, or in the same example, that the concentration at the water supply will remain below the MCL. It is reiterated that in accordance with convention, the burden of truth falls on the alternative hypothesis, while the null is assumed to be true until convincing evidence shows otherwise.

The evidence used to test the null hypothesis comes in the form of field data. After the field data is collected, processed, and used to inform modeling predictions, the decision to either accept or reject the null hypothesis is made. The criteria for this decision are based on the probability that the null hypothesis is true, and how that probability compares to a predefined level of significance. The level of significance is subjective and should be determined by regulators and policy makers based on the relative consequences associated with the null hypothesis.

This decision making process leads to four possible outcomes: correctly accepting or rejecting the null, and erroneously accepting or rejecting the null (see Figure 3.2). The possible errors are classified as Type I, or alpha, error which is erroneously rejecting the null, while Type II, or beta, error which is erroneously accepting the null. Since the null hypothesis represents the non-desirable, or risk-posing state in the context of risk assessment, Type II error represents being overly conservative, while Type I error represents assuming safety when the system is in reality not safe. Without sufficient evidence to reject the null hypothesis, the fallback assumption is to accept it.

3.4 Field Campaign Design

Field data plays a critical role in characterizing the site under consideration. In general, field data provides information about the conceptual model, boundary conditions, the physical parameters of the model and their statistical distributions and spatial structures. Here,

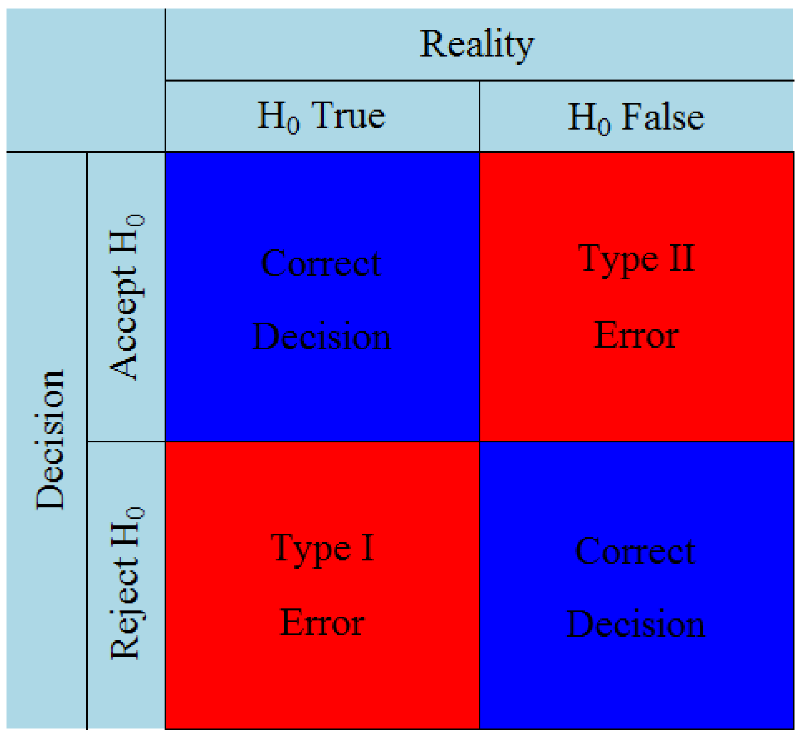

Figure 3.2: Four possible outcomes of hypothesis testing. These four possibilities stem from the binary nature of both the true state of the of the system being modeled and the conclusion which is made after making inferences and modeling.

we use the term field data to refer to results from *in situ* procedures as well as laboratory procedures using samples obtained *in situ*. Information from e.g. maps and satellite data can also be useful in many applications, but in this context is considered prior information. The processes of assimilating field data into models can be complex, and the resulting uncertainty can be difficult to predict because it stems from measurement errors and uncertain parameter estimates and manifests in the resulting EPM prediction. Another challenge is that different types of measurements provide different types of information and at different scales. Furthermore, is has been shown that information needs vary with different EPMs (*de Barros et al.*, 2012). These challenges obfuscate the process of designing field campaigns that best balance cost with prediction accuracy.

Optimal field campaign design is the design which can best balance economic, logistical, and practical constraints with modeling and prediction accuracy. The former can depend on many factors and is not the focus of the present study. With the framework of hypothesis testing, we can objectively and quantitatively assess and predict the effectiveness of field data in meeting decision making goals. In theory, uncertainty would decrease with increasing field data. However, this relationship is neither linear nor continuous, due to practical constraints.

Thus, given some proposed field campaign design, our aim is to test the design, where the criteria is simple: will the proposed design enable a justifiable decision? Again, we conceptualize this as a test of hypotheses. The undesirable scenario is our null hypothesis: the field campaign is inadequate to enable such a decision. Our desirable scenario is the alternative hypothesis: the field campaign design will enable a defensible decision. Again, without convincing evidence we accept the null hypothesis, since the burden of proof rests on the alternative, which in this context means that we assume the field campaign is inadequate unless we can convincingly demonstrate that it is adequate.

It is emphasized that this design framework is not intended to replace the need for locally experienced experts in field campaigns. Experience and intuition plays a vital role in the initial design of the proposed field campaigns since this framework does not actually design field campaigns. What this framework provides is a systematic, formal way to determine if any given campaign design will meet the needs of decision makers, better enabling practitioners to meet information needs while minimizing costs and meeting other constraints.

3.5 Communication of Uncertainty

There are multiple steps in the process by which field data is used to make predictions of EPMs. First, field data and prior information are used to estimate physical parameters (e.g. hydraulic conductivity), as well as parameters describing their patterns of spatial variability (e.g. autocorrelation length scales). With these parameter estimates, numerical models are used to predict the relevant EPMs. Depending on the specific application, either of these processes may need to be broken down further into subprocesses, all of which are subject to uncertainty. Significant research has been devoted to developing methods for quantification of this uncertainty (e.g. *Rubin* (2003)), and is not the focus here. As pointed out by *Höllermann and Evers* (2017), these measures of uncertainty are not always useful to decision makers. Instead, the focus of decision makers tends to be on risks associated with the possible outcomes of decision making.

Risk in decision making is determined by the probability of making an erroneous decision and the consequences of such errors. Assessing the consequence of such errors can be a challenging task, highly dependent on the specific application, and not the focus of this paper. What the framework presented in this paper provides is a means of aggregating uncertainty from each step in producing predictions into a simple, quantitative description of the probability of erroneous decisions. In this way, the framework simplifies communication of uncertainty between hydrogeologists, water resources managers, and other stakeholders. Instead of focusing on process uncertainty, model uncertainty, or parametric uncertainty, these uncertainties are aggregated into a description of risk.

3.6 Defensibility of Decisions

Goode and Evans (2007) describe the benefits of using the hypothesis testing framework in the context of biotechnology development and Food and Drug Administration regulations. While complete certainty regarding the safety and efficacy of a new product is impossible to obtain, approvals must be granted or denied nevertheless. Formulating and testing hypotheses regarding these considerations provides a rational method for communication of these uncertainties and making a decision in light of them.

The challenges in water resources management are similar: complete certainty is impossible to obtain, but decisions must still be made and, ultimately, justified and defended. Further complicating this requirement, as noted above, is that decisions must be defended to various stakeholders, many of whom are not trained in hydrogeological sciences. While use of the hypothesis testing framework does not guarantee a correct decision 100% of the time, what it does enable is transparency in the decision making process. Instead of making a decision regarding water supply while speculating about how parameter or model uncertainty may affect the outcome, decision makers are able to make decisions while knowing precisely the probability that they may be wrong. In other words, use of hypothesis testing shifts discussion and debate from whether or not the correct decision was made to with what level of certainty was the decision made.

The use of hypothesis testing does not serve as a substitute for judicious design of field campaigns, formulation of conceptual models, and validation of assumptions. Nor does it prevent the documentation and review of all assumptions made, models used, and calculations executed throughout the process. Utilization of hypothesis testing allows all of this but in addition enables simple description of the aggregated uncertainty.

Given regulations which mandate certain levels of acceptable risk for given applications, practitioners and managers have a straightforward way of ensuring and demonstrating compliance with such regulations. Therefore, even if a decision is made which is later shown to be erroneous, the decision is defensible if it can be shown that it was made under acceptable conditions.

described in subsection 3.1. Presented first is formal definition and testing of hypotheses in the context of water resources decision making, followed by the testing of a field campaign design to ensure that it enables such a test.

3.7 Hypothesis Testing in Water Resources Decision Making

This subsection describes how defensible water resources management decisions can be made in light of uncertainty of all kinds: measurement error, parametric uncertainty, statistical model uncertainty, and conceptual model uncertainty. The framework here can be used in a situation when any EPM is being predicted, using any statistical formulation as well as any conceptual model or combination of potential conceptual models.

Environmental Performance Metrics and the "Critical Range"

We start by considering a scenario where some water resources management, regulation, or policy decision must be made based on some binary outcome, or the answer to a yes/no question. Examples could be: will contaminant concentration at a water supply well exceed the MCL? Will a contaminant arrive at a water supply well before it degrades? Answering each of these questions hinges on predicting an EPM but in a binary state, i.e. is the EPM in the "critical range". The "critical range" is the range of values of the EPM which would pose a problematic or dangerous condition, e.g. concentration above the MCL or arrival before degradation. With this, we define the risk indicator variable I as 1 if the EPM is within the critical range (i.e. the risky scenario), and 0 outside of the critical range (i.e. the desirable scenario):

$$I = \begin{cases} 1 & \text{if } EPM \in EPM_{critical} \\ 0 & \text{if } EPM \notin EPM_{critical} \end{cases} \tag{3.1}$$

I being equal to one represents the scenario which poses danger. With this indicator variable we define the null and alternative hypotheses:

$$H_0^I : \; I = 1 \tag{3.2}$$

$$H_a^I : \; I = 0 \tag{3.3}$$

where H_0^I is the null hypothesis, and H_a^I is the alternative hypothesis, with the superscript I indicating that risk indicator I is the subject of these hypotheses. Again, it is emphasized that H_0^I represents the scenario which poses danger (e.g. water supplies are exposed to contamination), while H_a^I represents the "desirable" scenario (e.g. water supplies are safe from contamination). In conformity with the classical hypothesis testing, H_0^I is the fallback assumption, which is accepted by default in absence of evidence pointing to it being false. The burden of proof falls on H_a^I, which must be supported by convincing evidence before it is accepted. This aligns with safe water resources management–if the safety of our water supply is in question, we remain suspicious until we can reliably demonstrate that it is in fact safe.

The binary nature of the hypotheses and the decision making leads to four possibilities based on both what is actually true as well as what we determine to be true: 1) correctly accepting H_0^I, 2) correctly rejecting H_0^I, 3) erroneously rejecting H_0^I (Type I Error), and 4) erroneously accepting H_0^I (Type II Error), as shown in Figure 3.2.

3.8 Level of Significance

Given the hypotheses above and the idea that we only reject H_0 if we have convincing evidence otherwise, the question arises: how "convincing" must the evidence be? This question leads us to defining a level of significance, denoted by α. Given α, we reject H_0 only if the probability of H_0 being true is less than α, and accept it otherwise:

$$\text{accept } H_0^I \text{ if } Pr[I = 1] \geq \alpha \tag{3.4}$$

$$\text{reject } H_0^I \text{ if } Pr[I = 1] < \alpha \tag{3.5}$$

It is worth noting that α is not determined by any engineering calculation or modeling prediction. Rather, α should be determined by regulation or policy to strike a balance between accepted levels of uncertainty and characterization costs, and is often defined as 0.01 or 0.05. If the consequences of making Type I error are relatively high, then a relatively low value of α should be used, and vice versa.

3.9 Field Data

After field data g is collected we decide to accept or reject the null hypothesis based on the probability that the null hypothesis is true, $Pr[I = 1|g] = \langle I|g\rangle$ and the acceptance criterion, α.

$$D^g = \begin{cases} 1 & \text{if } \langle I|g\rangle \geq \alpha \text{ (accept } H_0^I) \\ 0 & \text{if } \langle I|g\rangle < \alpha \text{ (reject } H_0^I) \end{cases} \tag{3.6}$$

The field data vector g can contain measurements of any type. Depending on the EPM under consideration, the amount of information available about the site, and the type of measurements contained in g, computation of $\langle I|g\rangle$ can involve several steps. The steps can include distinguishing between alternative conceptual models, inferring geostatistical parameters of the relevant variables (e.g. hydraulic conductivity) as well as simulation of whichever processes are necessary, determined by the EPM under consideration. If uncertainty about $\langle I|g\rangle$ is too high, more field information can be obtained to reduce such uncertainty.

The main difficulty in planning field campaigns, then, is evaluating the effectiveness of measurements before they are even taken. In light of this, we must treat the measurement values probabilistically. We consider N_G proposed field campaign designs $G_j;\ j = 1, ..., N_G$, where each of the N_G alternatives specifies the quantity, types, and spatial locations of measurements to be taken. Thus, the field data g_j is considered a realization of the random variable G_j. To test the adequacy of any given campaign design G, we start by defining the indicator variables ϕ_α^g and ϕ_β^g as

$$\phi_\alpha^g = \begin{cases} 1 & \text{if } D^g = 0 \cap I = 1 \\ 0 & \text{otherwise} \end{cases} \tag{3.7}$$

$$\phi_\beta^g = \begin{cases} 1 & \text{if } D^g = 1 \cap I = 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.8}$$

where $\phi_\alpha^g = 1$ indicates the occurrence of Type I error, and $\phi_\beta^g = 1$ indicates the occurrence of Type II error. We can begin to summarize the effectiveness of G using the probabilities of occurrence of each type of error, $Pr[\phi_\alpha^G = 1] = \langle\phi_\alpha^G\rangle$ and $Pr[\phi_\beta^G = 1] = \langle\phi_\beta^G\rangle$. With these values, we can define the total decision risk

$$R^G = w_\alpha\langle\phi_\alpha^G\rangle + w_\beta\langle\phi_\beta^G\rangle \tag{3.9}$$

where w_α and w_β are coefficients selected to quantify the relative significance of Type I and Type II errors. With the decision risk R^G, we can express our null and alternative hypotheses regarding the field campaign design G

$$H_0^G:\ R^G \geq R_{crit} \tag{3.10}$$

$$H_a^G:\ R^G < R_{crit} \tag{3.11}$$

where R_{crit} is the maximum allowable decision risk and the superscript G indicates that these are hypotheses regarding the field campaign design G. In words, H_0^G indicates that G is not adequate to ensure an appropriate test of H_0^I on which our water resources management decision depends. The alternative hypothesis H_a^G indicates that the design G is indeed adequate to enable defensible decision making.

In general, Type I error is more consequential than Type II error because Type I error indicates falsely assuming safety of water supplies while Type II error indicates being overly conservative. Thus in many cases we may only be concerned with Type I error, and wish to show that the probability of Type I error is below α. In this case, which is the focus of the present work, w_α takes a value of 1, w_β takes a value of 0, and R_{crit} is equal to α. Thus, 3.10 can be expressed as

$$H_0^G:\ \langle\phi_\alpha^G\rangle \geq \alpha \tag{3.12}$$

$$H_a^G:\ \langle\phi_\alpha^G\rangle < \alpha \tag{3.13}$$

where α is the level of significance defined above, and $\langle \phi_\alpha^G \rangle$ is the same as previously defined. In words, this null hypothesis H_0^G states that the field campaign design G is inadequate to ensure a test where the probability of Type I error is less than α. The alternative hypothesis, for which we hope to provide evidence, states that G will indeed provide enough information to ensure an adequate test of H_0^I.

The framework above formalizes the goal of our field campaign design: to enable defensible decisions. As mentioned in previously, making decisions based on predictions of complete certainty is impossible. In absence of complete certainty, a formal method such as this is beneficial for water resources managers and regulators, as it allows for transparent handling of uncertainty: the regulators can define the accepted level of uncertainty (α) for any number of scenarios, and practitioners can formally justify decisions made in terms of how much field data to collect and decisions ultimately made.

3.10 General Procedure

Presented in this subsection is the general method to test the hypotheses presented by 4.4. Simply put, this is done by simulating an ensemble of physically plausible baseline fields, and then testing the ability of G to enable successful decisions.

Simulation of Baseline Fields

We start by generating an ensemble of baseline fields, $\tilde{Y_i^b}, \ i = 1, ..., N_Y$, where each $\tilde{Y_i^b}$ is a field of all parameters necessary to compute the EPM of concern. The N_Y fields together form an ensemble considered to represent the entire range of physically plausible possibilities of the site under consideration. One example could be simulating spatially variable hydraulic conductivity and/or porosity in order to simulate flow and transport in an aquifer. The baseline fields can be generated conditional to any knowledge state, ranging from uninformative prior distributions based on *ex situ* information to conditional distributions based on *in situ* data. The framework, in general, allows for consideration of competing conceptual models, which may necessitate the generation of multiple ensembles of baseline fields.

In many examples, prior information comes in the form of statistical distributions for parameters of Space Random Functions (SRFs). In this case, it is necessary to first generate an ensemble of these parameters, along with an ensemble of baseline fields for each of these parameter sets.

Given the ensemble of baseline fields $\tilde{Y^b}$, we use each baseline field $\tilde{Y_i^b}$ for two purposes: simulating the baseline EPM, and simulating the field campaign and resulting decision making process.

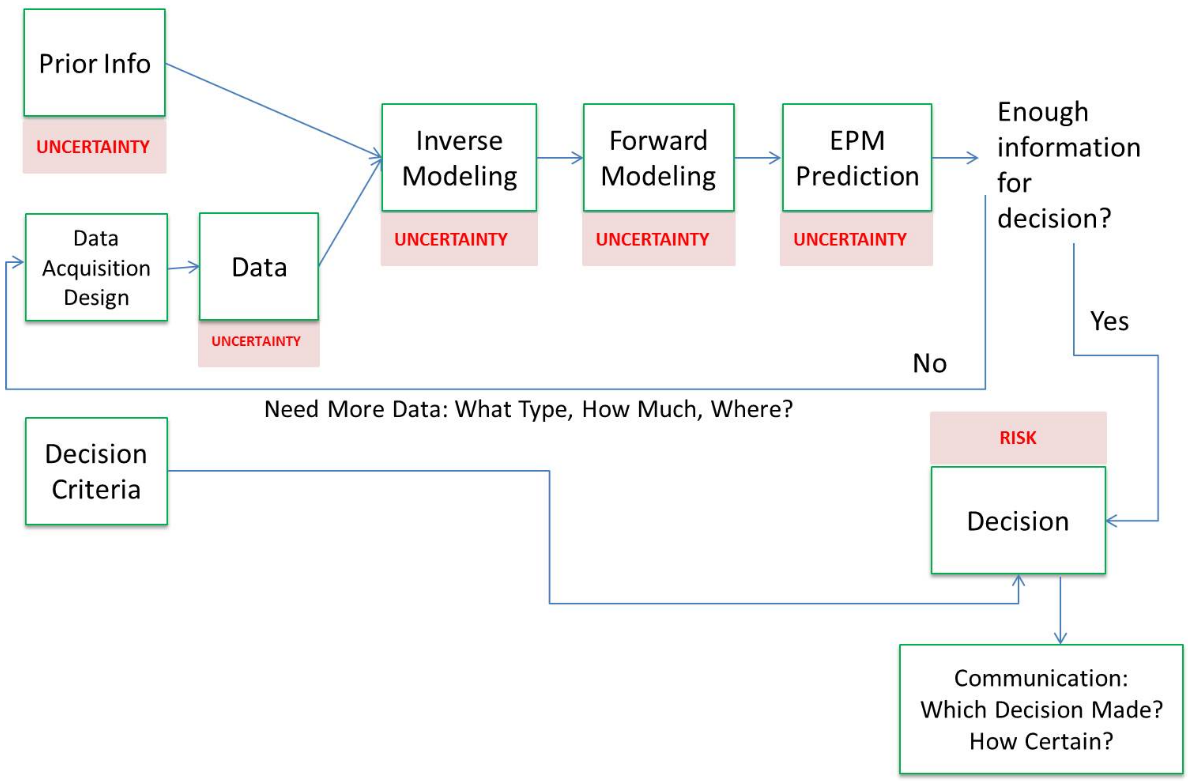

Figure 3.3: Conventional process for characterization, modeling, prediction, and decision making in hydrogeology and water resources management. Prior Info refers to geological descriptions of the site along with *ex situ* data from similar sites. Data Acquisition Design refers to the specification of the type, quantity, and location of field measurements to be taken and Data refers to the information obtained from such measurements. Decision criteria refers to the threshold value of the EPM being predicted (e.g. MCL), as well as the level of significance. Inverse modeling refers to whichever process by which the parameters of the site are inferred using the data. Forward modeling is the process by which numerical models are used to predict EPM(s) using the information obtained from Inverse Modeling. Finally, a decision regarding the management of water resources is made.

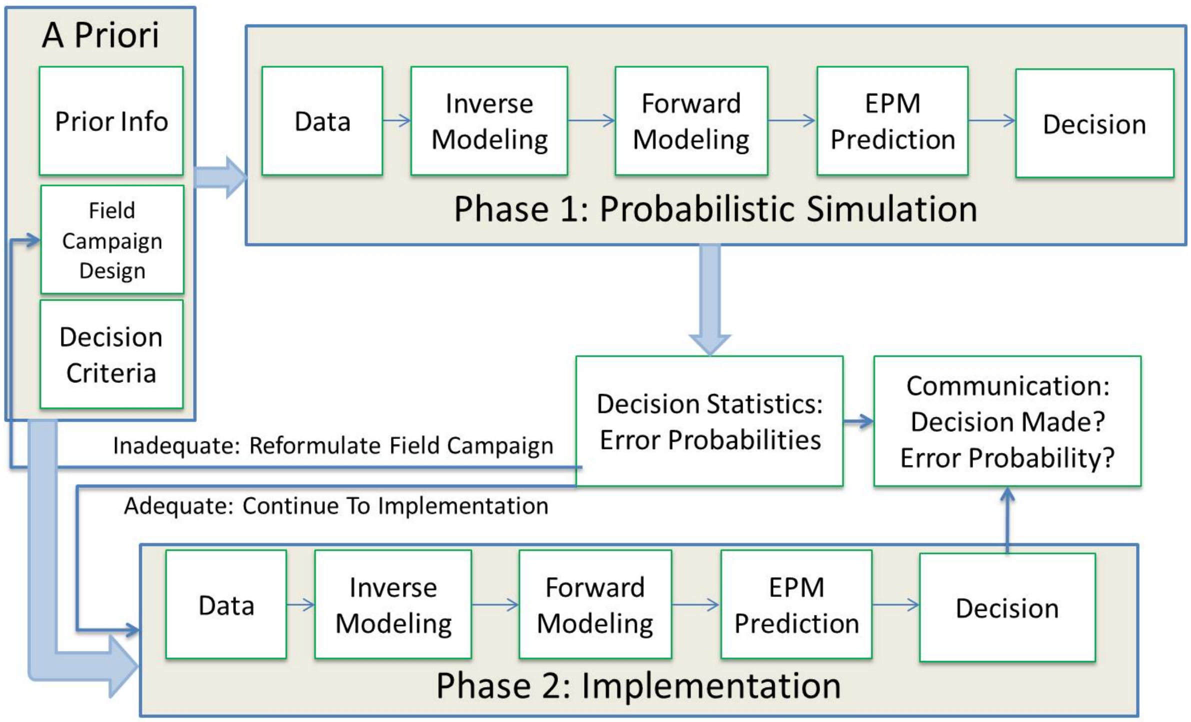

Figure 3.4: Process for characterization, modeling, prediction, and decision making using the framework presented in this paper. First, all steps leading up to and including final decision making are simulated with an ensemble of baseline fields (synthetic realities), which enables assessment of the field campaign design before data is collected. The criteria for assessment of the data is the probability that the data collected will lead to an erroneous decision. If this probability is low enough, then the steps are executed according to plan. This framework enables the simple communication of uncertainty by allowing decisions to be communicated in terms of the probability that an incorrect decision was made. While every step of the process is open for review and scrutiny, the framework allows the aggregation of uncertainty from each step into a simple description (risk) which is more useful to water resources managers than e.g. descriptions of uncertainty in variogram parameters. Such open communication of uncertainty in decisions improves the defensibility of decisions because it moves discussion from what may or may not be correct to what level of risk can be considered acceptable, which is defined outside the realm of engineering design.

Synthetic Truth

On each baseline field $\tilde{Y}_i^b$, the value of the EPM of concern is computed and denoted EPM_i. Depending on the application, computing the EPM may involve any number of hydrological, geochemical, biological, etc. models. In other words, EPM_i represents the value of the EPM which would occur if $\tilde{Y}_i^b$ were a true representation of the site under consideration. Thus, we can define for each baseline field the baseline indicator variable $I_i^b,\ i = 1, ..., N_Y$ as

$$I_i^b = \begin{cases} 1 & \text{if } EPM_i \in EPM_{critical} \\ 0 & \text{if } EPM_i \notin EPM_{critical} \end{cases} \tag{3.14}$$

which indicates, if baseline field $\tilde{Y}_i^b$ were a true representation of the site under consideration, if H_0^I or H_a^I would be the correct hypothesis.

Simulated Field Campaigns and Decision Making

The second use of each baseline field in the ensemble is to simulate the field campaign and the resulting decision making process. Simulating the field campaign G involves selectively collecting whatever information would be gathered from campaign G, assuming the baseline field $\tilde{Y}_i^b$ to be true, resulting in simulated field data g_i. In the case where the information measured is directly the quantity (or one of the quantities) simulated in Y_i^b, this process involves simply selecting the values corresponding to the grid points corresponding to the measurement locations specified by G. In cases where G calls for e.g. pumping tests, further modeling may be necessary to simulate the information that would be collected. After the simulated field data g_i is collected, we move on to simulated decision making that would result, *using only the information given by* g_i.

In most cases, simulated decision making based on g_i involves many steps, including inferring SRF parameters, distinguishing between conceptual models, forward modeling, etc. The framework allows for any type of conceptual model or combination of conceptual models, as well as any form of inverse modeling for model or parameter inference. The details of the inverse and forward modeling are to be selected for each application, and thus not the subject of discussion here.

After inferring any information about conceptual or statistical models (using only information from g_i), conditional simulation of EPMs is undertaken. Here, it is noted the double usage of field measurements: both for inferring geostatistical (global) parameters, as well as conditioning points in forward modeling. In most cases, this involves simulating an ensemble of conditional fields $\tilde{Y}_k^c,\ k = 1, ..., N_c$, where the superscript c denotes that the field is conditional to both the parameters inferred using g_i, as well as the measured values themselves. Numerical models are then executed with this ensemble of conditional fields. Again, depending on the application, any number of hydrological, geochemical, biological, etc. models may be necessary to compute the relevant EPM. With this ensemble of conditional EPM predictions, we obtain an ensemble of conditional risk indicator variables $I_{ik}^c;\ k = 1, ..., N_c$, determined by the same relationship as equation (3.14). With this ensemble, we can compute the probability that the null hypothesis is true, conditional to the simulated field data as:

$$Pr[I = 1|g_i] = \langle I|g_i\rangle = \frac{\sum_{k=1}^{N_c} I_{ik}^c}{N_c} \tag{3.15}$$

After computation of $\langle I|g_i\rangle$, the decision making process can be simulated by the same criteria given by (3.6), resulting in D_i^g. This simulated decision D_i^g represents what decision we would have made, assuming that Y_i^b were the real baseline field, and that the only information we knew about it was g_i, or the measurements specified by G.

Since the true value of the EPM (and thus the truth regarding H_0^I and H_a^I) is known about this baseline field (equation 3.14), we can now compare the simulated decision about the field to the true nature of the site. Now, for each baseline field $\tilde{Y}_i^b$, $i = 1, ..., N_Y$, we compute $\phi_{\alpha i}^g$ and $\phi_{\beta i}^g$ as:

$$\phi_{\alpha i} = \begin{cases} 1 & \text{if } I_i^b = 1 \cap D_i^g = 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.16}$$

$$\phi_{\beta i} = \begin{cases} 1 & \text{if } I_i^b = 0 \cap D_i^g = 1 \\ 0 & \text{otherwise} \end{cases} \tag{3.17}$$

In other words, if $\tilde{Y}_i^b$ were the actual field, and the field data g_i were the only information we knew about it in addition to prior information, would we make the correct decision? $\phi_{\alpha i}^g$ is equal to one if Type I error would have occurred, and $\phi_{\beta i}^g$ is equal to one if Type II error would have occurred.

3.11 Calculation of Error Probabilities

After following the procedure described by 3.10, 3.10, and 3.10 on all N_Y baseline fields, we can compute $\langle \phi_\alpha^G\rangle$ by equation (3.16) as

$$\langle \phi_\alpha^G\rangle = \frac{\sum_{i=1}^{N_Y} \phi_{\alpha i}^g}{N_Y} = q \tag{3.18}$$

Where $q = \langle \phi_\alpha^G\rangle$ is used as the test statistic to test the null hypothesis presented by 4.4 as

$$\text{reject } H_0^G \text{ if } q \leq \alpha \tag{3.19}$$

$$\text{accept } H_0^G \text{ if } q > \alpha \tag{3.20}$$

where rejection of H_0^G indicates that the field campaign design G is sufficient for testing the hypothesis H_0^I, and acceptance of H_0^G indicates the opposite. This process is summarized by the Figures 3.5 and 3.6. Figure 3.6 describes the treatment of a single baseline field, while 3.5 describes the treatment of the ensemble of baseline fields.

3.12 Conditional Error Probabilities

Other metrics for analyzing the effectiveness of G is the conditional error probabilities, P_α^G and P_β^G. P_α^G is the probability that Type I error occurs, conditional to H_0^I being true. Similarly, P_β^G is the probability that Type II error occurs, conditional to H_a^I being true. These quantities can be calculated by

$$P_\alpha^G = \frac{\langle \phi_\alpha^G \rangle}{\langle I \rangle} \tag{3.21}$$

$$P_\beta^G = \frac{\langle \phi_\beta^G \rangle}{1 - \langle I \rangle} \tag{3.22}$$

and also provide information regarding the relationship between field data and decision making. While conditional error probabilities (P_α^G and P_β^G) are usually the focus in classical hypothesis testing, in water resources management it makes sense to focus on the error occurrence probabilities ($\langle \phi_\alpha^G \rangle$ and $\langle \phi_\beta^G \rangle$). This is because in some cases, it may be practically impossible to predict an event of extremely low probability (e.g. a very early arrival time). In this scenario, the probability of occurrence would be very low, but the conditional error probability would be very high due to its conditional nature. In other words, if the risk-posing event has an exceptionally low probability of occurrence, no amount of field data will enable the managers to predict this event, which would prevent any course of action from being acceptable as indicated by conditional probabilities. This effect is demonstrated in section 4.4.

Analogy to Classical Hypothesis Testing

An analogy to hypothesis testing in quality control helps to demonstrate the similarities and differences between this framework and classical hypothesis testing. A manufacturer buys chips from a supplier, who claims that no more than 10% of the chips are defective. To test the supplier's claim the manufacturer defines the null hypothesis to be the supplier's claim, with the alternative hypothesis indicating that the claim is false. This definition of hypotheses follows convention as the manufacturer seeks to test the supplier's claim, as classical hypothesis testing is a test of the null. The hypotheses are defined as below, in terms of the proportion q of defective chips and $q_0 = 0.1$, the value claimed by the supplier.

$$H_0 : \ q \leq q_0 \tag{3.23}$$

$$H_a : \ q > q_0 \tag{3.24}$$

The manufacturer tests n chips, of which m are found to be defective. The test statistic $\hat{q} = m/n$, is computed, and H_0 is either accepted or rejected based on q_0 and $\hat{q}$, along with whatever acceptance criteria the manufacturer selects.

In water resources management, the "chip" to be tested is the site being modeled, and the chip is defective if the EPM is in the critical range, i.e. the "risky" scenario exists ($I = 1$).

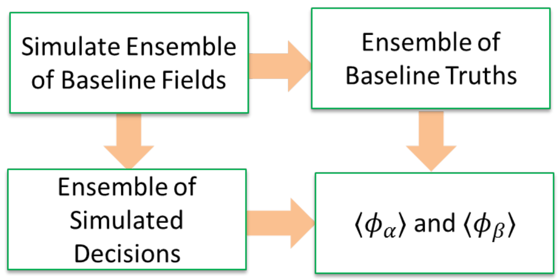

Figure 3.5: Flowchart of overall procedure computing error probabilities starting with prior information about geostatistical parameters, simulation of baseline fields and decisions, as well as computation of error probabilities. In many cases, simulation of an ensemble of geostatistical parameters is necessary in order to create an ensemble of baseline fields which accoutns for this variability. A more detailed description of the treatment of a single baseline field is provided by Figure 3.6.

The difference is that in water resources management, there is only one chip. Through the field and modeling campaigns we seek to determine if it is defective or not, rather than determining a proportion of many chips which are defective. The challenge, however, is that the chip cannot be directly tested for defectiveness. Instead, we collect information about the production of the chip (field data) and simulate the production of many chips (fields) matching the characteristics determined from the data. Then, the simulated chips (fields) are tested for defectiveness (via numerical modeling). Then, the test statistic is the proportion of simulated chips which were defective. This test statistic can be interpreted as the probability that the one real chip (field) is defective.

The information about the production of the chip corresponds to the collected field data, and the simulated production and testing of chips corresponds to conditional simulation of the site and solving the relevant models for the EPM of concern.

Another difference is that there is no supplier with a claim about the proportion, q_0, but rather we have the critical value of the EPM, which is defined by considerations external to the system being modeled. For example, maximum allowable concentration would be defined by regulation. Since field data can be costly to acquire, the goal of the manager is then to carefully select the appropriate amount and type of information which will most effectively enable decision making about the site.

3.13 Summary

This chapter has introduced a framework for rational, risk-based decision making in water resources management, policy, and regulation and field campaign design. The framework explicitly acknowledges that absolute certainty is impossible and enables straightforward management of uncertainty and risk in water resources decision making. Regulators and

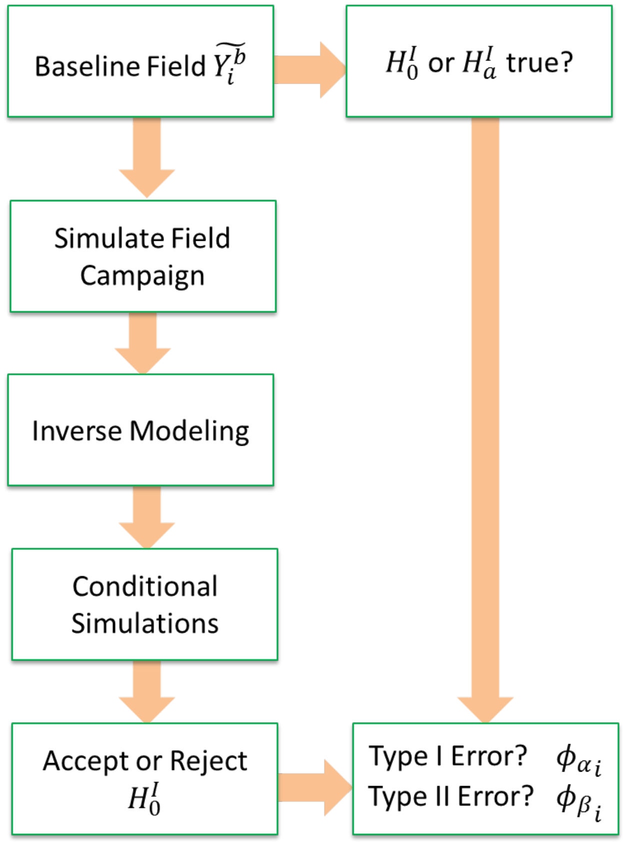

Figure 3.6: Flowchart describing the computational procedure for each baseline field. For a descripiton of the ensemble baseline fields, see Figure 3.5.

policy makers are able to define, in whatever way is deemed appropriate, an acceptable level of risk in management decisions. Managers and practitioners are able to simply demonstrate, even if the wrong decision was reached, that all necessary steps were taken to make a decision meeting the uncertainty levels deemed acceptable.

The framework presented here is general–it can accommodate any number of hydrogeological, biological, geochemical, etc. conceptual models and be used with any type of field data acquisition methods and inverse modeling methods. It is reiterated that a *correct* decision is not guaranteed, but rather enables demonstration that a decision is *justifiable*. Similarly, the method does not design a field campaign–rather, that is left to practitioners with experience in field methods and local hydrogeological conditions. What the framework does is allows practitioners to take a proposed field campaign design and determine whether or not this design will provide enough information to enable successful decision making.

Another beneficial feature of the framework is that it allows for simple communication of uncertainty and risk. Instead of focusing on geostatistical model uncertainty or parametric uncertainty or other concepts unfamiliar to stakeholders outside of hydrologists, the framework allows for simple communication of uncertainty in the ultimate EPM predictions on which the management decision depends.

Chapter 4

Case Study: Contaminant Arrival Time

The framework was applied in a synthetic case study where the goal is to predict the arrival time, τ, of a contaminant plume in an aquifer. Specifically, the risky scenario would arise if the contaminant plume arrives at a target before a critical amount of time, τ_{crit}, passes. This scenario could arise in many applications, such as evaluating locations for waste disposal sites or in assessing the risk posed by a plume to a nearby supply well. Since early arrivals are of concern, the indicator variable is

$$I = \begin{cases} 1 & \text{if } \tau \leq \tau_{crit} \\ 0 & \text{if } \tau > \tau_{crit} \end{cases} \tag{4.1}$$

which allows us to define the null and alternative hypotheses:

$$H_0^I : \ I = 1 \tag{4.2}$$

$$H_a^I : \ I = 0 \tag{4.3}$$

as indicated in Chapter 3. In this case study, only Type I error is of concern, so the hypotheses regarding the field campaign designs are given by

$$H_0^G : \ \langle \phi_\alpha^G \rangle \geq \alpha \tag{4.4}$$

$$H_a^G : \ \langle \phi_\alpha^G \rangle < \alpha \tag{4.5}$$

and the level of significance α is 0.05.

4.1 Statistical & Physical Setup

Aquifer flow was simulated in a 2-D planar (x, y) rectangular domain, with constant head boundary conditions along the two boundaries parallel to the y-axis, and no flow conditions

Scenario	Mean of Log-conductivity μ_Y	Variance σ_Y^2	Integral Scale I_Y
#1 (deterministic)	-5.5	0.55	4.5
#2	Uniform[-6, -5]	Uniform[0.1, 1]	Uniform[3,6]
#3	Uniform[-7, -4]	Uniform[0.1, 1]	Uniform[3,6]

Table 4.1: Three prior information scenarios for the geostatistical parameters μ_Y, σ_Y^2, and I_Y used in the case study.

along the boundaries parallel to the x-axis. Flow is uniform-in-the-average in the positive x-direction. Porosity, $n = 0.10$ was assumed known and homogeneous. We assume the contaminant to have originated from an instantaneous point release, where the time and location of the release are assumed to be known.

Steady-state groundwater flow was modeled by

$$\nabla \cdot [K(\mathbf{x})\nabla H(\mathbf{x})] = 0 \tag{4.6}$$

$$v(\mathbf{x}) = \nabla H(\mathbf{x})/n \tag{4.7}$$

and unconditional contaminant arrival time was computed with the methods of *Schlather et al.* (2017) and *Pollock* (1988). Conditional contaminant arrival times were computed with the methods described by *Rubin* (1990) and *Rubin* (1991).

A set of field measurements were taken to characterize the natural logarithm of hydraulic conductivity, $Y = ln(K)$, which was modeled as a SRF with a multivariate Normal distribution and exponential covariance. The measurements were used to estimate the parameters $\theta = (\mu_Y, \sigma_Y^2, I_Y)$ where μ_Y, σ_Y^2, and I_Y represent the mean, variance, and integral scale, respectively. The measurements were also used for conditioning points in forward modeling.

To investigate the effect of prior information, the case study was executed with three alternative scenarios regarding prior distributions. In scenario one, the SRF parameters θ were assumed to be deterministically known with $(\mu_Y, \sigma_Y^2, I_Y) = (-5.5, 0.55, 4.5)$. In this scenario, no inference of θ was necessary, and the measurements served only as conditioning points. In scenarios two and three, all three parameters were assumed to be distributed uniformly and independently of each other. In scenario two, μ_Y is distributed uniformly in the range $[-6, -5]$. In the second scenario, μ_Y is distributed uniformly in the range $[-7, -4]$. In both scenarios 2 and 3, σ_Y^2 and I_Y are distributed uniformly in the ranges $[0.1, 1]$ and $[3, 6]$, respectively. The three prior information scenarios were chosen to range from deterministic knowledge (first scenario), to a relatively informative probabilistic description (second scenario), to a relatively uninformative probabilistic description (third scenario) of the SRF parameters. This is to allow a closer examination of the relationship between parametric, travel time prediction, and decision making accuracy.

4.2 Field Campaign Setup

Field campaigns to be tested were designed with $n = 4, 8, 16, 32$ measurements. For each n, two alternative designs were tested. One configuration had measurements spread throughout the domain and covering various lag distances with the idea of improving estimates of the SRF parameters. The other configuration had all measurements located in the likely area of the travel path. The likely area of the travel path was determined by simulating an ensemble of particle paths conditional only to the prior information, and lateral displacement was plotted versus distance from the point source, and results are shown in Figure 4.1. The locations of the measurements for all field campaign designs $G_j,\ j = 1, ..., 8$ are shown in Figure 4.2.

4.3 Monte Carlo Methodology

In accounting for uncertainty in θ, Latin Hypercube integration was used in order to reduce the computational burden. For each realization of θ, traditional Monte Carlo sampling was used to simulate the baseline fields from the distribution $f(Y|\theta)$.

Baseline Fields and Simulated Field Campaigns

For each baseline field, the travel time was computed deterministically and N_Y values of I^b were computed via (5.1). After recording the deterministically known travel time, the field campaign was simulated by recording the values of hydraulic conductivity at the locations specified by the field campaign design. After this simulated field data g was collected, the measurements were used to compute the *maximum a posteriori* value for the geostatistical parameters, similar to the maximum likelihood method presented by *Kitanidis and Lane* (1985), but with bounds provided by the prior distributions. After θ_{MAP} was computed, the conditional distribution of travel time $f(\tau^c|\theta_{MAP}, g)$ was computed using semi-analytical particle tracking (*Rubin*, 1991).

Simulated Decision Making

After computing N_c realizations of τ^c, N_c realizations of I^c were computed by (5.1) and in turn, $\langle I|g\rangle$ was computed via (3.15). In turn, H_0^I was accepted or rejected by (3.6). After repeating on all baseline fields and for all field campaign designs, H_0^G was accepted or rejected for each campaign based on (3.19) and (3.20).

4.4 Case Study Results and Discussion

This section presents results of the case study described in this Chapter. The results focus on the effects of the critical value of travel time, measurement configurations, and

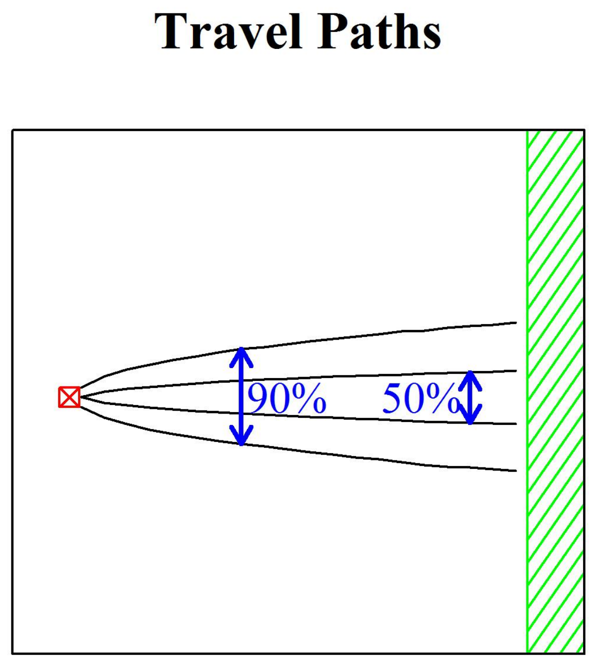

Figure 4.1: Area of travel paths illustrated by plotting the 50% and 90% quantiles of lateral displacement versus longitudinal displacement. The contaminant source is indicated by the red square, and the green rectangle indicates the control plane defining the environmentally sensitive target. The travel path quantiles were computed by simulating using unconditional particle trajectories.

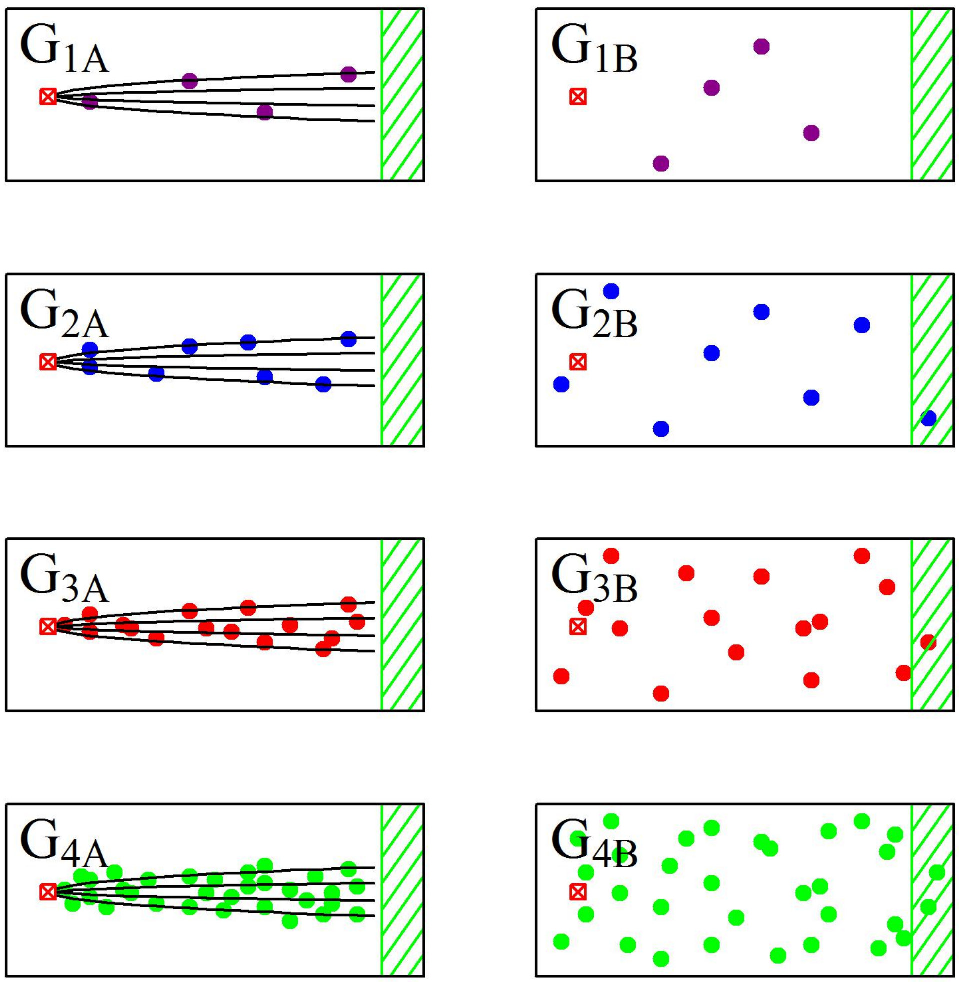

Figure 4.2: The 8 campaign designs G to be tested. Left-hand-side (subscript A): measurements concentrated along travel path. Right-hand-side (subscript B): measurements spread throughout the domain. $N = 4, 8, 16, 32$ corresponding to subscripts $1, 2, 3, 4$, respectively.

parametric uncertainty, both *a priori* and *a posteriori*.

Effect of τ_{crit}

The results from the case study are shown in several figures. Due to the reasons described above, we focus on the behavior of $\langle \phi_\alpha^G \rangle$ and P_α^G, though analogous conclusions can be made regarding $\langle \phi_\beta^G \rangle$ and P_β^G. The first thing to notice is that, regardless of the quantity or spatial configuration of measurements, $\langle \phi_\alpha^G \rangle$ and P_α^G are highly sensitive to τ_{crit}, as shown in Figures 4.4 and 4.5. In turn, whether or not H_0^G can be rejected and G deemed adequate is dependent on τ_{crit}. Figure 4.3 shows for each measurement configuration and each prior information scenario, the values of τ_{crit} for which H_0^G can be rejected.

For large and small values of τ_{crit}, we see $\langle \phi_\alpha^G \rangle$ approaching zero, for all measurement configurations, indicating very low occurrence of Type I error in these regions. To understand why, we examine the value of $\langle I \rangle$ for these regions, shown in Figure 4.5. By (5.1), $\langle I \rangle$ as a function of τ_{crit} corresponds directly to the cumulative distribution function of arrival time. Thus, Type I error is very unlikely for relatively small values of τ_{crit} due to the low probability of H_0^I being true in this region. For relatively large values of τ_{crit}, on the other hand, Type I error is unlikely due to the relatively small probability that H_0^I is true and is thus easier to correctly predict. Where Type I error is more likely, then, is where H_0^I is somewhat likely to be true, but more difficult to correctly predict–the intermediate portion of the cdf of arrival time.

The behavior of $\langle \phi_\alpha^G \rangle$ and $\langle I \rangle$ with varying τ_{crit} explain the behavior of P_α^G, recalling the definition (3.21). Figure 4.5 shows graphically the relationship between $\langle I \rangle$, $\langle \phi_\alpha^G \rangle$, and P_α and also highlights the difference between using occurrence probability and conditional probability to describe the effectiveness of field data. Using conditional probability, it is impossible to have a field campaign that would meet uncertainty criteria, while using occurrence probability takes into consideration the very unlikely nature of the earliest arrivals and does not penalize the field data for not enabling prediction of these events.

As we have seen, the effectiveness of a field campaign and the success of decision making is highly dependent on the value of τ_{crit}. The practical implication of this result is that for effective, goal-oriented characterization, the field campaign should be tailored to not just the EPM to ultimately be predicted but also to the critical value of this EPM on which decision making depends.

Effect of Measurement Configurations

Figures 4.7 and 4.8 show the root mean square error (RMSE) resulting from all of the SRF parameter estimates for all baseline fields for all measurement configurations. The parameter estimates improve with increasing quantity of measurements. For any given number of measurements, the error in parameter estimates was less for the measurements spread throughout the domain than for the measurements focused along the travel path.

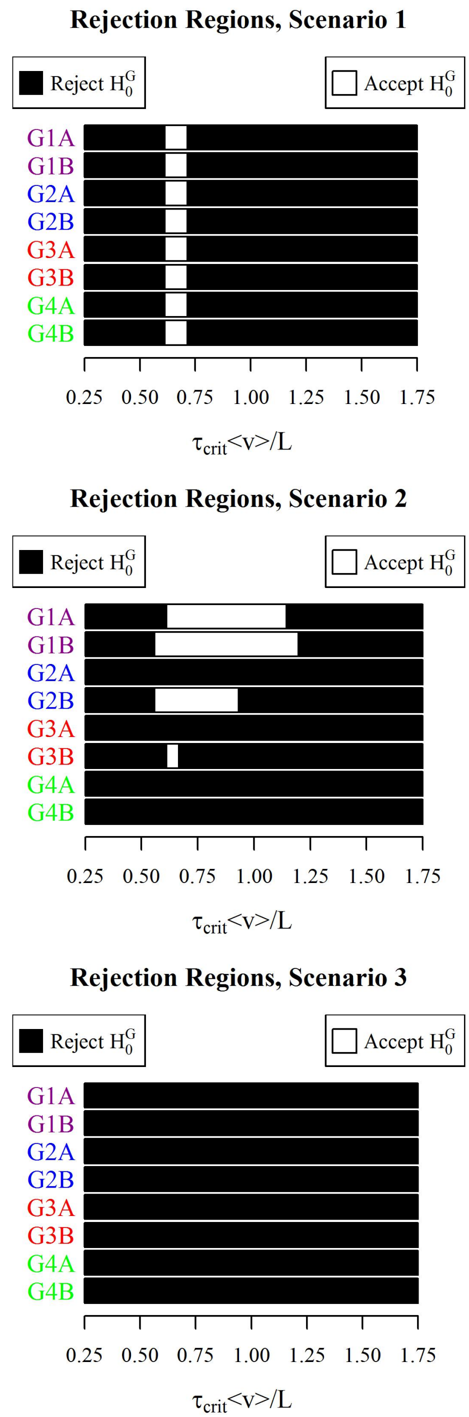

Figure 4.3: Rejection regions with respect to τ_{crit} (non-dimensionalized). "Reject" means that H_0^G was rejected in favor of H_a^G. In other words, rejection indicates that the specified field campaign design is sufficient. The rejection region for each campaign corresponds to the values of τ_{crit} for which $\langle \phi_\alpha^G \rangle$ exceeds α, as indicated by Figure 4.4. The region for which H_0^G is accepted is indicated by absence of shading. Prior Information Scenarios 1,2, and 3, respectively. (Scenarios described in section 4.1)

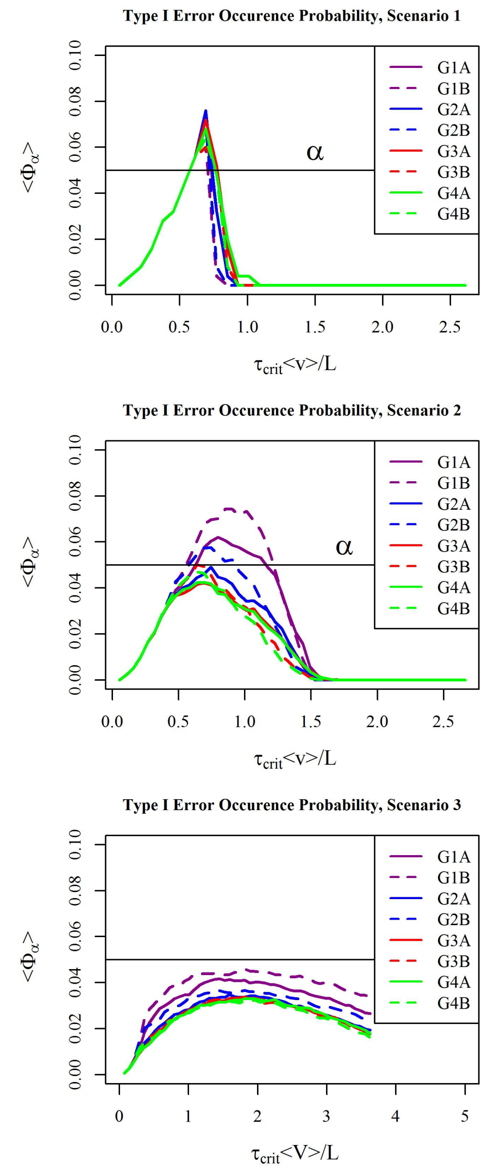

Figure 4.4: $\langle\phi_\alpha\rangle$ resulting from all 8 campaign designs, plotted against τ_{crit}, nondimensionalized by the travel length and the average velocity. As we can see, $\langle\phi_\alpha\rangle$ approaches zero as τ_{crit} appraoches zero, and also as τ_{crit} gets large, for all measurement configurations. We also see that increasing the quantity of measurements improves performance, with diminishing marginal returns. Furthermore, for a given quantity of measurements, the configuration with measurements focused along the travel path (A) performed better than the configuration with measurements spread throughout the domain (B). Prior Information Scenarios 1,2, and 3, respectively. (Scenarios described in section 4.1)

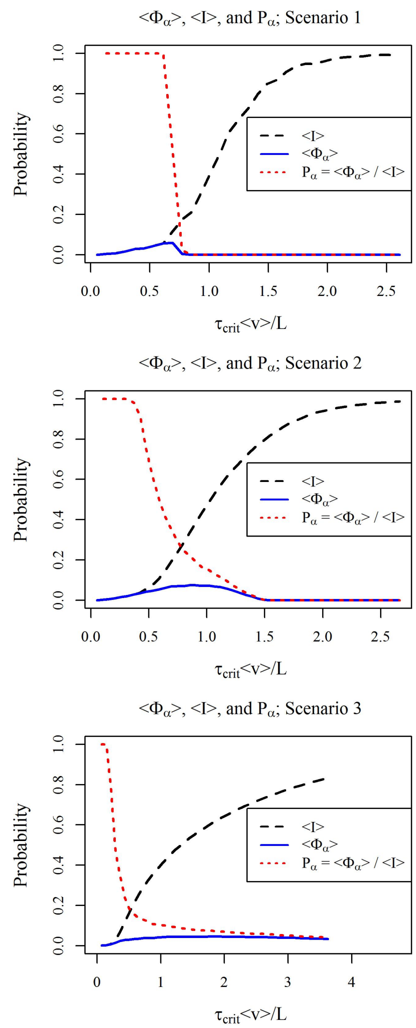

Figure 4.5: P_α^G, $\langle\phi_\alpha^G\rangle$, and $\langle I\rangle$ plotted against τ_{crit}, which is nondimensionalized by the travel path length and the average velocity. $\langle I\rangle$ coincides with the cumulative distribution function of τ due to the definition of I (equation 5.1). $\langle\phi_\alpha^G\rangle$ represents the probability that H_0^I is true and H_0^I is rejected, which happens most often when τ_{crit} is near $\langle\tau\rangle$. P_α^G, on the other hand, is defined as the probability that H_0^I is rejected, conditional to H_0^I being true, and is thus equal to $\frac{\langle\phi_\alpha\rangle}{\langle I\rangle}$. Due to this, P_α approaches one as τ_{crit} approaches zero. What this indicates is that as τ_{crit} decreases, the probability of H_0^I being true approaches zero. As this event becomes more unlikely, it becomes nearly impossible to predict. Thus, the occurrence probability $\langle\phi_\alpha^G\rangle$ remains small while the conditional probability P_α^G becomes large. Prior Information Scenarios 1,2, and 3, respectively. (Scenarios described in section 4.1)

Figure 4.3 shows the regions, with respect to τ_{crit}, in which H_0^G is rejected for the $N_G = 8$ measurement configurations. Focusing for the moment on prior information scenario 2, we notice two clear patterns: 1) larger quantities of measurements are adequate (H_0^G is rejected) for a greater range of τ_{crit}, and 2) given a specific quantity of measurements, the configuration with measurements focused along the travel path (A) outperforms the configuration with measurements spread throughout the domain, despite worse performance in estimating the SRF parameters.

While summarizing a field campaign design in terms of rejection of H_0^G or not is a useful tool for managers and practitioners, a more thorough description of the performance of a campaign can be provided by analyzing $\langle \phi_\alpha^G \rangle$, which indicates the probability that Type I error will occur. For prior information scenarios 1 and 3, Figure 4.3 does not provide much information regarding the relative performance of the different measurement configurations, so we instead look to Figure 4.4. For scenario 1, there is little difference between the behavior of the different configurations, stemming from the unrealistic assumption that the SRF parameters are known deterministically. For the relatively uninformative prior information (scenario 3), we see the same patterns as for scenario 2: increasing quantity of measurements improves performance, and measurements focused along the travel path are better for predicting earlier arrivals, despite worsened performance in estimating SRF parameters.

The practical implications of these results, again, are emphatic of the need for goal-oriented characterization design. Designing field campaign strategies to optimize performance in estimating SRF parameters is clearly not the best approach, as it is shown that improved parameter estimates do not necessarily indicate improved decision making performance. In addition to designing field campaigns tailored to predicting a specified EPM (e.g. arrival time), it is also necessary to take into consideration the critical value of this EPM, the threshold for decision making.

Effect of Parametric Uncertainty

The differing results from the three scenarios described in section 4.1 highlight the effect of prior information regarding the SRF parameters in predicting early arrival times. Assuming the prior information to be correct, a more diffuse prior (Scenario 3) allows a wider range of τ, which thus increases the range of τ_{crit} for which $\langle \phi_\alpha^G \rangle$ is non-zero. Figure 4.5 shows how the different amounts of prior information affect the behavior of $\langle I \rangle$, $\langle \phi_\alpha^G \rangle$, and in turn P_α^G. What we see is that a more diffuse knowledge of the SRF parameters (particularly μ_Y), leads to a more diffuse $f_\tau(\tau)$, and in turn, a more gradual change in $\langle I \rangle$ with respect to τ_{crit}. This in turn leads to different behavior of $\langle \phi_\alpha^G \rangle$, which is better shown by Figure 4.4, which shows a much wider range of non-zero $\langle \phi_\alpha^G \rangle$ for scenario 3, followed by scenario 2 and scenario 1.

One interesting result is that the peak values of $\langle \phi_\alpha \rangle$ are higher for the more informative prior distributions, despite having much smaller ranges of non-zero values. One cause of this is the bounded nature of the *maximum a posteriori* parameter estimation method.

Early (relative to $F_\tau(\tau)$) arrivals are associated with high permeability (relative to $f_{\mu_Y}(\mu_Y)$ values). With a very informative, or deterministically known, prior distribution for μ_Y, the conditional distribution $f_\tau(\tau|g,\hat{\theta})$ will change relatively little compared to the unconditional distribution. With a more diffuse prior $f_{\mu_Y}(\mu_Y)$, the inferred value $\hat{\mu_Y}$ can be estimated to higher values, thus allowing greater divergence between the unconditional and conditional distributions of τ, in turn allowing for improved prediction of early arrivals associated with high-K measurements.

In other words, early arrivals are most often associated with fields which have high permeability values along the travel path. If measurements indicate high permeability values, a less informative prior will cause a greater estimate of μ_Y (perhaps even overestimating), which improves the chances of correctly predicting the early arrival. The *maximum a posteriori* method may not always be the best method for inverse modeling, but it was selected for this case study due to its ability to be executed in a completely automated fashion, enabling computation on such a large number of baseline fields. What this result highlights, however, is the value of subjective interpretation of measurements and manual involvement in inverse modeling. In other words, a fully automated *maximum a posteriori* inversion fails to subjectively analyze the field data or perhaps reconsider prior information if field data casts doubt on its validity. The failure of the method to do this may contribute to these results. A question for future research would be to investigate how this effect changes with more sophisticated inversion methods.

Uncertainty in Results

For a sense of the uncertainty in the results, we look at the standard deviation of $\langle\phi_\alpha^G\rangle$, $\langle I\rangle$, and P_α^G, shown in Figure 4.6. What we see is that the standard deviation of both $\langle\phi_\alpha^G\rangle$ and $\langle I\rangle$ are maximal near the mean arrival time (when $\tau_{crit}\langle v\rangle/L$ is near one). The reason for this is that the majority of actual arrival times is near this value, making it more difficult to predict the binary outcome given by equation (5.1). On the other hand, the standard deviation of P_α^G is at its peak when τ_{crit} is equal to roughly $0.5\frac{\langle v\rangle}{L}$, meaning that the conditional error probability is most uncertain when τ_{crit} is about half of the expected value of arrival time. The practical implication of this result is that even before simulation is executed, a rough idea of how difficult the predictions will be can be provided by τ_{crit}, $\langle v\rangle$, and L. If $\tau_{crit}\langle v\rangle/L$ is close to zero, then it can be reasoned that H_0^I is unlikely to be true, but if it is, it will be difficult to detect. Conversely, if $\tau_{crit}\langle v\rangle/L$ is much greater than one, it can be reasoned that H_0^I is likely to be true, and it will be easy to detect.

4.5 Summary

The framework presented in Chapter 3 was demonstrated using a case study in predicting early contaminant arrival time in an aquifer, and several conclusions were drawn. It was shown that improved estimates of geostatistical parameters are not necessarily associated with improved water resources decision making, demonstrating the importance of designing

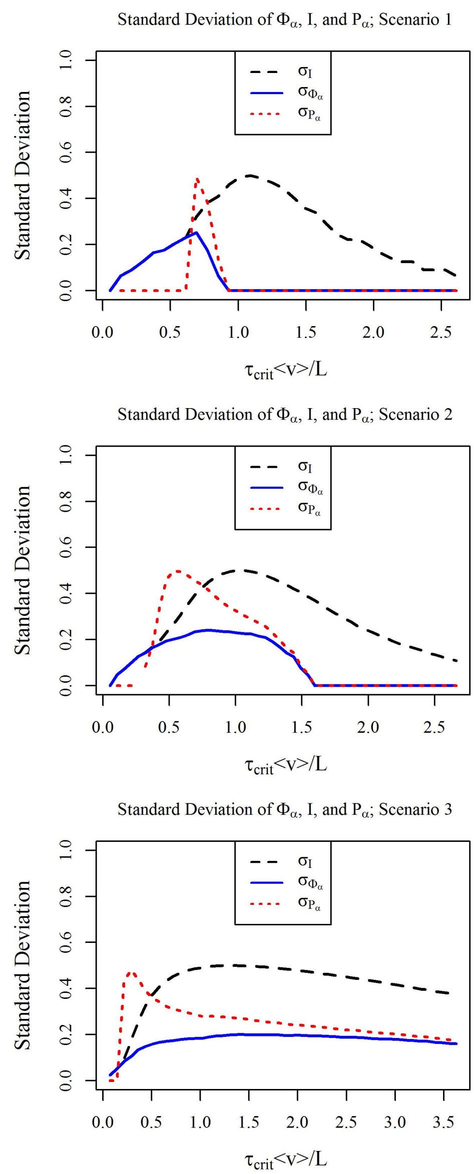

Figure 4.6: Standard deviation of I, ϕ_α, and P_α, plotted against non-dimenisonalized τ_{crit}. The figure shows that uncertainty in I and ϕ_α is highest when $\tau_{crit}\langle v\rangle/L$ is near one, while uncertainty in P_α is highest for lower values of τ_{crit}. Prior Information Scenarios 1,2, and 3, respectively. (Scenarios described in section 4.1)

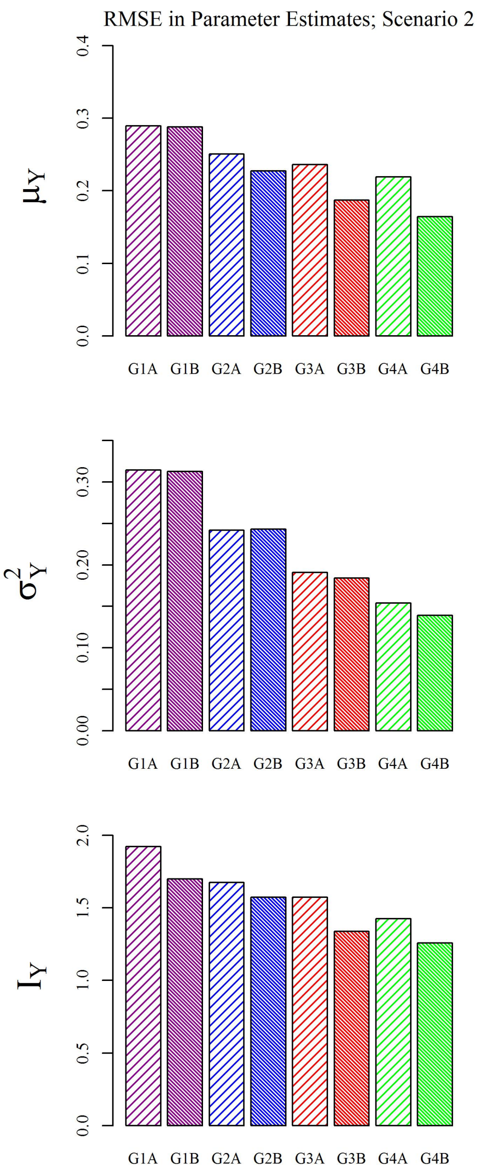

Figure 4.7: Root mean square error in estimating μ_Y, σ_Y^2, and I_Y resulting from each measurement configuration, for Scenario 2. As we can see, the estimates improve with increasing number of measurements, and measurements spread throughout the domain (denoted B) performed better than the configurations with measurements concentrated along the travel path (denoted A).

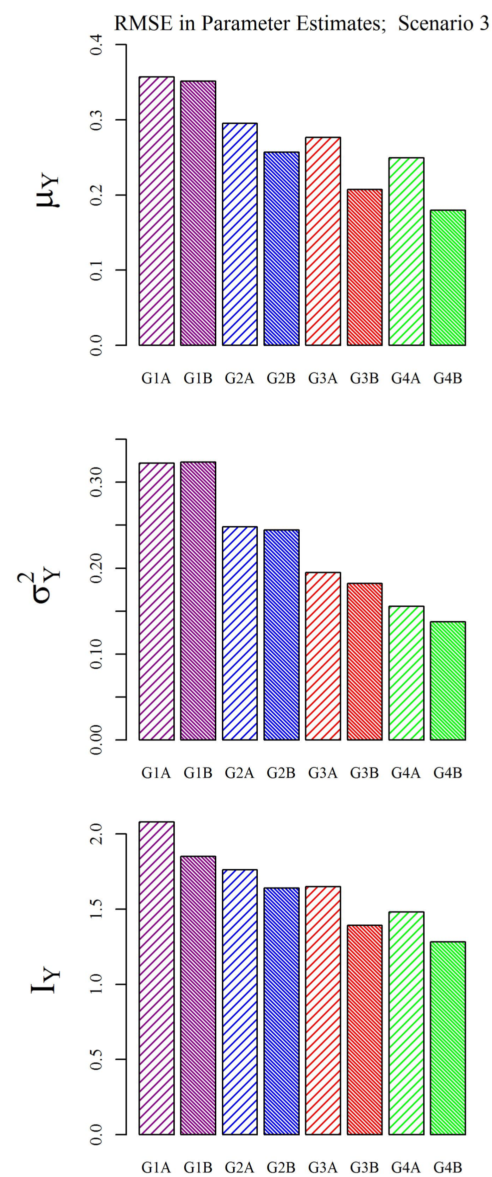

Figure 4.8: Root mean square error in estimating μ_Y, σ_Y^2, and I_Y resulting from each measurement configuration, for Scenario 3. As we can see, the estimates improve with increasing number of measurements, and measurements spread throughout the domain (denoted B) performed better than the configurations with measurements concentrated along the travel path (denoted A).

field campaigns with the goal of making defensible management decisions, as opposed to optimal parameter estimates.

It was shown that the amount of field data necessary to make a decision must be determined on a case-by-case basis. The critical value of the EPM on which the decision depends, and also the amount of prior information available about the site can significantly affect the amount of field data which is necessary. This further highlights the importance of goal-oriented characterization design, which is important in light of the costs associated with site characterization methods. The methods presented here can be utilized by managers to prevent over-spending on unnecessary amounts of field data and also ensure that measurements are strategically placed in order to ensure the maximum benefit of the data.

Chapter 5

Case Study: Increased Cancer Risk

This chapter presents a case study in which the decision making framework presented in Chapter 3 is demonstrated in an application. The application is chosen to emulate a scenario where a water resources decision must be made based on whether or not a contaminated groundwater site poses a risk to the health of a nearby population. The EPM of concern in this case study is Increased Lifetime Cancer Risk (ILCR). Prediction of ILCR requires knowledge of the hydrogeological characteristics of the site, toxicological characteristics of the contaminant, and physiological characteristics of the population.

5.1 Introduction

One reason for concern regarding groundwater contamination is its potential to affect nearby populations. As noted in Chapter 1, numerous people around across the United States and around the world rely on groundwater for drinking water. Contaminants can come from many sources, natural and anthropogenic, and if exposed to humans, can affect health in many different ways. Many organic contaminants can have carcinogenic effects on humans (*U.S. Environmental Protection Agency*, 1997). Guidelines for predicting the extent to which these effects can increase the risk of a human developing cancer throughout a lifetime have been presented by the United States Environmental Protection Agency (*U.S. Environmental Protection Agency*, 1989). The extent to which a carcinogenic contaminant increases the health risk of an individual is a function of the amount of the contaminant to which an individual is exposed and also the toxicity of the contaminant (e.g. *de Barros and Rubin* (2008). Groundwater contaminants can be exposed to humans via water pumped from groundwater wells. In turn, individuals can be exposed to contaminants by either dermal contact with the contaminated water, inhalation of volatilized contaminants, or by ingestion (*Maxwell et al.*, 1998a). Thus, the exposure to carcinogenic contaminants is determined by both the concentration in the water supply wells and also various behavioral components of the individuals, such as the amount of water ingested. The toxicity of the contaminant is quantified by a dose-response curve and a slope factor. The slope factor represents an estimate of the probability of an individual developing cancer over a lifetime of exposure to a given contaminant at a given level. These quantities must be estimated for each contaminant, and are determined based on a combination of data sources, depending on availability. Types of information which can aid the estimation of the dose-response curve or the slope factor can be information from epidemiological studies with human data, or toxicological studies with animal testing or other types of information such as research with cell cultures (*U.S. Environmental Protection Agency*, 2000).

In light of threats from groundwater contamination (*DeSimone et al.*, 2014), water resources managers may be tasked with e.g. identifying water sources which will not increase the local population's risk of developing cancer or determining whether or not to remediate a contaminated site. Often these decisions will be made based on whether or not the enhanced cancer risk of the population will exceed some threshold value, which are determined based on health effect and established by regulations (*U.S. Environmental Protection Agency*, 1989). Given the numerous components that comprise estimation of this risk, computing enhanced cancer risk can be a challenging task, prone to significant uncertainty (*Maxwell et al.* (1998b), *de Barros and Rubin* (2008)). To reduce the uncertainty in cancer risk predictions, many types of information can be obtained. For example, more information could be gathered regarding the hydrogeological characteristics, or more research could be devoted to epidemiological or toxicological laboratory studies to reduce uncertainty in the dose-response estimations. Given the time and costs associated with this information gathering however, some strategy which determines the most effective type of information (or combination of multiple types) can help managers prioritize the information which will best enable successful decision making. *de Barros and Rubin* (2008), *de Barros et al.* (2009), and *de Barros et al.* (2012) explored this topic and presented methods of determining marginal relative uncertainty reduction caused by additional hydrogeological and physiological information. This was done by comparing the reductions of uncertainty in final prediction of enhanced cancer risk, quantified by either variance or entropy. As an alternative, this Chapter presents a framework for evaluating proposed data acquisition strategies by evaluating their ability to enable successful decision making, as presented in Chapter 3.

5.2 Decision Making

The case study is intended to simulate a scenario where some water resources management decision must be made based on whether or not increased lifetime cancer risk (ILCR), denoted r, will exceed some threshold value, denoted r_{crit}. This scenario could arise, for example, when determining whether or not a contaminated site near a water supply well should be remediated, or in determining if alternate water supplies should be procured due to threats from contamination.

To illustrate the use of the proposed decision making framework, we investigate the risk of exposure to a carcinogenic substance. Therefore, the EPM of interest is ILCR (*U.S. Environmental Protection Agency*, 1989). Based on ILCR, r, and its threshold value r_{crit}, the risk indicator variable I is defined as

$$I = \begin{cases} 1 & \text{if } r > r_{crit} \\ 0 & \text{if } r \leq r_{crit} \end{cases} \tag{5.1}$$

where a value of one for I indicates that ILCR indeed exceeds the threshold value, and a value of zero indicates the contrary.

With the indicator risk variable, the null and alternative hypotheses are defined as before.

$$H_0^I : \; I = 1 \tag{5.2}$$

$$H_a^I : \; I = 0 \tag{5.3}$$

ILCR at levels below the critical value is the desirable scenario. In line with the conventions discussed in Chapter 3, this then forms the alternative hypothesis.

To aid in decision making via reduction in uncertainty, some amount of information is collected, which is denoted $g = \{g^H, g^\beta\}$ where g^H indicates hydrogeological information and g^β indicates information regarding toxicity of the contaminant or behavioral or physiological characteristics of the local population. Provided with this information, the decision to be made can be formulated in terms of accepting or rejecting the null hypothesis H_0^I. Recalling the decision criteria from Chapter 3, the definition of D^g remains the same:

$$D^g = \begin{cases} 1 & \text{if } \langle I|g \rangle \geq \alpha \text{ (accept } H_0^I) \\ 0 & \text{if } \langle I|g \rangle < \alpha \text{ (reject } H_0^I) \end{cases} \tag{5.4}$$

where a value of 1 for D^g indicates that the null hypothesis H_0^I was accepted based on the information contained in g, and 0 indicates rejection of H_0^I.

Data Acquisition

In this scenario, the goal is to test H_0^I with the defined level of significance α. The question, as before, is to determine how much information is necessary to appropriately test H_0^I with sufficiently low risk of making an incorrect decision. To begin, some number of data acquisition strategies $G_i; \; i = 1, ..., N_G$ are proposed, where upper case $G = \{G^H, G^\beta\}$ is used before the data is collected as it is treated probabilistically. The goal is then to determine, for each of these strategies, whether the information provided will enable successful decision making, defined by probability of Type I error being below the level of significance. While *de Barros and Rubin* (2008) and *de Barros et al.* (2012) presented a means for evaluating the relative effects of hydrogeological versus non-hydrogeological data acquisition, uncertainty in model predictions was used as a means for comparison. As the previous chapter showed, reduced uncertainty does not necessarily imply improved decision making so the goal in this chapter is to evaluate the relative impact of alternative types of information with respect to decision making.

The goal, then, is to design a data acquisition strategy which results in a sufficiently low probability of erroneous decisions. We recall from Chapter 3 the definition of decision risk

$$R^G = w_\alpha \langle \phi_\alpha^G \rangle + w_\beta \langle \phi_\beta^G \rangle \tag{5.5}$$

where w_α and w_β are weights to be defined based on economic, and political considerations. The expected values $\langle \phi_\alpha^G \rangle$ and $\langle \phi_\beta^G \rangle$ indicate the probability of occurrence of Type I and Type II errors, respectively. These indicator variables are defined as

$$\phi_\alpha^g = \begin{cases} 1 & \text{if } D^g = 0 \cap I = 1 \\ 0 & \text{otherwise} \end{cases} \tag{5.6}$$

$$\phi_\beta^g = \begin{cases} 1 & \text{if } D^g = 1 \cap I = 0 \\ 0 & \text{otherwise} \end{cases} \tag{5.7}$$

and can be computed for every baseline field. When averaged over an ensemble of baseline fields, they indicate the probability of occurrence of these types of errors.

5.3 ILCR Definition

For a carcinogenic contaminant, the ILCR posed to a population r can be modeled as (*U.S. Environmental Protection Agency* (1989), *Maxwell et al.* (1999), *de Barros and Rubin* (2008)):

$$r = 1 - Exp[-ADD_M \times CPF_M] \tag{5.8}$$

which, for small values of r can be approximated as

$$r \approx ADD_M \times CPF_M \tag{5.9}$$

where CPF_M is the metabolized cancer potency factor and ADD_M is the average daily dose (*Maxwell et al.* (1998a), *Maxwell and Kastenberg* (1999), *de Barros and Rubin* (2008)). In general, the average daily dose is the sum of the quantities metabolized via inhalation, ingestion, and dermal exposure.

The focus of the present work is ingestion, and thus inhalation and dermal exposure are neglected and ADD_M is given by

$$ADD_M = f_{mo} \times C_f \times \frac{IR}{BW} \frac{ED \times EF}{AT} \tag{5.10}$$

where C_f is the flux-averaged concentration at the control plane defining the water supply area, and is determined by hydrogeological processes. The other parameters represent some quantity related to the toxicity of the contaminant, the physiological characteristics of the population, or the behavioral characteristics of the population. These terms are defined in

Non-hydrogeological parameters	
IR	Ingestion Rate (l/d)
ED	Exposure Duration (y)
EF	Daily Exposure Frequency (d/y)
BW	Body Weight (kg)
AT	Average Expected Lifetime (y)
Sf_0	Metabolized fraction of carcinogen
f_{mo}	Cancer Potency Factor

Table 5.1: Non-hydrogeological quantities necessary to predict enhanced cancer risk to a population from a nearby groundwater contaminant plume.

Table 5.1. Each of these parameters may be variable among individuals in the population and may be uncertain due to lack of characterizing information. C_f can be evaluated as

$$C_f(\mathbf{x}_{cp}, t) = \frac{Q(\mathbf{x}_{cp}, t)}{Q_w(\mathbf{x}_{cp})} \tag{5.11}$$

where $\mathbf{x}_{cp}$ defines the location of the control plane, $Q(\mathbf{x}_{cp}, t)$ is the solute flux at the control plane, and $Q_w(\mathbf{x}_{cp})$ is the hydraulic flux at the control plane (e.g. *Andricevic and Cvetkovic* (1998)). Thus, ILCR can be computed as

$$r(t) = \beta \frac{Q(t)}{Q_w} \tag{5.12}$$

where the location $\mathbf{x}_{cp}$ of r, Q, and Q_w are assumed and not explicitly stated. β is a lumped parameter of all non-hydrogeological variables (*de Barros and Rubin*, 2008),

$$\beta = CPF \frac{IR \times ED \times EF}{BW \times AT} f_{mo} \tag{5.13}$$

where the various terms are defined in Table 5.1.

5.4 Case Study Setup

Hydrogeological Description

This case study focuses on a scenario where a contaminant plume of total mass $M_0 = 0.1kg$ is released over a duration $T_0 = 0.1d$ from a point source of known location and migrates towards a control plane defining a water supply located at $\mathbf{x}_{cp}$. Aquifer flow is uniform in the average. The control plane location $\mathbf{x}_{cp}$ is a distance of $L = 40m$ from the point source location, along the x-axis which defines the mean flow direction. This travel length

Mean of Log-conductivity μ_Y	Variance σ^2_Y	Integral Scale I_Y
Uniform[-6, -5] ($\ln(m/min)$)	Uniform[0.1, 1]	Uniform[3,6] (m)

Table 5.2: Prior information for the SRF parameters for $Y = \ln K$ used in the case study.

corresponds to a range of 6.67 to 13.3 integral scales. The natural logarithm of hydraulic conductivity $Y = \ln K$ was modeled as a Gaussian SRF with an exponential covariance. The independent prior distributions for the SRF parameters are shown in Table 5.2. As in Chapter 4, baseline fields were simulated using the method of *Schlather et al.* (2017)

Calculation of Solute Flux

The solute flux and in turn, ILCR were assumed to be lognormally distributed, and were calculated at the time of maximum solute flux,

$$Q(\mathbf{x}_{cp}, t_{peak}) = \max_t\{Q(\mathbf{x}_{cp}, t))\}. \tag{5.14}$$

which occurs at $t = t_{peak}$. The pdf of ILCR is defined by its first two moments, which can be defined by the pdf of contaminant arrival time using the expressions

$$\mu_R = \frac{\beta M_0}{Q_w T_0} \Delta F_\tau(t_{peak}) \tag{5.15}$$

$$\sigma^2_R = \left[\frac{\beta M_0}{Q_w T_0}\right]^2 \left[\Delta F_\tau(t_{peak}) - (\Delta F_\tau(t_{peak}))^2\right] \tag{5.16}$$

which are provided in *de Barros and Rubin* (2008). The hydraulic flux is assumed deterministic and given by

$$Q_w = n L_y \langle U \rangle \tag{5.17}$$

where L_y is the length of the control plane orthogonal to mean flow direction, and $\langle U \rangle$ is the mean velocity. $\Delta F_\tau(t_{peak})$ is determined by

$$\Delta F_\tau(t_{peak}) = \int_{t_{peak}-T_0}^{t_{peak}} f_\tau(\tau) d\tau \tag{5.18}$$

where f_τ is the pdf of arrival time.

Baseline Fields

On each baseline field Y^b_i; $i = 1, ..., N_Y$, $f_\tau(t)$ was a Dirac delta function with peak calculated by deterministic particle tracking using the method of *Pollock* (1988). In turn, the moments of the ILCR pdf $f_R(R)$ were computed using the expressions above.

A key difference in this case study from the one presented in Chapter 4 is that modeling the response of interest on baseline fields can only be done probabilistically. In predicting contaminant arrival time from a point source, a single baseline field produces a single deterministically known value, which simplifies the definition of I and in turn, computation of $\langle\phi_\alpha^G\rangle$ and $\langle\phi_\beta^G\rangle$. In this case study, the hydrogeological response to be predicted is solute flux, which is predicted probabilistically, even when modeling with a deterministically known baseline field. In light of this difference, the definition of I_i^b for each baseline field Y_i^b must be defined based on a probability:

$$I_i^b = \begin{cases} 1 & \text{if } Pr[R > r_{crit}|Y_i^b] \geq \alpha \\ 0 & \text{if } Pr[R > r_{crit}|Y_i^b] < \alpha \end{cases} \tag{5.19}$$

.

Conditional Simulations

With each baseline field $Y_i^b;\ i = 1, ..., N_Y$, data acquisition was simulated according to the quantity and spatial locations of measurements specified by the design G^H, which is defined below. With these measurements, the *maximum a posteriori* SRF parameters θ_{MAP} were estimated. The measurements and parameter estimates were used to semi-analytically compute the arrival time pdf $f_\tau(\tau|g^H, \theta_{MAP})$ using the method of *Rubin* (1991) and kernel density estimation (*R Core Team*, 2015). These pdfs were in turn utilized to calculate the moments of the ILCR, conditional to the hydraulic conductivity measurements and the inferred parameters.

At this point, decision making was simulated based on the pdf of ILCR and the criteria defined above.

$$D_i^g = \begin{cases} 1 & \text{if } Pr[R > r_{crit}|g^H, \theta_{MAP}] \geq \alpha \text{ (accept } H_0^I) \\ 0 & \text{if } Pr[R > r_{crit}|g^H, \theta_{MAP}] < \alpha \text{ (reject } H_0^I) \end{cases} \tag{5.20}$$

D_i^g was defined on each baseline field Y_i, which in turn allowed definition of $\phi_{\alpha i}$ and $\phi_{\beta i}$. Lastly, $\langle\phi_\alpha^G\rangle$ and $\langle\phi_\beta^G\rangle$ were computed by the summations

$$\langle\phi_\alpha^G\rangle = \frac{1}{N_Y}\sum_{i=1}^{N_Y}\phi_{\alpha i} \tag{5.21}$$

$$\langle\phi_\beta^G\rangle = \frac{1}{N_Y}\sum_{i=1}^{N_Y}\phi_{\beta i} \tag{5.22}$$

which indicate the probabilities of Type I and Type II error, respectively.

Contaminant, Physiological, and Behavioral Description

Many of the parameters lumped into β were treated deterministically, and are summarized in Table 5.3. The other parameters which were treated stochastically are summarized

Parameter	Deterministic Value
IR/BW	0.033 L/d-kg
EF	350 d/y
AT	22550 d
ED	30 d

Table 5.3: Values of non-hydrogeological parameters which are assumed deterministically known for the enhanced cancer risk case study.

Parameter	Probabilistic Description
CPF	Uniform[0.045, 0.175]
f_{mo}	Uniform[0.2, 0.7]

Table 5.4: Probabilistic description of non-hydrogeological parameters which were stochastically characterized for the enhanced cancer risk case study.

in Table 5.4. These parameters were chosen to simulate a PCE contaminant, in line with previous work (e.g. *McKone and Bogen* (1991), *Maxwell et al.* (1998a), *Maxwell and Kastenberg* (1999), *de Barros and Rubin* (2008)). For each baseline field, a realization of β was simulated from the prior pdfs given in Table 5.4 and $f_R(R)$ was computed using the solute flux on this baseline field and this realization of β. For the sake of simplicity, the conditional simulations of $f_R(R|G)$ were executed using only G_4^β for the simulated physiological characterization.

Data Acquisition Design

The case study aimed to compare the effectiveness of different forms of data acquisition: hydrogeological and non-hydrogeological data. For hydrogeological data acquisition, measurements of hydraulic conductivity were simulated. Four different quantities of measurements were taken, $N = 4, 8, 16, 32$. The measurements were clustered around the travel path, corresponding to hydrogeological campaigns G_{1A}, G_{2A}, G_{3A}, and G_{4A} shown in Figure 4.2. In this case study, these four hydrogeological campaigns are denoted G_1^H, G_2^H, and G_3^H, G_4^H, respectively.

To model non-hydrogeological data acquisition, four alternative strategies were compared, each of which defined the mean to be the baseline truth, corresponding to a realization of β for each baseline field Y_i^b. G_1^β involved reducing the variance of each of the prior pdfs by a factor of 2, while G_2^β and G_3^β reduced the variance in the prior pdfs by a factor of 4 and 10, respectively. G_4^β assumed perfect deterministic knowledge of β. Future research will aim to more realistically simulate the effects of physiological and behavioral data acquisition.

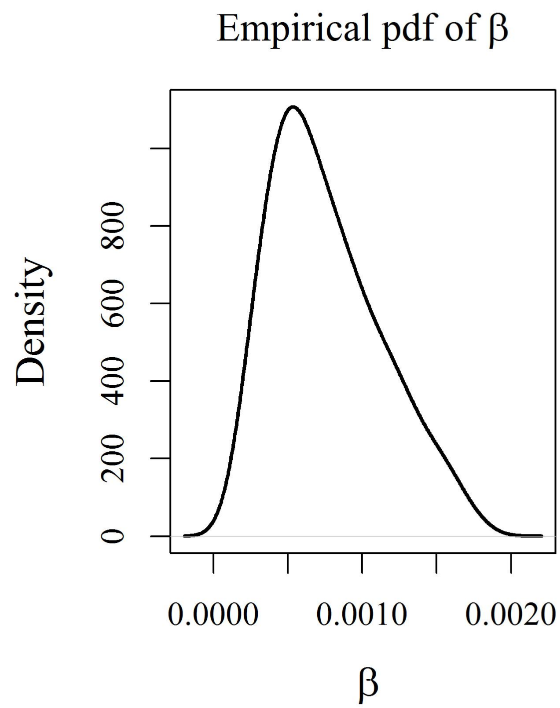

Figure 5.1: Empirical pdf showing uncertainty in lumped parameter β which results from uncertainty in its components (see Table 5.4). Since β varies over only a few orders of magnitude, the uncertainty in β contributes relatively little to uncertainty in predicting enhanced cancer risk.

5.5 Results and Discussion

This section presents and discusses results from the case study described earlier in this Chapter. The results focus on uncertainty in β and ILCR as well as the associated decision risk resulting from such uncertainty.

Uncertainty in β

The effect of uncertainty in CPF and f_{mo} is shown by the empirical pdf of β in Figure 5.1. The uncertainty in these two parameters cause a range of nearly two orders of magnitude in β, indicating the wide range of cancer risk to be experienced throughout the population, and enabling a preliminary assessment how much non-hydrogeological characterization may be necessary. As a precursor to examining the effect of physiological information on predictions of cancer risk, we focus for the moment on the effect of physiological information the prediction of the distribution of β alone. To compare between the effects of G_1^β, G_2^β, and G_3^β, we use the Kullback-Leibler convergence, defined as (*Kullback and Leibler* (1951), *Tang*

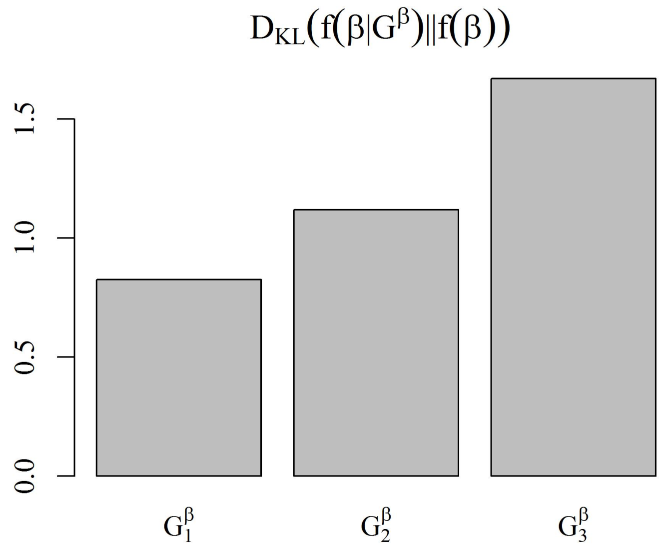

Figure 5.2: Barplot showing the Kullback-Leibler Divergence resulting from the simulated physiological information G_1^β, G_2^β, G_3^β, respectively.

et al. (2016))

$$D_{KL}(P||Q) = \int P(x) log \frac{P(x)}{Q(x)} dx \tag{5.23}$$

where $Q(x)$ is taken to be the prior distribution for β (see Figure 5.1), and $P(x)$ is taken to be the posterior pdf $f(\beta|G^\beta)$ for G_1^β, G_2^β, and G_3^β. D_{KL} was calculated using a conditional distribution for each unconditional realization of β, and the average values are presented in Figure 5.2. Larger values of D_{KL} indicate a greater "distance" between the prior pdf and the posterior pdf, i.e. a greater effect of the information contained in G^β. As the Figure shows, G_3^β has the greatest informative effect, followed by G_2^β and in turn G_1^β.

Error Probabilities

The resulting probability of Type I error for hydrogeological campaigns G_1^H, G_2^H, G_3^H, and G_4^H were all zero, regardless of the value of r_{crit}. The reasoning can be understood by examining the relationship between exceedance probabilities $Pr[R > r_{crit}|Y_i]$ and $Pr[R > r_{crit}|g^H]$, which are plotted in Figure 5.3 for three individual baseline fields. For Type I error to occur, the baseline exceedance probability would need to be greater than α in a region where the conditional exceedance probability is less than α. This would indicate a scenario where the null hypothesis H_0^I is actually true but the measurement information would lead us to reject it. However, this is not the case for any value of r_{crit} or for any measurement

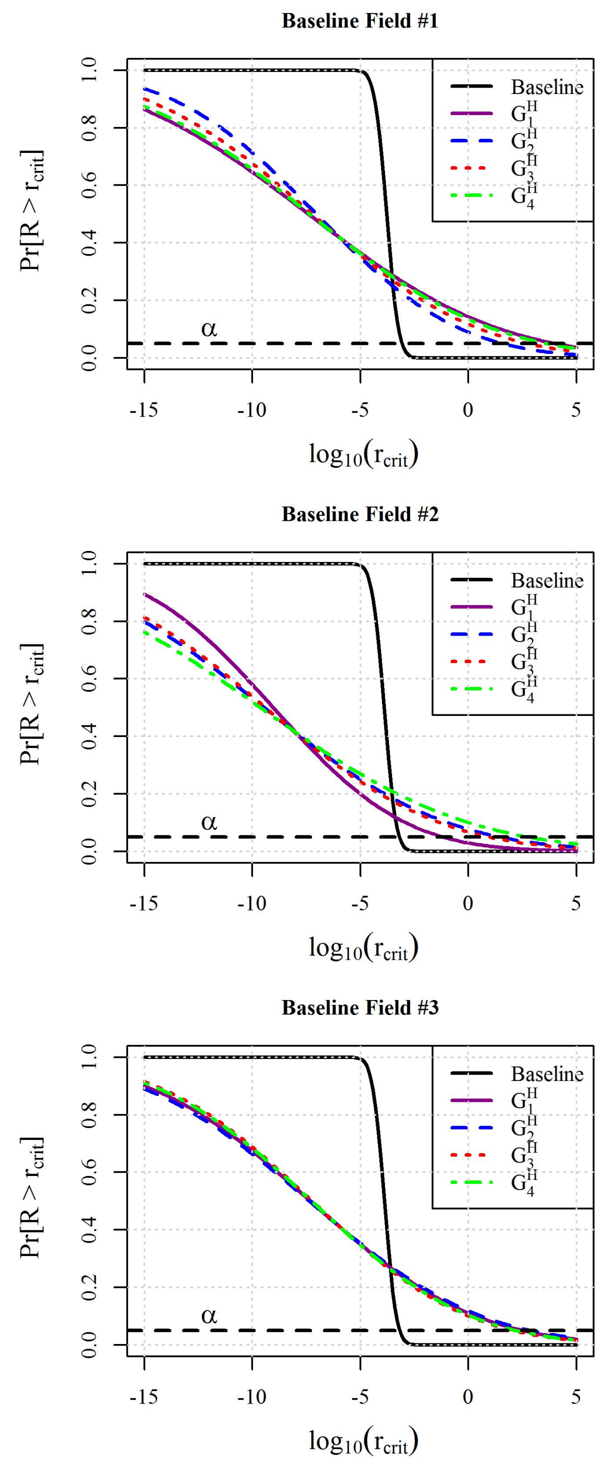

Figure 5.3: This figure shows how the probability $Pr[R > r_{crit}]$ varies with respect to r_{crit} for three selected baseline fields. The solid black line indicates the synhetic truth associated with each baseline field, and the colored lines indicate the predicted response conditional to the four different measurement configurations. Type I error would occur for values of r_{crit} at which the black line is greater than α but the colored lines are less than α, which never occurs in this scenario. Type II error would occur if the inverse was true.

configuration. Type II error, analogously, occurs when the baseline exceedance probability is less than α in a region where the conditional exceedance probability is greater than α. This occurs for all measurement configurations for a small range of r_{crit}. The probability of Type II error is plotted against r_{crit} in Figure 5.5. As indicated by the Figure, the response is quite insensitive to the quantity of hydraulic conductivity measurements taken. A likely explanation for this might be the lognormal assumption for the pdf of risk, which takes into account only the first two moments of solute flux, rather than the entire distribution.

The aggregated probability of the null hypothesis being true conditional to baseline fields as well as to the four hydrogeological campaigns are plotted versus r_{crit} in Figure 5.4. This plot shows the same effect but instead averaged over the entire ensemble of baseline fields. We see a much more diffuse behavior of the conditional distributions than we do for the baseline fields. The reasoning for this is the point source nature of the contaminant source. In the baseline fields, the advection dominated transport model produces an arrival time pdf which emulates a Dirac delta, and therefore has a very steep descent as can be seen on Figures 5.3 and 5.4. The conditional distributions of arrival times, on the other hand, are based on an entire pdf of arrival time, which causes a much greater variance in the solute flux pdf and in turn the risk pdf.

Future Work

Future research will aim to more realistically model uncertainty in physiological and behavioral parameters to more closely examine their effect on decision making. A more realistic method of simulating data acquisition aiming to characterize these parameters would also be helpful.

As mentioned above, the lognormal assumption regarding the solute flux pdf can be rather limiting. The effect of conditioning on hydraulic conductivity measurements mainly manifests itself by altering the shape of the arrival time pdf, which would in turn affect the solute flux pdf and the distribution of ILCR. Using the lognormal approximation only takes into account a limited portion of these distributions near the time of peak solute flux. Accounting for the entire arrival time pdf may provide greater insight into the relationship between hydrogeological field data and decision making with respect to ILCR.

Finally, more research is needed into how the dimensions of the contaminant source affect decision making. The present case study focused on a point source, which caused the rather steep descent of the exceedance probability with respect to r_{crit}, which can be very difficult to reproduce, regardless of the amount of hydrogeological data available. Further research into varying source dimensions can indicate under what circumstances is hydrogeological information more important than physiological information and vice versa.

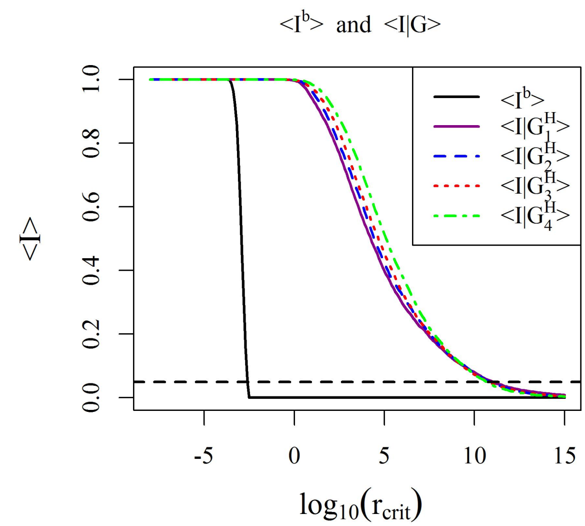

Figure 5.4: This figure shows the probability of the null hypothesis being true, conditional to five information states. The black line is conditional only the prior distributions $f_\beta(\beta)$ and $f_\theta(\theta)$. The colored lines are conditional the four different measurement configurations. This plot relates to Figure 5.3 by indicating whether or not the probabilities exceed α, and averaged over all baseline fields.

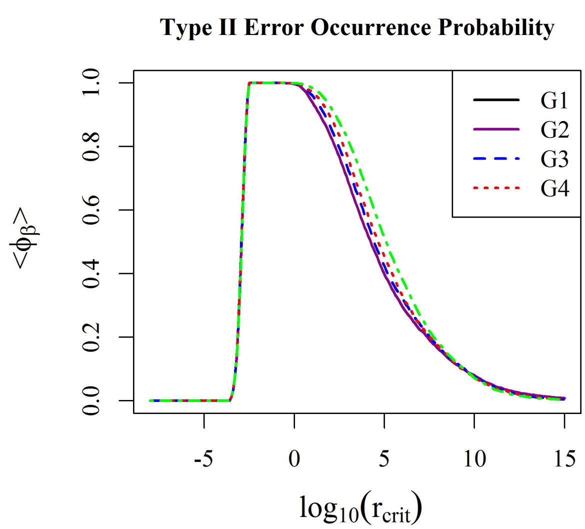

Figure 5.5: This figure shows the occurrence probability for Type II error for the four measurement configurations G_1^H, G_2^H, G_3^H, and G_4^H.

Chapter 6

Conclusions

This dissertation has discussed the role of stochastic hydrogeological modeling in successful, sustainable, groundwater resources management. While widespread adoption of stochastic hydrogeology has been relatively slow, it has great potential to improve decision making in water resources management, regulation, and policy making. This potential is due to the ability to make predictions regarding quantities which inform the most appropriate course of action in e.g. selection of water supply and allocation of limited remediation resources. In addition to making these predictions, stochastic methods can improve decision making by quantifying uncertainty stemming from all sources, which enables risk-based decisions. However, challenges remain due to the complicated relationship between site characterization, modeling, prediction, decision making, and the role uncertainty plays in each of these steps. Further complications arise when needing to communicate uncertainty, which stems from many sources and requires hydrogeology-specific training to understand, to stakeholders outside of the hydrogeological community.

This dissertation presented a framework which addresses many of these challenges. The benefits of the framework are simplified decision making, simplified communication of uncertainty, direct translation from uncertainty to risk at the knowledge-decision interface, and straightforward evaluation of data acquisition strategies *before* data is collected. Simplified decision making and translation of uncertainty to risk is enabled via utilization of the hypothesis testing framework, which aggregates uncertainty stemming from all sources into a single quantity representing the decision making risk. Evaluation of data acquisition strategies is enabled by using stochastic simulation and a holistic approach to planning of field campaigns, inverse modeling, forward modeling, and decision making, which enables prediction of decision risk resulting from the data, providing the ultimate measure of efficacy of the data. The framework is able to accommodate any state of information, ranging from relatively uninformative priors to possessing in situ data and planning of iterative field campaigns.

The framework was demonstrated in two case studies, where the goal was predicting 1) contaminant arrival time and 2) enhanced cancer risk. The case study predicting contaminant arrival time had a few key conclusions:

1. The amount of field data necessary to predict early arrivals is highly dependent on threshold value of time which defines "early" arrivals.

2. Improved estimation of geostatistical parameters does not necessarily improve decision making.

3. A more informative prior distribution does not always lead to improved decision making.

The first two conclusions strongly reinforce the notion that field campaigns should be designed in a goal-oriented manner where successful decision making is defined as the goal, as opposed to simply optimizing field campaign design with respect to improved inverse modeling. The third conclusion is somewhat surprising and indicates that further research is required to specifically investigate the role of prior information and how it may change with alternative inverse modeling.

The case study in predicting enhanced cancer risk aimed to explore the relationship between hydrogeological characterization and physiological and behavioral characterization in the context of decision making based on increased lifetime cancer risk caused by groundwater contamination. More specifically, the aim was to explore the relationship between different types of information and the probability of erroneous decision making. A point source contaminant plume was modeled with lognormally distributed solute flux at a control plane defining the contaminant receptor. Under these conditions, predictions are relatively insensitive to hydrogeological measurement quantity and Type I error is virtually impossible, though further research is needed for more general conclusions.

Bibliography

Abellan, A., and B. Noetinger (2010), Optimizing Subsurface Field Data Acquisition Using Information Theory, *Mathematical Geosciences*, *42*(6), 603–630, doi:10.1007/s11004-010-9285-6.

Anderson, M. P., W. W. Woessner, and R. J. Hunt (2015), *Applied Groundwater Modeling*, second ed., Elsevier Inc., San Diego, California.

Andricevic, R., and V. Cvetkovic (1998), Relative dispersion for solute flux in aquifers, *Journal of Fluid Mechanics*, *361*, 145–174.

Baker, V. R. (2017), Debates - hypothesis testing in hydrology: Pursuing certainty versus pursuing uberty, *Water Resources Research*, *53*(3), 1770–1778, doi: 10.1002/2016WR020078.

Beven, K. J. (2002), Towards a coherent philosophy for environmental modelling., *Proceedings of the Royal Society A Methematical, Physical & Engineering Sciences*, *458*, 2465–2484, doi:10.1098/rspa.2002.0986.

Bloschl, G. (2017), Debates Hypothesis testing in hydrology : Introduction, *Water Resources Research*, *53*(3), 1767—-1769.

Butler, J. J., J. M. Healey, G. W. McCall, E. J. Garnett, and S. P. L. II (2002), Hydraulic Tests with Direct-Push Equipment, *Ground Water*, *40*(1), 25–36.

Butler, J. J., P. Dietrich, V. Wittig, and T. Christy (2007), Characterizing hydraulic conductivity with the direct-push permeameter, *Ground Water*, *45*(4), 409–419, doi: 10.1111/j.1745-6584.2007.00300.x.

Carrera, J., A. Alcolea, A. Medina, J. Hidalgo, and L. J. Slooten (2005), Inverse problem in hydrogeology, *Hydrogeology Journal*, *13*(1), 206–222, doi:10.1007/s10040-004-0404-7.

Cherry, J. A., and R. A. Freeze (1979), *Groundwater*, Prentice-Hall, Englewood Cliffs, New Jersey.

Christakos, G. (2004), A sociological approach to the state of stochastic hydrogeology, *Stochastic Environmental Research and Risk Assessment*, *18*(4), 274–277, doi: 10.1007/s00477-004-0197-1.

Cirpka, O. A., and A. J. Valocchi (2016), Debates - stochastic subsurface hydrology from theory to practice: Does stochastic subsurface hydrology help solving practical problems of contaminant hydrogeology?, *Water Resources Research*, *52*(12), 9218–9227, doi: 10.1002/2016WR019087.

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

Dagan, G. (2004), On application of stochastic modeling of groundwater flow and transport, *Stochastic Environmental Research and Risk Assessment*, *18*(4), 266–267, doi: 10.1007/s00477-004-0191-7.

de Barros, F. P. J., and Y. Rubin (2008), A risk-driven approach for subsurface site characterization, *Water Resources Research*, *44*(1), 1–14, doi:10.1029/2007WR006081.

de Barros, F. P. J., and Y. Rubin (2011), Modelling of block-scale macrodispersion as a random function, *Journal of Fluid Mechanics*, *676*, 514545, doi:10.1017/jfm.2011.65.

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 Resources Research*, *45*(6), doi:10.1029/2008WR007324, w06401.

de Barros, F. P. J., S. Ezzedine, and Y. Rubin (2012), Impact of hydrogeological data on measures of uncertainty, site characterization and environmental performance metrics, *Advances in Water Resources*, *36*, 51–63, doi:10.1016/j.advwatres.2011.05.004.

DeSimone, L., P. McMahon, and M. Rosen (2014), The quality of our Nation's waters–Water quality in Principal Aquifers of the United States, 1991-2010, *U.S. Geological Survey Circular*, *1360*, 1–151.

Dietrich, P., J. J. Butler, and K. Faiß (2008), A rapid method for hydraulic profiling in unconsolidated formations, *Ground Water*, *46*(2), 323–328, doi:10.1111/j.1745-6584.2007.00377.x.

Farber, D. A., and R. W. Findley (2010), *Environmental Law in a Nutshell, 8th*, West Academic, St. Paul, Minnesota.

Fetter, C. (2001), *Applied Hydrogeology*, 4th ed., Prentice-Hall, Upper Saddle River, New Jersey.

Fiori, A., V. Cvetkovic, G. Dagan, S. Attinger, A. Bellin, P. Dietrich, A. Zech, and G. Teutsch (2016), Debates - stochastic subsurface hydrology from theory to practice: The relevance of stochastic subsurface hydrology to practical problems of contaminant transport and remediation. what is characterization and stochastic theory good for?, *Water Resources Research*, *52*(12), 9228–9234, doi:10.1002/2015WR017525.

Floris, F. J. T., M. D. Bush, M. Cuypers, F. Roggero, and A.-R. Syversveen (2001), Methods for quantifying the uncertainty of production forecasts : a comparative study, *Petroleum Geoscience*, *7*, S87–S96, doi:10.1144/petgeo.7.S.S87.

Fogg, G. E., and Y. Zhang (2016), Debates - stochastic subsurface hydrology from theory to practice: A geologic perspective, *Water Resources Research*, *52*(12), 9235–9245, doi:10.1002/2016WR019699.

Freeze, R. A. (2004), The role of stochastic hydrogeological modeling in real-world engineering applications, *Stochastic Environmental Research and Risk Assessment*, *18*(4), 286–289, doi:10.1007/s00477-004-0194-4.

Freeze, R. A., and S. M. Gorelick (1999), Convergence of Stochastic Optimization and Decision Analysis in the Engineering Design of Aquifer Remediation, doi:10.1111/j.1745-6584.1999.tb01193.x.

Ginn, T. R. (2004), On the application of stochastic approaches in hydrogeology, *Stochastic Environmental Research and Risk Assessment*, *18*(4), 282–284, doi:10.1007/s00477-004-0199-z.

Goode, R., and K. D. Evans (2007), The Scientific Method : Application to FDA-Regulated Industry, *Journal of Validation Technology*, *13*(4), 329–335.

Harrell, D. R., and T. L. Gardner (2003), Significant Differences in Proved Reserves Volumes Estimated Using SPE / WPC Reserves Compared to United States Securities and Exchange Commission (SEC) Definitions, *SPE Annual Technical Conference and Exhibition*, doi:10.2118/84145-MS.

Höllermann, B., and M. Evers (2017), Environmental Science & Policy Perception and handling of uncertainties in water management A study of practitioners ' and scientists ' perspectives on uncertainty in their daily decision-making, *Environmental Science and Policy*, *71*, 9–18, doi:10.1016/j.envsci.2017.02.003.

Intergovernmental Panel on Climate Change (2010), *Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties*, url: https://www.ipcc.ch/pdf/supporting-material/uncertainty-guidance-note.pdf.

Jonkman, R. M., C. F. M. Bos, J. N. Breunese, D. T. K. Morgan, J. A. Spencer, and E. Sondena (2000), Best Practices and Methods in Hydrocarbon Resource Estimation , Production and Emissions Forecasting , Uncertainty Evaluation and Decision Making, *Proceedings of the SPE European Petroleum Conference*.

Kass, R. E., and L. Wasserman (1996), The Selection of Prior Distributions by Formal Rules, *Journal of the American Statistical Association*, *91*(435).

Kitanidis, P. (1997), *Introduction to Geostatistics*, Cambridge University Press, New York, New York.

Kitanidis, P. K., and R. W. Lane (1985), [1] Maximum Likelihood Parameter Estimation of Hydrologic Spatial Processes by the Gauss-Newton Method, *Journal of Hydrology*, *79*, 53–71.

Köber, R., G. Hornbruch, C. Leven, L. Tischer, J. Großmann, P. Dietrich, H. Weiß, and A. Dahmke (2009), Evaluation of combined direct-push methods used for aquifer model generation, *Ground Water*, *47*(4), 536–546, doi:10.1111/j.1745-6584.2009.00554.x.

Kullback, S., and R. A. Leibler (1951), On information and sufficiency, *The annals of mathematical statistics*, *22*(1), 79–86.

Liu, N., and D. S. Oliver (2003), Evaluation of Monte Carlo Methods for Assessing Uncertainty, *Society of Petroleum Engineers Journal*, (November 2001), doi:10.2118/84936-PA.

Maxwell, R. M., and W. E. Kastenberg (1999), Stochastic environmental risk analysis: an integrated methodology for predicting cancer risk from contaminated groundwater, *Stochastic Environmental Research and Risk Assessment*, *13*(1-2), 27–47, doi: 10.1007/s004770050030.

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

Maxwell, R. M., S. D. Pelmulder, A. F. B. Tompson, and W. E. Kastenberg (1998b), On the development of a new methodology for groundwater-driven health risk assessment, *Water Resources Research*, *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 Resources Research*, *35*(9), 2841, doi:10.1029/1999WR900103.

Mays, L. W., and D. K. Todd (2005), *Groundwater Hydrology*, 3rd ed., John Wiley & Sons, Inc.

McKnight, D. M. (2017), Debates - hypothesis testing in hydrology: A view from the field: The value of hydrologic hypotheses in designing field studies and interpreting the results to advance hydrology, *Water Resources Research*, *53*(3), 1779–1783, doi: 10.1002/2016WR020050.

McKone, T. E., and K. T. Bogen (1991), Predicting the uncertainties in risk assessment, *Environmental science & technology*, *25*(10), 1674–1681.

Misund, B., and P. Osmundsen (2015), Probable Oil and Gas Reserves and Shareholder Returns : The Impact of Shale Gas, *CESifo Working Paper Series*, *5687*.

Molz, F. (2004), A rational role for stochastic concepts in subsurface hydrology: a personal perspective, *Stochastic Environmental Research and Risk Assessment*, *18*(4), 278–279, doi:10.1007/s00477-004-0195-3.

Mukhopadhyay, N. (2000), *Probability and Statistical Inference*, Marcel Dekker, Inc., New York, New York.

Navidi, W. (2015), *Statistics for Engineers and Scientists*, fourth ed., McGraw-Hill Education, New York, New York.

Neuman, S. P. (2004), Stochastic groundwater models in practice, *Stochastic Environmental Research and Risk Assessment*, *18*(4), 268–270, doi:10.1007/s00477-004-0192-6.

Neuweiler, I., and R. Helmig (2017), Debates - hypothesis testing in hydrology: A subsurface perspective, *Water Resources Research*, *53*(3), 1784–1791, doi:10.1002/2016WR020047.

Nowak, W., Y. Rubin, and F. P. J. de Barros (2012), A hypothesis-driven approach to optimize field campaigns, *Water Resources Research*, *48*(6), W06,509, doi: 10.1029/2011WR011016.

Oliver, D. S., and Y. Chen (2010), Recent progress on reservoir history matching : a review, *Computational Geosciences*, *15*, 185–221, doi:10.1007/s10596-010-9194-2.

Oreskes, N., K. Shrader-Frechette, and K. Belitz (1994), in Models Numerical and the Confirmation of Earth Sciences, *Science*, *263*(5147), 641–646, doi:10.1126/science.263.5147.641.

Pfister, L., and J. W. Kirchner (2017), Debates - hypothesis testing in hydrology: Theory and practice, *Water Resources Research*, *53*(3), 1792–1798, doi:10.1002/2016WR020116.

Pollock, D. W. (1988), Semianalytical Computation of Path Lines for Finite - Difference Models, doi:10.1111/j.1745-6584.1988.tb00425.x.

R Core Team (2015), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Rajaram, H. (2016), DebatesStochastic subsurface hydrology fromtheory to practice: Introduction, pp. 9215–92,177, doi:10.1002/2013WR014979.Reply.

Rubin, Y. (1990), Stochastic modeling of macrodispersion in heterogeneous porous media, *Water Resources Research*, *26*(1), 133–141.

Rubin, Y. (1991), Prediction of Tracer Plume Migration in Disordered Porous Media by the Method of Conditional Probabilities, *Water Resources Research*, *27*(6), 1291–1308.

Rubin, Y. (2003), *Applied Stochastic Hydrogeology*, Oxford University Press, New York, New York.

Rubin, Y. (2004), Stochastic hydrogeology - challenges and misconceptions, *Stochastic Environmental Research and Risk . . .*, pp. 280–281, doi:1029.2003/WR002420.

Rubin, Y., and S. S. Hubbard (2006), *Hydrogeophysics*, vol. 50, Springer Science & Business Media.

Rwechungura, R., M. Dadashpour, and J. Kleppe (2011), Advanced History Matching Techniques Reviewed, *Society of Petroleum Engineers*, doi:10.2118/142497-MS.

Sanchez-Vila, X., and D. Fernndez-Garcia (2016), Debates - stochastic subsurface hydrology from theory to practice: Why stochastic modeling has not yet permeated into practitioners?, *Water Resources Research*, *52*(12), 9246–9258, doi:10.1002/2016WR019302.

Schlather, M., A. Malinowski, M. Oesting, D. Boecker, K. Strokorb, S. Engelke, J. Martini, F. Ballani, O. Moreva, J. Auel, P. J. Menck, S. Gross, U. Ober, Christoph Berreth, K. Burmeister, J. Manitz, P. Ribeiro, R. Singleton, B. Pfaff, and R Core Team (2017), *RandomFields: Simulation and Analysis of Random Fields*, r package version 3.1.50.

Smalley, J. B., B. S. Minsker, and D. E. Goldberg (2000), Risk-based in situ bioremediation design using a noisy genetic algorithm, *Water Resources Research*, *36*(10), 3043–3052, doi:10.1029/2000WR900191.

Smith, M., K. Cross, M. Paden, and P. Laban (2016), *Spring - Managing groundwater sustainably*, IUCN, Gland, Switzerland.

Sudicky, E. (2004), On certain stochastic hydrology issues, *Stochastic Environmental Research and Risk Assessment*, *18*(4), 2004, doi:10.1007/s00477-004-0196-2.

Tang, Y., L. Marshall, A. Sharma, and T. Smith (2016), Tools for investigating the prior distribution in Bayesian hydrology, *Journal of Hydrology*, *538*, 551–562, doi: 10.1016/j.jhydrol.2016.04.032.

Ulrych, T. J., M. D. Sacchi, and A. Woodbury (2001), Tutorial A Bayes tour of inversion : A tutorial, *Geophysics*, *66*(1), 55–69.

U.S. Environmental Protection Agency (1989), *Risk Assessment Guidance for Superfund Volume I: Human Health Evaluation Manual (Part A)*, Report Number EPA/540/1-89/002.

U.S. Environmental Protection Agency (1997), *Health Effects Assessment Summary Tables*, Report Number EPA/540/R/97/036.

U.S. Environmental Protection Agency (2000), *Supplementary Guidance for Conducting Health Risk Assessment of Chemical Mixtures*, Report Number EPA/630/R-00/002.

U.S. Environmental Protection Agency (2014), *Risk Assessment Forum White Paper: Probabilistic Risk Assessment Methods and Case Studies*, no. July in Report number EPA/100/R-14/004., 98 pp.

Winter, C. L. (2004), Stochastic hydrology: practical alternatives exist, *Stochastic Environmental Research and Risk Assessment*, *18*(4), 271–273, doi:10.1007/s00477-004-0198-0.

Yeh, W. W.-G. (1986), Review of Parameter Identification Procedures in Groundwater Hydrology ', *Water Resources Research*, *22*(2), 95–108.

Zhang, Y. K., and D. Zhang (2004), Forum: The state of stochastic hydrology, *Stochastic Environmental Research and Risk Assessment*, *18*(4), 265, doi:10.1007/s00477-004-0190-8.

Zhou, H., J. J. Gómez-Hernández, and L. Li (2014), Advances in Water Resources Inverse methods in hydrogeology : Evolution and recent trends, *Advances in Water Resources*, *63*, 22–37, doi:10.1016/j.advwatres.2013.10.014.