A Least-Cost Strategy for Evaluating a Brownfields Redevelopment Project Subject to Indoor Air Exposure Regulations

by

Xiaomin Wang

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Earth Sciences

Waterloo, Ontario, Canada, 2012

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Over the course of the past several decades the benefits of redeveloping brownfields have been widely recognized. Actions have been taken to foster sustainable redevelopment of brownfields by government, policy makers and stakeholders across the world. However, redevelopments encounter great challenges and risks related to environmental and non-environmental issues. In this work, we intend to build a comprehensive and practical framework to evaluate the hydrogeological and financial risks involved during redevelopment and to ensure developers reserve sufficient capital to cover unexpected future costs within the guarantee period. Punitive damages, which contribute to these costs, are in this thesis solely associated with the cost of repossessing a house within a development should the indoor air concentration of TCE exceed the regulatory limit at a later time.

The uncertainties associated with brownfield remediation have been among the barriers to brownfield redevelopment. This is mainly caused by the lack of knowledge about a site's environmental condition. In order to alleviate uncertainties and to better understand the contaminant transport process in the subsurface, numerical simulations have been conducted to investigate the role of controlling parameters in determining the fate and transport of volatile organic compounds originating from a NAPL source zone located below the water table in the subsurface. In the first part of this thesis, the numerical model *CompFlow Bio* is used on a hypothesized three-dimensional problem geometry where multiple residential dwellings are built. The simulations indicate that uncertainty in the simulated indoor air concentration is sensitive to heterogeneity in the permeability structure of a stratigraphically continuous aquifer with uncertainty defined as the probability of exceeding a regulatory limit. Houses which are laterally offset from the groundwater plume are less affected by vapour intrusion due to limited transverse horizontal flux of TCE within the groundwater plume in agreement with the ASTM (2008) guidance. Within this uncertainty framework, we show that the Johnson and Ettinger (1991) model generates overly-conservative results and contributes to the exclusion zone being much further away from the groundwater plume relative to either *CompFlow Bio* or ASTM (2008). The probability of failure (or the probability of exceedence of the regulatory limit) is defined and calculated for further study.

Due to uncertainties resulting from parameter estimation and model prediction, a methodology is introduced to incorporate field measurements into the initial estimates from the numerical model in order to improve prediction accuracy. The principle idea of this methodology is to combine the geostatistical tool kriging with the statistical data assimilation method Kalman filter to evaluate the worth and effectiveness of data in a quantitative way in order to select an optimal sampling scenario.

This methodology is also used to infer whether one of the houses located adjacent to affected houses has indoor air problems based on the measurements subject to the observation that the affected house is monitored and has problems and developers have liability if a problem occurs. In this part of the study, different sampling scenarios are set up in terms of permeability (1 – 80 boreholes) and soil gas concentration (2, 4 and 7 samples) and three metrics are defined and computed as a criterion for comparison.

Financing brownfield redevelopment is often viewed as a major barrier to the development process mainly due to risks and liabilities associated with brownfields. The common way of managing the risk is to transfer it to insurers by purchasing insurance coverage. This work provides two different strategies to price the risk, which is equivalent to an insurance premium. It is intended to give an instructive insight into project planning and feasibility studies during the decision-making process of a brownfield project. The two strategies of risk capital valuation are an actuarial premium calculation principle and a martingale premium calculation principle accounting for the hydrogeological and financial uncertainties faced in a project. The data used for valuation are the posterior estimates of data assimilation obtained from the results of different sampling scenarios. The cost-benefit-risk analysis is employed as a basis to construct the objective function in order to find the least cost among sampling scenarios for the project. As a result, it shows that drilling seven boreholes to extract permeability data and taking soil gas samplings in four locations or seven locations alternatively give the minimum total cost. Sensitivity analysis of some influential parameters (the safety loading factors and the possible methods to calculate the probability of failure) is performed to determine their roles of importance in the risk capital valuation. This framework can be applied to provide guidance for other risk-based environmental projects.

Acknowledgements

Dedication

To my parents Zhongliang Wang and Muhua Tang, for their great support.

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

1.1 Background Research

Brownfields are defined as "abandoned, idled, or under-used industrial and commercial facilities where expansion or redevelopment is complicated by real or perceived environmental contamination" by the U.S. Environmental Protection Agency (US EPA). Brownfields are usually unintended by-products of industrial practices of the last several decades in which measures were not taken to ensure that industrial operations did not harm the natural environment. In brownfields, there exist known and potential contaminants, such as volatile organic compounds (VOCs) or semi-volatile organic compounds (SVOCs) which over time, may be exposed to the air and potentially threaten public health and have negative impacts on ecological systems. Redevelopment of brownfields is beneficial for the environment as well as communities. It is an efficient, effective, and environmentally-friendly way to encourage development using existing infrastructures, services, and resources. Redeveloping brownfields can also generate great economic benefits if appropriate and reasonable methodologies are designed. There are strong environmental, social, and economic grounds to redevelop and utilize brownfields. Redevelopment provides a means of creating jobs, increasing the federal and local tax base, increasing the attractiveness of neighbourhoods, and protecting natural ecosystems (US EPA, 1999; NRTEE, 2003; UK Environment Agency, 2003).

Since the 1980s, North America and Europe have made moves to foster sustainable redevelopment of brownfields. The United States of America enacted the Liability Act Superfund to reclaim 1,410 heavily contaminated sites and the National Brownfields Association has estimated that 600,000 brownfield sites exist across the nation (Mueller, 2005). Canada also has paid a great deal of attention to brownfield redevelopment projects (aboutRemediation^{TM}, 2012). Policy makers and developers have focused their efforts on transforming brownfields into industrial areas, commercial areas, and residential areas depending on the characteristics of the community and the site itself. At the same time, brownfield redevelopment projects also harbour great risks, which can impede the decision-making process. Major risks faced during brownfield redevelopment associated with costs are possible cost overruns in clean-up operations (the technologies for remediation of brownfields summarized by Reddy et al., (1999)), possible liability claims from accidents or contaminant exposure spilled in the past or during the clean-up and uncertainty about future community acceptance (Wade VanLandingham et al., 2002). All of these risks mainly arise from imperfect knowledge of the complexity of brownfield sites. Complications include complex geological formations and physical properties of the subsurface, the uncertainty of the source zone, the number

and features of existing and potential contaminants, and the unreliability of detecting and contouring of the contaminant plume. In addition, cleanup standards have also changed over time and vary depending on the use of a property creating further uncertainty and impediments to development.

In order to reduce the risks and uncertainties of brownfield site assessment, geologists and hydrogeologists have endeavoured to create effective methods and invent innovative tools to help improve site characterization. Limiting factors such as cost, time, and accuracy may not always permit further sampling. Empirical and comprehensive analysis of field data is one option and construction of screen-level or more sophisticated models is another option. The typical screen-level model to evaluate the soil vapour transport to indoor air pathway was built by Johnson and Ettinger (J&E) (1991) which incorporates analytical and semi-analytical solutions for steady-state one-dimensional vapour transport. Despite numerous simplifying assumptions used to develop the closed-form J&E model, limited site characterization budgets have motivated its wide-spread use. Alternative models (Abreu and Johnson, 2005, 2006; Bozkurt et al., 2009; Yu et al., 2009; Ozgur et al., 2009), which include additional science-based mechanisms controlling the fate and transport of contaminants from the subsurface into buildings accounting for advection, diffusion, and biodegradation processes, could also be used. These models have been applied to investigate the impact of soil type, physical properties of aquifers, vapour source concentration, vapour source location relative to the house, the air exchange rate in the house, building construction, the chemical degradation rate, the anatomy of the contaminant source on vapour pathways, and the distribution of indoor air concentration. However, at present, we are unaware of any previous efforts to simulate the fate and transport of contaminants from groundwater through the vadose zone and into indoor air for multiple residential dwellings on a three-dimensional problem geometry. The criterion of the influence of neighbouring houses on the soil vapour distribution is essential when attempting to use a numerical model to help delineate an exclusion zone (defined relative to the geometry of the groundwater plume, providing the rationale to exclude the need to conduct further pathway assessment and monitoring and hence used to characterize the house as being unaffected by the subsurface contamination).

The problem of uncertainty reduction in site characterization, subject to a limited amount of data collection, has also been addressed using geostatistical tools, inverse modeling, Monte Carlo simulation, and optimization techniques by many researchers. James and Gorelick (1994) built a Bayesian data worth framework for an optimal remediation design given the high cost of data collection and found out that the optimum mainly depended on the mean and variance of hydraulic conductivity. Zhang et al. (2005) applied the Kalman filter algorithm to determine the least cost

design of groundwater quality monitoring networks. No researcher has adopted these approaches for presenting the worth of data in a risk-based brownfield redevelopment project. Comprehensive algorithms are needed to reduce uncertainty and improve accuracy in order to provide better measures of the effectiveness of data and more insight into the complexities of making a project management decision.

The feasibility of brownfield redevelopment projects do not only rely on environmental and engineering advancement, but also need to consider the economic ramifications. The net cash flow, which is the stream of present value, is the economic measure of income and capital investment project selection. Massmann et al. (1987) applied a risk-cost-benefit analysis for waste management facilities to look at the variation of risk, cost, and benefits with different risk-based engineering designs. Also, they introduced and defined the risk capital that quantified the risk associated with the waste management project. Yu et al. (2012) derived and summarized five different approaches to estimate the risk capital of turning brownfields into residential buildings. Risks associated with these approaches include hydrogeological risks and financial risks (discount rate and the US national house price index). However, they failed to link actuarial and financial approaches to define safety loadings to compensate against uncertainties from hydrogeological and financial aspects in an objective and practical way. Moreover, they ignore the value of collecting site information to incorporate into the cost function.

Risk management options are self-insurance, avoidance, risk control/engineering, contractual transfer, and risk transfer. Insurance is considered as the primary risk management tool in this thesis. For the last decade environmental insurance has become one of the few tools capable of mitigating and transferring risks, which also has been welcomed by regulators and public officials. The protection afforded by environmental insurance is available in a number of different policy terms and types of coverage (Yount and Meyer, 2005; Wernstedt et al., 2003), but there are no standard policies and actual rates of utilization are below what insurers might expect. The availability of insurance coverage is low in the insurance market and the fact that the total number of insurance companies that offer these products is declining indicates that a more disciplined and professional underwriting is required. The risk capital introduced by Yu et al. (2009) was equivalent to the net insurance premium in the traditional insurance industry. In the practical world, insurance underwriters make insurance policies to adjust the premium for a given ground-up loss claimed by insured parties. It is needed to build a more sophisticated structure of insurance product for brownfield redevelopment claims to help stabilize a developer's long term environmental cost exposure.

1.2 Research Objectives

The main objectives of this thesis are to improve the understanding of the physical process of the contaminated soil vapour exposure pathways from a residual source zone located under the water table into multiple residential buildings; provide an awareness of the importance of site characterization in a risk-based brownfield redevelopment project and determine the effort required to collect data; and build a framework for evaluating the risks associated with imperfect knowledge of the subsurface and the financial market, in a quantitative way, in order to develop an optimal financial strategy for the project. Specific objectives of this study were:

- To examine the fate and transport of volatile organic contaminants originating from a NAPL source zone located below the water table and their potential impact on multiple residential dwellings located down-gradient of the source zone using the numerical model *CompFlow Bio*,

- To compare *CompFlow Bio* and J&E models, as well as the ASTM (2008) guidance in terms of delineating the exclusion zone,

- To use the Kalman filter method to quantify the worth of data based on statistical measures in order to demonstrate the importance of site characterization and long-term monitoring to reduce parameter uncertainty and model prediction uncertainty for the decision making process,

- To study the reliability of using sparse datasets to infer whether one of the houses located adjacent to affected houses has indoor air problems given that the affected houses are monitored,

- To further develop two approaches of pricing the risk capital (hydrogeological and financial risks) based on the actuarial premium and martingale premium calculation principles with an innovative way of defining the safety loading terms to avoid developers' insolvency,

- To construct an objective cost function to consider the cost of data and the risk capital in order to obtain the least cost and the best sampling scheme.

1.3   Thesis Scope

The thesis has been organized into 5 chapters. Chapter 2, 3, and 4 constitute the main body of this thesis, and each chapter was completed as an independent manuscript for publication containing its own abstract, introduction, and conclusion.

Chapter 2 summarizes the simulation results and findings of the fate and transport of a contaminant from groundwater to indoor air using the numerical model *CompFlow Bio* with variations of model parameters. A three-dimensional problem geometry, reminiscent of the Rivett (1995) field experiment in the Borden aquifer and motivated by the work of Yu et al. (2009), is used, which contains houses located both above and adjacent to the groundwater plume in order to define an exclusion zone. In this chapter, the impact of the controlling parameters on the indoor air concentration is observed and results are compared with the conservative J&E model and ASTM (2008) in terms of the exclusion zone.

Chapter 3 presents the methodology to evaluate data worth for vapour intrusion into buildings under uncertainty. The geostatistical kriging algorithm and the statistical Kalman filter data assimilation method are combined to improve the estimation of indoor air concentration. The number of sampling data acquired in the estimation is selected from a practical perspective and used to examine the effectiveness of data collection. Eventually, an optimal number of samplings is selected according to three statistical metrics based on different measures of optimization results.

Chapter 4 presents an optimal design for a risk-based site characterization of a brownfield redevelopment project using the statistics obtained from the results of incorporating the hydrogeological model with the estimation made through Kalman filter based on given sampling information. This chapter proposes two major methods to price the risk capital modified from Yu et al. (2012). They both take consideration of hydrogeological risk and financial risk in the form of a safety loading term which represents the level of risk aversion by developers. The safety loading term reflects the accuracy (the expected value) and the uncertainty (the standard deviation) resulting from the estimation using Kalman filter. The optimal least cost including the risk capital and the cost of data is then calculated and selected.

Chapter 5 summarizes the findings of the individual chapters and gives a broader picture and insights into potential applications of this research in a risk-based environmental project. This chapter also provides recommendations for future studies.

1.4 Contributions

This thesis provides a comprehensive framework for evaluating a risk-based project from different points of view, mainly focusing on environmental and financial aspects. It acknowledges uncertainties a project may face during developers' liability period and raises the question of how to optimize the trade-off between increasing site-investigation budgets and overall cost reductions.

The objective of this thesis is not only to alleviate the risk exposure during brownfields redevelopment projects by improving the prediction of contaminated indoor air concentrations and to search for the optimal balance between increasing site characterization budgets and decreasing contingency measure to cover the risk exposure, but also to quantify and manage the risk. The foundation of this thesis is to provide a practical guideline for developers and decision makers to look upon when benefits of a project are substantial in comparison with risks from both environmental/engineering and financial perspectives. This thesis shows that minimizing damage and potential risks to meet environmental regulatory standards and maximizing the investment capital may not be in an agreement with each other. Site characterization can help to better capture and understand a site and the surrounding environmental influences so that effective remedial action can be enforced and accurate prediction of a possible future failure can be made. At the same time, high cost of sampling and monitoring makes it impossible for developers to increase the budget infinitely and to ignore probability principles. This thesis presents a balance between the two and shows that both needs can be relatively satisfied.

This thesis helps hydrogeologists comprehend the risk from a different point of view and to understand what important factors affect the project during the decision making process. Influential parameters like discount rate and real estate value could be dominant factors contributing to risk capital and ultimately determine the feasibility of an environmental project. This study highlights the important role site characterization plays, although, excessive information is not necessary. Different techniques are combined to improve model prediction and parameter estimation, which are quite advanced. The numerical model used to simulate TCE soil gas transport is a three-dimensional multi-phase multi-compositional model with all three phases (water, gas, and oil) active and is able to simulate the complex physics involved in the process. Even though it is computationally expensive, it gives a better interpretation of the physics.

There are several insurance products with different types of coverage in the market for environmental risk-based projects, but there is no standard policy. The market has been relatively small because of lack of information and experience despite the high risks. Therefore, environmental insurance markets need scientific support to better understand loss occurrence intensity and loss

frequency (the loss is defined when environmental standards are not met) for professional insurance policy underwriting. This thesis proposes a risk capital valuation methodology, which is equivalent to an insurance premium calculation for environmental projects. The way insurance premiums are calculated follows the classic actuarial insurance premium calculation principle and the martingale insurance premium calculation principle, which are commonly seen in the insurance industry. This thesis builds a bridge between insurance and hydrogeology. Also, analysis of different sampling scenarios can draw developers' attention to risk management since the cost of risk could be substantial in some cases in an environmental project. In a nutshell, understanding of risk capital valuation in a risk-based environmental project will facilitate developers to make an optimal investment strategy.

Chapter 2 Simulating factors contributing to an exclusion zone for vapour intrusion of TCE from groundwater into indoor air

Chapter Summary

This chapter is an extension of the work by Yu et al. (2009) to examine exposure pathways of volatile organic compounds (VOCs) originating from a NAPL source zone located below the water table, and their potential impact on multiple residential dwellings down-gradient of the source zone. We use three-dimensional problem geometry reminiscent of the Rivett (1995) field experiment in the Borden aquifer, which contains houses located both above and adjacent to the groundwater plume in order to define an exclusion zone. Simulation results using the numerical model *CompFlow Bio* indicate that houses which are laterally offset from the groundwater plume are less affected by vapour intrusion due to limited transverse horizontal flux of TCE within the groundwater plume, in agreement with the ASTM (2008) guidance. Uncertainty in the simulated indoor air concentration is sensitive to heterogeneity in the permeability structure of a stratigraphically continuous aquifer, with uncertainty defined as the probability of exceeding a regulatory limit. Within this uncertainty framework, we show that the Johnson and Ettinger (1991) model generates overly-conservative results and contributes to the exclusion zone being much further away from the groundwater plume relative to either *CompFlow Bio* or ASTM (2008).

2.1   Introduction

Numerous sites with monitored shallow groundwater plumes originating from volatile organic contaminant source zones also exhibit deleterious impacts on the vadose immediately above the plume (Rivett, 1995; Smedes et al., 1993). This issue is problematic given the potential for vapour intrusion of the volatile organic contaminants to degrade the indoor air quality of houses with foundation slabs located within the area of impacted vadose zone. To address this issue, two layers of regulatory guidance have been developed. In the first level, the concepts of "exclusion criteria" and an "exclusion distance" are defined relative to the geometry of the groundwater plume and other characteristics of the site, and provide the rationale to exclude the need to conduct further pathway assessment and monitoring and hence characterize the house as being unaffected by the subsurface contamination (Luo et al., 2010). For the scenario where the vapour source is a non-biodegradable volatile organic groundwater plume, ASTM (2008) itemizes the critical distance demarking the

exclusion zone as being 100 feet (approximately 30 m) from the perimeter of the plume. This critical distance is adjusted whether, for instance, the source of the vapours is a dissolved and biodegradable hydrocarbon plume, or even a NAPL source zone. In the second level, science-based regulatory criteria are used to evaluate the fate and transport of volatile organic contaminants from the subsurface into the indoor air of residential dwellings. In particular, the Johnson and Ettinger (1991) heuristic model (the J&E model) has become the de-factor standard. While designed to be conservative (USAF, 2006; Hers et al., 2003; Schreuder, 2006; Fitzpatrick and Fitzgerald, 2002; Mills et al., 2007; Yu et al., 2009; Provoost et al., 2010; Johnston and Gibson, 2011), the J&E model was also found to be the most accurate and least conservative in comparison to six other methodologies used in Europe (Provoost et al., 2009). Despite numerous simplifying assumptions used to develop the closed-form analytical solution constituting the J&E model, limited site characterization budgets have motivated its wide-spread application as a preliminary screening tool to virtually all sites in the US and Canada. Given a regulatory limit defining the maximum permissible indoor air concentration (OEHHA-CEPA, 2007; NYSDOH, 2005), the J&E model can then be as an alternative to ASTM (2008) to identify those residential dwellings that are unlikely to be adversely affected by subsurface contamination and hence are within the exclusion zone.

While the use of the J&E model to define an exclusion zone is a straight-forward extension of its current scope of application, alternative models could also be used to include additional science-based mechanisms controlling the fate and transport of contaminants from the subsurface into the indoor air further refining the geometry of the exclusion zone. Yu et al. (2009) used the numerical model *CompFlow Bio* to simulate the fate and transport of TCE from groundwater to indoor air for the problem geometry reproduced here in Figure 2.1. This problem geometry was developed under the assumption that the pathway for the contamination from the source zone into the indoor air is: advective and diffusive transport in the groundwater from the NAPL source zone to beneath the foundation slab of the residential dwelling, upward diffusive transport across the capillary fringe into the vadose zone, and advective and diffusive transport in the gas phase within the vadose zone through a crack in the foundation slab and into the residential dwelling. This problem geometry was motivated by a field experiment conducted by Rivett (1995) in the Borden aquifer, who observed that weak vertical transverse transport resulting from hydrodynamic dispersion below the water table allows the groundwater plume to advect contaminants a significant distance without mass transport across the capillary fringe. McCarthy and Johnson (1993) conducted a laboratory experiment and observed that vertical mass transport of contaminants across the capillary fringe was controlled by aqueous phase diffusion. Yu et al. (2009) used a two-dimensional numerical mesh to represent the

problem geometry and hypothesized that the resulting indoor air concentrations would be conservative given that the groundwater plume was forced to advect beneath the foundation slab of the house, and the lack of transverse horizontal transport to the groundwater plume could not reduce groundwater concentrations within the plume. Given this limitation in the numerical mesh, they were unable to conclude whether the J&E model generated conservative indoor air concentrations relative to *CompFlow Bio* for the problem geometry shown in Figure 2.1.

Within the last two decades, numerical modeling tools have evolved significantly to help interpret and parameterize the physical processes controlling the fate and transport of contaminants from groundwater into the vadose zone. Early models such as those by Mendoza and Frind (1990a, 1990b), Culver et al. (1991), Celia and Binning (1992) and Thomson et al. (1997) focused on mathematical formulation and model development issues to adequately represent this transport mechanism. Later models evolved to focus on soil vapour intrusion pathways, including: the effect of atmospheric pressure (Massmann and Farrier, 1992), and the effect of wind speed and direction (Riley et al., 1999). Parker (2003) and DeVaull et al. (2002) emphasized the significance of biodegradation, while Abreu and Johnson (2005, 2006) then included biodegradation processes into a three-dimensional numerical model to investigate the impact of building construction, degradation rate, and the source zone architecture on indoor air concentrations. Robinson and Turczynowicz (2005) and Bozkurt et al. (2009) also use fully three-dimensional problem geometries, with the latter examining the impact of multiple stratigraphic units on the fate and transport of contaminants through the vadose into the indoor air. At present, we are unaware of any previous efforts to simulate the fate and transport of contaminants from groundwater, through the vadose zone and into the indoor air for multiple residential dwellings. The last criterion is essential when attempting to use a numerical model to help delineate an exclusion zone.

The objective of this chapter is to use the multi-phase compositional numerical model *CompFlow Bio* to examine the fate and transport of volatile organic contaminants originating from a NAPL source zone located below the water table, and their potential impact on multiple residential dwellings located down-gradient of the source zone. While the problem geometry is based directly on that of Yu et al. (2009) and hence is reminiscent of the Rivett (1995) field experiment in the Borden aquifer, we extend the problem geometry here to be fully three-dimensional. This enhancement was motivated by the need to accommodate the multiple houses, with some laterally offset from the direction of groundwater plume advection, in an attempt to delineate an exclusion zone. In addition, we are now in a position to evaluate the impact of the transverse horizontal flux of contaminant (to the direction of the groundwater plume advection) on reducing contaminant concentrations within the plume and

ultimately within the indoor air. Following the discussion in Yu et al. (2009) detailing the methodology for comparing the *CompFlow Bio* and J&E models for the problem geometry shown in Figure 2.1, we can then further refine statements in Yu et al. (2009) regarding the conservative nature of the J&E model. We have three fundamental questions that we wish to explore through the use of the conceptual and numerical model. First, what is the relative importance of variations in the conceptual model such as recharge, inserting a clay layer with and without recharge, extending the thickness of the vadose zone, and heterogeneity in the subsurface permeability structure of the stratigraphically continuous aquifer, on indoor air concentrations within houses located directly above and adjacent to the groundwater plume? Second, how do the *CompFlow Bio* and J&E models as well as the ASTM (2008) guidance compare in terms of delineating the exclusion zone for the conceptual model shown in Figure 2.1? Third, by using statistical methods to calculate the probability of indoor air concentrations exceeding the NYSDOH (2005) regulatory limit for all houses within the *CompFlow Bio* model domain, can we determine with confidence the degree to which the exclusion zone can be delineated, and furthermore the degree to which the J&E model provides conservative estimates of indoor air concentration for the conceptual model shown in Figure 2.1?

2.2   Conceptual and Numerical Model

The conceptual and numerical model used in this chapter is largely an extension of that initially presented in Yu et al. (2009). As such, development of the conceptual and numerical model consists of two stages. First, we outline the formulation of the three-phase multi-component compositional numerical model *CompFlow Bio* to solve the relevant governing equations and constitutive relationships to simulate the fate and transport of TCE emanating from a non-aqueous phase source zone located in a variably saturated aquifer. In contrast to Yu et al. (2009), we only focus on the scenario where the source zone is located below the water table. In this case, the pathway includes dissolution of TCE into the ambient groundwater, mass transfer across the capillary fringe, and subsequent advective-dispersive transport in the mobile soil gas and groundwater towards the foundation slab of multiple structures located below grade. Second, we develop a conceptual model of the variably saturated aquifer, which is: (1) extension of the Yu et al. (2009) model to become fully three-dimensional, (2) inclusion of six regularly-spaced structures within the expanded domain, with two located directly above the groundwater plume, and the remaining four laterally offset from the plume, (3) a statistical analysis of the distribution of indoor air concentrations and attenuation coefficients within all six houses, resulting from Monte Carlo simulations using alternative

11

heterogeneous Borden-like permeability aquifer distributions, and (4) inclusion of discrete stratigraphic features into the three-dimensional simulation domain in order to assess their impact on indoor air concentrations. We reiterate from Yu et al. (2009) that the intent of this conceptual model is to serve as a process-based model to assess indoor air concentrations within prototypical dwellings located in a heterogeneous sandy variably-saturated aquifer with a shallow water table. Once again, the intended outcome of this model (i.e., the *CompFlow Bio* model) is to act as a benchmark to compute the impact of factors controlling the attenuation of indoor air concentrations for direct comparison to the *de facto* industry standard approach developed by Johnson and Ettinger (1991) (i.e., the J&E model). However, here we compare and contrast the merits of the *CompFlow Bio* and the J&E models for use in defining an exclusion zone. The reader is referred to Yu et al. (2009) for a discussion of differences between these two conceptual and mathematical/numerical models.

2.2.1 Formulation

The numerical model presented here follows directly from Yu et al. (2009), and was initially developed by Forsyth (1993) and then further upgraded by Unger et al. (1996), Forsyth et al. (1998) and Slough et al. (1999). Here, we only present those equations from Yu et al. (2009) needed for this discussion, and reference the omitted equations from Yu et al. (2009) where relevant. *CompFlow Bio* is a multi-phase multi-component compositional model that considers three mobile phases, namely; the aqueous (q), non-aqueous (n) and gas (g) phases. Components are summarized as water(w), air (a) and TCE (c_{TCE}). Assuming equilibrium partitioning of components $p = \{w, a, c_{\text{TCE}}\}$ between phases $l = \{q, n, g\}$ under isothermal conditions, the conservation of moles for each component p is written as:

$$\frac{\partial}{\partial t}\left[\sum_l (\phi S_l M_l X_{pl}) + \rho_r K_{dp} M_q X_{pq}\right] = -\sum_l \nabla \cdot \left(M_l X_{pl} \mathbf{V}_l\right) + \sum_l \nabla \cdot \left(\phi S_l \mathbf{D}_l M_l \nabla X_{pl}\right) + Q_p \quad (2.1)$$

where the Darcy flux of each phase is given by:

$$V_l = -\mathbf{K} \cdot \frac{k_{rl}}{\mu_l} \left(\nabla P_l + \rho_l g \nabla z \right) \quad . \tag{2.2}$$

The hydrodynamic dispersion tensor $\mathbf{D}_l$ includes both mechanical dispersion and molecular diffusion in porous medium and is given by:

$$\phi S_l \mathbf{D}_l = \alpha_T^l V_l \mathbf{I} + \left(\alpha_L^l - \alpha_T^l \right) \frac{V_l V_l}{V_l} + \phi S_l \tau d_l \mathbf{I} \quad . \tag{2.3}$$

A summary of the nomenclature is as follows:

$S_l =$ saturation of phase l $[-]$

$P_l =$ pressure of phase l [kPa]

$M_l =$ molar density of phase l [moles/m^3]

$X_{pl} =$ mole fraction of component p in phase l $[-]$

$\rho_r =$ mass density of porous media (rock) r [kg/m^3]

$\rho_l =$ mass density of phase l [kg/m^3]

$\mu_l =$ viscosity of phase l [kPa $\cdot$ day]

$\phi =$ porosity of porous media $[-]$

$\mathbf{K} =$ Intrinsic permeability of porous media [m^2]

$K_{dp} =$ sorption coefficient of component p onto porous media [m^3/kg]

$k_{rl} =$ relative permeability of phase l $[-]$

$z =$ depth [m]

$d_l =$ molecular diffusion of all components in phase l [m^2/day]

$\tau =$ tortuosity of porous media $[-]$ from Millington (1959)

$\alpha_L^l =$ longitudinal dispersivity of phase l [m]

$\alpha_T^l =$ transverse dispersivity of phase l [m]

$Q_p =$ source (+ve) or sink (-ve) term for component p [mole/(m^3 · day)].

The four terms in Equation (2.1) each have a physically simple interpretation in the context of this work. The first term represents the accumulation of moles of a single component (i.e. TCE where $p = \{c_{\text{TCE}}\}$) in all phases $l = \{q, n, g\}$ within a single control volume node of the domain. The second term represents the advective molar flux of the single component within all phases into and out of the node. The third term represents the dispersive Fickian-like molar flux of component p within all phases. Finally, the fourth term represents the molar rate by which the single component is either injected into or removed from the node through a physical process such as an injection well, or even biological decay.

2.2.2   Problem Geometry

In this part of study, we wish to use the conceptual model to examine the issue of defining an exclusion zone for residential houses located adjacent to a groundwater plume. However, definition of an exclusion zone is based on fundamental knowledge of processes governing the fate and transport of TCE from the source zone beneath the water table to the indoor air of each house. In particular, the two processes described in Yu et al. (2009) also apply in this study, and are summarized here for completeness. First, heterogeneity in the permeability of the aquifer may cause the plume to be deflected away from the centre of the groundwater plume and vertical diffusion process may cause the plume to be transferred upwards and approaching the water table while the flow of water that has infiltrated from precipitation will act to push the plume downward. Once the plume is sufficiently near the water table, hydrodynamic dispersion (including molecular diffusion) will then cause the upward transport of TCE across the capillary fringe and into the vadose zone. Second, after TCE has reached the vadose zone, transport of TCE towards the foundation slab of the residential dwelling will be driven by advection and diffusion. Operation of a high-efficiency furnace (or alternatively heating/ventilation/HVAC system) will induce a "stack effect" causing the basement to be slightly under-pressurized with respect to the main floor and the ambient atmospheric pressure. Assuming there is a crack in the foundation slab in the basement, slight depressurization of the basement may

induce advective flow of the soil gas beneath the foundation slab into the indoor air. Bearing in mind the above two processes, the notion of an exclusion zone can be examined by introducing multiple residential dwellings within the conceptual model, with some directly overlying the groundwater plume, and others adjacent to the plume. The proximity of the foundation slab of each house to the TCE source zone, groundwater plume, and resulting TCE soil gas, will affect the balance of advective and dispersive soil gas flux of TCE through the foundation slab into the indoor air.

Figure 2.2a and 2.2b depict the geometry of the base scenario numerical model, listed as "scenario 1" on Table 2.1. Dimensions of the problem geometry are $160\ m$ (length) $\times 50\ m$ (width) $\times 8\ m$ (depth). A total of six houses are included within the domain, each with a foundation slab area of $150\ m^2$. Each house is assumed to consist of a basement and a main floor at ground level. Assuming each floor is $3\ m$ tall, the volume of the house V_H is $900\ m^3$. In addition, the living space of each house given by the above-ground floor is $150\ m^2$. This is somewhat smaller than the average size of a single family house from the US Census Bureau given as $201\ m^2$. The houses are arranged in two rows of three houses perpendicular to the direction of groundwater flow (i.e. the x-axis), with the center houses directly over the groundwater plume. The houses are arranged with $10\ m$ gaps between them. The foundation slab for each house is located $2\ m$ below grade and is $0.2\ m$ thick. The domain is discretized with $41 \times 70 \times 35$ nodes in the $x-$, $y-$ and $z-$ directions, respectively. Along the $x-$ direction, the maximum mesh size is $5\ m$ and is reduced to $2.5\ m$ in the vicinity of the houses. Along the $y-$ direction, the maximum mesh size is $2\ m$ between the houses, and is reduced to $0.5\ m$ beneath the houses. Finally, in the $z-$ direction the discretization follows from Yu et al. (2009) with a mesh size of $0.2\ m$ assigned from the ground surface to a depth of $6\ m$ below grade. Thereafter, the mesh size increases to $0.4\ m$ until the bottom of the domain. This discretization was established during a grid convergence test conducted to balance the need to: minimize numerical dispersion, retain realistic simulation times, ensure that the groundwater plume of TCE could travel the entire length of the domain with minimal impact to the vadose zone in keeping with Rivett (1995), and accommodate details of alternative heterogeneous permeability realizations. Figure 2.3 depicts a single realization of a heterogeneous and statistically isotropic permeability field with properties listed on Table 2.2. Values on Table 2.2 are characteristics of the Borden aquifer as measured by Woodbury and Sudicky (1991) and Conant et al. (1996), with the variance increased substantially to be similar to that of the more heterogeneous Cape Cod aquifer (Hess et al., 1992). The permeability realization was generated using the algorithm described by Robin et al. (1993). Note that the base scenario simulation does not include the clay layer depicted n Figure 2.2b and Figure 2.3. Instead, inclusion of this discrete stratigraphic feature into the three-dimensional simulation domain will be

15

used in an alternative scenario (see scenarios 4 and 5 in Section 2.3.2) in order to assess its impact on indoor air concentrations. Finally, we explore the impact of the thickness increasing of the vadose zone in an alternative scenario (see scenario 6 in Section 2.3.2) with the revised problem geometry depicted in Figure 2.2c.

A critical element of this chapter is to examine the impact of the dimensionality of the problem geometry on the simulated indoor air concentrations. This is motivated by the statement in Yu et al. (2009) that their use of a two dimensional model would create conservative results by forcing the groundwater plume emanating from the TCE source zone to advect beneath the foundation slab, and by not permitting horizontal diffusion (in their y − direction) to diminish TCE concentrations before they reach the vadose zone. We examine this issue of dimensionality by adding the y-dimension while retaining as many features of the original Yu et al. (2009) problem geometry as reasonable. Specifically, all boundary conditions, physical and chemical hydrologic properties, foundation slab, basement wall/interior properties, and indoor air concentrations are accommodated within and post-processed from the base scenario in a manner identical to that described by Yu et al. (2009). In summary, boundary conditions for the base scenario model are as follows: the ground surface has a constant pressure gas phase boundary and the ambient atmosphere pressure is $100\ kPa$; the bottom and side boundaries parallel to the groundwater flow direction are impermeable; a constant aqueous phase head is imposed at the groundwater inflow and outflow boundaries resulting in a groundwater flow velocity of 0.09 m/day (Woodbury and Sudicky, 1991) along the x-direction; 0.18 m^3 of TCE is injected 55 m up-gradient of the first row of houses below the water table; and, the houses are depressurized by 10 Pa relative to ambient atmosphere pressure. For the base scenario model, we assume there is no infiltration from precipitation to generate conservative indoor air concentrations. However, in an alternative scenario (see scenario 2 of Section 2.3.2) an annual infiltration rate of 0.2 $m/year$ is assigned (Solomon et al., 1992). All relevant physical and chemical hydrologic properties of the system follow directly from Tables 2 to 5 of Yu et al. (2009). The soil gas emanating from the subsurface can only pass through the foundation slab to reach the basement. The basement walls of all houses are impermeable, and each house has a single shrinkage crack along the entire perimeter of the basement floor. The aperture of the crack for the base scenario is $2b = 100\ \mu m$ yielding a permeability of $K_{slab_C} = 8.33\ \times 10^{-10}\ m^2$, while the foundation slab which is made of concrete is assigned a permeability of $K_{slab_M} = 1\ \times 10^{-20}\ m^2$. The resultant bulk permeability of the foundation slab is $K_{slab_B} = 1.83\ \times 10^{-13}\ m^2$ following Equation (11) of Yu et al. (2009), with all basement and foundation slab properties summarized here on Table 2.3. The methodology for

calculating the indoor air concentration of TCE in unit of parts per million by volume $ppmV$ follows from Equations (12) to (14) of Yu et al. (2009), with a base scenario air exchange rate of $0.5/hr$.

Simulations using the *CompFlow Bio* model consist of three main stages. First, we establish an appropriate initial condition for steady-state aqueous and gas phase flow in the subsurface using the boundary conditions itemized above. This initial condition includes the presence of all six houses. Second, we introduce a TCE source zone within the model by injecting $0.18 \ m^3$ of TCE over a period of one day forming a DNAPL source zone of volume $0.18 \ m^3$ below the water table. The start of TCE injection is denoted as $t = 0 \ days$. Third, the TCE source zone is then allowed to redistribute itself as necessary, and TCE contamination then emanates from the DNAPL source zone and is transported within the steady-state aqueous and gas phase flow fields towards the basements of the six houses which are all located down-gradient of the source zone.

2.3   Results

Use of the *CompFlow Bio* numerical model for the purposes of defining an exclusion zone for the six houses located within the problem geometry is based on a three-step process, each with an established body of science-based literature. First, previous laboratory and field experiments (Rivett and Cherry, 1991; Rivett, 1995; McCarthy and Johnson, 1993) have established a process-based hydrological conceptual model defining the mechanisms for the fate and transport of volatile organic contaminants from the groundwater, through the capillary fringe, and into the vadose zone. Section 2.2.2 of this chapter outlines our efforts to ensure that the simulation domain used as the basis for this work is faithful to these previous efforts. Second, each house within the simulation domain is in effect a sensitive receptor based on monitored indoor air concentrations. The indoor air concentration is impacted via the mass flux of TCE (the volatile organic contaminant simulated in this chapter) in the soil gas across the foundation slab and into the basement of each house. As such, the foundation slab is in effect a "flux fence" reminiscent of field experiments conducted by Mackay et al. (1986), Leblanc et al. (1991), Conant et al. (1996), Devlin and Barker (1996), Ezzedine and Rubin (1997), Schirmer and Barker (1998), Zheng and Jiao (1998), King et al. (1999), Rivett et al. (2001), Hess et al. (2002), Bockelmann et al. (2003), Mackay et al. (2006), Cai et al. (2011) and Béland-Pelletier et al. (2011), and further utilized in the body of work related to partial source zone depletion (Blum and Annable, 2008; Sale et al., 2008; Brooks et al., 2008). However, the fate and transport processes between the TCE source zone and foundation slab fence are significantly more complicated than

previously examined given that the foundation slab is in the vadose zone rather than beneath the water table. Also, while the foundation slab is parallel to the advection of the groundwater plume (and of course, the ground surface), it is perpendicular to soil gas flow induced by the occupancy of each house. Yu et al. (2009) and Yao et al. (2011) both discuss this idea of the soil gas flux as a contribution to the indoor air concentration of TCE. The efforts of Johnson et al. (2011) to monitor soil gas concentrations beneath the foundation slab is reminiscent of efforts to monitor flux fences for groundwater plumes. Third, a given house is defined as being either within (i.e., unaffected house fit for occupancy) or outside (i.e., and an affected house unfit for occupancy) the exclusion zone based on whether its indoor air concentration exceeds a regulatory limit during the entire transient period over which the TCE source zone is depleted. Here we use an indoor air regulatory limit for TCE of 5 $\mu g/m^3$ which can also be expressed as $8.95 \text{X} 10^{-4}\ ppmV$ (NYSDOH, 2005).

2.3.1 Base Scenario Exclusion Zone

The fate and transport of TCE from groundwater to the indoor air as simulated by *CompFlow Bio* for each of the six houses for the base scenario (scenario 1) is shown in Figure 2.4a. This figure shows the mole fraction of TCE in the gas phase within the aquifer 1800 days after TCE was introduced. Contours of TCE gas phase mole fraction are provided along a transect through the source zone and the middle of houses 2 and 5, and parallel to groundwater advection (i.e. along the x-z plane). In addition, contours are also shown on transects through the rows of houses 1-2-3 as well as houses 4-5-6, perpendicular to groundwater advection (i.e. along the y-z plane). We remind the reader that all physical dispersion below the water table and in the vadose zone is entirely due to molecular diffusion as all dispersivities α_L^l and α_T^l $l = \{q, n, g\}$ (resulting from hydrodynamic dispersion) were set to zero. Also, an extensive grid refinement was performed to minimize numerical dispersion within the region of contaminant transport as exhibited by Figure 2.4a. We note that below the water table, the groundwater plume exhibits a sharp contrast from high to low concentrations transverse to the direction of groundwater advection in keeping with observations at the Borden aquifer (Sudicky, 1986). Perturbations in the shape of the groundwater plume are a consequence of the impact that heterogeneity in the permeability structure of the aquifer has on groundwater flow, as well as the induced macro-dispersivity that these perturbations in groundwater flow will have at the mesh discretization scale. Given the close proximity of the TCE source zone to the water table, we can see that for the particular permeability realization shown in Figure 2.3, once the TCE plume has advected about 30 m down-gradient from the source zone, it transports upwards by advection to the capillary

fringe due to the structure of the heterogeneous permeability field. Once TCE diffuses through the capillary fringe and reaches the vadose zone, the resulting TCE vapour plume rapidly spreads across the vadose zone. This spread is largely due to the vacuum induced by the heating/ventilation/HVAC system within each house which creates an advective flux of the gas phase from the vadose zone, across the foundation slab, and into the indoor air. Molecular diffusion of TCE in the gas phase is also a substantial transport mechanism, particularly in regions distant from the foundation slab. Houses 2 and 5, which reside directly over the groundwater plume, exhibit the highest mole fraction of TCE in the gas phase beneath their foundations slab. All of the houses which are laterally offset from the groundwater plume exhibit lower soil gas concentrations beneath their foundation slabs, with houses 1 and 4 being less than houses 3 and 6 due to the structure of the particular permeability realization.

The mass flux of TCE in the soil gas across the foundation slab and into the indoor air for scenario 1 is shown in Figure 2.5a. This flux is the product of the TCE concentration in the soil gas as given by Equation (12) of Yu et al. (2009) and the volumetric flow rate of soil gas into the house as shown in Figure 2.5b, with this flux of TCE representing F_1 in Equation (13) of Yu et al. (2009). Note that each foundation slab was discretized using $6 \times 20 \times 1$ nodes of dimension $2.5\ m \times 0.5\ m \times 0.2\ m$ in the $x-$, $y-$ and $z-$ directions, respectively, in an attempt to resolve spatial variations in the soil gas composition of TCE beneath the foundation slab induced by the heterogeneous aquifer permeability structure in keeping with the field observations of Johnson et al. (2011). The mass flux of TCE across the foundation slab for all houses generally increases in a transient manner over the simulated 4800-day simulation period as the groundwater plume advects beneath the development. For houses 2 and 5, the mass flux diminishes somewhat after it reaches a peak at about 2000 days due to partial dissolution of the TCE source zone. At 4800 days, the mass flux for houses 1 and 4 is less than that for houses 3 and 6, although they are all within an order-of-magnitude of one another. However, the mass fluxes of TCE into houses 2 and 5 are roughly one to two orders-of-magnitude greater. In contrast, the volumetric flow of soil gas into all houses is relatively similar between 3.96 to 5.74 m^3/day (see Figure 2.5b and Table 2.3) over the entire simulation period, with the fluctuations due to the numerical difficulty of imposing the penalty sink term given by Equation (14) of Yu et al. (2009).

Indoor air concentrations of TCE for scenario 1 within all six houses are shown in Figure 2.6. These concentrations were calculated using Equation (13) of Yu et al (2009) and closely mimic the mass flux of TCE across the foundation slab as is expected. We note two key observations in the indoor air concentrations. First, the indoor air concentrations for houses 2 and 5 which are directly above the groundwater plume are about an order-of-magnitude lower than those presented by Yu et

al. (2009 − see "no recharge" in Figure 13). We endeavored to retain an identical problem geometry and parameterization between the models, with the notable difference of this model containing six houses and also being three-dimensional. Clearly, dimensionality of the problem geometry has a significant influence on the simulation results as hypothesized in Section 2.2.2. Second, the indoor air concentrations for houses 1, 3, 4, and 6 which are adjacent to the groundwater plume are significantly lower than houses 2 and 5. Using the NYSDOH (2005) indoor air regulatory limit for TCE of 5 $\mu g/m^3$, houses 2 and 5 are adversely affected and are outside the exclusion zone, while houses 1, 3, 4 and 6 are all below the regulatory limit and are within the exclusion zone. Given the narrow spacing between the houses of $10\ m$, we conclude that the weak transverse macro-dispersion seen in groundwater plumes generates narrow exclusion zone bands that effectively trace the projection of the groundwater plume onto the ground surface. This occurs despite the fact the region transverse to the plume that exhibits measurable soil gas impacts may extend beyond the boundaries of the exclusion zone. In the United States, the average air exchange rate during the seasons when houses are kept closed is within the range of $0.1 − 2.0\ hr^{-1}$ (ASHRAE, 2009; Ek et al., 1990), and a minimum air exchange rate of $0.35\ hr^{-1}$ is also recommended by the American Society of Heating, Refrigerating, and Air Condition Engineers (ASHRAE). We conclude that the dilution that the indoor air exchange rate offers for a modest-sized house confers a certain degree of protection against subsurface contamination issues, even when weighed against the relatively strict NYSDOH (2005) indoor air regulatory limit.

Both ASTM (2008) and the J&E model provide alternative strategies to delineate the exclusion zone. Specifically, ASTM (2008) indicates that the critical distance beyond which a site may be excluded from further assessment when the vapour source is a dissolved non-degradable plume is 100 feet (or approximately $30\ m$). Given that the centroid of the foundation slab of all houses is within 20 m of the centre-line of the plume along the ground surface, all houses would fall outside the exclusion zone by about $10\ m$. Therefore, in the case of the base scenario (scenario 1), ASTM (2008) provides slightly more conservative guidance than *CompFlow Bio*. As part of additional scenarios in Section 2.3.2, we attempt to generalize this observation by comparing *CompFlow Bio* against ASTM (2008) using alternative problem geometries. Parameterization of the J&E model for direct comparison to *CompFlow Bio* for the base scenario follows from the extensive discussion in Yu et al. (2009), with relevant parameters summarized on Table 2.4 and using Q_{soil} values from Table 2.3. Indoor air concentrations calculated using Equation (19) of Johnson and Ettinger (1991) for each of the six houses in the base scenario are shown in Figure 2.6. Results indicate that the indoor air concentration calculated using the J&E model for each of the six houses is greater than the NYSDOH (2005)

20

regulatory limit indicating that all houses would fall outside the exclusion zone. This result is identical to ASTM (2008) and is more conservative than *CompFlow Bio*. Common use of the J&E model is by means of the "attenuation coefficient" denoted here as $\alpha_{J\&E}$ representing the ratio of the indoor air concentration to that of the source, as defined by Equations (20) and (21) of Johnson and Ettinger (1991). For the *CompFlow Bio* model, α_{CFB} is computed when the TCE indoor air concentration reaches its peak value within each house individually. Comparison of $\alpha_{J\&E}$ and α_{CFB} is provided on Table 2.5. For the base scenario, $\alpha_{J\&E}$ is between 14 to 30 times greater than α_{CFB} for houses 2 and 5 which are both directly over the groundwater plume. For houses 1, 3, 4 and 6 which are adjacent to the groundwater plume, $\alpha_{J\&E}$ is between 617 to 6239 times greater than α_{CFB}. Consequently, we speculate that while the agreement between the exclusion zones defined by ASTM (2008) and the J&E model are reassuring for the base scenario problem geometry, the J&E model appears to be overly conservative when predicting indoor air concentrations relative to the *CompFlow Bio* model. We will revisit this issue later in Section 2.3.3 when we quantify the uncertainty that the heterogeneous permeability field impacts on the α_{CFB} values.

2.3.2 Variations in the problem geometry

A thorough site investigation is a necessary prerequisite, and also part of the professional engineers or geoscientists due diligence, before making any assessments regarding the fate and transport of TCE from an observed source zone to the indoor air of a house. Yu et al. (2009) observed that imperfect knowledge of subsurface conditions can have a significant impact on uncertainty in predicting, or interpreting, indoor air concentrations arising from subsurface conditions. In this context, we investigate the impact of a series of modifications to the base scenario problem geometry that could potentially impact the fate and transport of TCE from the source zone below the water table, across the capillary fringe, into the vadose zone, through the foundation slab, and into the indoor air. These scenarios are itemized as scenario $2 - 6$ as listed on Table 2.1, and are described below.

Following the arguments outlined in Section 3.3.5 of Yu et al. (2009), we begin by introducing scenario 2 which involves modifying the base scenario by increasing infiltration to $0.2\ m/yr$. The rationale for this scenario being that the aqueous phase diffusive flux of TCE across the capillary fringe is the limiting hydrologic process controlling indoor air exposure. Increasing the infiltration rate diminishes the diffusive flux of TCE across the capillary fringe by effectively pushing the groundwater plume downward as the additional water recharges the water table. In a manner

consistent with Yu et al. (2009), we observed in Figure 2.6 that the indoor air concentration within all six houses decreases for scenario 2 relative to the base scenario. In fact, house 2 no longer exhibits indoor air concentration in excess of the regulatory limit and hence should be part of the exclusion zone. We note that neither ASTM (2008) nor the J&E model (see Table 2.4) accommodates recharge in their guidance towards delineating the exclusion zone. Comparison of α_{CFB} for scenarios 1 and 2 (as listed on Table 2.5) provides a convenient methodology to quantify the impact of this particular variation on the problem geometry on reducing indoor air concentration. Recharge is able to reduce the attenuate factor (i.e. decrease indoor air concentrations) by a factor from between 7 and 21 for houses 2 and 5 which are directly above the groundwater plume, to between 43 and 105 for houses 1, 3, 4, and 6 which are laterally offset. Note that volumetric rate of air flow into the houses as listed on Table 2.3 is only marginally impacted by infiltration. Figure 2.4b shows that infiltration is able to generate a greater reduction in TCE soil gas concentrations beneath the foundation slab of houses that have lateral offset relative to those over the groundwater plume, resulting in a corresponding decrease in the flux of TCE across the foundation slab and into the indoor air. We conclude that increased recharge serves to diminish the weak lateral transport of TCE from the groundwater plume into the vadose zone, further tightening the boundaries of the exclusion zone to the perimeter of the groundwater plume. For comparative purposes, we note that Yu et al. (2009) observed that adding recharge (see scenario 8 on Table 8 of Yu et al. (2009)) increased α_{CFB} by a factor of 8 for the single house in the domain which was located in the same position as house 2 for this base scenario.

   With or without recharge, TCE does exhibit low but significant mole fractions along the impermeable side boundaries of the base scenario domain, constraining the further lateral spread of TCE in the vadose zone and effectively increasing the TCE soil gas concentration beneath the foundation slab of all houses. We examine the undue influence of the lateral boundary conditions on the simulated indoor air concentrations by introducing scenario 3 which involves modifying the base scenario by reducing the pressure drop within each house to zero, i.e.: $\Delta P_g^{basement} = 0$ Pa . Therefore, there is no advection of the gas phase in the vadose zone to transport TCE to houses that are laterally offset from the groundwater plume and adjacent to the side boundaries. Diffusive flux of TCE in the gas phase is the only transport mechanism to reach these houses. Figure 2.4c shows little change in the mole fraction of TCE in the gas phase along the side boundaries between scenarios 1 and 3. However, the mole fraction of TCE in the gas phase is much greater for scenario 3 beneath the foundation slab of houses 2 and 5, and between all houses just beneath the ground surface. For scenario 1, the decrease in soil gas concentration is a consequence of clean atmospheric air which is drawn downward from the ground surface and then across the foundation slab when $\Delta P_g^{basement} =$

10 Pa. Figure 2.6 shows that indoor air concentration in all houses for scenario 3 are considerably lower than those for scenario 1. In fact, they are all lower than the regulatory limit. Therefore, we conclude that while the lateral extent of the domain is insufficient to prevent lateral diffusion of TCE in the gas phase within the vadose zone from impacting the side boundaries, the advective flux of TCE in the gas phase across the foundation slab is the dominant mechanism impacting the indoor air quality. To this end, extending the domain further in the y-direction would serve to allow additional clean atmospheric air to enter the basements of the lateral houses and further dilute and constrain the lateral migration of the TCE in the vadose zone.

A significant simplification inherent in the base problem geometry is that the domain consists of a single stratigraphic unit, albeit with a heterogeneous permeability structure. Scenarios 4 and 5 involve adding a low permeability clay layer between the water table and foundation slab within the vadose zone, as depicted in Figure 2.2b and Figure 2.3. Otherwise, the permeability structure of the domain remains unchanged like the base scenario. Scenarios 4 and 5 differ in that the infiltration rate is increased from 0 to 0.2 m/yr to investigate the impact of water saturation within the clay layer on the fate and transport of TCE through the vadose zone. For scenario 4 when infiltration is 0 m/yr, the water saturation of the clay layer is $S_q = 0.90$ due to its close proximity to the capillary fringe and high air entry pressure $P_{cgq}(S_q = 1) = 10.11kPa$. At this water saturation, the Darcy flux of the gas phase (see Equation (2.2)) is reduced as $k_{rg}(S_g = 0.10) = 0.43$ effectively diminishing the advective flux of TCE in the gas phase between the top of the capillary fringe and the base of the foundation slab. In addition, the dispersive flux of TCE in the gas phase is diminished given that it is proportional to S_g (see Equation (2.1)). For scenario 5, the water saturation of the clay layer increases to $S_q = 0.999$ as recharge is added. At this water saturation, both the advective and diffusive fluxes of TCE in the gas phase are non-existent. The only transport mechanism for TCE through the clay layer is diffusive flux in the aqueous phase. Figure 2.4d shows contours of the mole fraction of TCE in the gas phase for scenario 5 at 1800 days after emplacement of the TCE source zone. The clay layer has completely prevented TCE from impacting any of the houses despite the fact the vadose zone below the clay layer exhibits significant TCE soil gas concentrations. Figure 2.6 indicates that while the clay layer with no infiltration (scenario 4) does reduce indoor air concentrations, the degree of protection is not nearly as significant as when the clay layer is fully water saturated (scenario 5). The indoor air concentrations computed using the J&E model for scenarios 4 and 5 are provided in Figure 2.6, and are indistinguishable from those calculated for scenarios 1 and 2. Clearly, the process of identifying stratigraphic units as part of a site investigation is a prerequisite step when quantifying the fate and transport of contaminants from groundwater to the indoor air. For scenario 5, the

CompFlow Bio model indicates that all houses should be part of the exclusion zone whereas both J&E and ASTM (2008) provide the exact opposite guidance. As above, comparison of α_{CFB} for scenarios 4 and 5 (as listed on Table 2.5) provides a convenient methodology to quantify their impact on reducing indoor air concentration relative to the base scenario. For the clay layer without recharge, the attenuation factor for all houses was only decreased by a factor between 1.3 and 10.5. In contrast, with recharge the attenuation factor was decreased by a factor between 1.8×10^4 and 1.7×10^5.

The thickness of the vadose zone has a direct impact on the fate and transport of TCE from groundwater to the indoor air given that it alters the offset between the source zone and center of the foundation slab. Scenario 6 involves increasing the thickness of the vadose zone to approximately 10 m as shown in Figure 2.2c. We note that the structure of the heterogeneous permeability field for the bottom 8 m is identical to that of the base scenario to ensure that the groundwater plume was not perturbed relative to the base scenario. The additional 7 m contains a continuation in the heterogeneous permeability structure. Figure 2.4e shows contours of the mole fraction of TCE in the gas phase for scenario 6 at 1800 days after emplacement of the TCE source zone. Surprisingly, the soil gas composition of TCE in the vadose zone actually increased relative to the base scenario, particularly for houses 1, 3, 4, and 6 which are laterally offset from the groundwater plume. This is reflected in Figure 2.6 which also shows: a minor decrease in the peak indoor air concentration for house 2, and a minor increase for house 5, with both houses being directly over the groundwater plume; and an increase in peak indoor air concentration for the remaining houses by a factor between 4.5 and 9.3. This increase in indoor air concentration is a product of the increased flow of soil gas into each house across the foundation slab (see Table 2.3), and the increased TCE soil gas composition in the vadose zone, effectively increasing the flux of TCE into the indoor air. Note that by increasing L_T for scenario 6 in Table 2.4, the J&E model calculates a decrease in the indoor air concentrations (see Equation (19) of Johnson and Ettinger (1991)) for all houses. This decrease is insignificant relative to the J&E indoor air concentration for the base scenario as shown in Figure 2.6. Most importantly, both the *CompFlow Bio* and J&E models indicate that the thickness of the vadose zone within the constraints of this particular problem geometry is not a significant factor when attempting to define an exclusion zone.

2.3.3   Aquifer heterogeneity

Imperfect knowledge of the point-to-point structure of subsurface properties is a fundamental issue when assessing the fate and transport of contaminants from the groundwater to the indoor air. Heterogeneity in the permeability structure of the aquifer will impart uncertainty in indoor air concentrations and consequently the perimeter of the exclusion zone. We address this issue by generating 50 alternatives, but statistically equivalent, heterogeneous permeability realizations to that shown in Figure 2.3. Each realization is unconditioned in the sense that we assumed no permeability measurements were available to constrain the realizations, beyond implicit knowledge of the statistical structure itself as provided on Table 2.2. Next, we used *CompFlow Bio* to compute the resulting indoor air concentration in each of the six houses for the base scenario using each permeability realization in a Monte Carlo approach. These results are shown in Figure 2.7. Comparison of the indoor air concentrations for the base scenario as shown in Figure 2.6 to those calculated using the alterative permeability realizations indicates that the majority of the realizations yield lower indoor air concentrations in all six houses. As discussed in Yu et al. (2009), these situations arose when the groundwater plume was deflected downward effectively diminishing the ability of aqueous phase diffusion to transport TCE across the capillary fringe. We remind the reader that Rivett (1995) observed that groundwater could transport contaminants a significant distance without impacting the vadose zone in the Borden aquifer. In contrast, the base scenario using the permeability realization shown in Figure 2.3 exhibits a strong impact on the vadose zone and hence is overly conservative.

Comparison of aquifer heterogeneity to the alterative problem geometries (scenarios 2 to 6) discussed in the previous section indicates that only scenario 5, which includes the clay layer and infiltration, has the greatest impact on the indoor air concentrations. Consequently, we suggest that the professional engineer or geoscientist focus their efforts first on characterizing the stratigraphy, and then on obtaining permeability measurements and other hydrogeological properties (i.e. infiltration rate) of the aquifer. Yu et al. (2009) provide an extensive analysis of the impact of alternative properties, including; lateral offset of the source zone, barometric pressure fluctuations, aperture of the foundation crack, pressure drop within the house, and thickness of the capillary fringe. In the context of this study, these same alterative properties have an identical impact on indoor air concentrations relative to aquifer heterogeneity for identical reasons as discussed in Yu et al. (2009) and hence are not repeated here for brevity.

Definition of an exclusion zone based on the Monte Carlo indoor air concentrations generated using the 50 permeability realizations follows two strategies. In the first strategy, we assume that the developer monitors the indoor air quality periodically every $\Delta t_{monitor} = 100$ days for the first $T = 1800$ days (approximately 5 years). For a given monitoring interval $(t_{i-1}, t_i]$ where $t_i = i \times \Delta t_{monitor}$ with $i = \{1, 2, \ldots, n_{times}\}$ and $n_{times} = T/\Delta t_{monitor}$, the probability of exceedence $P_{i,j}$ for the jth house $j = \{1, 2, \ldots, n_{houses}\}$ and $n_{houses} = 6$ is calculated as: the sum of the number of Monte Carlo realizations for which the indoor air concentration for TCE exceeds the NYSDOH (2005) indoor air regulatory limit c_{regu} of 5 $\mu g/m^3$ at the time interval $(t_{i-1}, t_i]$ for the jth house defined as $n_{c_{i,j} > c_{regu}}$; followed by dividing by the total number of Monte Carlo realizations $n_{realizations}$; yielding

$$P_{i,j} = \frac{n_{c_{i,j} > c_{regu}}}{n_{realizations}} \ . \tag{2.4}$$

The exclusion zone is then demarked by the line which separates those houses for which $P_{i,j} = 0$ for $i = \{1, 2, \ldots, n_{times}\}$ from the remainder of the houses within the development that exhibit $P_{i,j} > 0$. The deficiency in this approach is that 50 Monte Carlo realizations may not be a sufficient sample to accurately demark the perimeter of the exclusion zone to the satisfaction of regulatory guidance specifying an acceptable probability of failure (exceedence of the regulatory limit). In the second strategy, we begin with the hypothesis that the indoor air TCE concentrations for all Monte Carlo realizations $\mathbb{C}_{c_{TCE}IA}|_{i,j,l}$ with $l = \{1, 2, \ldots, n_{realizations}\}$ for a given monitoring interval $(t_{i-1}, t_i]$ and the jth house are log-normally distributed. In contrast, Schwede et al. (2008) suggest average concentrations within a large sampling volume follow a beta distribution. However, we note their analysis is specific to a groundwater plume whereas we focus on that of house located above the water table. To test our hypothesis, we used a one-sample Kolmogorov-Smirnov test (Massey, 1951; Miller, 1956) with the null hypothesis being that the vector of $\ln \mathbb{C}_{c_{TCE}IA}|_{i,j,l}$ follows a normal distribution. The alternative hypothesis is that the vector of $\ln \mathbb{C}_{c_{TCE}IA}|_{i,j,l}$ does not follow a normal distribution at the 1% significance level. The benefit of "fitting" a log-normal distribution to $\mathbb{C}_{c_{TCE}IA}|_{i,j,l}$ is that we are able to extrapolate the heavy tail of the log-normal distribution to determine the probability $P_{i,j}$ by which indoor air concentrations exceed the regulatory limit, and conservatively demark the perimeter of the exclusion zone.

The temporal transition of $\mathbb{C}_{c_{TCE}IA}|_{i,j,l}$ to being log-normally distributed as the increment of the monitoring interval $(t_{i-1}, t_i]$ increases is best shown in Figure 2.7 when the mean concentration changes from a dashed to a solid line. The dashed line indicates the null hypothesis is not accepted, while the solid line indicates that it is accepted. Indoor air concentrations appear to become log-normally distributed only once a sufficient number of realizations l in a given monitoring interval $(t_{i-1}, t_i]$ for house j are impacted. This occurs at latest by about 700 days when the bulk of the groundwater plume has advected at least 63 m given that the groundwater velocity is on average $0.09 \, m/day$. We note that the distance between the TCE source zone and the front of house 2 is 55 m (see Figure 2.2a). The upper 95%, 99% and 99.9% confidence intervals are also shown in Figure 2.7 and are computed using the mean and standard deviation of indoor air concentration $\ln \mathbb{C}_{c_{TCE}IA}|_{i,j,l}$. Once all houses exhibit log-normal concentrations, the upper 95%, 99% and 99.9% confidence intervals all decrease sharply to a minima demarked by the vertical gray line. These minima occur once all Monte-Carlo groundwater plumes have impacted the indoor air concentrations of a given house. In fact, this early time peak is caused by a large standard deviation in $\ln \mathbb{C}_{c_{TCE}IA}|_{i,j,l}$ which remains inflated by the low indoor air concentrations that occur before all realizations have impacted a given house. Following these minima, the mean and upper 95%, 99% and 99.9% confidence intervals all increase monotonically before they plateau at the end of the monitoring period at 1800 days. It is this statistically stable region following the vertical gray line (which is characterized by having a minimum and maximum p-statistic from the Kolmogorov-Smirnov test of 0.192 and 0.946, respectively) that will be used to demark the perimeter of the exclusion zone.

Using the first strategy defined above, Figure 2.7 indicates that the exclusion zone contains houses 1, 3, 4 and 6 which are all offset from the groundwater plume given that $P_{i,j} = 0$ for $i = \{1, 2, \dots, n_{times}\}$. Using the second strategy defined above, the exclusion zone at the 95% confidence interval is identical to the first strategy using the raw Monte Carlo simulations to calculate $P_{i,j}$. However, at the 99% confidence interval the exclusion zone contains only houses 1 and 4. This lack of symmetry whereby houses to the left of the groundwater plume are slightly less impacted than those to the right is a consequence of insufficient Monte Carlo realizations of the heterogeneous permeability field to resolve the ensemble statistics of the groundwater and vadose zone TCE distributions. At the 99.9% confidence interval, none of the houses are within the exclusion zone consistent with both the J&E and ASTM (2008).

Numerous studies have made statements regarding the conservative nature of the J&E model (USAF, 2006; Hers et al., 2003; Schreuder, 2006; Fitzpatrick and Fitzgerald, 2002; Mills et al., 2007; Yu et al., 2009; Provoost et al., 2010; Johnston and Gibson, 2011). We now attempt to statistically quantity the degree of this conservatism by comparing $\alpha_{J\&E}$ and α_{CFB} values using the 50 Monte Carlo permeability realizations for the base scenario problem geometry. We remind the reader that the base scenario problem geometry was motivated by the Rivett (1995) field experiment at the Borden site. As such, the degree of conservatism as expressed here cannot be generalized to other sites without following the methodology outlined in this thesis. We begin by computing $\alpha_{CFB}|_{j,l}$ for each realization $l = \{1,2,\dots,n_{realizations}\}$ taking note that $\alpha_{CFB}|_{j,l}$ occurs when the indoor air concentration $\mathbb{C}_{c_{TCE}IA}|_{i,j,l}$ is at its peak value which may occur at different monitoring intervals $(t_{i-1}, t_i]$ for the jth house and for all realizations. We then test the hypothesis that $\ln \alpha_{CFB}|_{j,l}$ is normally distributed using the Kolmogorov-Smirnov test at the 1% level of significance. We observed strong confirmation of this hypothesis given that the p-statistic had a minimum and maximum value of 0.602 (house 2) and 0.944 (house 3), respectively. Table 2.6 itemizes the statistical results for the $\ln \alpha_{CFB}|_{j,l}$ distribution for each house. In summary, only for house 2 which directly overlies the plume is *CompFlow Bio* more conservative than the J&E model at the 99.9% confidence level. Otherwise, J&E is more conservative than *CompFlow Bio*, with the degree of conservativeness expressed by the probability of $P(\alpha_{CFB} > \alpha_{J\&E})$. Clearly for houses 1, 3, 4 and 6 which are laterally offset from the groundwater plume, the degree to which the J&E model is conservative is rather onerous. Assuming the professional engineer or geoscientist will define the perimeter of the exclusion zone by adjusting L_T in the J&E model to decrease $\alpha_{J\&E}$ until the calculated indoor air concentrations are less than the regulatory limit, the resulting lateral offset of the exclusion zone will be much greater than what *CompFlow Bio* would otherwise suggest given the weak transverse dispersion of groundwater plumes observed at field sites, and as implicitly supported by the ASTM (2008) regulatory guidance.

2.4 Conclusions

The objective of this chapter is to use the multi-phase compositional numerical model *CompFlow Bio* to examine the fate and transport of volatile organic contaminants originating from a NAPL source zone located below the water table, and their potential impact on multiple residential dwellings located down-gradient of the source zone. While the problem geometry is based directly on that of Yu

et al. (2009) and hence is reminiscent of the Rivett (1995) field experiment in the Borden aquifer, the problem geometry is extended to be fully three-dimensional. This enhancement was motivated by the need to accommodate multiple houses, with some laterally offset from the direction of groundwater plume advection, in an attempt to delineate an exclusion zone. We note that the Rivett (1995) field experiment conducted in the Borden aquifer involved a DNAPL source zone located a short distance below the water table. Monitoring of both the groundwater plume and soil gas above the plume indicated that groundwater could transport contaminants a significant distance without transverse vertical mass transport of these contaminants across the capillary fringe. This key observation was replicated in the Yu et al. (2009) problem geometry, and hence is present in this work as well. By extending the Yu et al. (2009) problem geometry to be fully three-dimensional, we observed that the transverse horizontal flux of contaminant (to the direction of the groundwater plume advection) significantly reduces contaminant concentrations within the plume and ultimately within the indoor air relative to the equivalent two-dimensional problem geometry.

A key factor controlling the relative importance of variations in the conceptual model (such as recharge, inserting a clay layer with and without recharge, extending the thickness of the vadose zone, and heterogeneity in the subsurface permeability structure of the stratigraphically continuous aquifer) on indoor air concentrations is the degree to which these variations impact the flux of TCE in the soil gas across the foundation slab. We observed that the combination of inserting the clay layer with recharge created an impervious barrier for TCE beneath the foundation slab effectively negating the flux of TCE in the soil gas across the foundation slab. Next in importance is heterogeneity in the subsurface permeability structure of the stratigraphically continuous aquifer. We observed that this heterogeneity conferred a greater variability in indoor air concentrations than any other variations in the conceptual model, except for the clay layer with recharge which exhibited no indoor air impact whatsoever. We conclude that characterizing the site stratigraphy is a first-order priority when attempting to assess the impact of the fate and transport of TCE from an observed source zone to the indoor air. Having established the site stratigraphy, we then recommend detailed soil core permeability measurements to characterize the heterogeneity within each stratigraphic unit.

A central objective of this chapter is to compare the *CompFlow Bio* and J&E models as well as the ASTM (2008) guidance in terms of delineating the exclusion zone for the conceptual model shown in Figure 2.1. For the base scenario involving this conceptual model where the vapour source is a non-biodegradable volatile organic groundwater plume, ASTM (2008) itemizes the critical distance demarking the exclusion zone as being 100 $feet$ (approximately 30 m) from the perimeter of the groundwater plume. In general, *CompFlow Bio* exhibits an agreement with the ASTM (2008)

guidance on the basis that transverse horizontal macro-dispersivity induced by the heterogeneous permeability field is quite weak generating plumes with sharp lateral boundaries. However, the region of soil gas adversely impacted by the groundwater plume is extended in a lateral direction far beyond the boundaries of the plume due to a strong soil gas diffusion transport mechanism. Despite this, we observed that the dilution that the indoor air exchange rate offered for a modest-sized house confers a certain degree of protection against the extensive soil gas plume, even when weighed against the relatively strict NYSDOH (2005) indoor air regulatory limit. For the specific permeability realization used as part of the base scenario (and for variations in the conceptual model), we observed that the J&E model calculated conservative estimates for indoor air concentrations for all houses both directly above and laterally offset from the groundwater plume. Consequently, the J&E model generated an exclusion zone that would be significantly further away from the groundwater plume than what either *CompFlow Bio* or ASTM (2008) would predict.

Given limited site characterization budgets, there is a practical need to be able to assess the fate and transport of TCE from an observed source zone to the indoor air in a manner that is both economical and conservative. The J&E model is an excellent choice for this task which has motivated its widespread use in industry. We attempted to quantify the degree to which the J&E model is conservative by comparing J&E-calculated attenuation coefficients ($\alpha_{J\&E}$) against those computed using *CompFlow Bio* (α_{CFB}) using 50 alterative permeability realizations for the base scenario version of the problem geometry. *CompFlow Bio* was more conservative than the J&E model at the 99.9% confidence level only for house 2 which directly overlies the groundwater plume. Otherwise, J&E was more conservative than *CompFlow Bio*. For houses 1, 3, 4 and 6 which are offset from the groundwater plume, the degree to which the J&E model was conservative was unrealistic from a practical engineering design perspective.
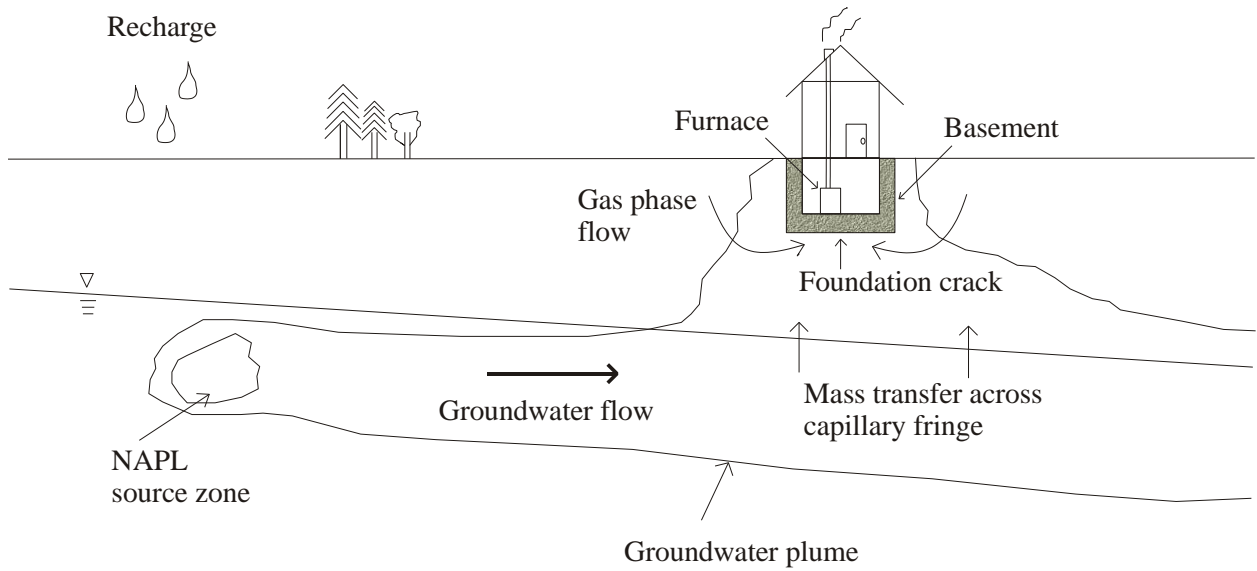
Figure 2.1: Conceptual model depicting the transport of contaminants from a NAPL source zone below the water table to a house.
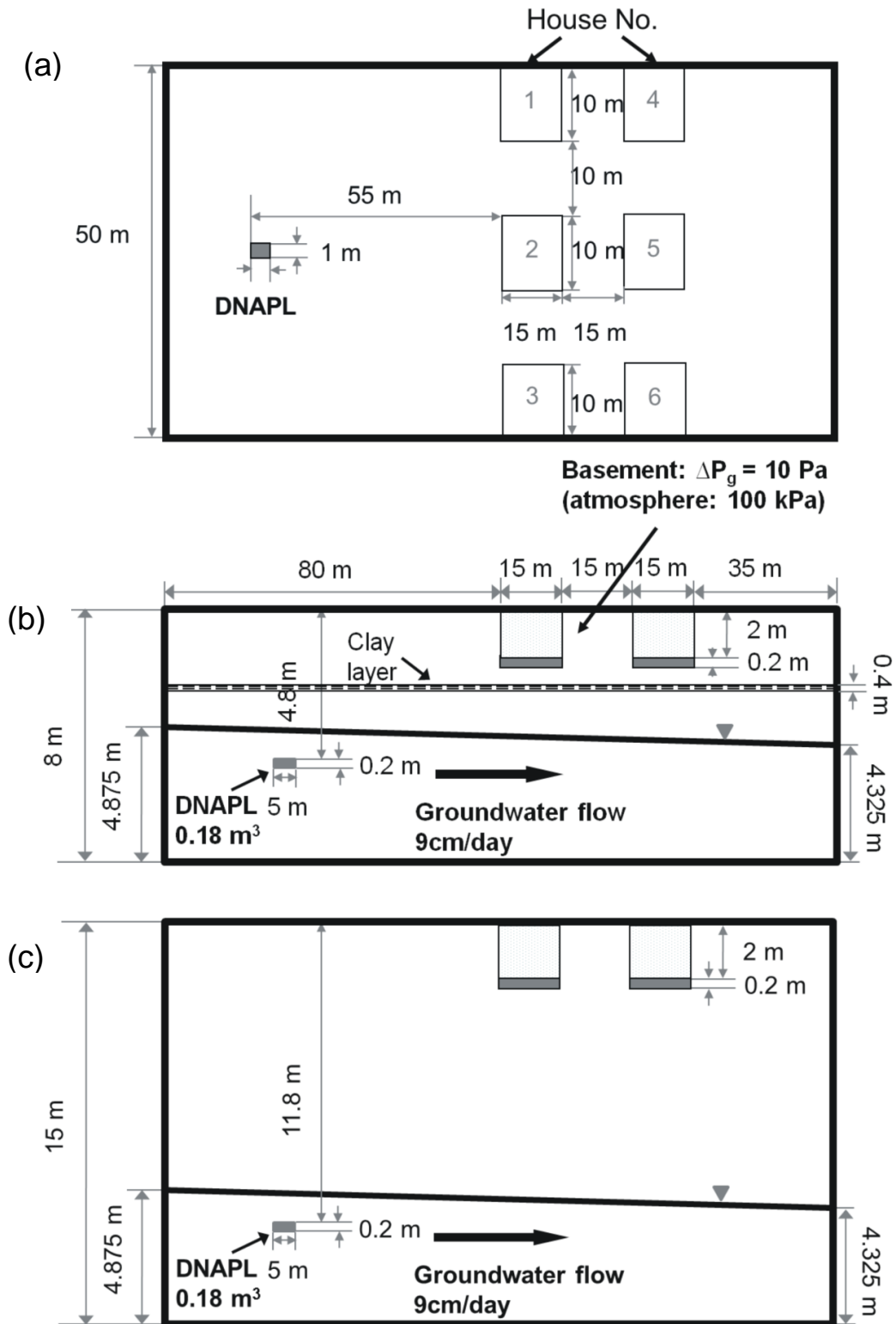
Figure 2.2: The 2D cross section views of the problem geometry: a), b) scenario 1 on Table 2.1, and c) scenario 6.
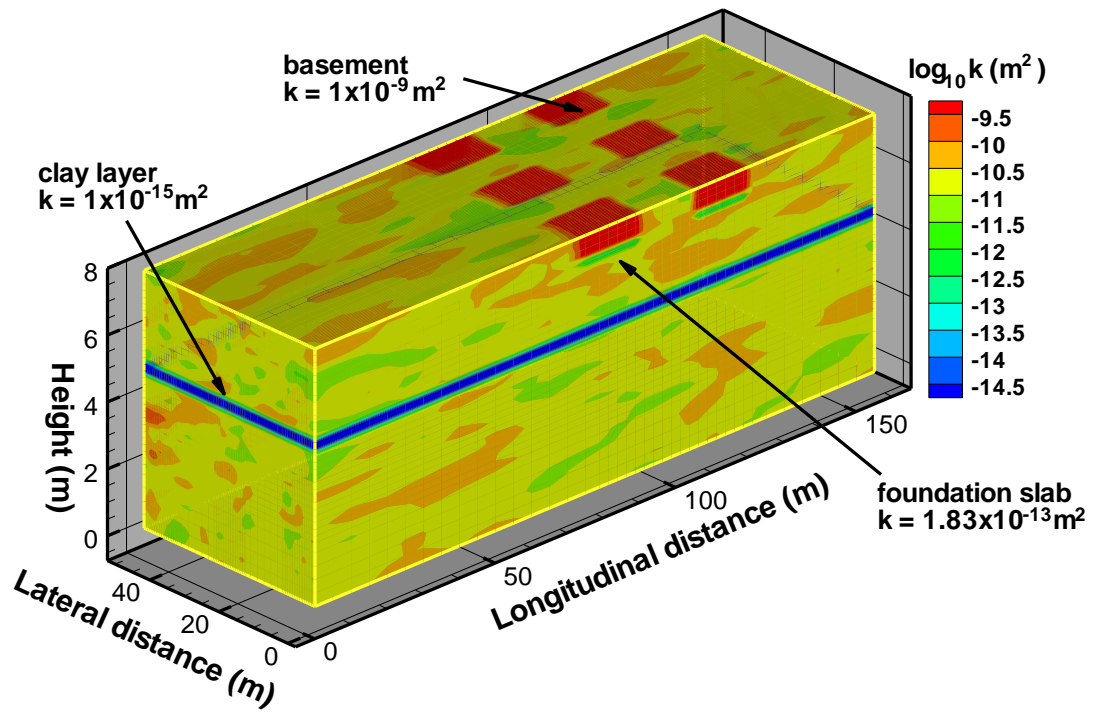
Figure 2.3: 3D plot of the heterogeneous permeability field with a clay layer.
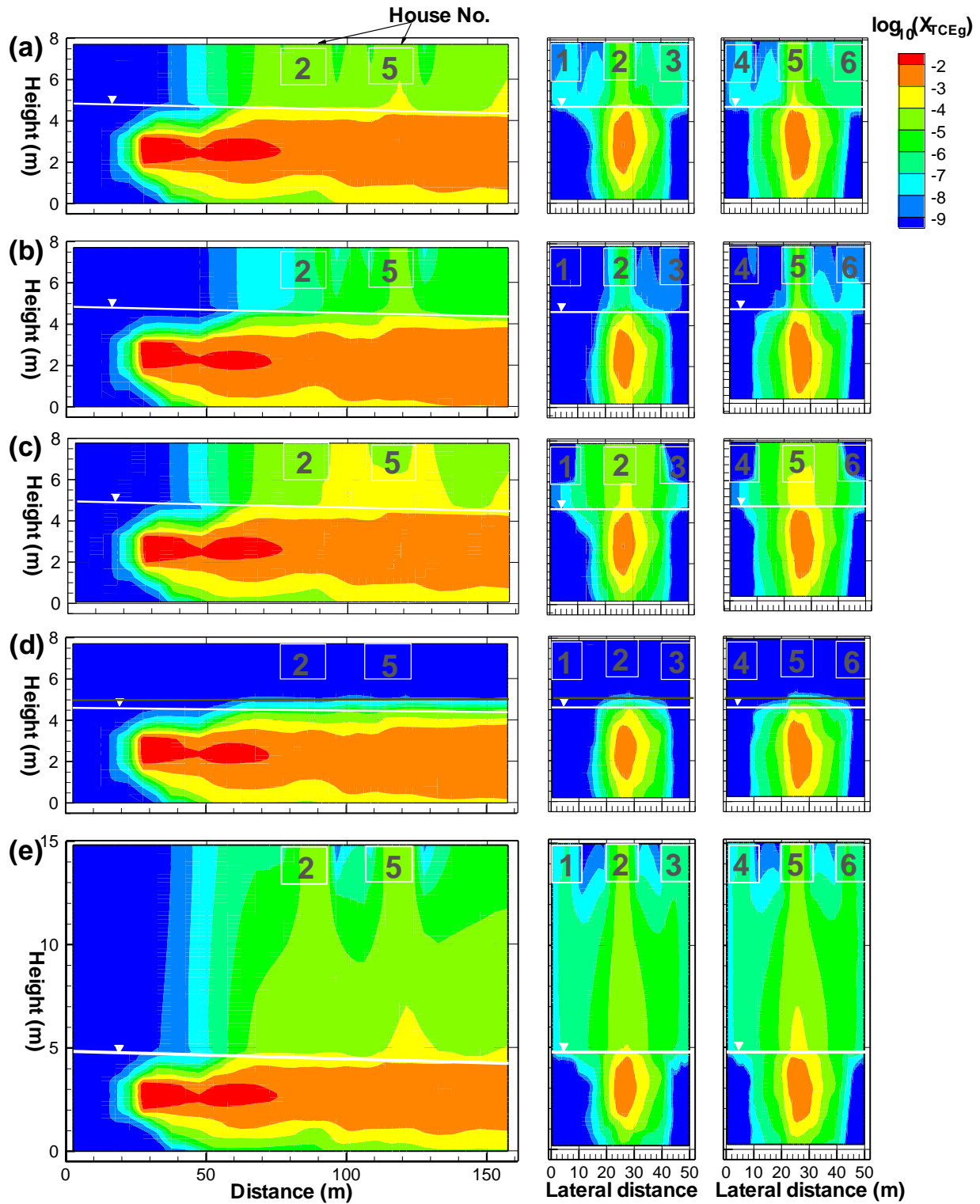
Figure 2.4: 2D cross section plots along x- axis and y- axis of mole fraction of TCE in gas phase 1800 days after TCE was injected into the aquifer: a) base scenario 1 on Table 2.1, b) scenario 2, c) scenario 3, d) scenario 5, and e) scenario 6.
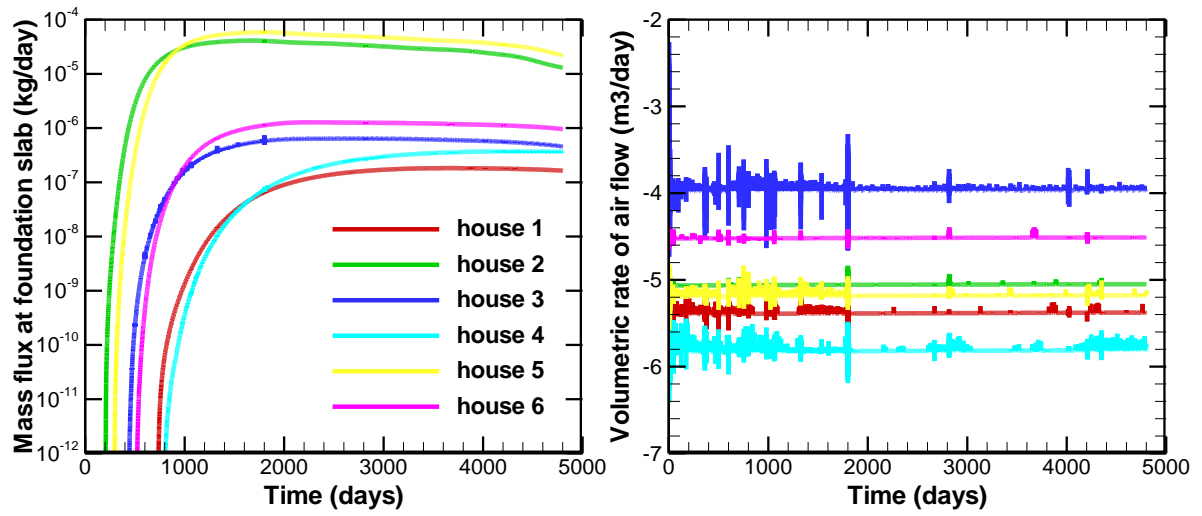
Figure 2.5: The mass flux of TCE in the soil gas and the volumetric flow rate of soil gas across the foundation slab into the houses for the base scenario, scenario 1.
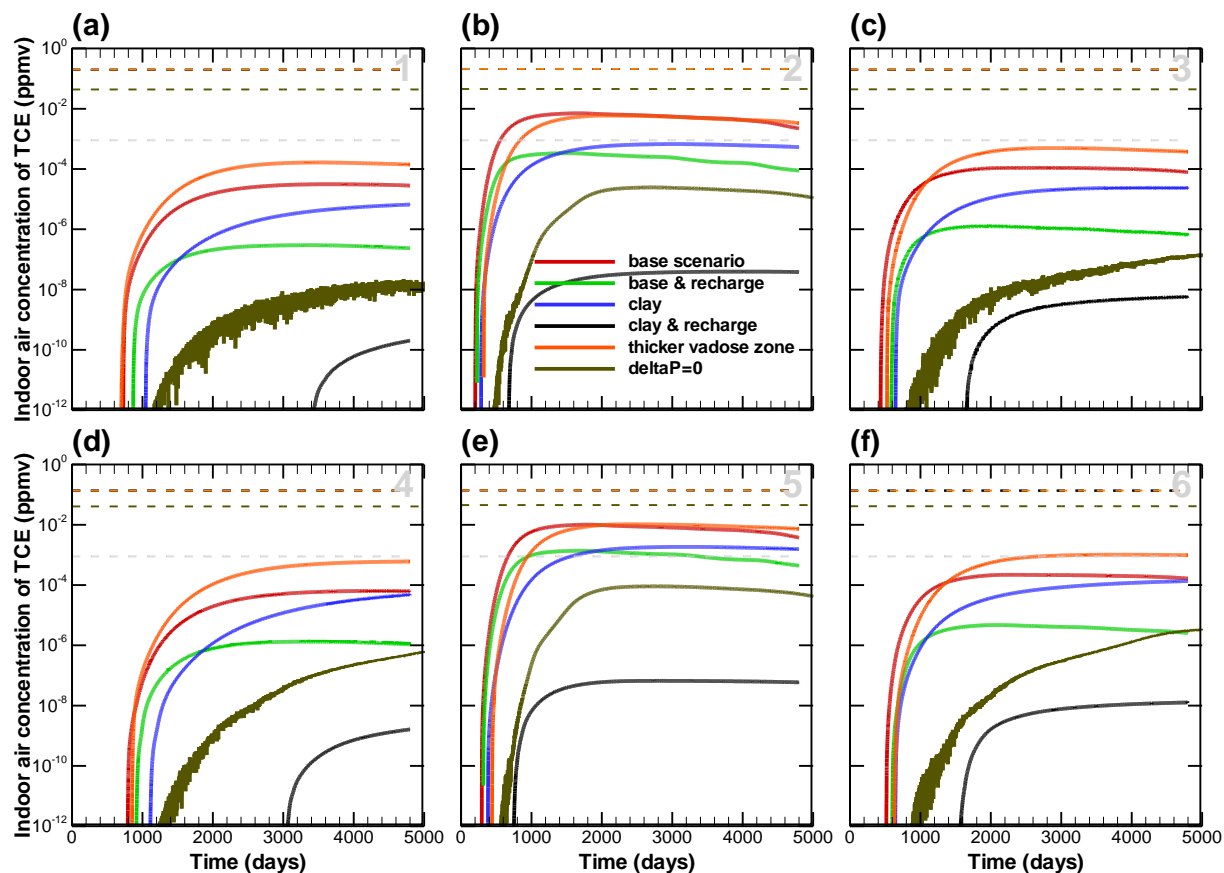
Figure 2.6: Indoor air concentration of TCE in logarithmic scale with the air exchange rate 0.5/hr for the six scenarios on Table 2.1 using *CompFlow Bio* model: a) House No. 1, b) House No. 2, c) House No. 3, d) House No. 4, e) House No. 5, and f) House No. 6.

Note: The number 1 – 6 represents the house number from Figure 2.2. The grey dash line represents the regulatory limit. The coloured dash lines represent the Indoor air concentrations of TCE in logarithmic scale using J-E model with parameters on Table 2.4.

36
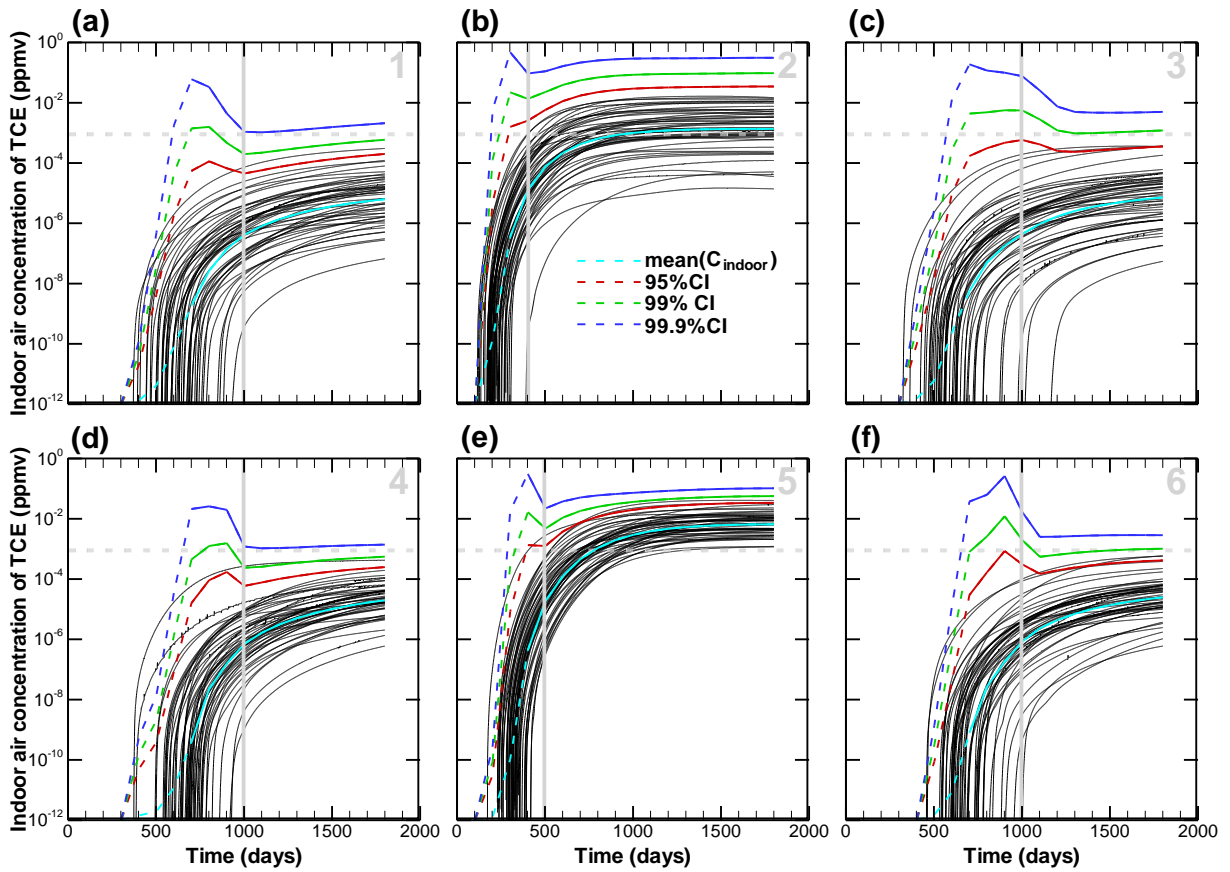
Figure 2.7: Indoor air concentration of TCE in logarithmic scale with the air exchange rate 0.5/hr for 50 unconditioned random permeability realizations with the mean value, the 95%, 99% and 99.9% confidence intervals (CI): a) House No. 1, b) House No. 2, c) House No.

Note: The number 1 – 6 represents the house number from Figure 2.2. The dashed horizontal line represents the regulatory limit. The dashed lines for the mean, the 95%, 99% and 99.9% confidence intervals of the indoor air concentration represent they do not follow a log-normal distribution; the solid lines represent the indoor air concentrations follow a log-normal distribution. The solid grey vertical line divides the strong acceptance of log-normally distributed indoor air concentrations (on the right) from the week acceptance or non-acceptance (on the left).

Table 2.1: Scenario description

No.	Scenario
1	Base scenario model $\Delta P_g^{basement} = 10$ Pa
2	Base scenario model with Infiltration = 20 cm/yr
3	Base scenario with $\Delta P_g^{basement} = 0$ Pa
4	Base scenario model with a clay layer
5	Base scenario model with a clay layer and Infiltration = 20 cm/yr
6	Base scenario with a 7 m thicker vadose zone

Table 2.2: Porous media properties

Property	Value
Porosity: ϕ [−]	0.33
Geometric mean permeability: $\mathbf{K}$ [m^2]	2×10^{-11}
Variance of $Y = \ln \mathbf{K}$: σ_Y [−]	1.0
Correlation length of $\mathbf{K}$ in principal bedding direction: λ_1 [−]	5.0
Correlation length of $\mathbf{K}$ in transverse bedding direction: λ_2 [−]	2.0
Correlation length of $\mathbf{K}$ perpendicular to bedding direction: λ_3 [−]	0.15
Dispersivity: α_L^l and α_T^l [m], $l = \{q, n, g\}$	0.0
Mass density of porous media: ρ_r [kg/m^3]	1810
Organic carbon content of porous media: f_{oc} [−]	0.0002
Sorption of TCE onto porous media: $K_{dc_{TCE}}$	1.1×10^{-5}

Table 2.3: Volumetric flow rate of soil gas for all houses

Scenario No.	Volumetric rate of air flow [m³/day] (negative sign means air flowing into houses)					
	House No.					
	1	2	3	4	5	6
1	-5.3774	-5.0492	-3.9574	-5.7368	-5.1776	-4.5111
2	-5.2627	-4.9160	-3.8057	-5.7472	-5.0975	-4.4256
3	-1.8052 X10⁻⁴	3.4629 X10⁻⁴	-4.0967 X10⁻⁴	-9.7921 X10⁻⁴	1.9232 X10³	-1.2368 X10³
4	-3.0586	-2.9350	-2.8919	-4.7474	-3.3595	-3.7493
5	-2.7764	-3.1373	-2.8316	-4.1534	-3.2409	-3.2617
6	-8.4244	-9.0287	-8.4527	-8.4324	-9.7370	-7.2889

Table 2.4: Parameters used to compute $\alpha_{J\&E}$ for the base scenario in *CompFlow Bio* model

Parameter		Value
D_T^{eff} [m²/day]	assuming $S_g\phi = S_q\phi = \frac{1}{2}\phi$	0.009 [a]
D^{crack} [m²/day]	assumed equal to d_g from Table 5 in Yu et al. (2009)	0.394
$2b$ [μm]	aperture of crack in foundation slab	100
A_B [m²]	15×10 (foundation slab) + 2×15×2 (two basement walls along x) + 2×10×2 (two basement walls along y)	190
A_{crack} [m²]	50 m (long) × 2b (wide), with 2b from Table 1	5×10⁻³
$Q_{building}$ [m³/day]	$V_H \times A_{ex}$ where $V_H = 900$m³ and $A_{ex} = 0.5/hr$	10800
L_T [m]	offset between source zone and the center of the foundation slab	
	house 1, 3:	63
	house 2:	60
	house 4, 6:	92
	house 5:	90
L_{crack} [m]	thickness of the foundation slab	0.2
K_g [m²]	permeability of foundation slab to gas phase	1.83×10⁻¹³
C_{source} [ppmV]	the TCE concentration at source	6.05×10⁴

[a] $D_T^{eff} = D_g^{eff} + D_q^{eff}/H = d_g(S_g\phi)^{3.33}/\phi^2 + d_q(S_q\phi)^{3.33}/\phi^2 H$ where ϕ is from the Table 2.2; d_g and d_q are from Table A.4; Henry constant $H = 6.44$ [m³ aq/m³ g] based on Table A.3.

Table 2.5: The attenuation coefficients for J&E model $\alpha_{J\&E}$ and *CompFlow Bio* model α_{CFB}. α_{CFB} is computed when indoor air concentrations reached their peak in the *CompFlow Bio* simulation scenario with an air exchange rate of $A_{ex} = 0.5/hr$. $\alpha_{J\&E}$ is calculated using J&E model parameters from Table 2.3 using Q_{soil} from *CompFlow Bio* at the time when α_{CFB} was obtained

| Scenario No. | $\alpha_{J\&E}$(X10^{-9}) | | | | | | α_{CFB} (X10^{-9}) | | | | | |
| | House No. | | | | | | House No. | | | | | |
	1	2	3	4	5	6	1	2	3	4	5	6
1	3269	3443	3269	2249	2303	2249	0.524	116.4	1.828	1.074	166.6	3.644
2	3269	3442	3269	2249	2302	2249	0.005	5.535	0.023	0.025	22.62	0.078
3	722	742	722	673	744	673	4x10^{-4}	0.407	0.003	0.010	1.505	0.056
4	3267	3443	3267	2249	2302	2249	0.109	11.09	0.413	0.803	30.55	2.247
5	3259	3438	3259	2248	2300	2248	3x10^{-6}	7x10^{-4}	1x10^{-4}	3x10^{-5}	1x10^{-3}	2x10^{-4}
6	3242	3413	3242	2241	2295	2241	2.733	98.24	8.192	9.957	171.1	16.83

Table 2.6: The statistics of the attenuation coefficients α_{CFB} computed using the *CompFlow Bio* model results from Monte Carlo realizations based on the base scenario

Parameter		House 1	House 2	House 3	House 4	House 5	House 6
$Max(\alpha_{CFB})$	(X 10^{-9})	5.0275	271.55	6.0189	7.0015	675.57	9.8156
$\mu(ln\alpha_{CFB})$	$(-)$	-22.99	-17.58	-22.84	-21.84	-16.03	-21.61
$\sigma(ln\alpha_{CFB})$	$(-)$	1.7671	1.6386	1.9857	1.2929	0.8380	1.4409
upper 95% CI α_{CFB}	(X 10^{-9})	3.3072	573.02	5.8730	4.1264	565.85	6.9068
upper 99% CI α_{CFB}	(X 10^{-9})	9.8195	1572.0	19.951	9.1488	948.10	16.775
upper 99.9% CI α_{CFB}	(X 10^{-9})	34.719	5070.6	82.470	23.049	1725.7	46.977
$P(\alpha_{CFB} > \alpha_{J\&E})$	$(-)$	2.28e-9	0.0011	1.35e-7	4.1e-12	1.39e-4	1.15e-9

Chapter 3 Risk-Based Site Characterization for Vapour Intrusion at a Brownfields Site: Data Worth and Prediction Uncertainty

Chapter summary

In this chapter, the data-worth evaluation methodology adopted for a risk-based characterizing for vapour intrusion at a Brownfield site is a combination of the use of the predictive physical model *CompFlow Bio*, the geostatistical tool Kriging and the statistical data assimilation method Kalman filter. The objective of this chapter is not only to assess the worth of soil core permeability and TCE soil gas concentration data contributing to the reduction of prediction uncertainty in quantifying indoor air concentration of TCE in residential dwellings, but also to use sparse data to infer whether one of the houses located adjacent to affected houses has indoor air problems subject to the observation that affected house is monitored and has problems and developers have liability if a problem occurs. The results which are measured by three metrics among different combinations of sampling strategies show that incrementally increasing sampling data incorporated into the model can improve the reduction of the prediction uncertainty; however, the types and quantities of data determine the level of reduction. In this case, 7K4C sampling scheme reaches an optimal balance between the two and should be favoured by developers.

3.1   Introduction

Redevelopment of brownfields land presents a significant challenge in that people may become exposed to legacy contamination at the site. This contamination may continue to exist even after extensive remediation efforts given the difficult nature of characterizing the subsurface hydrogeological properties (Turcke and Kueper, 1996; Hyndman et al., 2007), locating the sources of contamination (Lesage and Jackson, 1992; Laird et al., 2005), and subsequently removing them from the soil in compliance with regulatory criteria (Russell, 1992; Ward et al., 1997; ITRC, 2007a, b; Longino, 2005). The potential for people to suffer potential health impacts from exposure to legacy contamination could result in the developer receiving punitive damages of unlimited value. To promote the developers due diligence in remediating the site and simultaneously limit the potential for future punitive damages, government agencies provide liability protection beyond a specified monitoring period (NRTEE, 2003; UK Environmental Agency, 2003; US EPA, 1999).

Discovery of an oily-waste source zone beneath the water table after redevelopment of a site is usually followed from monitoring groundwater quality and observing a contaminant plume. Yu et al. (2009) and the previous chapter of this thesis discussed the hydrogeological process governing the fate and transport of TCE (a volatile chlorinated solvent) from groundwater to indoor air. As a first response, the developer would likely repurchase those houses directly overlying the plume given the likelihood that they will be adversely impacted, with the intent of minimizing their obvious liability exposure. Thereafter, regulatory guidance regarding an exclusion zone (ASTM, 2008) can be used to specify a critical distance beyond the lateral boundaries of the plume beyond which it is no longer necessary to conduct pathway assessment and monitoring and hence to characterize the house as being affected by subsurface contamination. While this clearly specifies the developers due diligence in managing the housing development upon discovery of a groundwater plume, it does not limit their liability to the houses within the exclusion zone before the end of the liability period. The developer faces an additional difficulty in that they may not actually even have access to those houses in order to monitor the indoor air quality without directly purchasing them. In order for the developer to minimize their liability for all houses within the development, they must be able to assess the probability that houses located within the exclusion zone may be adversely impacted given pathway assessment and monitoring data that they have collected within the exclusion zone as part of their due diligence. Consequently, they must reserve sufficient capital to be able to repurchase any adversely impacted house within the development provided that data become available to suggest that the indoor air quality exceeds a regulatory limit (OEHHA-CEPA, 2007; NYSDOH, 2005).

Massmann and Freeze (1987a, b), Freeze et al. (1992) and Yu et al. (2012) provide a cost-risk-benefit framework that is directly amenable to address capital reserve issues posed by the developer. Specifically, the framework combines both hydrogeological and economic information when pricing the reserve capital to cover the financial risk based on the probability that the indoor air concentrations of these houses may exceed a regulatory limit. The objective of this work is to build upon framework as adapted by Yu et al. (2012) to the problem at hand. To further introduce our objectives, we briefly review how Yu et al. (2012) describe the developer views on the balance of hydrogeological and financial uncertainties and their combined impact on the value of the risk capital.

The highly stylized brownfields project commences with the proposition that a company owns a property that contains legacy non-aqueous phase liquid (NAPL) contamination. They remediate the subsurface in compliance with governing regulatory requirements (Federal Contaminated Sites Action Plan, Canada; Superfund, or Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA), USA). However, there is a possibility that some source zones of NAPL

contamination may remain below the water table. They now wish to construct multiple residential dwellings on the site. As with the previous chapter, the central intent of this work revolves around the issue of assessing indoor air concentrations within a number of houses (six in total) that directly overlying and laterally offset from a groundwater plume emanating from a remnant TCE source zone. This TCE groundwater plume then results in soil gas concentrations of TCE in the vadose zone in the vicinity of the plume, causing indoor air impacts to the houses surrounding the plume. To entice buyers, the developers guarantee that they will immediately repurchase any affected house at the original sale price appreciated by the US national home price index in our highly stylized brownfields project. Indoor air concentrations are monitored every $\Delta t_{monitor}$ days, and their guarantee lasts for the first T days. Thereafter, the government indemnifies the developer (OME, 1997; OMMAH, 2007). In return for the repurchase agreement and continuous monitoring, the residents agree not to blame any reasonably foreseeable health issues arising from potential long-term exposure of indoor air contamination on the developer given their due diligence. To maintain the development, the developer then remediates the subsurface to resolve the source of contamination, and demolishes/rebuilds/renovates the house to prevent further intrusion of soil gas. The house is then resold.

In the first part of this two-chapter (Chapter 3 and 4) series, we build upon the work of Yu et al. (2012) by demonstrating a strategy to assimilate hydrogeological data in order to reduce the prediction uncertainty associated with forecasting the probability that indoor air concentrations will exceed a regulatory limit. Chapter 2 demonstrate that the lithological structure of the subsurface exerts a primary control on the fate and transport of TCE from groundwater to the indoor air, while the heterogeneous permeability structure of an individual lithological unit is of secondary but paramount importance. In this chapter, we focus exclusively on heterogeneity in the permeability structure as the primary contribution to prediction uncertainty. The conceptual model used by Chapter 2 is repeated here in Figure 3.1 and forms the foundation for this work, albeit with further modifications which are discussed below. With regards to Figure 3.1 and the highly stylized brownfields project as described above, we assume the developer has already purchased houses 2 and 5 given that they directly overlie the plume. Furthermore, houses 1, 3, 4 and 6 which are all laterally offset from the plume remain privately owned and inaccessible to the developer. The developer must reserve sufficient risk capital in order to purchase these houses should it become apparent that any one of them is impacted.

The problem of reducing prediction uncertainty in groundwater flow and solute transport models which primarily arises from the combination of the uncertainty in conceptual models and the

uncertainty in model parameters has been addressed by many researchers. In recent years, a number of authors have acknowledged that conceptual model uncertainty (Neuman, 2003; Bredehoeft, 2003, 2005; Refsgaard et al., 2006). However, it is not our focus in this chapter. Uncertainty about the spatial variability of aquifer (geologic natural heterogeneity) is a key component and is mostly discussed in the risk assessment process, and has been shown to influence the predicted flow field and contaminant exposure risk (Dagan, 1984; Robin and Dagan, 1992; Yu et al., 2009). This uncertainty can be reduced by performing more investigations and collecting more data. While the collection of new or additional data is generally the most direct way of reducing uncertainty, limiting factors such as cost, time, and accuracy may not always permit further sampling. These limitations, particular the most concerned limited site characterization budgets, have motivated this study for the purpose of uncertainty reduction and accuracy improvement with effective information. The problem of model parameter uncertainty has been addressed using inverse modeling and optimization. The inverse modeling (e.g., Yeh, 1986; Abbaspour, 1997) procedure involves minimization of the square difference function of some measured and simulated flow or transport variables, and is used to decrease the uncertainty in the initial estimates of the parameters, followed by a data worth analysis to show the effect of additional measurements on the confidence of the model prediction. The optimization technique (e.g., Marryott, 1996) has been commonly used for problems of groundwater remediation design, which combined with the hydrogeological model to build a framework during the decision-making process. Monte Carlo technique (e.g., Essaid and Hess, 1993) provides a stochastic method in recognition of uncertainty inherent in both conceptual models and model parameters. When additional information on model parameters or field measurements becomes available, it is used to update the prior estimates of the statistics to posterior estimates in order to reduce the estimation uncertainty. The prior estimate is generally based on limited early data or expert judgment or combination of both. Bayes theorem is as one of the methodologies used for this purpose. A Bayesian framework allows for a project to iterate among data sampling, data analyses, sampling network design and decision making (Abbaspour et al., 1998; Back et al., 2007; Freeze et al., 1992; James et al., 1996a; James and Freeze, 1993; James and Gorelick, 1994; Massman and Freeze, 1987a, 1987b; Norberg and Rosén, 2006; Wagner, 1999). One of the most important features of the Bayesian framework is data worth analysis which is very practical in an engineering project due to the fact that collection of field data is expensive and time consuming. Under linear quadratic Gaussian circumstance, Kalman filter can be derived within the Bayesian framework, which was first observed by Ho and Lee (1964). It consists of an iterative prediction-correction process. In the prediction step, the time update is taken where the one-step forward prediction of observation is calculated; in the correction step, the measurement update is taken where the correction to the estimate of current state

is calculated. Zhang et al. (2005) applied this algorithm to determine the least cost design of groundwater quality monitoring networks. McGrath and Pinder (2003) introduced the worth of sample data function (WSD) using Kalman filter technique to maximize the contaminant plume boundary uncertainty reduction based on updated field information to design the search strategy for plume delineation. It has also been used for the other studies, e.g. Andricevic (1990, 1993), Dokou and Pinder (2011), Graham and McLaughlin (1989, 1991), Herrera and Pinder (2005), and Loaiciga (1989).

Following the work of Freeze et al. (1990, 1992) and Zhang et al. (2005), we explore a methodology to assimilate soil core permeability data and TCE soil gas concentration data and to assess their worth in terms of reducing the prediction uncertainty. In particular, we introduce three metrics to quantify our notion of prediction uncertainty. First, the ability to correctly estimate the actual indoor air concentration within houses 1, 3, 4 and 6 at any point in time. This metric is quantified by conditioning the permeability structure with the available borehole data so that the expected value of the indoor air concentration arising from a series of Monte Carlo simulations involving alterative permeability realizations (see Figure 2.7) approaches that of the actual indoor air concentration. This latter point presumes that even though houses 1, 3, 4 and 6 are not monitored, we do have information at these points by way of reserving one permeability realization as "reality". This reality realization is used to provide the borehole sampling data, and also to transport TCE from the source zone to observe the actual indoor air concentrations. Second, the ability to use available soil gas concentration data to reduce the standard deviation of the indoor air concentration at unmonitored locations (i.e. houses 1, 3, 4 and 6) arising from the series of Monte Carlo simulations through time. Third, the ability to accurately forecast the probability of indoor air concentrations within houses 1, 3, 4 and 6 exceeding a regulatory limit. This third metric is directly related to the first two metrics, which rely on estimating the mean and variance of the distribution of simulated indoor air concentration arising from the Monte Carlo simulations, except that we are now required to evaluate the "tail" of the distribution. As part of this data worth question, we explore various combinations of permeability and soil gas concentration sampling strategies and assess the significance of their impact on reducing prediction uncertainty. Each sampling strategy uses progressively more data, and (except in one case) is consistent with private property access limitations that would follow from the configuration of houses shown in Figure 3.1. In the second part of this two-chapter series, we will describe our methodology for using these three metrics for pricing the risk capital, with contributions from: the expected value and the variance of all housing re-purchase expenditures arising from indoor air concentration exceeding a regulatory limit; the risk posed by imperfect data constraining the

ability of the developer to accurately forecast an exceedence event; and, the cost of the data collection.

Permeability data are collected in a series of boreholes located between the TCE source zone and the first row of houses 1, 2 and 3, along a presumed right-of-way which allows drilling without having to access private property. The data are collected at the beginning of the liability period t_0. Given that the aquifer representing "reality" consists of a heterogeneous permeability field reminiscent of the Borden aquifer, the resulting soil cores represent sparse data samples of the subsurface structure. In order to reduce prediction uncertainty the permeability data are used to condition point values and the general structure of multiple heterogeneous permeability realizations are generated based on a geostatistical analysis of the soil cores (Bianchi et al., 2011; Eggleston et al., 1996; Feehley et al., 2000; Guadagnini et al., 2004; Maji et al., 2006; Neuman et al., 2008). In particular, the permeability beneath the foundation slab exerts a significant control on the flow of soil gas across the foundation slab and into the indoor air. Consequently, the multiple heterogeneous permeability realizations generate a distribution of expected soil gas flow rates that should tend towards "reality" with the inclusion of more permeability data. Soil gas concentration data are collected at a series of boreholes and beneath the foundation slab of houses 2 (and sometimes 5) which directly overlies the plume. The concentration data are collected in each monitoring time interval $(t_{i-1}, t_i]$ where $i = \{1,2,...,n_{times}\}$ and $n_{times} = \Delta t_{monitor}/T$ at each monitoring location, and are generated via the numerical model *CompFlow Bio* (Forsyth, 1993; Unger et al., 1996) using the "reality" permeability realization. These data are also used to reduce prediction uncertainty via assimilation into a discrete static Kalman filter (Graham and McLaughlin, 1991; Herrera and Pinder, 2005; Zhou et al., 1991; Zou and Parr, 1995). In the case of this work, the multiple heterogeneous permeability realizations are used to build a covariance matrix for use within the Kalman filter to optimize the predicted soil gas concentration beneath houses 1, 3, 4 and 6 in each monitoring time interval $(t_{i-1}, t_i]$. Each of the multiple heterogeneous permeability realizations is an equally probable reconstruction of "reality" given that they are all conditioned to have the same values along the sampled boreholes. Consequently, each realization is used to perturb the Kalman filter and generate a distribution of predicted soil gas concentrations. This distribution of soil gas concentrations is then combined with the paired soil gas flow rate to compute the resulting distribution of indoor air concentrations. The three metrics defining prediction uncertainty are all quantified based on mean, variance and shape of the tail of the distribution of indoor air concentrations. As such, the question of data worth for reducing prediction uncertainty relates to changes in these statistical measures with various combinations of permeability and soil gas concentration sampling strategies.

The outline of this first part in the two-chapter series is as follows. In Section 3.2, we present and interpret results relating to the assimilation of permeability and soil gas concentration data for the purposes of estimating the reduction in prediction uncertainty in relation to the first two metrics. In Section 3.3, we present prediction uncertainty results for the third metric.

3.2   The mean and variance of the soil gas and indoor air concentrations

The problem geometry framing the data worth analysis presented in this work largely follows from Yu et al. (2009) and the previous chapter and is depicted here in Figure 3.1. In order to further frame the discussion in this work, we begin by discussion the data collection strategy for the problem geometry shown in Figure 3.1. Specifically, we begin under the assumption that a public right-of-way separates the TCE DNAPL source zone and the row of houses 1, 2 and 3. Furthermore, the developer has purchased house 2 and 5 given that it directly overlies the known TCE plume and directly faces the TCE source zone across the right-of-way. As part of the developers due diligence in assessing their liability, they obtain a number of soil cores along the edge of the right-of-way for the purpose of measuring the subsurface permeability structure. Despite the fact that the developer owns houses 2 and 5, it is impossible for them to drill vertically beneath the foundation without tearing them down (by way of preview, house 2 is not determinately impacted over the liability period and could result "as is" at a later date; however, house 5 is impacted above the regulatory limit). The previous chapter identifies that permeability variations in the subsurface lithology as well as heterogeneity within a single lithological unit can both impact the fate and transport of TCE from groundwater to indoor air. In the context of this work, we restrict our attention to permeability variations arising from heterogeneity within a single lithological unit. As such, we ignore the issue that depositional or erosional "holes" in a laterally pervasive clay layer (aquitard) may permit vertical contaminant transport analogous to Freeze et al. (1990).

A key issue in this work is our assumption that although the developer has liability for indoor air impacts to houses 1, 3, 4 and 6, the developer does not have access to these properties which is often the case at real sites. To constrain the prediction uncertainty for indoor air impacts to the remaining houses, the developer then uses; knowledge of the subsurface permeability structure obtained from the soil cores in combination with TCE soil gas concentration measurements within these same boreholes, as well as TCE soil gas concentration measurements from beneath the foundation slab of house 2. In this section, we define prediction uncertainty using two metrics. First, the ability to

correctly estimate the actual indoor air concentration within houses 1, 3, 4 and 6 at any point in time. This metric is quantified by conditioning the permeability structure with the available borehole data so that the expected value of the indoor air concentration arising from a series of Monte Carlo simulations (Figure 2.7) approaches that of the actual indoor air concentration which we denote as "reality". Reality will be defined shortly. Second, the ability to use available soil gas concentration data to reduce the standard deviation of the indoor air concentration at unmonitored locations (i.e. houses 1, 3, 4 and 6) arising from the series of Monte Carlo simulations through time. The methodology for quantifying these two metrics is described below.

3.2.1   Worth of permeability data

To evaluate the worth of permeability data to constrain the prediction uncertainty of indoor air concentrations arising from subsurface heterogeneity in the permeability structure of an aquifer, we begin by assigning the permeability realization denoted as "reality" in Figure 3.2 as a representation of the point-to-point permeability structure of an actual site. This permeability realization was generated using the FGEN software (see Robin et al. (1993)) adopting the direct Fourier transform method based on geostatistics of the Borden aquifer, and was not conditioned using actual field data. In particular, the geometric mean permeability is $K = 2 \times 10^{-11} \, m^2$, the variance of $Y = \ln K$ was assigned a value of $\sigma_Y = 1.0 \, [-]$, and the correlation lengths of K in the principal, transverse and perpendicular bedding direction are $\lambda_1 = 5.0 \, m$, $\lambda_2 = 2.0 \, m$, and $\lambda_3 = 0.15 \, m$, respectively (Sudicky, 1986; Woodbury and Sudicky, 1991; Turcke and Kueper, 1996). A Gaussian power spectral density function was used to distribute the geostatistical parameters over the spatial domain. The variance σ_Y was increased from that of the Borden aquifer to that of the more heterogeneous Cape Cod aquifer (see Hess et al. (1992)). In the context of this work, this change was motivated by our desire to pose a more difficult test (relative to the more homogeneous Borden aquifer) for a realistically limited permeability sampling strategy to be able to reconstruct the structure of the aquifer over the entire conceptual model domain and effectively constrain prediction uncertainty.

Aquifer reconstruction followed by first sampling the "reality" aquifer in a series of patterns (i.e. 1K, 3K, 5K, 7K and 80K) itemized on Table 3.1 as "K" with the actual borehole locations depicted in Figure 3.1 and Figure 3.2. Boreholes 1 to 7 used in the 1K, 3K, 5K and 7K sampling patterns are all located along the right-of-way with a minimum spacing of $5 \, m$ and a maximum spacing of $10 \, m$ between them. Each borehole is sampled for permeability along $0.1 \, m$ vertical increments across the

entire thickness of the problem geometry. An obvious difficulty with this particular pattern of boreholes for the purpose of aquifer reconstruction is that they do not capture information over much of the domain and in particular near all of the houses in the back row (i.e. houses 4, 5 and 6). Once again, we assume access limitations preclude mobilizing a drill rig onto these properties in order to obtain a core. Even though the developers own the houses 2 and 5, they do not drill on these properties in order to make the problem a more challenging test for kriging conditioned on the permeability data and assimilating concentration data. An extreme and entirely unrealistic sampling strategy involving 80 boreholes (i.e. 80K) is used as a benchmark to illustrate the value of detailed and widespread permeability information across the entire conceptual model domain. Furthermore, 80K provides some perspective on the balance between two contrasting strategies, namely: being data rich at the expense of significant data collection expenditures with the intent of minimizing prediction uncertainty; and, alternatively conducting ones professional due diligence in characterizing the site under modest data collection expenditure while accommodating a reasonable (sustainable) amount of prediction uncertainty. The terms involving reasonable and sustainable sampling design will be further investigated in Chapter 4.

Ordinary kriging is discussed in Freeze et al. (1990) as a means of conditioning a heterogeneous permeability relation using sample data within a data worth framework. Essentially, ordinary kriging in an interpolation/extrapolation methodology is used to estimate the value of a random value (in this study, permeability) at one or more non-sampled locations from sparse sample data, taking into consideration the Gaussian statistical structure of the aquifer (Olea, 1974; Royle et al., 1981; Lam, 1983; Heine, 1986; Davis, 1986; Deutsch and Journel, 1992; Journel and Huijbregts, 1992). We use ordinary kriging as implemented by S-GEMS (Stanford Geostatistical Earth Modeling Software) to generate multiple statistically-equivalent heterogeneous permeability realizations to that of "reality" as shown in Figure 3.2, where each realization is conditioned using the available borehole data from the 1K, 3K, 5K and 7K sampling patterns. Each realization is generated after S-GEMS which is first used to estimate a variogram based on the available sparse data with the prior knowledge that the variogram is Gaussian. Turcke and Kueper (1996) used a statistical goodness-of-fit test to fit Gaussian, exponential and spherical variograms to Borden data (see also: Barnes, 1991; van Groenigen, 2000; Rogerson et al., 2004). This approach would alleviate our need of the prior knowledge which we acknowledge will unduly influence the correlation between our "reality" aquifer and each synthetic realization. While a single vertical borehole (1K) may be sufficient to characterize the subsurface lithology (absent the issue of finding a hole a laterally continuous clay layer analogous to Scenarios 4 and 5 of Chapter 2), it is probably entirely insufficient for the purposes of kriging in

this context particular because it is impossible to estimate λ_1 and λ_2. Here we assume that the same correlation length scale applies to different coordinate directions, namely isotropic correlation length. As such, 1K should be viewed as the "least" effort that one should use to investigate a site albeit the most likely one from a typical field investigation perspective, although it is not appropriate for the purpose of aquifer reconstruction as we denote it here. In contrast, we postulate that 7K may be the "greatest" effort that one could reasonably use to investigate the site, although it is lack of aquifer information in the transverse bedding direction and also assumed to have isotropic correlation length. However, we anticipate that the worth of the permeability data in terms of aquifer reconstruction will substantially diminish as we move from the front row of the houses (1, 2, and 3) to the back row (4, 5 and 6) given that these houses are up to $40\ m$ away from the row of boreholes along the right-of-way, and hence multiple values of λ_1. The statistics of kriging variance $\sigma_k^2(x_0)$ which is also the variance of the kriged estimates error and variance of permeability sample data $\sigma_{k\,z}^2$ for the 1K, 3K, 5K, 7K and 80K sample patterns are provided on Table 3.2. The kriging variance does not give an accurate reflection of the local variation in that it is a function of the spatial distribution and configuration of sample data but not the values of the sample data. Kriging variance will decrease as sample data are closer to the location of estimation. As we can see, the mean value of the kriging variance decreases when additional samples are incorporated into the kriging system (which can also be seen in Gao et al. (1996)). The 80K shows the smallest mean value of the kriging variance even if the variability of the sample data is relatively high. Using the criterion of minimizing the average kriging variance, more boreholes of permeability data exhibit a more accurate interpretation of the aquifer structure.

To test whether our sampling and kriging strategy have any merit in using the permeability data to reduce prediction uncertainty of the indoor air concentrations, we generate $n_{realizations} = 50$ alterative permeability realizations for each of the testing scenarios. Monte Carlo method is the most widely used technique for capturing the uncertainty in predictions. Whether 50 Monte Carlo simulations would be sufficient and reliable for uncertainty analysis here is a question. A Monte Carlo realization size can be evaluated by running as many Monte Carlo runs as necessary to obtain a reasonable statistic and a reasonable statistic is obtained when the addition of more Monte Carlo runs does not change the first and second moment of the simulated indoor air concentration of TCE distributions. The results of the mean and the standard deviation of the indoor air concentrations for FGEN, 1K and 7K scenarios with the increasing number of Monte Carlo realizations are shown in Figure 3.3. The mean value tends to become stable after 10 realizations for FGEN and 7K scenarios and after approximately 20 - 30 realizations for 1K scenario depending on the location of the house. The standard deviation has the similar trend for each house under different scenarios. The FGEN and

50

7K scenarios obtain constant values within 10 Monte Carlo realizations; while 1K tends to get more stable with additional 10 realizations. We can conclude here that 50 Monte Carlo simulations would be sufficient and large enough to provide a statistically representative concentration distribution for each scenario.

At this point, we restrict our attention to 1K, 7K and 80K scenarios given that they provide a diverse spatial distribution of sparse data to test the kriging algorithm. We then use *CompFlow Bio* (see scenario 1 of Chapter 2) to simulate the fate and transport of TCE from groundwater to the indoor air of each of the six houses within the conceptual model domain in a Monte Carlo framework. Figure 3.4 shows the resulting indoor air concentrations (initially denoted as $\mathbb{C}_{c_{TCE}IA}$ in Yu et al. (2009) and Chapter 2, and here revised to $c_{IA} \equiv \mathbb{C}_{c_{TCE}IA}$ and furthermore $\hat{c}_{IA} = \ln c_{IA}$) for each of the permeability realizations, as well as that arising from reality (here denoted as $c_{IA_{reality}}$). Note that at this point, we have not used any TCE soil gas concentration data to reduce prediction uncertainty. Clearly, 1K does a poor job of capturing the temporal shape of the indoor air concentrations in all of the six houses relative to reality. In contrast, both 7K and 80K capture the temporal shape albeit with 7K exhibiting greater uncertainty than 80K. To this end, we feel satisfied that at least (in a subjective sense) the 7K scenario provides some value in using available permeability data to constrain prediction uncertainty. What remains is a more quantitative description of prediction uncertainty, especially with the inclusion of the TCE soil gas data. Note that Figure 2.7 displays about as much variability as the 80K scenario in Figure 3.4. This is the consequence of that fact FGEN was used in the previous chapter to generate all 50 unconditioned permeability realizations, with each realization generated using identical statistics to our reality. As such, perfect knowledge of the aquifer geostatistics is implied by Figure 2.7 and serves to constrain prediction uncertainty substantially. The reader should bear in mind that obtaining this perfect knowledge of the aquifer geostatistics implicitly requires obtaining many cores (see Sudicky (1986)), well beyond the 7K scenario.

3.2.2   Worth of soil gas concentration

In the previous section, we described how permeability data used to condition the heterogeneous permeability structure of an aquifer could influence the fate and transport of TCE from groundwater to the indoor air. Part of a professional engineers or geoscientists due diligence in managing such a site also involves collecting TCE soil gas and indoor air concentrations in a routine pattern over space and time at accessible and pertinent locations to assess the potential for indoor air concentrations to

exceed a regulatory limit. Therefore, these TCE soil gas and indoor air samples should serve to further constrain prediction uncertainty at locations where the developer has liability exposure for but limited access to monitor actual TCE concentrations. In the case of this study, the developer has access to all boreholes along the right-of-way as well as within and beneath the foundation slab of house 2. However, the developer has liability for, but no access to, houses 1, 3, 4 and 6.

The TCE soil gas sampling strategy involves collecting near-surface (at a depth of $0.1\ m$) samples within boreholes 2, 3, 4, 5 and 6, as well as beneath the foundation slab of houses 2 and 5. Various strategies are used to sample these locations and are itemized on Table 3.1 as 2C, 4C and 7C. For each sample location within a given sampling strategy, samples are regularly collected at each ith monitoring time interval $(t_{i-1}, t_i]$ over the duration of the liability period. For all sampling strategies, borehole 4 as well as beneath the foundation slab of house 2 are always monitored given that they directly overlie the plume, are also accessible, and would seemingly provide peak soil gas concentrations. The latter point is significant in that Chapter 2 established the idea that it is the mass flux of TCE crossing the foundation slab which largely controls the indoor air concentration, with the mass flux of TCE being the product of the TCE soil gas concentration beneath the foundation slab times the volumetric flow rate of soil gas into the house through the foundation slab (due to HVAC activities which ultimately depressurize the house relative to the vadose zone). Consequently, the combined worth of TCE soil gas data to estimate peak concentrations with permeability data to estimate the volumetric flow of soil gas across the foundation slab are essential to constrain prediction uncertainty of indoor air concentrations within houses 1, 3, 4 and 6. Sampling strategy 2C denotes the minimal effort of monitoring borehole 1 and house 2, and is combined with any of 1K, 3K, 5K, 7K and even 80K to evaluate the combined worth of both soil gas and permeability data. Sampling strategy 4C provides an incremental improvement over 2C by further sampling boreholes 3 and 5 which straddle the surface projection of the groundwater plume across the right-of-way (see Figure 2.4a). This strategy is motivated by McGrath and Pinder (2003) who concluded that sampling a plume along its lateral edges where the concentration gradient is large serves to provide a great deal of information regarding plume location uncertainty. This should in turn help us further reduce prediction uncertainty for houses 1, 3, 4 and 6 which are all laterally offset from the plume. Finally, sampling strategy 7C further samples boreholes 2 and 6 as well as beneath the foundation slab of house 5. This strategy is motivated by the previous study who observed that indoor air impacts to houses decrease precipitously as a function of their lateral distance from the groundwater plume, generating narrow exclusion zones. Boreholes 2 and 6 are included to further resolve the lateral decline in soil gas concentrations in the transverse direction to the groundwater plume along the right-

of-way. House 5 is included to provide TCE soil gas concentration information pertinent to the back row of houses. As with sampling strategy 2C, both 4C and 7C are combined with any of 1K, 3K, 5K, 7K and 80K to evaluate the combined worth of both soil gas and permeability data.

3.2.2.1   Application of the discrete static Kalman filter

Motivated by the work of Zhang et al. (2005), we adopt the discrete static Kalman filter (KF) as a tool to assimilate the spatial and temporal set of TCE soil gas data, arising from each sampling scenario identified on Table 3.1, to reduce prediction uncertainty of indoor air concentrations within houses 1, 3, 4 and 6. In general, the KF is used to estimate the state of a linear dynamic system based as simulated by a stochastic model in association with noisy data and appropriate initial and boundary conditions. The KF describes how the system would respond after processing the noisy data in order to achieve an optimal estimation of its current state. In essence, it is a recursive solution to least-squares problems. As data become available, the KF can then deduce a minimum error estimate to improve future model prediction. A complete presentation of the discrete static KF can be found in Gelb (1974). Within this framework, we use the numerical model *CompFlow Bio* (see Equations (2.1−2.2) and Equations (1−10) of Yu et al. (2009)) in combination with the heterogeneous permeability structure of the aquifer to represent the stochastic model. In this work, both groundwater and soil gas flow are steady-state processes. However, dissolution of TCE source zone with resulting transport of TCE from the groundwater, across the capillary fringe and into the vadose zone, and finally into the indoor air, is a transient process. Dissolution of the TCE source zone is the only process that imparts non-linearity to the transport of TCE and is isolated far from houses 1, 3, 4 and 6 where we desire to constrain prediction uncertainty. We remind the reader that permeability data serve to condition the heterogeneous permeability structure of the aquifer as described in Section 3.2.1. TCE soil gas concentrations constitute the noisy data, where the noise is imparted by; point-to-point variability in heterogeneous permeability structure of the aquifer as it impacts contaminant transport, potential field collection and laboratory analysis errors of the soil gas data.

An essential element in the application of the discrete static KF is that the noisy observed data conform to a Gaussian distribution. To confirm that this requirement is met, in Figure 3.5 we present the natural logarithm of simulated TCE soil gas concentrations (initially denoted as $\mathbb{C}_{c_{TCE}g}$ in Yu et al. (2009) and used in Chapter 2, and here revised to $c_{SG} \equiv \mathbb{C}_{c_{TCE}g}$ and furthermore $\hat{c}_{SG} = \ln c_{SG}$) for 1K and 7K, and before assimilation of data arising from the 2C, 4C or 7C sampling scenarios. The

notion of "before assimilation" for these concentrations is denoted as $c_{SG}{}^-$. To test whether $c_{SG}{}^-$ are Gaussian, we begin with the hypothesis that for all Monte Carlo realizations, the TCE soil gas concentration $c_{SG_{i,j,l}}{}^-$ at the ith monitoring time interval $(t_{i-1}, t_i]$ where $i = \{1, 2, \ldots, n_{times}\}$, the jth house, and with $l = \{1, 2, \ldots, n_{realizations}\}$, are log-normally distributed. To test our hypothesis, we use a one-sample Kolmogorov-Smirnov test (Massey, 1951; Miller, 1956) with the null hypothesis being that the vector of $\hat{c}_{SG_{i,j,l}}{}^-$ follows a normal distribution. The alternative hypothesis is that the vector of $\hat{c}_{SG_{i,j,l}}{}^-$ does not follow a normal distribution at the 5% significance level. The temporal transition of $\hat{c}_{SG_{i,j,l}}{}^-$ to being normally distributed as the increment of the monitoring interval $(t_{i-1}, t_i]$ increases is shown in Figure 3.5 when the mean concentration $\mu_{SG_{i,j}}{}^-$ calculated as:

$$\mu_{SG_{i,j}}{}^- = \frac{1}{n_{realizations}} \sum_{l=1}^{n_{realizations}} \hat{c}_{SG_{i,j,l}}{}^- \tag{3.1}$$

changes from a dashed to a solid line. The dashed line indicates the null hypothesis is not accepted, while the solid line indicates that it is accepted. In general, $\hat{c}_{SG_{i,j}}{}^-$ are not normally distributed at early time because TCE soil gas concentrations beneath the houses for numerous realizations are lower than the accuracy of what can be reliably resolved with the *CompFlow Bio* model. Despite this predicament, we are compelled to use $\hat{c}_{SG}{}^-$ within the KF over the entire liability period for reasons of continuity, as well as necessity as will be made clear shortly. We note that 3K, 5K and 80K conform to the same behaviour. Finally, TCE soil gas concentrations $\hat{c}_{SG}{}^-$ provided in Figure 3.5 are post-processed via Equations (12-14) of Yu et al. (2009) to yield the indoor air concentrations c_{IA} depicted here in Figure 3.4.

Following the methodology of Zhang et al. (2005), we now adopt the static discrete KF to the problem at hand. The involvement of the *CompFlow Bio* model as the stochastic model used to transport soil gas concentrations within a given heterogeneous permeability realization can be written succinctly as:

$$c_{SG_{i+1}}{}^- = \Phi\, c_{SG_i}{}^- + w_i \tag{3.2}$$

where $c_{SG_{i+1}}{}^-$, $c_{SG_i}{}^-$ are the vectors of discrete estimates of the state variable representing TCE soil gas concentrations at the monitoring time interval $i + 1$ and i respectively (before application of the KF), the matrix Φ is the state transition matrix (i.e. the *CompFlow Bio* model) from the ith to $(i + 1)$th monitoring interval, and w_i is the system noise (model error) at the ith interval, which has a zero mean, white sequence of covariance Q_i. Herrera (1998) showed that if the state vector is spatially correlated, then the resulting model error at some nodes at some times may be strongly correlated to the model error at some other nodes at some other time. Therefore, we also assume that the model error w_i is spatially correlated and its correlation Q_i is imbedded in the physical model. Also, the soil gas concentrations are obtained via the *CompFlow Bio* model with heterogeneous permeability fields based on a given spatial correlation and statistics, which would be different for different number of permeability data scenarios. From Equation (3.2), the first step in the application of discrete static KF is to propagate the state variables and error covariance of the system forward in time. As discussed in Zhang et al. (2005), the vector of discrete estimates of the state variable here includes both spatial and temporal TCE soil gas concentrations, with the static discrete KF not differentiating between the space and time dimensions. Therefore, we further simplify $c_{SG_i}{}^-$ in Equation (3.2) to simply $c_{SG}{}^-$.

We now show that the state matrix $\hat{c}_{SG}{}^-$ is built by the concentration vectors $\hat{c}_{SG_{i,j}}{}^-$ at location $j = \{1, 2, \ldots, n_{locations}\}$, the ith monitoring time interval $i = \{1, 2, \ldots, n_{times}\}$, and for a given permeability realization l. The dimension of the state matrix $\hat{c}_{SG}{}^-$ is $m \times 1$ and $m = n_{locations} \times n_{times}$. The variable $n_{locations}$ is the number of target locations to be investigated including measurement and non-measurement locations (6 houses and 5 boreholes, see Table 3.1) where $n_{locations} = 11$. The variable n_{times} is the maximum number of monitoring time intervals at every 100 days from TCE injection to the end of liability period 5 years with $n_{times} = 18$. It has to be noted that $\hat{c}_{SG}{}^-$ is independently reconstructed for each permeability realization l. The vector of discrete estimates of the state variable is given as:

$$\hat{c}_{SG}{}^- = [\hat{c}_{SG_{1,1}}{}^-, \ldots, \hat{c}_{SG_{n_{times},1}}{}^-, \hat{c}_{SG_{1,2}}{}^-, \ldots, \hat{c}_{SG_{n_t,2}}{}^-, \hat{c}_{SG_{1,n_{locations}}}{}^-, \ldots, \hat{c}_{SG_{n_{times},n_{locations}}}{}^-]^T. \tag{3.3}$$

Each realization of $\hat{\boldsymbol{c}}_{SG}{}^-$ for a given l is an initial optimal estimate of the state vector for either the 1K, 3K, 5K, 7K or 80K scenarios.

The prior estimate of the error covariance $\boldsymbol{P}^-$ before any soil gas samples are taken using one of 2C, 4C or 7C is assembled using the covariance matrix elements $\text{cov}(c_{SG_{i,j}}{}^-, c_{SG_{i,j+\tau}}{}^-)$ calculated by:

$$\text{cov}\left(\hat{c}_{SG_{i,j}}{}^-, \hat{c}_{SG_{i,j+\tau}}{}^-\right) =$$

$$\frac{1}{n_{realizations} - 1} \sum_{l=1}^{n_{realizations}} \left(\hat{c}_{SG_{i,j,l}}{}^- - \mu_{SG_{i,j}}{}^-\right)\left(\hat{c}_{SG_{i,j+\tau,l}}{}^- - \mu_{SG_{i,j+\tau}}{}^-\right) \tag{3.4}$$

where τ is the lag between two $\hat{c}_{SG_{i,j}}{}^-$ separated by both location j and monitoring time i in the vector $\hat{\boldsymbol{c}}_{SG}{}^-$, and is of dimension $m \times m$. Given that $\hat{\boldsymbol{c}}_{SG}{}^-$ varies for each of the 1K, 3K, 5K, 7K and 80K scenarios, so does the covariance matrix. We reiterate our earlier hypothesis that the progressive addition of permeability data to condition the heterogeneous structure of the aquifer should serve to increase the correlation between two values of $\hat{c}_{SG_{i,j}}$ separated by τ. Furthermore, the combined worth of permeability and concentration data should serve to reduce prediction uncertainty of the TCE soil gas concentrations beneath the foundation slab, and the flow of soil gas across the foundation slab into the indoor air of houses 1, 3, 4 and 6. The net result should be a reduction in the prediction uncertainty of indoor air concentrations within these same houses.

The rationale for each TCE soil gas sampling strategy (i.e. 2C, 4C and 7C) was previously described in Section 3.2.2. These samples all arise from use of the "reality" aquifer shown in Figure 3.2, with the resulting TCE soil gas concentrations beneath all six houses provided in Figure 3.5 and denoted as $\hat{\boldsymbol{c}}_{SG,reality}$. These measurements are uniquely defined and known a priori everywhere within the computational domain in space and time, and are assumed to be linearly related to the system and are employed for the filtering process. A vector of measurement data $\boldsymbol{z}$ that is corrupted by non-negligible measurement errors $\boldsymbol{v}$ is extracted from $\hat{\boldsymbol{c}}_{SG,reality}$. These measurement errors are assumed to be uncorrelated in time and independent of the system estimation. Marrin and Thompson

(1987) suggested that the average error for different sampled soil gas data in shallow aquifer is $\pm 21\%$ which reflects both human/instrument performance and the effect of horizontal heterogeneities. We assume that the acceptable relative error level for measurements varies within $\pm 25\%$ of the measured values and this error does not vary between different measurements. The measurement equation describing how the estimated $\hat{c}_{SG}{}^{-}$ are related to the actual system state z is:

$$z = H \times \hat{c}_{SG}{}^{-} + v \tag{3.5}$$

where

$z =$ the vector of g noise-corrupted measurements, with dimension $g \times 1$, where in our case $g = n_{sample\ locations} \times n_{times}$ with $n_{sample\ locations} \in n_{locations}$, such that $n_{sample\ locations}$ includes only that subset of locations where concentration measurements are taken (i.e. where a 'C' is listed on Table 3.1);

$H =$ the measurement matrix, dimension $g \times m$;

$v =$ the vector of random measurement noise, $v \sim N(0, R)$, assumed to be uncorrelated with previous measurement errors, dimension $g \times 1$;

$R =$ the measurement noise covariance matrix, independent of the state variable, with dimension $g \times g$.

Elements of the measurement matrix H are assigned a value of 1 if a sample is taken at location j at the ith time interval; 0 otherwise. The diagonal elements of R are assigned a value of $(25\%\ z)^2$.

As measurement data z become available, the concentration and its associated error covariance are updated by assimilating the data and its uncertainty. First, we compute the Kalman gain matrix K of dimension $m \times g$ which contains the model and measurements bias as:

$$K = P^{-}\ H^{T}\ (H\ P^{-}\ H^{T} + R)^{-1}\ . \tag{3.6}$$

Next, the posterior estimate of $\hat{c}_{SG}{}^+$ can be updated using a linear function of the prior estimate and the measurement $\mathbf{z}$ as:

$$\hat{c}_{SG}{}^+ = \hat{c}_{SG}{}^- + K\,(\mathbf{z} - H \times \hat{c}_{SG}{}^-) \ . \tag{3.7}$$

Finally, the corresponding error covariance $\mathbf{P}^+$ is updated as:

$$\mathbf{P}^+ = \mathbf{P}^- - K\,H\,\mathbf{P}^- \ . \tag{3.8}$$

The optimal estimation for $\hat{c}_{SG}{}^+$ is obtained when the Kalman gain K is maximized and the sum of the diagonal of the covariance matrix $\mathbf{P}^+$ is reduced.

3.2.2.2 Interpretation of the discrete static Kalman filter

A critical assumption in the presentation of the discrete static KF in Section 3.2.2.1 was that an optimal estimate of the state vector of soil gas concentrations could be obtained without simultaneously estimating permeability within the state vector. This strategy is possible via the ensemble KF (Evensen, 2003) given that the resulting dynamic system would become non-linear. The disadvantage to the ensemble KF approach for this application is that a significantly greater number of forward simulations (see Equation (3.2)) would be required. This is not possible given our current computational resources.

We remind the reader that although there is only one set of permeability and soil gas concentration data that we can actually observe in reality (as itemized on Table 3.1), we never know with certainty what the permeability and soil gas concentrations are in the remaining of the domain. In other words, during the application of the Kalman filter, we must turn a blind eye towards "reality" as depicted in Figure 3.2 (except at the boreholes sampled by one of 1K, 3K, 5K, 7K and 80K) as well as $c_{SG,reality}$ in Figure 3.5 for houses 1, 3, 4, 5 (for 2C and 4C) and 6. Bearing this in mind, for a given 1K, 3K,

5K, 7K or 80K, all permeability realizations l have identical values at those points where we actually take measurements and hence are equally probable versions of reality. Therefore, we reiterate that each solution of $c_{SG}{}^-$ obtained via Equation (3.2) for every permeability realizations l is an initial (prior) optimal estimate of the state vector for either the 1K, 3K, 5K, 7K or 80K scenarios. Application of the Kalman gain matrix via Equation (3.7) then serves to correct the prior estimate of $\hat{c}_{SG}{}^-$ at those locations where to not have data (i.e. houses 1, 3, 4 and 6) to $\hat{c}_{SG}{}^+$. This correction should force $\hat{c}_{SG}{}^+$ to tend towards reality as progressively more data are assimilated.

Following the above interpretation of the Kalman filter, we are then left with a distribution of $\hat{c}_{SG}{}^+$ arising from each permeability realizations l for a given 1K, 3K, 5K, 7K or 80K. This distribution must then be interpreted with regards to our notion of prediction uncertainty. Once again, we define prediction uncertainty using two metrics: first, the ability to correctly estimate the actual indoor air concentration within houses 1, 3, 4 and 6 at any point in time; second, the ability to use available soil gas concentration data to reduce the standard deviation of the estimated indoor air concentration within houses 1, 3, 4 and 6 at any point in time. We investigate these metrics by introducing the following moments of the posterior soil gas and indoor air concentration distributions:

$$\mu_{SG_{i,j}}{}^+ = \mathrm{E}\left[\hat{c}_{SG_{i,j}}{}^+\right] = \frac{1}{n_{realizations}} \sum_{l=1}^{n_{realizations}} \hat{c}_{SG\,i,j,l}{}^+$$

$$\mu_{IA_{i,j}}{}^+ = \mathrm{E}\left[\hat{c}_{IA_{i,j}}{}^+\right] = \frac{1}{n_{realizations}} \sum_{l=1}^{n_{realizations}} \hat{c}_{IA\,i,j,l}{}^+$$

(3.9)

$$\sigma_{SG_{i,j}}{}^+ = \sqrt{\mathrm{E}\left[\left\{\hat{c}_{SG_{i,j}}{}^+ - \mu_{SG_{i,j}}{}^+\right\}^2\right]}$$

$$\sigma_{IA_{i,j}}{}^+ = \sqrt{\mathrm{E}\left[\left\{\hat{c}_{IA_{i,j}}{}^+ - \mu_{IA_{i,j}}{}^+\right\}^2\right]}\,.$$

While the indoor air concentrations are no directly estimated as part of the KF, values of the posterior $\hat{c}_{SG_{i,j}}{}^+$ are post-processed as outlined by Yu et al. (2009) and the previous chapter to yield $\hat{c}_{IA_{i,j}}{}^+$.

3.2.3 Results

Figure 3.6 and Figure 3.7 depict the mean and standard deviation (see Equation (3.9)) of the posterior soil gas and indoor air concentrations within all six houses over the liability period. These results are shown for four separate permeability and soil gas sampling combinations which would appear to have common use in practice, namely; 1K2C, 1K4C, 7K4C and 7K7C. They span the range of the least and greatest reasonable effort that one would consider when characterizing a site. The transition from 1K2C to 1K4C is meant to evaluate the worth of moderately increasing the amount of soil gas concentration data given limited permeability data. The transition from 1K4C to 7K4C is meant to evaluate the worth of significantly increasing the amount of permeability data given moderate amount of soil gas concentration data. Finally, the transition from 7K4C to 7K7C is meant to demonstrate the value of a detailed site investigation. For each scenario, the solid line is the mean value while the error bars provide the standard deviation. The reality soil gas concentration $\hat{c}_{SG,reality_{i,j}}$ is also provided for reference.

To demonstrate the contribution of the KF, consider the soil gas concentration beneath the foundation slab of house 2 (see Figure 3.6b) which is monitored for each of 1K2C, 1K4C, 7K4C and 7K7C. Despite the fact the prior $\bar{\hat{c}_{SG_{i,j}}}$ (see Figure 3.5) shows a great deal of variability for each permeability realization l, with this variability for 1K being unreasonably large relative to 7K and reality, the KF is able to reduce prediction uncertainty drastically. However, the indoor air concentration within house 2 (see Figure 3.7b) shows unreasonably large prediction uncertainty for both 1K2C and 1K4C, particularly at early time. In contrast, 7K4C and 7K7C show near identical prediction uncertainty demonstrating the value of characterizing the subsurface heterogeneous permeability structure in order to correctly estimate the flow of soil gas into the house across the foundation slab and hence the mass flux of TCE into the indoor air. This same pattern is repeated for house 5 (see Figure 3.6e), except that the soil gas is only sampled for 7K7C. As a consequence, soil gas concentration prediction uncertainty for 7K4C increases slightly relative to 7K7C, while it visibly deteriorates for both 1K2C and 1K4C. The resulting pattern is further compounded for the indoor air concentrations shown in Figure 3.7e.

Of particular interest is the prediction uncertainty within houses 1, 3, 4 and 6 for which the developer does not possess but has liability. Soil gas concentrations within each of these four houses show a great deal of prediction uncertainty (see Figures 3.6a, 3.6c, 3.6d and 3.6f), to the degree that it

seems impossible to visibly note any trend that would differentiate between any of 1K2C, 1K4C, 7K4C and 7K7C. In contrast, a trend becomes apparent for prediction uncertainty of the indoor air concentrations (see Figures 3.7a, 3.7c, 3.7d and 3.7f). For 1K2C and 1K4C, the KF significantly underestimates $\mu_{IA_{i,j}}^{+}$ relative to reality with very large values of $\sigma_{IA_{i,j}}^{+}$. This problem appears far worst for houses 1 and 4 which are both on one side of the groundwater plume relative to houses 3 and 6. While 7K7C appears to do a slightly better job at reducing prediction uncertainty than 7K4C, both 7K7C and 7K4C are vast improvements over 1K2C and 1K4C. The trend where reduction in prediction uncertainty is favored for houses 3 and 6 over houses 1 and 4 still appears to be present, but is relatively subdued.

To quantify the prediction uncertainty, we initially focus only on the first metric, namely; the ability to correctly estimate the actual indoor air concentration within houses 1, 3, 4 and 6 at any point in time. As such, we use the root-mean-square-error ($RMSE_{c_j}$) defined as:

$$RMSE_{c_j} = \sqrt{\frac{\sum_{i=1}^{n_{times}} \left| \mu_{IA_{i,j}}^{+} - \hat{c}_{IA,reality_{i,j}} \right|^2}{n_{times}}} \tag{3.10}$$

where j is one of houses 1, 3, 4, 5 and 6, and for all sampling scenarios listed on Table 3.1. Results for all possible permeability and concentration sampling combinations are shown in Figure 3.9a. With 7C, a clear trend develops for the progression of 1K, 3K, 5K, 7K and 80K where the $RMSE_{c_j}$ decreases, and reaches a minimum at 80K, for all houses. This same trend holds as the concentration sampling strategy is decreased to 4C. This trend is in support of the first metric. As the concentration sampling strategy is decreased to 2C, this trend becomes erratic. The KF does not appear to be able minimize the $RMSE_{c_j}$ based on the combined worth of the permeability and concentration data when only two concentration sampling points are used.

3.2.3.1 Assessing the significance in improved predictions of indoor air concentrations at unmonitored houses

The monetary commitment required to collect increasing amounts of data could, in part, be motivated by the need to demonstrate a statistically significant improvement in the reduction of prediction uncertainty at the locations where the developer has liability but no access (i.e. houses 1, 3, 4 and 6). In this section once again we focus on the first metric measuring prediction uncertainty, namely; the ability to correctly estimate the actual indoor air concentration within houses 1, 3, 4 and 6 at any point in time. Here we apply a statistical test to infer whether there is a reduction in the absolute difference between $\mu_{SG_{i,j}}^{+}$ and $\hat{c}_{SG,reality_{i,j}}$ as well as between $\mu_{IA_{i,j}}^{+}$ and $\hat{c}_{IA,reality_{i,j}}$ for the pairs of incrementally increasing sampling data strategies represented by; 1K2C to 1K4C, 1K4C to 7K4C, and finally 7K4C to 7K7C. Within the context of the statistical inference test, each pair of sampling strategies represents a population of $\hat{c}_{SG_{i,j}}^{+}$ and $\hat{c}_{IA_{i,j}}^{+}$ at each houses $j = 1, 3, 4$ and 6, within time interval $(t_{i-1}, t_i]$. These populations are generated by the l permeability realizations.

The test statistic used here is based on the standard normal distribution, which requires the use of the natural logarithm of the soil gas concentration and indoor air concentrations. For brevity, the presentation which follows is for indoor air concentrations but is equally applicable to soil gas concentrations. Let $\mu_{IA,1_{i,j}}^{+}$ and $\mu_{IA,2_{i,j}}^{+}$ represent the means of two populations "1" and "2", $\left|\mu_{IA,1_{i,j}}^{+} - \hat{c}_{IA,reality_{i,j}}\right|$ and $\left|\mu_{IA,2_{i,j}}^{+} - \hat{c}_{IA,reality_{i,j}}\right|$ are the absolute errors between the mean and the actual concentrations at location j and the ith time interval, $\sigma_{IA,1_{i,j}}^{2}{}^{+}$ and $\sigma_{IA,2_{ij}}^{2}{}^{+}$ are their variances, and each population is of size $n_{IA,1_{i,j}}$ and $n_{IA,2_{i,j}}$, respectively. Let Δ_0 be the hypothetical difference in the absolute error, while $\left|\hat{c}_{IA,1_{i,j}}^{+} - \hat{c}_{IA,reality_{i,j}}\right|$ and $\left|\hat{c}_{IA,2_{i,j}}^{+} - \hat{c}_{IA,reality_{i,j}}\right|$ are the absolute errors obtained from two samples drawn randomly from each populations. The statement for the hypothesis H_0 test is as follows:

Null hypothesis: $H_0 : \left|\mu_{IA,1_{i,j}}^{+} - \hat{c}_{IA,reality_{i,j}}\right| - \left|\mu_{IA,2_{i,j}}^{+} - \hat{c}_{IA,reality_{i,j}}\right| = \Delta_0$ \hfill (3.11)

Test statistic: $\quad z_0 = \dfrac{\left| \hat{c}_{IA,1_{i,j}}^{\;+} - \hat{c}_{IA,reality_{i,j}} \right| - \left| \hat{c}_{IA,2_{i,j}}^{\;+} - \hat{c}_{IA,reality_{i,j}} \right| - \Delta_0}{\sqrt{\dfrac{\sigma_{IA,1_{i,j}}^{2\;+}}{n_{IA,1_{i,j}}} + \dfrac{\sigma_{IA,2_{i,j}}^{2\;+}}{n_{IA,2_{i,j}}}}}$ (3.12)

Alternative hypotheses: Rejection criterion:

$H_1: \left| \mu_{IA,1_{i,j}}^{\;+} - \hat{c}_{IA,reality_{i,j}} \right|$

$\qquad - \left| \mu_{IA,2_{i,j}}^{\;+} - \hat{c}_{IA,reality_{i,j}} \right| \neq \Delta_0$ $z_0 > z_{\alpha/2}$ or $z_0 < -z_{\alpha/2}$

$H_2: \left| \mu_{IA,1_{i,j}}^{\;+} - \hat{c}_{IA,reality_{i,j}} \right|$

$\qquad - \left| \mu_{IA,2_{i,j}}^{\;+} - \hat{c}_{IA,reality_{i,j}} \right| > \Delta_0$ $z_0 > z_{\alpha}$ (3.13)

$H_3: \left| \mu_{IA,1_{i,j}}^{\;+} - \hat{c}_{IA,reality_{i,j}} \right|$

$\qquad - \left| \mu_{IA,2_{i,j}}^{\;+} - \hat{c}_{IA,reality_{i,j}} \right| < \Delta_0$ $z_0 < -z_{\alpha}$

and the parameter α denotes the level of statistical significance.

The results of this statistical test are reported on Table 3.3 for the pairs of incrementally increasing sampling data strategies represented by; 1K2C to 1K4C, 1K4C to 7K4C, and finally 7K4C to 7K7C. A value of "T" denotes acceptance of hypothesis H_2, "F" denotes acceptance of hypothesis H_3, and "0" denotes acceptance of hypothesis H_0, all at a level of significance of $\alpha = 0.05$ or 5%. In other words, "T" denotes that the first prediction uncertainty metric is achieved via the worth of adding more data, "F" denotes the data have the opposite impact (which would not be anticipated), and "0" denotes that no statement of significance can be made. A dash "-"denotes that the test could not be conducted because one or both of the populations do not conform to a normal distribution.

For the pair of sampling strategies 1K2C to 1K4C, Table 3.3 indicates that there appears to be a consistent reduction in prediction uncertainty in the soil gas concentration beneath houses 3 and 4 only. However, there is no consistent reduction in prediction uncertainty for the indoor air concentrations for any of the houses. These results follow from the $RMSE_{c_j}$ analysis, and imply only one permeability core is insufficient data for the KF to assimilate the available concentration data in a useful manner. For sampling strategies 1K4C to 7K4C, houses 3, 5 and 6 (and possibly 1) all show a

significant reduction in the prediction uncertainty for soil gas concentrations. More important is that all houses show a significant improvement for the indoor air concentrations. These results imply that there is a great deal of value in collection permeability data as part of a detailed site characterization (i.e. 7K) when attempting to use the KF to assimilate even a modest amount of soil gas concentration data (i.e. 4C) to reduce the prediction uncertainty of indoor air concentrations. Finally, for sampling strategies 7K4C to 7K7C, there is a significant reduction in prediction uncertainty for soil gas concentrations beneath houses 1, 4 and 5. However, only house 5 shows a reduction in the prediction uncertainty of the indoor air concentrations. This occurs because 7C is the only strategy that samples the soil gas beneath house 5. These results imply that once a detailed site investigation has been conducted (i.e. 7K), there is a point of diminishing returns where the further addition of soil gas concentration data (i.e. 4C to 7C) does little via the KF to significantly reduce the prediction uncertainty of indoor air concentrations. In summary, there is a clear balance between the types and quantities of data that should be obtained as part of a site characterization effort in order to maximize the combined data worth. In the case of the problem at hand, 7K4C appears to provide the optimal balance.

3.2.3.2 Assessing the significance in uncertainty reduction of predicted indoor air concentrations at unmonitored houses

Following from Section 3.2.3.1, we now focus on the second metric measuring prediction uncertainty, namely; the ability to use available soil gas concentration data to reduce the standard deviation of the estimated indoor air concentration within houses 1, 3, 4 and 6 at any point in time. We apply a statistical test to infer whether there is a reduction in $\sigma_{SG_{i,j}}{}^+$ and $\sigma_{IA_{i,j}}{}^+$ for the pairs of incrementally increasing sampling data strategies represented by; 1K2C to 1K4C, 1K4C to 7K4C, and finally 7K4C to 7K7C. As before, each pair of sampling strategies represents a population of $\hat{c}_{SG_{i,j}}{}^+$ and $\hat{c}_{IA_{i,j}}{}^+$ at each house $j = 1, 3, 4$ and 6, within time interval $(t_{i-1}, t_i]$. These populations are generated by the l permeability realizations.

We apply a statistical test based on the inference of variances of two normal populations "1" and "2", with variances $\sigma_{IA,1_{i,j}}^2{}^+$ and $\sigma_{IA,2_{i,j}}^2{}^+$. For brevity, the presentation which follows is for indoor air concentrations but is equally applicable to soil gas concentrations. The development of this test hypothesis requires the introduction of the $\mathcal{F}$-distribution, which is defined as follows:

64

$$\mathcal{F} = \frac{W/u}{Y/v} \tag{3.14}$$

where W and Y are independent chi-square random variables with u and v degrees of freedom, respectively. It has the probability density function:

$$f(x) = \frac{\Gamma\left(\frac{u+v}{2}\right)\left(\frac{u}{v}\right)^{u/2} x^{u/2-1}}{\Gamma\left(\frac{u}{2}\right)\Gamma\left(\frac{v}{2}\right)\left[\left(\frac{u}{v}\right)x + 1\right]^{(u+v)/2}} \qquad 0 < x < \infty \tag{3.15}$$

where Γ is the gamma function. The statement for the hypothesis test is as follows:

Null hypothesis: $\quad H_0 : \sigma_{IA,1_{i,j}}^{2}{}^{+} = \sigma_{IA,2_{i,j}}^{2}{}^{+}$ $\hfill$ (3.16)

Test statistic: $\quad f_0 = \dfrac{\sigma_{IA,1_{i,j}}^{2}{}^{+}}{\sigma_{IA,2_{i,j}}^{2}{}^{+}}$ $\hfill$ (3.17)

Alternative hypotheses: $\hspace{4cm}$ Rejection criterion:

$H_1 : \sigma_{IA,1_{i,j}}^{2}{}^{+} \neq \sigma_{IA,2_{i,j}}^{2}{}^{+} \qquad f_0 > f_{\frac{\alpha}{2}, n_{IA,1_{i,j}}-1, n_{IA,2_{i,j}}-1}$ or $f_0 < f_{1-\frac{\alpha}{2}, n_{IA,1_{i,j}}-1, n_{IA,2_{i,j}}-1}$ $\hfill$ (3.18)

$H_2 : \sigma_{IA,1_{i,j}}^{2}{}^{+} > \sigma_{IA,2_{i,j}}^{2}{}^{+} \qquad\qquad\qquad f_0 > f_{\alpha, n_{IA,1_{i,j}}-1, n_{IA,2_{i,j}}-1}$

$H_3 : \sigma_{IA,1_{i,j}}^{2}{}^{+} < \sigma_{IA,2_{i,j}}^{2}{}^{+} \qquad\qquad\qquad f_0 < f_{1-\alpha, n_{IA,1_{i,j}}-1, n_{IA,2_{i,j}}-1}$

$$f_{1-\alpha, n_{IA,1_{i,j}}-1, n_{IA,2_{i,j}}-1} = F^{-1}(\alpha)_{n_{IA,1_{i,j}}-1, n_{IA,2_{i,j}}-1}$$

the parameter α denotes the level of statistical significance and $F^{-1}(\alpha)$ is the quantile function of the $\mathcal{F}$-distribution.

The results of this statistical test are reported on Table 3.4 for the pairs of incrementally increasing sampling data strategies represented by; 1K2C to 1K4C, 1K4C to 7K4C, and finally 7K4C to 7K7C. A value of "T" denotes acceptance of hypothesis H_2, "F" denotes acceptance of hypothesis H_3, and "0" denotes acceptance of hypothesis H_0, all at a level of significance of $\alpha = 0.05$ or 5%. In other words, "T" denotes that the second prediction uncertainty metric is achieved via the worth of adding more data, "F" denotes the data have the opposite impact (which would not be anticipated), and "0" denotes that no statement of significance can be made. A dash "-"denotes that the test could not be conducted because one or both of the populations does not conform to a normal distribution.

Comparison of Table 3.4 to Table 3.3 indicates that the reduction in prediction uncertainty as represented by the second metric follows the same pattern as that of the first, albeit with more "noise" (i.e. tests of "0" or "F"). The two patterns typically become aligned at late time near the end of the liability period. We surmise that fifty Monte Carlo simulations based on alternative permeability realizations may not be a sufficient number to reliably conduct the statistical test represented by Equations (3.14) - (3.18) and test the second metric. In summary, we reiterate that there is a clear balance between the types and quantities of data that should be obtained as part of a site characterization effort in order to maximize the combined data worth. Once again, 7K4C appears to provide the optimal balance.

3.3 Probability of indoor air concentrations exceeding a regulatory limit

As discussed previously, the key issue in this work is our assumption that although the developer has liability for indoor air impacts to houses 1, 3, 4 and 6, the developer does not have access to these properties which is often the case at the real sites. Liability would occur should the indoor air concentration in one of these houses exceed the regulatory limit, at which point (upon receiving notification) the developer would need to purchase the house. As before, to constrain the prediction uncertainty for indoor impacts on these houses, the developer then uses; knowledge of the subsurface permeability structure obtained from the soil cores in combination with TCE soil gas concentration measurements within these same boreholes, as well as TCE soil gas concentration measurements from beneath the foundation slab of house 2. Therefore, we define the third metric associated with prediction uncertainty, namely; the ability to accurately forecast the probability of indoor air

concentrations within houses 1, 3, 4 and 6 exceeding a regulatory limit. This third metric is directly related to the first two which relies on the first and second moments of $\hat{c}_{SG_{i,j}}{}^{+}$ and $\hat{c}_{IA_{i,j}}{}^{+}$ (see Equation (3.9) and Sections 3.2.3.1 and 3.2.3.2) except that we are now required to evaluate the "tail" of the distribution.

The definition of probability of failure (also called the probability of exceedence) follows from Chapter 2 who uses the criterion to delineate an exclusion zone. Failure occurs when the indoor air concentration of TCE at a single house exceeds the NYSDOH (2005) regulatory limit of $5 \ \mu g/m^3$. The probability of failure $P_{i,j}$ for the jth location (i.e. one of houses 1, 3, 4, 5 and 6) is calculated as the sum of the number of Monte Carlo realizations for which the indoor air concentration of TCE $c_{IA_{i,j}}{}^{+}$ exceeds $c_{regu} = 5 \ \mu g/m^3$, defined as $n_{c_{IA_{i,j}}>c_{regu}}$, divided by the total number of Monte Carlo realizations $n_{realizations}$:

$$P_{i,j} = \frac{n_{c_{IA_{i,j}}>c_{regu}}}{n_{realizations}} \ . \tag{3.19}$$

In Figure 3.8, we provide $P_{i,j}$ for $c_{IA_{i,j}}{}^{+}$ for sampling data strategies represented by; 1K2C, 1K4C, 7K4C and 7K7C. For reference, the actual indoor air concentration $c_{IA,reality_{i,j}}$ is provided in Figure 3.4 for reference as the solid red line. Because house 2 is monitored for all sampling scenarios, there is no uncertainty regarding its indoor air concentration. Houses 1, 3, 4 and 6 which are all laterally offset from the groundwater plume, and for which the developer has liability, exhibit indoor air concentrations that in reality never exceed the regulatory limit. In contrast, house 5 which the developer already owns exceeds the regulatory limit only after 1100 days. Figure 3.8 shows that 1K2C yields a probability of exceedence (using Equation (3.19)) for houses 1, 3 and 6 which over-predicts reality, and correctly estimates $P_{i,j}$ for house 4. In addition 1K2C underestimates $P_{i,j}$ for house 5. Surprisingly, the transition to 1K4C estimated the correct $P_{i,j}$ for all houses except for; one event in house 6 at 600 days, and house 5 for the entire monitoring period. We remind the reader that 1K4C performs poorly based on the first two performance metrics regarding prediction uncertainty (see Sections 3.2.3.1 and 3.2.3.2). As we transition to 7K4C, $P_{i,j}$ is now correctly estimated for houses 1, 3 (except for one minor occurrence at 1100 days), 5 (after 1100 days) and 6. House 4 shows some minor values of probability of exceedence between 1100 and 1800 days. Finally, as we

67

transition to 7K7C, the values of $P_{i,j}$ in house 4 are reduced as well as the values occurring before 1100 days in house 5. We note that the soil gas beneath house 5 is monitored for 7K7C, and yet we do not get the exact probability of exceedence outcome as occurs in reality. This is a due to 7K permeability sampling strategy incorrectly estimating the permeability below house 5, leading to an incorrect estimate in the flux of TCE across the foundation slab and into the indoor air.

We anticipate that through the process of collecting progressively more permeability and concentration data, the difference between the estimated probabilities of exceedence $P_{i,j}$ and whether actual indoor air concentration exceeds the regulatory limit in reality $P_{reality_{i,j}}$ should tend to zero. We quantify this as:

$$RMSE_{P_j} = \sqrt{\frac{\sum_{i=1}^{n_{times}} \left| P_{i,j} - P_{reality_{i,j}} \right|^2}{n_{times}}} \qquad (3.20)$$

where;

$$P_{reality_{i,j}} = \begin{cases} 1, & \text{if } c_{IA,reality_{i,j}} > c_{regu} \\ 0, & \text{otherwise} \end{cases} \qquad (3.21)$$

and $RMSE_{P_j}$ denotes the root mean square error of the probability of failure for the jth house. Figure 3.9b shows results for $RMSE_{P_j}$ with $P_{i,j}$ calculated using Equation (3.19). The results do not clearly show our anticipated pattern which is largely due to insufficient Monte Carlo realizations in order to adequately represent the "tail" of the distribution via Equation (3.19). For instance, 1K seems to exhibit a lower $RMSE_{P_j}$ for various concentration sampling combinations relative to 3K, 5K and 7K. However, we know from Figure 3.7 and Figure 3.9a that this is resulted from 1K erroneously underestimating indoor air concentrations. This in turn leads to a smaller tail on the probability distribution. In general, the anticipated pattern is apparent for 7C with the progression from 1K to 80K. Despite the "noise" in $RMSE_{P_j}$ where lines of different K samples cross, it does appear that for a given sample line K there is a general downward slope (representing a decline in $RMSE_{P_j}$) with

increasing values of C. This supports the idea that the combined worth of permeability and concentration data contributes to reduction in prediction uncertainty as it applies to the third metric.

In an alternative strategy to define $P_{i,j}$, we follow the previous work "fitting" a log-normal as well as a beta distribution to $c_{IA_{i,j}}^+$ in order to better capture the tail of the distribution with the limited number of Monte Carlo simulations. Both the log-normal and beta distributions make use of the first two moments of $\hat{c}_{IA_{i,j}}^+$ given by $\mu_{IA_{i,j}}^+$ and $\sigma_{IA_{i,j}}^+$ in Equation (3.9). The log-normal distribution is characterized as being left-skewed with a heavy tail, and will therefore yield conservative estimates for the probability of exceedence relative to Equation (3.19) for extreme (low probability) events. In contrast, the unimodal form of the beta distribution with a bounded domain does not exhibit a heavy tail, and the normalized beta distribution has a range from 0 to 1. Therefore, it will therefore yield less conservative estimates for the probability of exceedence relative to the log-normal distribution and possibly even Equation (3.19). Figure 3.10 and Figure 3.11 provide $P_{i,j}$ for $c_{IA_{i,j}}^+$ for sampling data strategies 1K2C, 1K4C, 7K4C and 7K7C using the log-normal and beta distributions, respectively.

For 1K2C, the $P_{i,j}$ pattern using both the log-normal (Figure 3.10) and beta (Figure 3.11) distributions is identical to that established using Equation (3.19) (Figure 3.8). Surprisingly, the log-normal distribution yields a lower $P_{i,j}$ at houses 3 and 6 relative to Equation (3.19). The $P_{i,j}$ at house 5 is poorly predicted using all three methods. As we transition to 1K4C, both the log-normal and beta distributions now correctly estimate $P_{i,j}$ for all houses except for house 5. We reiterate that 1K4C performs poorly based on the first two metrics regarding to the prediction uncertainty (see Sections 3.2.3.1 and 3.2.3.2). In summary, we see little value in using either the log-normal or beta distributions to extrapolate the probability distribution for extreme events when $\mu_{IA_{i,j}}^+$ and $\sigma_{IA_{i,j}}^+$ are poorly constrained by insufficient data. As we transition to 7K4C, our anticipated pattern for $P_{i,j}$ emerges. Using $P_{i,j}$ as calculated by Equation (3.19) and shown in Figure 3.8 as a benchmark, for houses 1, 3, 4 and 6, the log-normal distribution does provide slightly greater estimates of $P_{i,j}$ while the beta distribution provides slightly lower estimates. For house 5, all three distributions appear to predict $P_{i,j}$ equally well. Finally, as we transition to 7K7C, the same pattern established by 7K4C is retained albeit with slightly lower values of $P_{i,j}$ relative to 7K4C. Of particular note is that the beta distribution for 7K7C yields $P_{reality_{i,j}}$ exactly for houses 1, 3 and 6, and exhibits only minor deviations between 1600 and 1800 days for house 4. As such, the beta distribution combined with sufficient data yields the most accurate prediction uncertainty estimate based on the third metric for those houses offset from the groundwater plume. We conclude that extensive site characterization of

69

the subsurface permeability structure is an essential prerequisite for estimating the probability of indoor air concentrations exceeding a regulatory limit, where the exceedence occurs as an extreme event.

Figure 3.9 shows the $RMSE_{P_j}$ with $P_{i,j}$ now calculated using the log-normal (Figure 3.9c) and the beta (Figure 3.9d) distributions. In general, the pattern of $RMSE_{P_j}$ for both the log-normal and beta distributions follows that of Equation (3.19) and is subject to the same interpretation. However, one additional feature becomes evident; that is, once we transition to a 7K sampling strategy, the $RMSE_{P_j}$ for the log-normal distribution appears greater than either that from Equation (3.19) or the beta distribution, for any given C sampling strategy. Furthermore, 7K4C and 7K7C appear to perform equally well at positioning the $RMSE_{P_j}$ for the beta distribution as being greater than that of Equation (3.19) but less than that of the log-normal distribution. Given the definition of the $RMSE_{P_j}$ given by Equation (3.20), we conclude that for the lateral houses offset from the groundwater plume (i.e. houses 1, 3, 4 and 6) the log-normal distribution provides the most conservative estimates for prediction uncertainty as characterized by the probability of exceedence. In addition, the beta distribution and Equation (3.19) provide intermediate and least conservative estimates.

3.4   Conclusions

A methodology was developed to assimilate soil core permeability data and TCE soil gas concentration data and to assess their worth for the purpose of reducing the numerical model prediction uncertainty. We proposed three metrics to quantify the prediction uncertainty, namely, the ability to estimate the actual indoor air concentration within houses 1, 3, 4 and 6 at any point in time; the ability to use available soil gas concentration data to reduce the standard deviation of the indoor air concentration within these houses arising from the series of Monte Carlo simulations through time; the ability to accurately forecast the probability of indoor air concentrations within these four houses exceeding a regulatory limit.

Due to the fact that the developer does not possess but has liability of houses 1, 3, 4 and 6, to evaluate the prediction uncertainty within houses 1, 3, 4 and 6 is of our particular interest. The results of soil gas concentrations and indoor air concentrations within each of these four houses after assimilating the soil core permeability and soil gas concentration data showed a great deal of prediction uncertainty in terms of the expected value relative to reality and the standard deviation.

70

Both 7K4C and 7K7C had a trend of improvements over 1K2C and 1K4C in reducing the prediction uncertainty; while 7K7C appeared to do an even better job. We used the root-mean-square-error ($RMSE_{c_j}$) and the statistical test (Equation (3.11) – (3.13)) to quantify the prediction uncertainty (the difference between the expected and the real values) associated with the first metric, and learnt that with the incremental increasing number of permeability and concentration samplings, the uncertainty would be reduced. Based on the second metric, another statistical test (Equation (3.14) – (3.18)) was conducted and the same trend was discovered: the standard deviation of the estimated indoor air concentration would decrease with the increasing number of data. Moreover, different measures of the prediction uncertainty all implied that once a detailed site investigation has been conducted (i.e. 7K), further additional soil gas concentration data did little improvement via the KF (i.e. 4C to 7C) and 7K4C appeared to provide the optimal balance between the types and quantities of data.

This third metric was presented by three different evaluations of the probability of exceedence (Equation (3.19) and fitting the tail of the indoor air concentration distribution to a log-normal and beta distribution) using $RMSE_{P_j}$ as a criterion in terms of the reduction of the prediction uncertainty. We concluded that the probabilities of 7K4C and 7K7C provided the closest trends to the reality, of which the Figures 3.8, 3.10 and 3.11 gave a good visual representation, and 7K4C and 7K7C appeared to perform equally well at positioning the $RMSE_{P_j}$ as well. In addition, the log-normal distribution provided the most conservative estimates for prediction uncertainty as characterized by the probability of exceedence, and the beta distribution and Equation (3.19) provided intermediate and least conservative estimates.

As we have seen, the methodology we developed, which is the combination of stochastic simulation in a geostatistical context with conditioned permeability data and Kalman filter – a special case of Bayesian filtering with respect to the indoor air concentration based on updating soil gas samplings, is a useful tool for improving the model prediction in space and time for the selected sample points problem considered here. Despite the fact that only limited number of samplings in terms of permeability and concentration are assimilated due to the availability of samplings in residential dwellings and high sampling costs, the uncertainty reduction is significantly improved compared with no measurements conditioned. It is very important for decision maker to choose the number of samplings, the location and time of samplings and the category of samplings as well. In our case, the more data are acquired, the better estimations of indoor air concentration and probability of exceedence we obtained. However, with the increasing number of permeability data applied in the geostatistical tool, the model prediction uncertainty reduces much more than the one with the

increasing number of soil gas sampling data. In another word, the value of permeability data is worth more than the value of concentration data under this framework; there is a clear balance between the types and quantities of data that should be obtained in order to maximize the combined data worth (i.e. 7K4C). In an environmental project, the site characterization of a geological aquifer is the most important phase.

In general, the use of Monte Carlo simulation and the KF requires a great number of groundwater flow and solute transport simulations per optimal sampling design which could be very computationally expensive. Therefore, the efficiency of this methodology depends on the type of application and the computation speed of the model. The judgement from experienced professionals for an initial sampling design is needed.
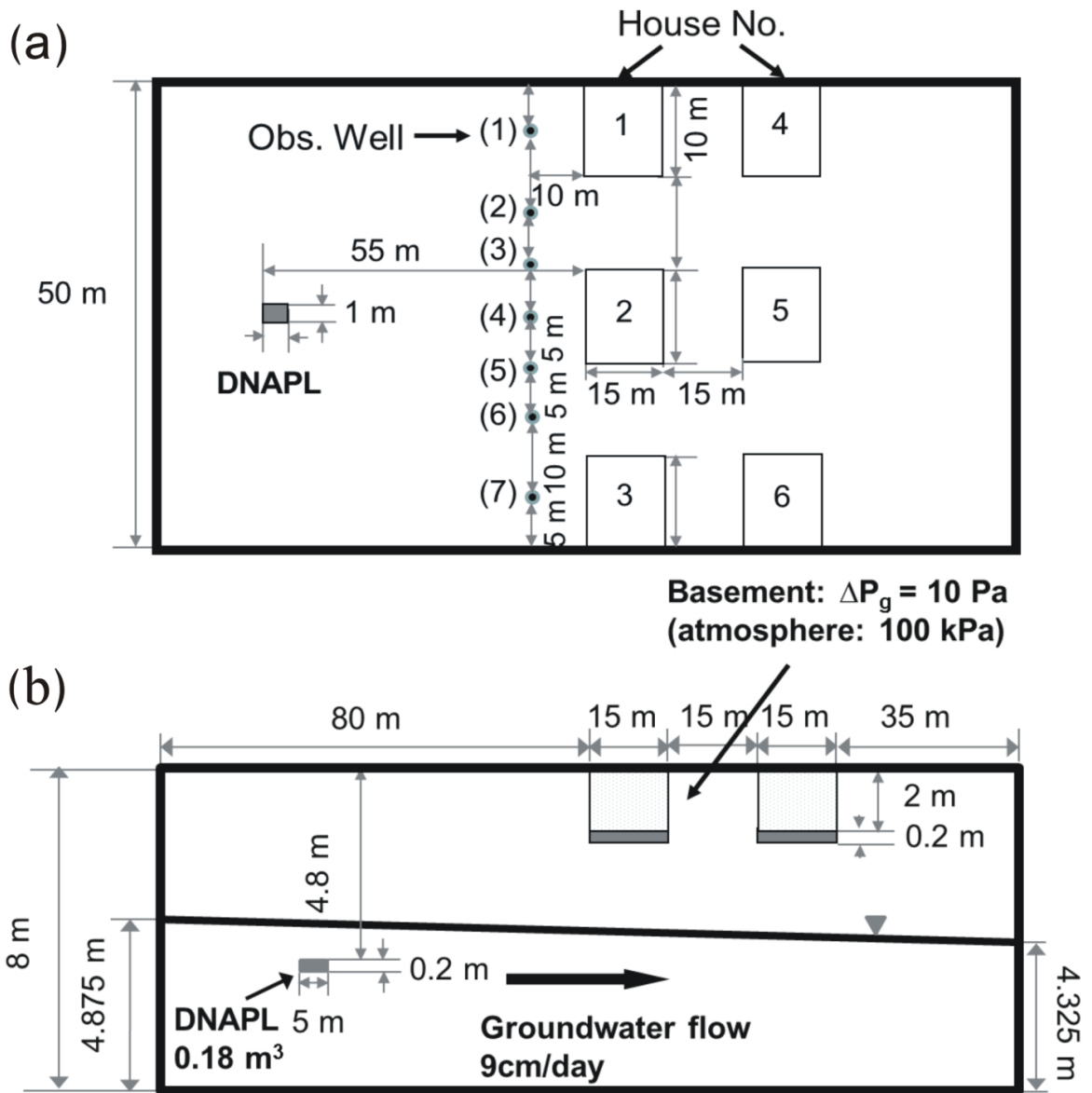
Figure 3.1: The 2D cross section views of the problem geometry: a) top view, and b) vertical cross section.
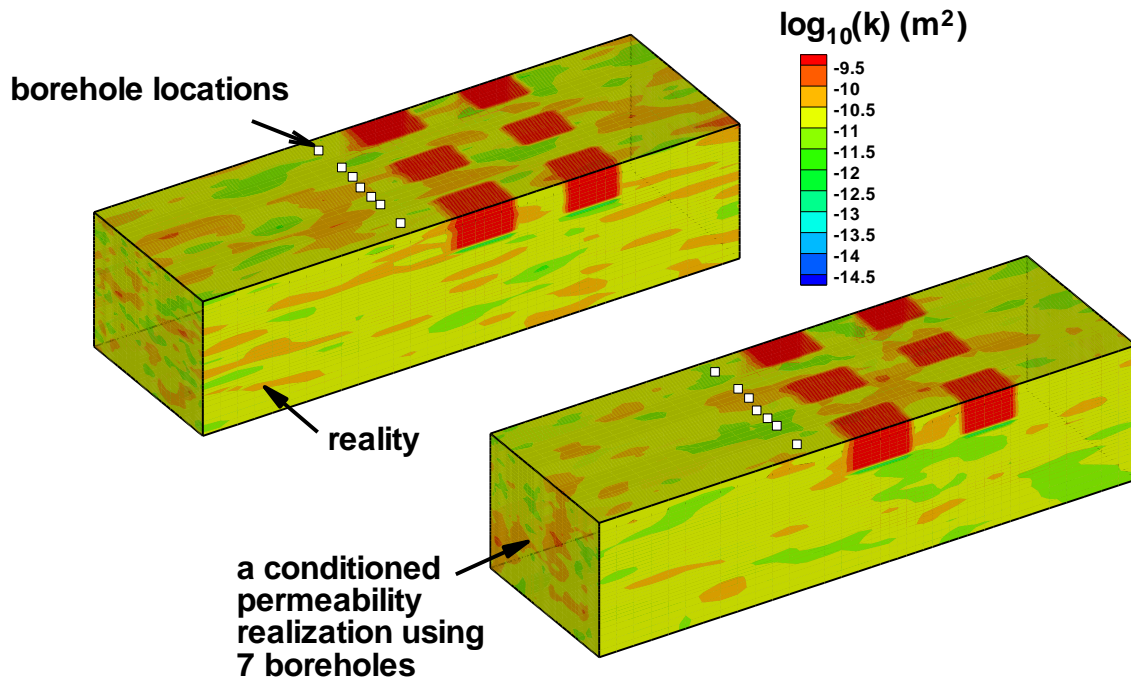
Figure 3.2: One interpolated random permeability field realization of the reality aquifer conditioning 7 boreholes of data statistics via kriging.
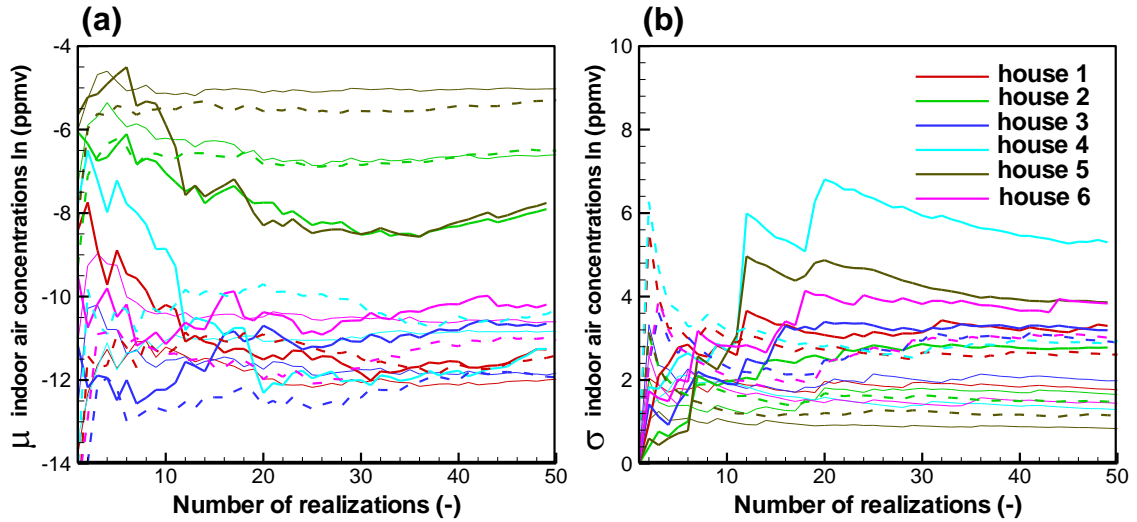
Figure 3.3: The mean and the standard deviation of the natural logarithmic values of indoor air concentration of TCE for different numbers of Monte Carlo realizations.

Note: The thin solid lines represent the indoor air concentrations of TCE from the permeability realizations incurred from FGEN data; the thick solid lines represent the indoor air concentrations of TCE from the permeability realizations incurred from 1 core data; and the dashed lines represent the indoor air concentrations of TCE from the permeability realizations incurred from 7 cores data.
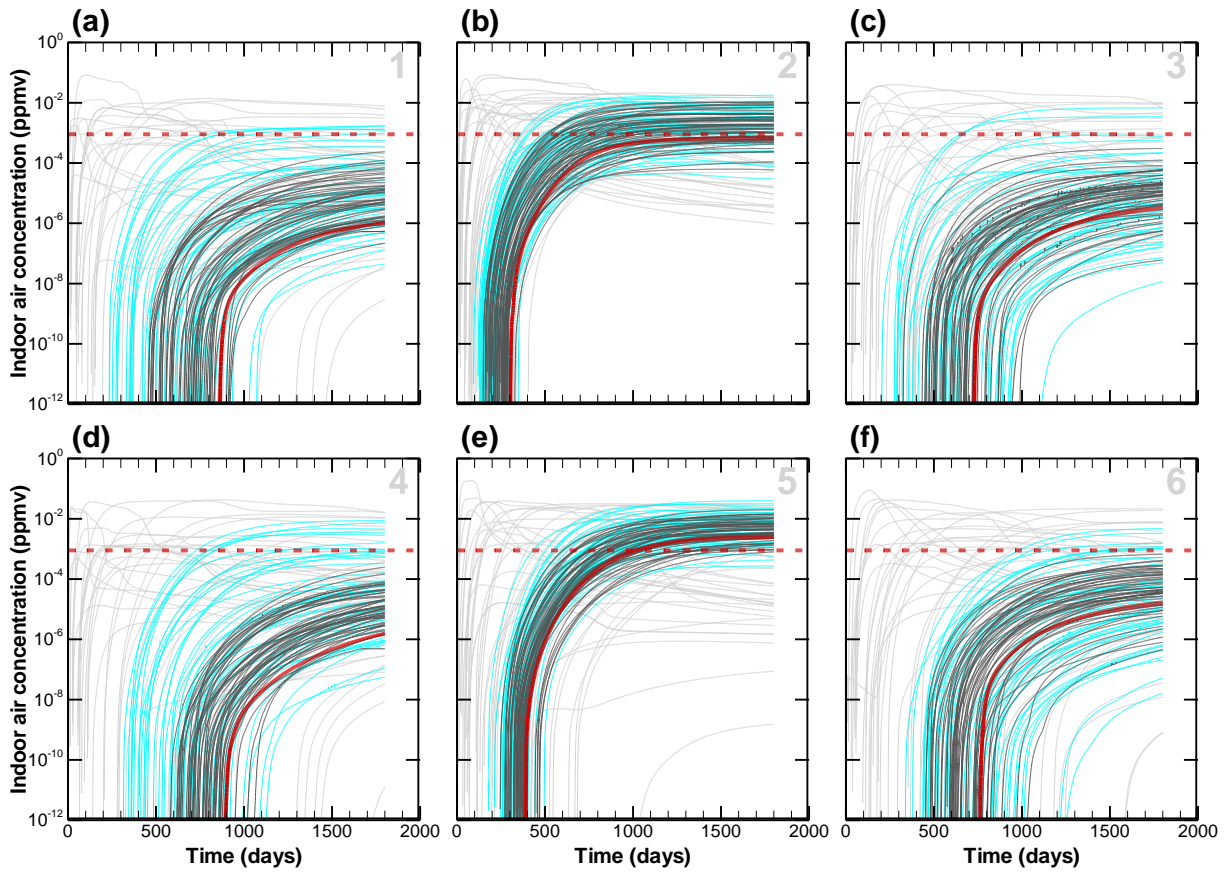
Figure 3.4: The indoor air concentration of TCE in logarithmic scale with air exchange rate $0.5/hr$ till 1800 days after TCE injection using multiple alternative permeability realizations via *CompFlow Bio* model: a) House No. 1, b) House No. 2, c) House No. 3, d) House No. 4, e) House No. 5, and f) House No. 6.

Note: The light grey lines represent the indoor air concentrations of TCE from the permeability realizations incurred from 1 core data; the cyan lines represent the indoor air concentrations of TCE from the permeability realizations incurred from 7 cores data; and the dark grey lines represent the indoor air concentrations of TCE from the permeability realizations incurred from 80 cores data. The red solid line represents the concentrations in reality. The red dashed line represents the regulatory limit for indoor air concentration of TCE.

76
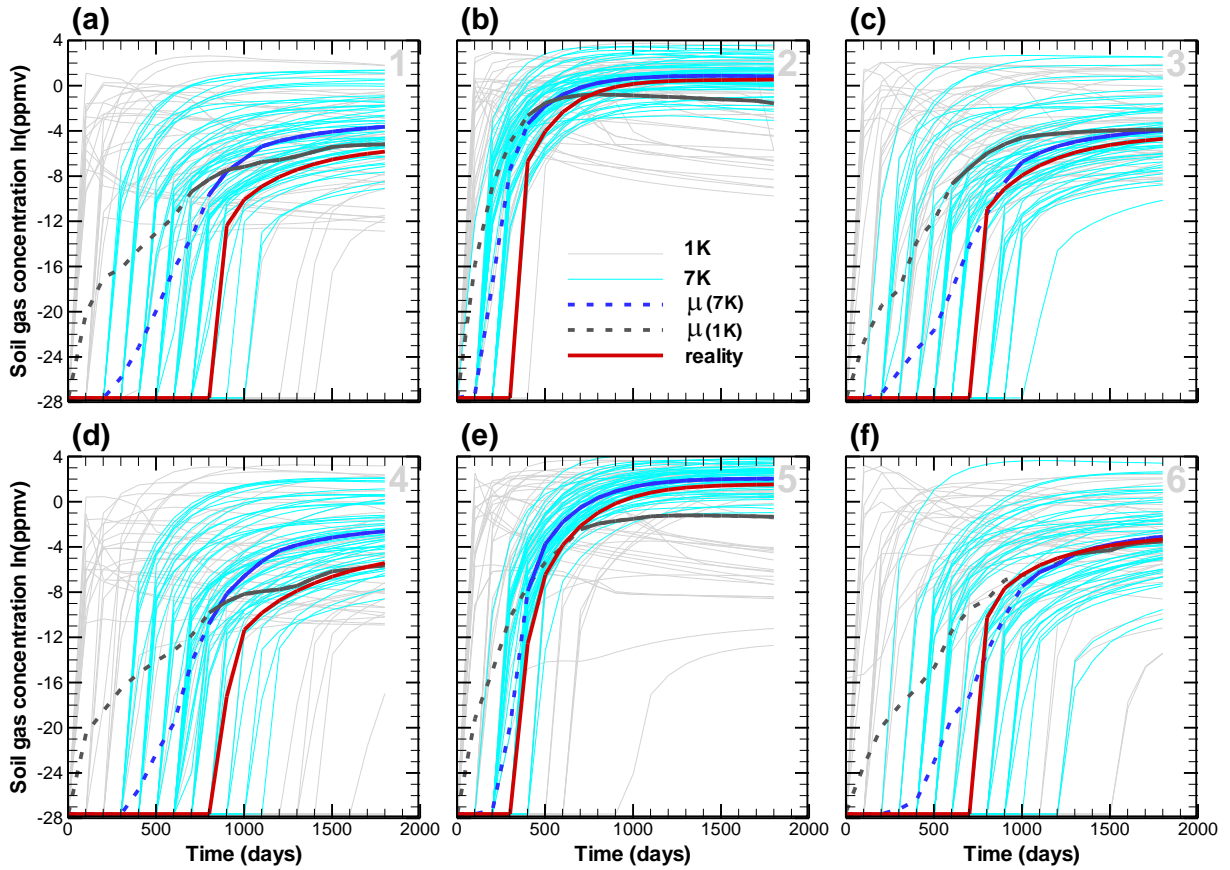
Figure 3.5: The natural logarithmic values of soil gas concentration of TCE from *CompFlow Bio* model results before assimilating data $\hat{c}_{SG}^{-}$ in light grey and cyan and their mean μ_{SG}^{-} in dark grey and dark blue: a) House No. 1, b) House No. 2, c) House No. 3, d) House No. 4, e) House No. 5, and f) House No. 6.

Note: The light grey lines represent the soil gas concentration of TCE from the permeability realizations incurred from 1 core data; the cyan lines represent the soil gas concentration of TCE from the permeability realizations incurred from 7 cores data. The red solid line represents the concentrations in reality $\hat{c}_{SG,reality}$. The dashed lines represent the concentration population do not follow a log-normal distribution verified using the Kolmogorov-Smirnov test; while the solid lines represent the acceptance of the log-normally distributed concentrations.
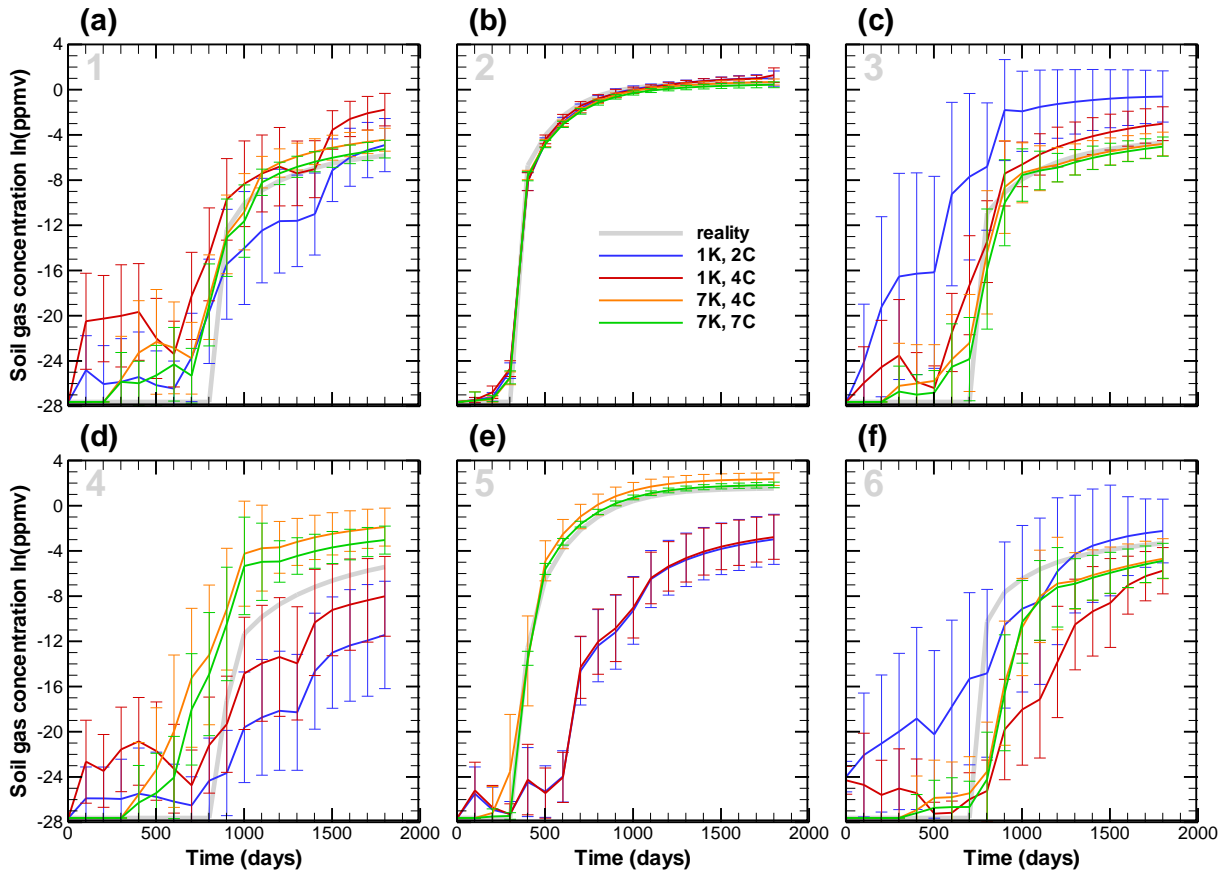
Figure 3.6: The mean $\mu_{SG}{}^+$ and the standard deviation $\sigma_{SG}{}^+$ of natural logarithmic values of soil gas concentration of TCE after updating different numbers of concentration measurements for different scenarios: a) House No. 1, b) House No. 2, c) House No. 3, d) House No. 4, e) House No. 5, and f) House No. 6.

Note: The solid lines are the mean values and the error bars represent the standard deviations of the Monte Carlo simulations.
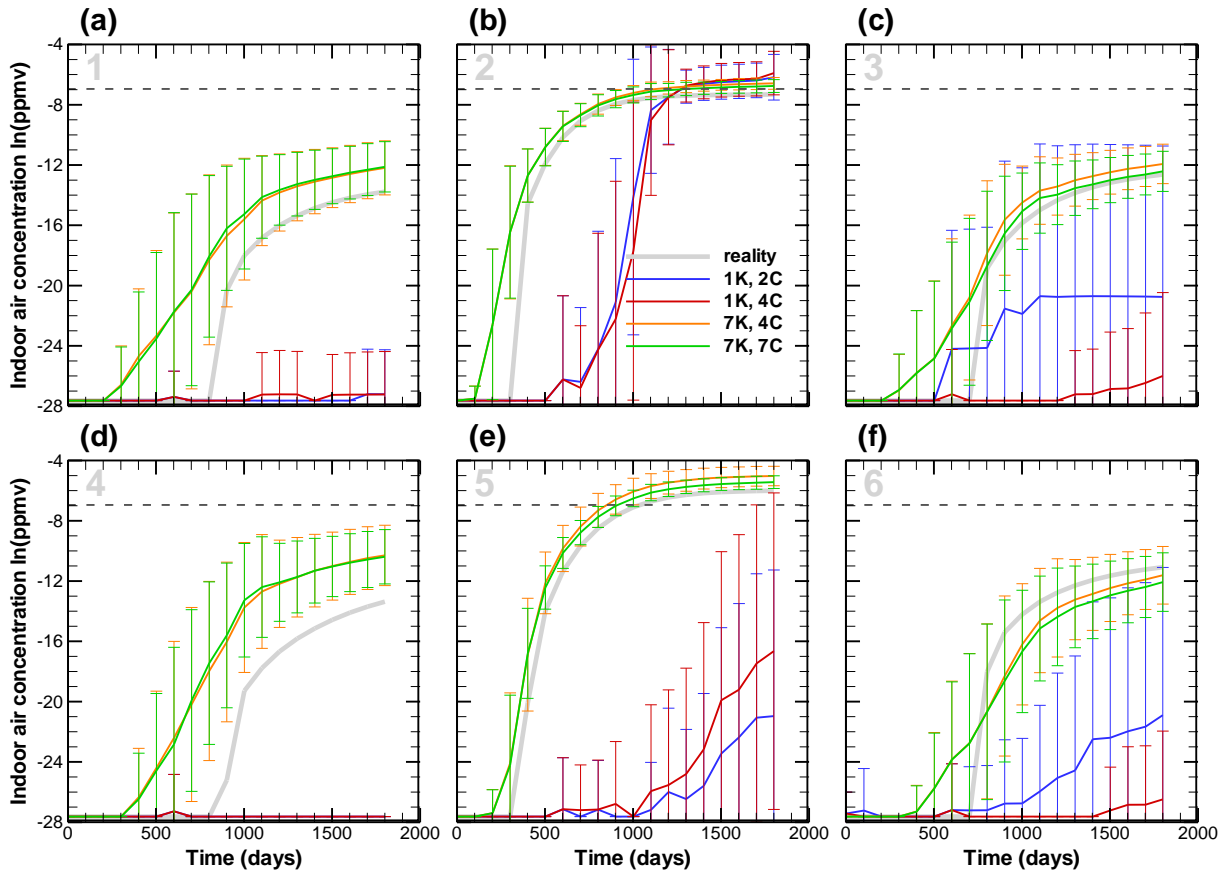
Figure 3.7: The mean μ_{IA}^+ and the standard deviation σ_{IA}^+ of natural logarithmic values of indoor air concentration of TCE after updating different numbers of concentration measurements for different scenarios: a) House No. 1, b) House No. 2, c) House No. 3, d) House No. 4, e) House No. 5, and f) House No. 6.

Note: The dashed line represents the regulatory limit for indoor air concentration of TCE. The solid lines are the mean values and the error bars represent the standard deviations of the Monte Carlo simulations.
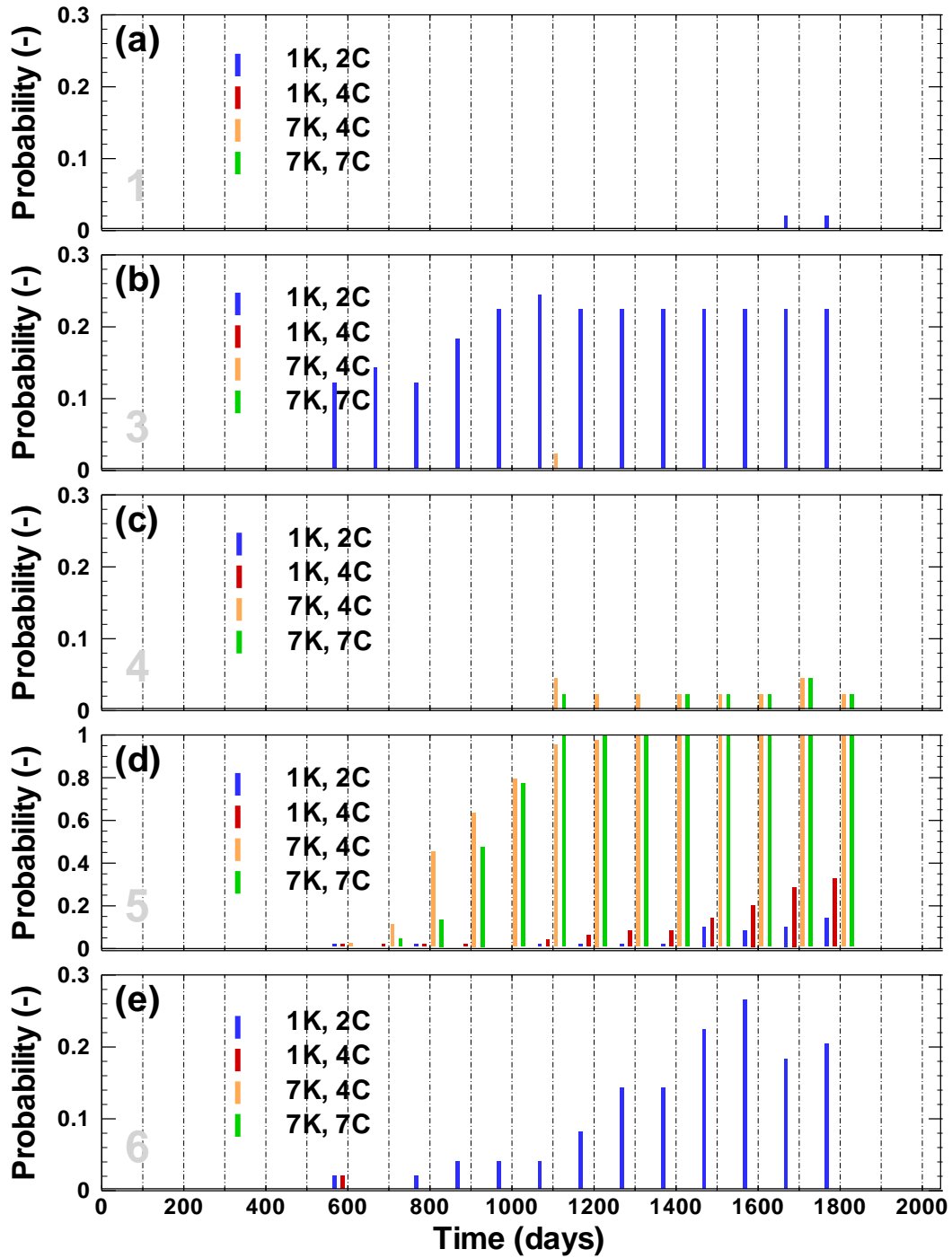
Figure 3.8: The probability of exceedence from the Monte Carlo realizations for the six houses for the cases 1)1K, 2C; 2) 1K, 4C; 3) 7K, 4C; and 4) 7K, 7C: a) House No. 1, b) House No. 3, c) House No. 4, d) House No. 5, and e) House No. 6. Note: d) has different scale in *y*- axis from the other figures.
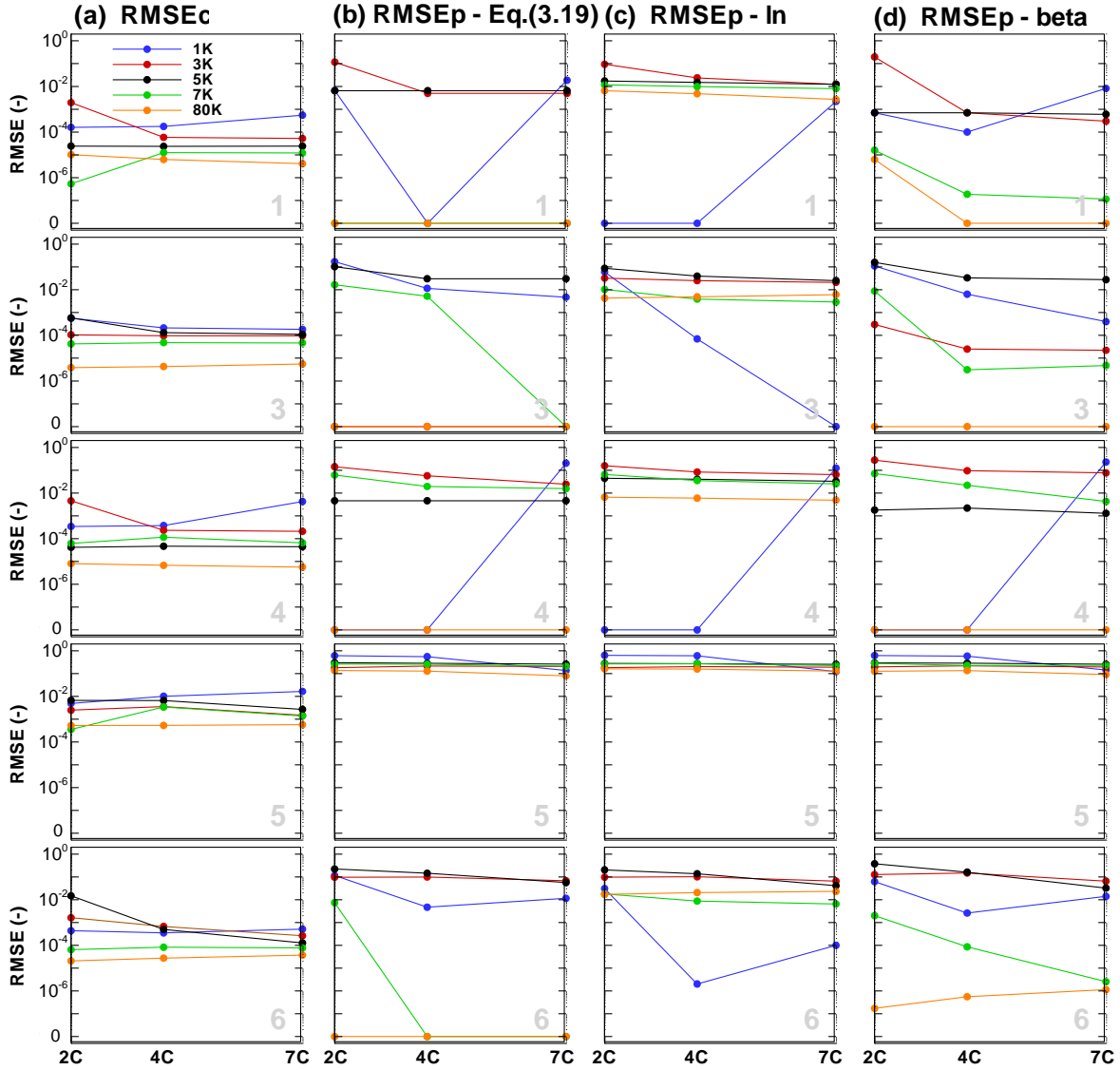
Figure 3.9: a) The Root Mean Square Error of the estimation of indoor air concentration of TCE $RMSE_{c_j}$ for different number of permeability and concentration data; b) The Root Mean Square Error of the estimation of probability of failure $RMSE_{p_j}$ for different number of permeability and concentration data using definition by Eq. (3.19); c) $RMSE_{p_j}$ using log-normal distribution; d) $RMSE_{p_j}$ using beta distribution.

Note: Number 1, 3, 4, 5 and 6 represent house No. 1, 3, 4, 5, and 6. In figures b), c) and d), if $RMSE_{p_j}$ value is not shown, it implies that $RMSE_{p_j} = 0$.
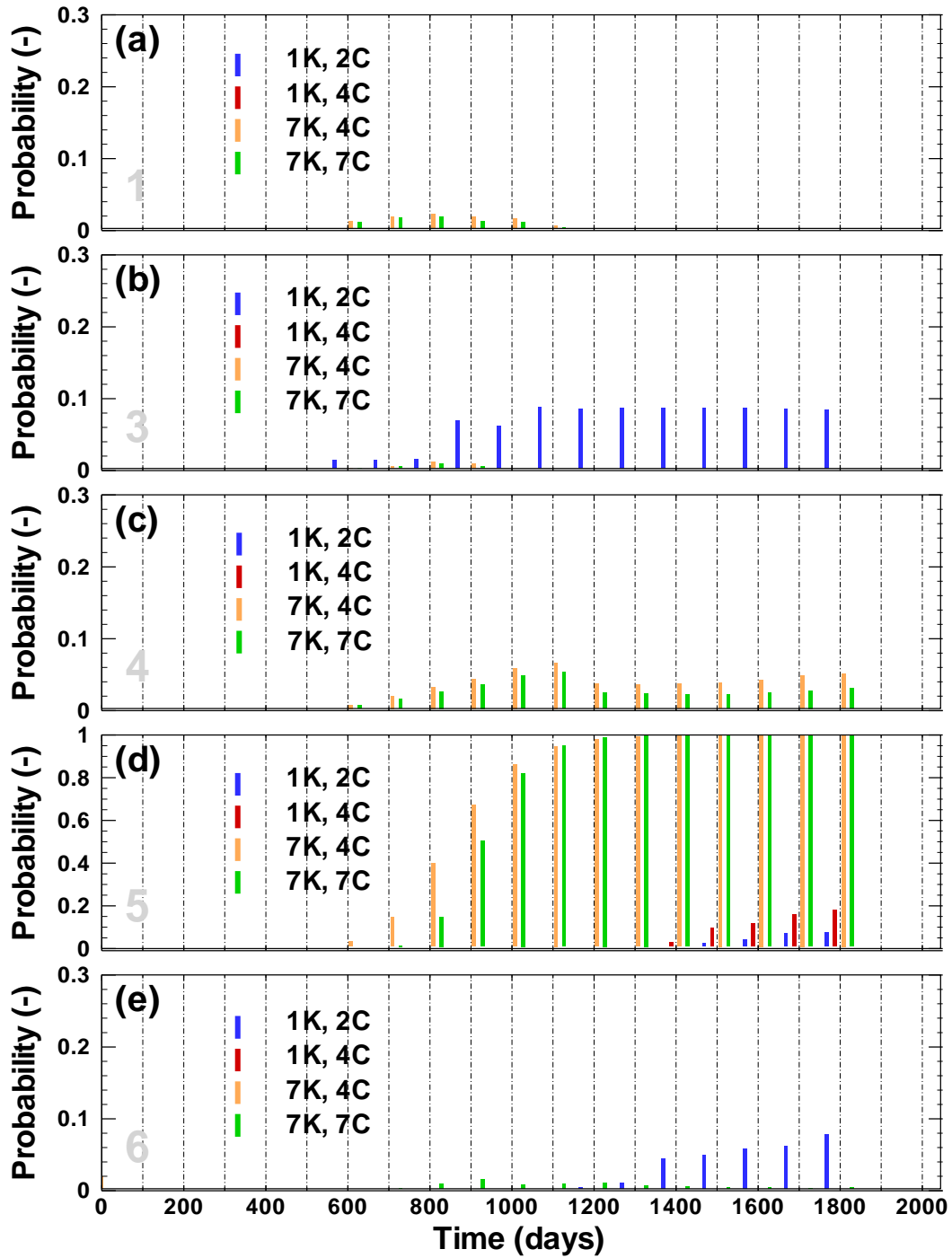
Figure 3.10: The probability of exceedence from the Monte Carlo realizations fitting a log-normal distribution for the six houses for the cases 1)1K, 2C; 2) 1K, 4C; 3) 7K, 4C; and 4) 7K, 7C: a) House No. 1, b) House No. 3, c) House No. 4, d) House No. 5, and e) House No. 6. Note: d) has different scale in *y*- axis from the other figures.
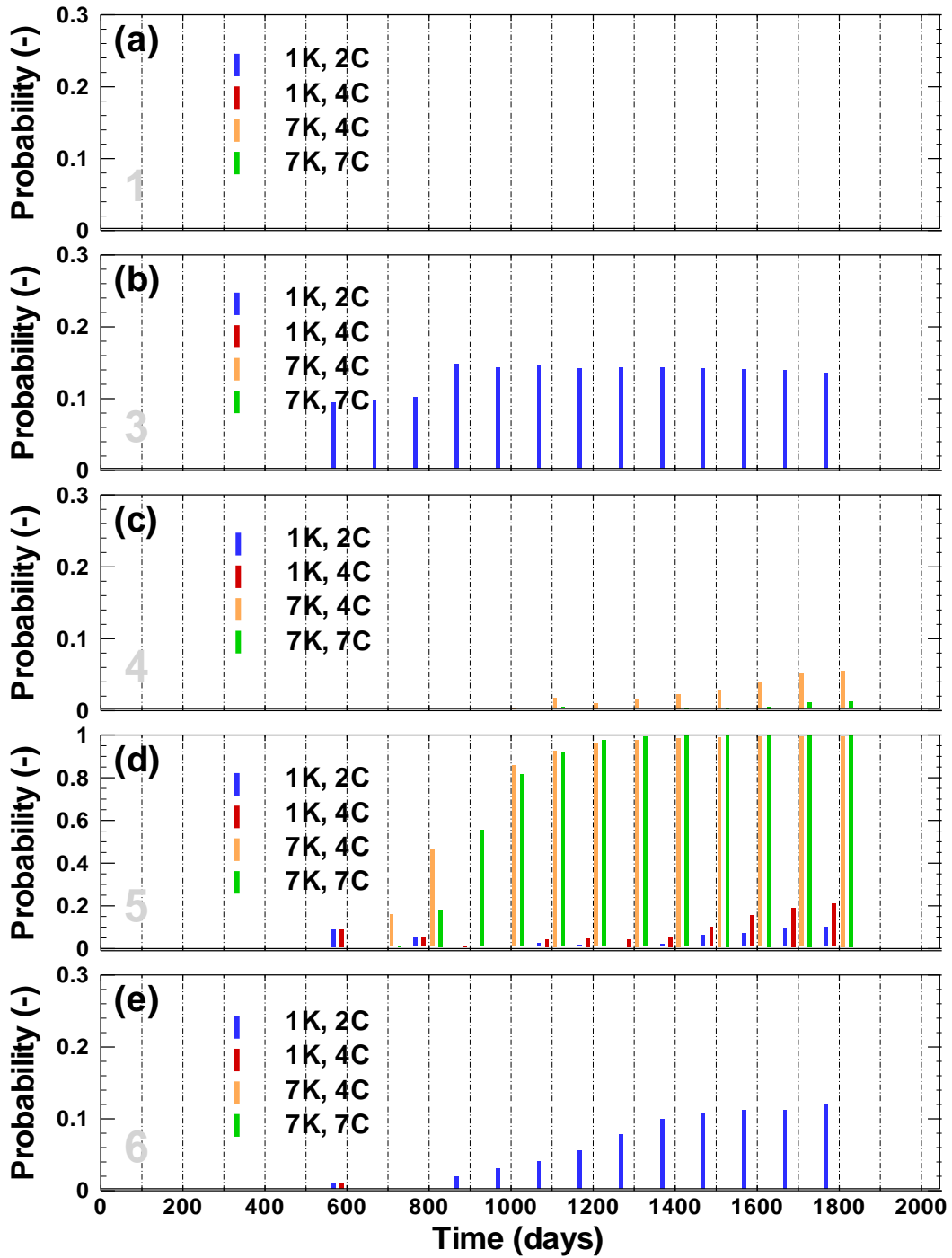
Figure 3.11: The probability of exceedence from the Monte Carlo realizations fitting a beta distribution for the six houses for the cases 1)1K, 2C; 2) 1K, 4C; 3) 7K, 4C; and 4) 7K, 7C: a) House No. 1, b) House No. 3, c) House No. 4, d) House No. 5, and e) House No. 6. Note: d) has different scale in *y*- axis from the other figures.

Table 3.1: Locations where the permeability and concentration data are obtained for different scenarios

Locations	Scenarios														
	1K 2C	1K 4C	1K 7C	3K 2C	3K 4C	3K 7C	5K 2C	5K 4C	5K 7C	7K 2C	7K 4C	7K 7C	80 K2 C	80 K4 C	80 K7 C
House 1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
House 2	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
House 3	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
House 4	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
House 5	–	–	C	–	–	C	–	–	C	–	–	C	–	–	C
House 6	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
BH 1	–	–	–	–	–	–	–	–	–	K	K	K	K	K	K
BH 2	–	–	C	–	–	C	K	K	K, C	K	K	K, C	K	K	K, C
BH 3	–	C	C	K	K, C	K, C	K	K, C	K, C	K	K, C	K, C	K	K, C	K, C
BH 4	K, C	K, C	K,C	K, C	K, C	K, C	K, C	K, C	K, C	K, C	K, C	K, C	K, C	K, C	K, C
BH 5	–	C	C	K	K, C	K, C	K	K, C	K, C	K	K, C	K, C	K	K, C	K, C
BH 6	–	–	C	–	–	C	K	K	K, C	K	K	K, C	K	K	K, C
BH 7	–	–	–	–	–	–	–	–	–	K	K	K	K	K	K
⋮	–	–	–	–	–	–	–	–	–	–	–	–	⋮	⋮	⋮
BH 80	–	–	–	–	–	–	–	–	–	–	–	–	K	K	K

  –   denotes no sample data at the location

  K   denotes permeability measurements obtained from boreholes (BH)

  C   denotes soil gas concentration sample obtained from either boreholes or residential houses

Table 3.2: The statistics of the kriging variance inferred from different boreholes of permeability data

	1K	3K	5K	7K	80K
mean($\sigma_k^2(x_0)$)	1.2727	0.8929	0.8878	0.8418	0.8264
mean($\sigma_{k\,z}^2$)	1.0127	0.7434	0.7140	0.7509	0.8439
max($\sigma_k^2(x_0)$)	1.4265	0.9	0.8959	0.8543	0.9300
min($\sigma_k^2(x_0)$)	9.9984e-5	1.0000e-4	1.0000e-4	1.0000e-4	1.0000e-4
95% CI	(1.2090, 1.3364)	(0.8267, 0.9591)	(0.7947, 0.9809)	(0.7438, 0.9398)	(0.4809, 1.1719)

Note: $\sigma_k^2(x_0)$ represents the kriging variance; $\sigma_{k\,z}^2$ represents the variance of samplings; CI represents confidence interval.

Table 3.3: The differences in the absolute error of soil gas and indoor air concentration of TCE between the mean and the real value at 5% confidence level in each 100 days interval $(t_{i-1}, t_i]$ till 1800 days after TCE injection: 'T' – hypothesis is true; '0' – equality; 'F' – false; '-' – normal distribution does not apply

Time (days)	$\begin{array}{c}\lvert\mu_{1K2C}^+ - \hat{c}_{reality}\rvert >\\ \lvert\mu_{1K4C}^+ - \hat{c}_{reality}\rvert\\ \text{SG/IA}\end{array}$					$\begin{array}{c}\lvert\mu_{1K4C}^+ - \hat{c}_{reality}\rvert >\\ \lvert\mu_{7K4C}^+ - \hat{c}_{reality}\rvert\\ \text{SG/IA}\end{array}$					$\begin{array}{c}\lvert\mu_{7K4C}^+ - \hat{c}_{reality}\rvert >\\ \lvert\mu_{7K7C}^+ - \hat{c}_{reality}\rvert\\ \text{SG/IA}\end{array}$				
	$m \in$ House No.					$m \in$ House No.					$m \in$ House No.				
$(t_{i-1}, t_i]$	1	3	4	5	6	1	3	4	5	6	1	3	4	5	6
(0,100]	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-
(100,200]	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-
(200,300]	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-
(300,400]	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	0/0	-/-
(400,500]	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	T/0	-/-
(500,600]	-/-	T/T	-/-	-/-	-/-	-/-	-/	-/-	-/-	-/-	-/-	-/-	-/-	T/0	-/-
(600,700]	F/0	T/T	-/-	0/0	-/-	-/-	-/-	-/-	T/T	-/-	-/-	-/-	-/-	T/0	-/-
(700,800]	F/0	0/F	F/0	0/0	-/-	T/F	-/-	F/F	T/T	-/-	0/0	-/-	0/0	T/T	-/-
(800,900]	0/0	T/F	T/0	0/0	-/-	T/T	0/T	F/F	T/T	-/-	0/0	0/0	0/0	T/T	-/-
(900,1000]	T/0	T/F	T/0	0/0	-/-	0/T	0/T	F/T	T/T	-/-	0/0	0/0	0/0	T/T	0/0
(1000,1100]	T/0	T/F	T/0	0/0	-/-	0/T	T/T	F/T	T/T	-/-	T/0	0/0	0/0	T/T	0/0
(1100,1200]	T/0	T/F	T/0	0/0	-/-	0/T	T/T	0/T	T/T	-/-	T/0	0/0	T/0	T/T	0/0
(1200,1300]	T/0	T/F	T/0	0/0	F/F	F/T	T/T	0/T	T/T	T/T	T/0	0/0	T/0	T/T	0/0
(1300,1400]	T/0	T/F	T/0	0/0	F/F	F/T	T/T	0/T	T/T	T/T	T/0	0/0	T/0	T/T	0/0
(1400,1500]	F/0	T/F	T/0	0/0	F/F	T/T	T/T	F/T	T/T	T/T	T/0	0/0	T/0	T/T	0/0
(1500,1600]	F/0	T/F	T/0	0/0	F/F	T/T	T/T	F/T	T/T	T/T	T/0	0/0	T/0	T/T	0/0
(1600,1700]	F/0	T/F	T/0	0/0	F/F	T/T	T/T	0/T	T/T	T/T	T/0	0/0	T/0	T/T	0/0
(1700,1800]	F/0	T/F	T/0	0/0	F/F	T/T	T/T	0/T	T/T	T/T	T/0	0/0	T/0	T/T	0/0

Table 3.4: The differences in variances of soil gas and indoor air concentration of TCE at 5% confidence level in each 100 days interval $(t_{i-1}, t_i]$ till 1800 days after the TCE injection: 'T' – hypothesis is true; '0' – equality; 'F' – false; '-' – normal distribution does not apply

Time (days) $(t_{i-1}, t_i]$	$\sigma_{1K2C}{}^+ > \sigma_{1K4C}{}^+$ SG/IA $m \in$ House No.					$\sigma_{1K4C}{}^+ > \sigma_{7K4C}{}^+$ SG/IA $m \in$ House No.					$\sigma_{7K4C}{}^+ > \sigma_{7K7C}{}^+$ SG/IA $m \in$ House No.				
	1	3	4	5	6	1	3	4	5	6	1	3	4	5	6
(0,100]	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-
(100,200]	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-
(200,300]	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-
(300,400]	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	T/0	-/-
(400,500]	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	T/T	-/-
(500,600]	-/-	T/T	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	T/T	-/-
(600,700]	0/0	T/T	-/-	0/F	-/-	-/-	-/-	-/-	T/T	-/-	-/-	-/-	-/-	T/T	-/-
(700,800]	0/0	T/T	0/0	0/0	-/-	0/F	-/-	F/F	T/T	-/-	0/0	-/-	0/0	T/T	-/-
(800,900]	T/0	T/T	0/0	0/F	-/-	0/F	F/F	0/F	T/T	-/-	0/0	0/0	0/0	T/T	-/-
(900,1000]	0/0	T/T	0/0	0/0	-/-	0/F	F/F	0/F	T/F	-/-	0/0	0/0	0/0	T/T	0/0
(1000,1100]	T/F	T/T	0/0	0/F	-/-	T/0	0/F	T/F	T/T	-/-	0/0	0/0	0/0	T/T	0/0
(1100,1200]	0/F	T/T	0/0	0/0	-/-	T/0	0/F	T/F	T/T	-/-	0/0	0/0	0/0	T/T	0/0
(1200,1300]	T/F	T/T	0/0	0/F	0/T	T/0	0/T	T/F	T/T	T/F	0/0	0/0	0/0	T/T	0/0
(1300,1400]	T/0	T/T	0/0	0/F	0/T	T/F	0/T	T/F	T/T	T/F	0/0	0/0	0/0	T/T	0/0
(1400,1500]	T/F	T/T	0/0	0/0	0/T	T/T	0/T	T/F	T/T	T/0	0/0	0/0	0/0	T/T	0/0
(1500,1600]	T/F	T/T	0/0	0/0	T/T	T/T	T/T	T/F	T/T	0/T	0/0	0/0	0/0	T/T	0/0
(1600,1700]	T/0	T/T	T/0	0/0	T/T	T/T	T/T	T/F	T/T	0/T	0/0	0/0	0/0	T/T	0/0
(1700,1800]	T/0	T/T	T/0	0/0	T/T	T/T	T/T	T/F	T/T	0/T	0/0	0/0	T/0	T/T	0/0

Chapter 4 Risk-Based Site Characterization for Vapour Intrusion at a Brownfields Site: Pricing the Risk Capital

Chapter Summary

A framework for assessing an optimal sampling design for a brownfields redevelopment project is presented. The cost-benefit-risk analysis is employed as a basis to construct the objective function which is the sum of the cost of risk (also defined as risk capital) and the cost of data in order to facilitate the optimal sampling design decision making. The alternative sampling strategies result from a series of sampling cases used in the previous chapter for progressively reducing the prediction uncertainty which are the product of the numerical model *CompFlow Bio*, geostatistical Kriging and data assimilation tool Kalman filter. Two methods are developed to price the risk accounting for the hydrogeological and financial uncertainties possessed in the brownfields redevelopment project in an actuarial premium calculation principle and a martingale premium calculation principle. The optimal sampling design is then selected based on the objective function showing that drilling 7 boreholes to extract permeability data and taking soil gas samplings in 4 locations give the least total cost. The sensitivity analysis of some influential parameters (the long-term interest rate spread over which engineering companies' shares trade in excess of appreciation in the US housing market α_S and the evaluation methods of the probability of failure) is performed to determine their roles of importance in the risk capital valuation. This framework can be applied to provide guidance for a risk-based environmental project.

4.1 Introduction

Since the 1980's, both North America and Europe have recognized the social, environmental, and economic benefits of fostering the redevelopment of brownfields sites. The USA enacted the Superfund Liability Act to reclaim 1,410 heavily contaminated sites across the nation. However, the complex and uncertain nature of the subsurface hydrogeology and its impact on financial liabilities, benefits and risks inherent in the redevelopment project often serve to stifle any redevelopment plans. Yu et al. (2012) identify two hurdles that face the developer contemplating the purchase, remediation, and redevelopment of the brownfields land. First, the cost of redevelopment plus the initial purchase price may be greater than the value of its end use. Second, the possibility that people occupying the redevelopment site may suffer adverse health impacts from lingering subsurface contamination could

expose the developer to punitive damages of unlimited value. This latter issue is the focus of Yu et al. (2009) and Wang et al. (2012) who explore the impact of multiple hydrogeological factors on the fate and transport of TCE from a DNAPL source zone located below the water table, with dissolution into the groundwater and vaporization into the indoor air of a residential dwelling located above the groundwater plume. Health impacts are assumed to arise should the indoor air concentration exceed a regulatory limit (OEHHA-CEPA, 2007; NYSDOH, 2005).

A typical brownfields project undergoes numerous stages before development is complete (Meyer et al., 1997; McCarthy, 2001; aboutRemediation[TM], 2012). For the purposes of this work, we aggregate them into five stages as depicted in Figure 4.1. Stage 1 involves site purchase at time t_0. Stage 2 follows with environmental site assessment, evaluation and planning at time t_I. Stage 3 includes site remediation and additional risk assessment at time t_{II}. Stage 4 begins the process of construction which we assume involves a residential development, and the sale of the houses, at time t_{III}. Stage 5 operates until the termination of the project and involves site operation and monitoring until time $t_{III} = T$. The worth of the project to the developer can be valued by discounting all cash flows arising from the above five stages to a present value (Pratt and Grabowski, 2008), and is expressed as:

$$PV(\mathcal{C}_{total}) = PV(\mathcal{C}_0) + PV(\mathcal{C}_{I}) + PV(\mathcal{C}_{II}) + PV(\mathcal{C}_{III}) + PV(\mathcal{C}_{IV}) \qquad (4.1)$$

where PV denotes present value, $\mathcal{C}_{total}$ is the total cost of the project, and $\mathcal{C}_0$, $\mathcal{C}_{I}$, $\mathcal{C}_{II}$, $\mathcal{C}_{III}$ and $\mathcal{C}_{IV}$ are the costs of each of the five stages.

Following Yu et al. (2012), this chapter focuses on Stage 5 otherwise known as the guarantee period. This stage arises because hydrogeological site complexities often prevent complete remediation of legacy DNAPL source zones in Stage 3, providing the potential for future degradation of the indoor air quality within the residential houses within the development. Our interpretation of the guarantee is as follows. To entice buyers, the developer guarantees that they will immediately repurchase any affected house at the original sale price appreciated by the US national home price index. Indoor air concentrations are monitored every $\Delta t_{monitor}$ days, and the guarantee lasts from time t_{III} until time T. Thereafter, the government indemnifies the developer (OME, 1997; OMMAH, 2007). In return for the repurchase agreement and continuous monitoring, the residents agree not to

blame any reasonable foreseeable health issues arising from potential long-term exposure of indoor air contamination on the developer given their due diligence. To maintain the development, the developer then remediates the subsurface to resolve the source contamination, and demolishes/rebuilds/renovates the house to prevent further intrusion of soil gas. The house is then resold. While this description of the guarantee period is identical to that of Yu et al. (2012), central questions regarding its implementation by the developer in the context of this study are substantially different and are outlined below.

Wang et al. (2012a) explore the impact of various hydrogeological factors on defining an exclusion zone for indoor air contamination around the lateral edges of a groundwater plume. This same problem geometry is directly amenable to exploring questions related to the guarantee period in this study. Specifically, we assume the developer has already repurchased the two houses located directly over the plume (see houses 2 and 5 in Figure 4.1 of Wang et al. (2012b)) as part of their due diligence and is faced with dilemma of assessing their liability with regards to the houses adjacent (see houses 1, 3, 4 and 6 in Figure 1 of Wang et al. (2012b)) to those directly over the plume. The dilemma arises from the fact that although they have liability should the indoor air concentrations in these houses exceed the regulatory limit, they do not necessarily have access to monitor the indoor air quality without purchasing the houses a priori. In part 1 of this two-chapter series (Chapter 3), we explore the worth of collecting permeability and soil gas concentration data for reducing the prediction uncertainty of whether the TCE concentration within the indoor air of the laterally offset houses will exceed a regulatory limit. Without direct access, we presume the developer indicates their due diligence by using the data to simultaneously reduce prediction uncertainty while minimizing costs associated with the guarantee period and in effect promoting sustainable development. A specific objective of part 1 was to quantify predication uncertainty using three metrics: first, the ability to correctly estimate the actual indoor air concentrations at unmonitored locations (i.e. within houses 1, 3, 4 and 6) at any point in time; second, the ability to use available soil gas concentration data to reduce the standard deviation in indoor air concentration at unmonitored locations; and third, to accurately forecast the probability that indoor air concentrations will exceed a regulatory limit at unmonitored locations. The third metric is in fact the probability of failure.

Costs associated with the guarantee period C_{IV} are assumed to arise from two categories; the cost of the monitoring data C_{data} , and the cost of failure arising from the need to repurchase and eventually resell an affected house $C_{failure}$. This is expressed as:

$$PV(\mathcal{C}_{\text{IV}}) = PV(\mathcal{C}_{failure}) + PV(\mathcal{C}_{data}) \ . \tag{4.2}$$

Yu et al. (2012) focus entirely on the cost of failure term as the central question facing the developer. Specifically, they state that the cost of failure is essentially a contingency fee that the developer must set aside from the sale of each house to cover the "risk" that they may need to repurchase it and maintain the development at a future date. Alternatively, the contingency fee could be viewed as the value of an insurance policy with the "risk" transfer to an insurance company following from the upfront payment of the premium. Yu et al. (2012) further clarify their definition of "risk" as "risk capital", which includes a surcharge in excess of the expected present value of the probability of failure times the cost of failure. This surcharge represents the value that a rational and risk adverse person places on investing in the brownfields project due to their anticipated loss of investment principal. The objective of part 2 of this two-chapter series (Chapter 4) is to expand upon the work of Yu et al. (2012) to value the risk capital as $PV(\mathcal{C}_{\text{IV}})$. As such, we seek to minimize an objective function based largely upon Equation (4.2) that balances the trade-offs between collecting progressively more data to reduce project uncertainty and hence the cost of failure, but at an ever increasing site characterization cost.

Massmann and Freeze (1987a, b), Massmann et al. (1991) and Freeze et al. (1992) proposed and applied a general cost-benefit-risk framework that is directly amenable to our brownfields redevelopment project in that it combines hydrogeological and economic information into an estimate of the projects value. Barnes and McWhorter (2000) used the same framework for the design of soil vapor extraction systems. Wong and Yeh (2002) developed a cost-benefit uncertainty analysis for solving a contaminated aquifer management problem focusing on the relationship between increased management costs and the desired level of protection. Additional cost-benefit analysis has been used to examine risk assessment strategies for analyzing human health risks due to chemical exposure arising from hydrogeological projects (Reichard and Evans 1989; Pelmulder et al. 1996), and to make remedial action decisions in an optimization framework (Smalley et al. 2000). A central principle in the above (and our) cost-benefit-risk analysis is that all risks with an engineering project can be defined as the product of the probability that the engineered system will fail to meet the intended goal (probability of failure) and the cost of not reaching that goal (James et al. 1996b; Khadam and Kaluarachi 2003).

Yu et al. (2102) indicate that a common element in the Massmann and Freeze (1987a, b), Massmann et al. (1991) and Barnes and McWhorter (2000) cost-benefit-risk framework is the application of the classical economic utility function which takes into account the risk aversion that a decision maker has to accept. The utility function is an abstract measure of risk-averse tendencies capturing the subjective perception of risk based on prevailing social and economic factors. It is difficult to both conceptualize and define. However, empirical studies by Neufville and King (1991) and Laryea and Hughes (2008) indicate that common practice by contractors is to surcharge their bid on a project by a factor of 1%-3% to accommodate construction risk. Balatbat et al. (2010) and Unger et al. (2012) further indicate that this surcharge is then expressed as an interest rate spread over and above the consumer price index, and in fact can be inferred from appreciation of the stock of construction companies trading on the market. A novel contribution of this work is to adapt two risk capital valuation methods from Yu et al. (2012) to unambiguously parameterize and value the risk capital problem at hand. Specifically, we show how the hydrogeological and financial (market) data can be used to inform; a classical P&C insurance valuation involving safety loading (see Method 2 as described in Section 3.2 of Yu et al. (2012)) which we term an "actuarial" approach, and a risk-neutral valuation that is based on implied loss distributions from market prices (see Method 5 as described in Section 3.5 of Yu et al. (2012)) which we term a "financial" approach. By equating and then comparing and contrasting these two risk capital valuation approaches, we relate both hydrogeological and financial uncertainties to the developer's needs to minimize an objective function based largely upon Equation (4.2).

4.2 The Guarantee Period

4.2.1 Probability of Failure

The probability of failure follows directly from Section 3.3, and is reviewed briefly here for completeness. The key issue in this work is our assumption that although the developer has liability for indoor air impacts to houses 1, 3, 4 and 6, the developer does not have access to these properties which is often the case at the real sites. To constrain the prediction uncertainty for indoor impacts on these houses, the developer uses; knowledge of the subsurface permeability structure obtained from the soil cores in combination with TCE soil gas concentration measurements within these same boreholes, as well as TCE soil gas concentration measurements from beneath the foundation slab of house 2. The In Chapter 3, we define the third metric associated with prediction uncertainty, namely;

the ability to accurately forecast the probability of indoor air concentrations within houses 1, 3, 4 and 6 exceeding a regulatory limit. This third metric is directly related to the first and second moments of $\hat{c}_{IA_{i,j}}^{+}$ (see Equation (3.9)) which represent indoor air concentrations conditioned using kriging and the static Kalman filter to assimilate the permeability and soil gas data.

The probability of failure during the guarantee period is calculated in the following two steps. First, we assume that the developer monitors the indoor air quality periodically every $\Delta t_{monitor} = 100$ days for the duration of the guarantee period which lasts for $T - t_{\mathrm{III}} = 1800$ days (5 years). For a given monitoring interval $(t_{i-1}, t_i]$ where $t_i = i \times \Delta t_{monitor}$ with $i = \{1, 2, \dots, n_{times}\}$ and $n_{times} = (T - t_{\mathrm{III}})/\Delta t_{monitor}$, we calculate the probability of exceedence $P_{i,j}$ for the jth location (i.e. one of houses 1, 3, 4 and 6) in one of two methods. First, as the sum of the number of Monte Carlo realizations for which the indoor air concentration first exceeds the regulatory criterion, defined as $n_{c_{IA_{i,j}} > c_{regu}}$, divided by the total number of Monte Carlo realizations $n_{realizations}$:

$$P_{i,j} = \frac{n_{c_{IA_{i,j}} > c_{regu}}}{n_{realizations}} \ . \tag{4.3}$$

The indoor air concentration of TCE in each single house would first increase, reach a plateau and then tend to decline (see Figure 2.6). The probability of failure is independent within each monitoring interval and not based on prior values. Here, it assumes that occurrence of failure in one house can be more than one in different time intervals. In other words, if failure occurs in one time interval it can also occur in another time interval at a later time. Therefore the evaluation of the probability of failure using this method represents its conservative nature.

Second, we fit a log-normal as well as a beta distribution to $c_{IA_{i,j}}^{+}$ in order to better capture the tail of the distribution with the limited number of Monte Carlo simulations. Both the log-normal and beta distributions make use of the first two moments of $\hat{c}_{IA_{i,j}}^{+}$ given by $\mu_{IA_{i,j}}^{+}$ and $\sigma_{IA_{i,j}}^{+}$ in Equation (3.9). All values of $P_{i,j}$ are shown in Figures 3.8, 3.10 and 3.11.

4.2.2 Cost of Failure

The cost of failure is largely adapted and abbreviated from Yu et al. (2012) with modifications to accommodate the optimization problem need to value $PV(C_{IV})$ as expressed by Equation (4.2). The cost itself follows from the developer's guarantee presented in Section 4.1, and its value $C_{failure_i}(t_i)$ at time t_i in the time interval $(t_{i-1}, t_i]$ is:

$$C_{failure_i}(t_i) = \eta \, S_i \tag{4.4}$$

where S_i is the value of one of the lateral houses offset from the groundwater plume (i.e. houses 1, 3, 4 and 6) at time t_i as appreciated by the US national home price index, and η is a scalar $0 < \eta < 1$ and is used to represent the lost revenue arising from the purchase, remediation, demolition, rebuilding and stigma associated with the resale value of the house. The present value of the cost of failure within time interval $(t_{i-1}, t_i]$ is:

$$C_{failure_i}(t_0) = \eta \, S_i \exp\left\{ \int_{t_0}^{t_i} -\rho \, dt \right\} \tag{4.5}$$

where ρ is the discount rate for the brownfields redevelopment project, and will be discussed in greater details in Section 4.2.4.

Following from Yu et al. (2012), the US national home price index S (see Figure 1a of Yu et al., 2012) is assumed to follow a stochastic process defined by Geometric Brownian Motion (GBM) with drift given by:

$$dS = \alpha_S \, S \, dt + \sigma_s \, S \, dZ_s \tag{4.6}$$

where α_S is the annual rate of appreciation in the US housing market, σ_s is the volatility, and dZ_s is a Wiener process. Estimates of α_S and σ_s obtained using the maximum likelihood estimation methodology are provided on Table 4.1. A sample realization of $C_{failure_i}(t_i)$ based on market movements in S_i follows as:

$$C_{failure_i}(t_i) = \eta \, S_0 \exp\left\{\left(\alpha_S - \frac{\sigma_S^2}{2}\right) t_i + \sigma_s Z_s(t_i)\right\} . \tag{4.7}$$

Because $C_{failure_i}(t_i)$ appears frequently in this work, we will simplify its notation to $C_{failure_i}$. Equation (4.7) provides an estimate of $C_{failure_i}$ with the expectation:

$$\mathrm{E}\left[C_{failure_i}\right] = \eta \, S_0 \exp\{\alpha_S \, t_i\} \tag{4.8}$$

and variance:

$$\mathrm{Var}\left[C_{failure_i}\right] = \eta^2 \, S_0^2 \, \exp\{2 \, \alpha_S \, t_i\} \left(\exp\{\sigma_S^2 \, t_i\} - 1\right) . \tag{4.9}$$

4.2.3 Cost of data

Within Equation (4.1), the present value of the cost of data is expressed as $PV(C_{data})$ and is further defined as:

$$C_{data}(t_0) = \sum_{i=1}^{n_{times}} n_C C_C exp\left\{\int_{t_0}^{t_i} r_J \, dt\right\} \exp\{-\rho \, t_i\} + n_K C_K \tag{4.10}$$

95

where $\mathcal{C}_C$ is the cost of obtaining an single concentration sample and $\mathcal{C}_K$ is the cost of obtaining a single permeability measurement, while n_C and n_K denote the number of concentration (i.e. 0C, 2C, 4C and 7C) and permeability data (i.e. 1K, 3K, 5K, 7K and 80K) collected according to the sampling strategies outlined on Table 3.1. The cost of individual concentration and permeability measurements is provided here on Table 4.2. Aller et al. (1989) reviews numerous factors governing the borehole drilling strategy, including; depth, cost, reliability for obtaining samples, availability of drilling equipment, site accessibility, well installation and development time, and minimizing damage to the site and the subsurface. We chose auger drilling to obtain the samples at 10 cm spacing throughout the 8 m aquifer thickness. The main itemized cost for obtaining the permeability data includes; operation of the drill rig, mobilization and demobilization, and sample analysis (Kram et al., 2002; US EPA, 1998). Soil gas concentration sampling strategies are either passive or active. We chose passive soil gas sampling via a buried collector at an overall cost of $250 per sample (Korte, 1992). While all permeability data are collected at time t_0, the concentration data all collected in the time intervals $(t_{i-1}, t_i]$. The cost of all data is appreciated from their collection time to t_i at the inflation rate r_J.

The annual rate of inflation r_J using the Consumer Price Index as a proxy (see Figure 1b of Yu et al. (2012)) is assumed to follow a stochastic process defined by the mean-reverting arithmetic Ornstein-Uhlenbeck model given by:

$$dr_J = \kappa_J \left(\theta_J - r_J\right) dt + \sigma_J \, dZ_J \tag{4.11}$$

where κ_J denotes the speed of adjustment, θ_J is the reversion level, σ_J is the volatility, and dZ_J is a Weiner process. Estimates of κ_J, θ_J, and σ_J obtained using the maximum likelihood estimation method are provided on Table 4.1. The following equation provides an estimate of $\mathcal{C}_{data}$ Equation (4.10) with the expectation:

$$\mathrm{E}[\mathcal{C}_{data}] = \sum_{i=1}^{n_{times}} n_C \mathcal{C}_C \mathrm{E}\left[\exp\left\{\int_{t_0}^{t_i} r_{\mathcal{J}}\, dt\right\}\right] \exp\{-\rho\, t_i\} + n_K \mathcal{C}_K \tag{4.12}$$

Section 3.7 of Yu et al. (2012) provides a discussion of how to evaluate $\mathrm{E}\left[\exp\left\{\int_{t_0}^{t_i} r_{\mathcal{J}}\, dt\right\}\right]$ when $r_{\mathcal{J}}$ is a stochastic process as expressed by Equation (4.11).

4.2.4   The Risk Free and Project Discount Rates

The conventional approach in the engineering literature for estimating the discount rate ρ for cost-benefit-risk projects is to use a range of values to reflect the decision-makers' subjective evaluation of risk. For instance, Massman and Freeze (1987a) use $\rho = 0.05$ to 0.20 per annum with a base case of $\rho = 0.1$ per annum. Schwartz (1994) advocates an approach based on arbitrage or equilibrium in the financial markets in order to estimate the market value of the project. Insley and Wirjanto (2010) further modify this approach with information from the Capital Asset Pricing Model (CAPM) in order to relate ρ to the risk free rate, with their approach then adopted by Yu et al. (2012) for application to a brownfields project. Here, we briefly review Yu et al. (2012) to ensure consistency and clarity of notation in this work.

The risk free rate in the USA follows from movements in the price of US Treasury securities that, by their nature, are risk-free and represent the nominal annual interest rate $r_{\mathcal{N}}$. Typically, $r_{\mathcal{N}}$ is assumed to follow a stochastic process defined by the mean-reverting Cox-Ingersoll-Ross (CIR) model (Cox et al., 1985) given by:

$$dr_{\mathcal{N}} = \kappa_{\mathcal{N}}(\theta_{\mathcal{N}} - r_{\mathcal{N}})\, dt + \sigma_{\mathcal{N}}\, dZ_{\mathcal{N}} \tag{4.13}$$

where $\kappa_{\mathcal{J}}$ denotes the speed of adjustment, $\theta_{\mathcal{J}}$ is the reversion level, $\sigma_{\mathcal{J}}$ is the volatility, and $dZ_{\mathcal{N}}$ is a Weiner process. Duan and Simonato (1999) provide estimates of $\kappa_{\mathcal{N}}$, $\theta_{\mathcal{N}}$, and $\sigma_{\mathcal{N}}$, with values provided on Table 4.1.

Insley and Wirjanto (2010) demonstrate an approach to estimate the discount rate that uses a risk premium consistent with the CAPM. Specifically, the discount rate ρ is given as:

$$\rho = r_{\mathcal{N}} + \hat{q}_s \sigma_S \tag{4.14}$$

where $\hat{q}_s$ is the market price of risk for a contract (i.e. the guarantee) that depends linearly on the stochastic underlying variable S. Finally, $\hat{q}_s$ is calculated as:

$$\hat{q}_s = \frac{(E[\mu_M] - r_{\mathcal{N}})\beta_{SM}}{\sigma_S} \quad . \tag{4.15}$$

where Siegel and Schwartz (2004) estimate the expected return on the market portfolio $E[\mu_M] = 0.12$ per annum, Yu et al. (2012) estimate $\beta_{SM} = 0.0197$, yielding $\hat{q}_s = 0.03160$ and $\rho = 0.0612$ per annum. These results are summarized on Table 4.1.

4.3 Risk Capital Valuation

In the context of the guarantee period for this brownfields problem involving houses 1, 3, 4 and 6, H is the total claim or risk at time t_0 and is calculated as the product of the probability and cost of failure within a given time interval $(t_{i-1}, t_i]$ as:

$$H = \sum_{i=1}^{n_{times}} \sum_{j=1}^{n_{houses}} H_{i,j}(t_0) = \sum_{i=1}^{n_{times}} \sum_{j=1}^{n_{houses}} C_{failure_i} \exp(-\rho t_i) \mathbf{1}_{c_{IA_{i,j}} > c_{regu}} \tag{4.16}$$

where: $n_{houses} \in \{1,3,4,6\}$ so that $n_{houses} = 4$; and $\mathbf{1}_{c_{IA_{i,j}} > c_{regu}}$ is a Bernoulli random variable of the value 1 with the probability of $P_{i,j}$ in time interval $(t_{i-1}, t_i]$ for the jth house when the indoor air concentration $c_{IA_{i,j}}$ exceeds the regulatory limit c_{regu}, and has a value 0 with the probability of $1 - P_{i,j}$. The Bernoulli probability distribution of variable $\mathbf{1}$ is formulated as

$$f(\mathbf{1}, p) = \begin{cases} P_{i,j}, & \text{if } \mathbf{1} = 1 \text{ when } c_{IA_{i,j}} > c_{regu} \\ 1 - P_{i,j}, & \text{if } \mathbf{1} = 0 \text{ otherwise .} \end{cases} \tag{4.17}$$

Yu et al. (2012) indicate that H is not a complete estimate of the risk capital in that it does not cover the developers' preference for risk aversion. Yu et al. (2012) then adapt five methodologies from the actuarial and financial literature to price the risk capital in the guarantee period that directly take consideration of the developer's risk aversion within their formulation. In this study, we further adapt two of these methodologies to account for the worth of hydrogeological data in reducing prediction uncertainty as to whether the indoor air concentrations within houses 1, 3, 4 and 6 exceed the regulatory limit. First, we focus on method 2 as described in Section 3.1 of Yu et al (2012) that follows classical P&C insurance policies and uses the standard deviation of the cost of failure as a safety loading factor (Moller, 2002; Sondermann, 1991). The worth of the hydrogeological data is to reduce the standard deviation in the cost of failure and hence minimize the safety loading factor. Here, we denote this methodology as the "actuarial" premium calculation principle. Second, we focus on method 5 as described in Section 3.5 of Yu et al. (2012) which follows that of empirical studies that seek implied loss distributions from market prices (Torresetti et al., 2007; Hardle and Lopez Cabrera, 2007; Hayek and Ghanem, 2002; Christensen, 2001). Yu et al. (2012) demonstrate that the safety loading factor can be expressed as an interest rate surcharge in excess of the risk free rate. Once again, the worth of the hydrogeological data is to reduce this interest rate surcharge. Here, we denote this methodology as the "financial" premium calculation principle. Below, we present our adaptations of these two methodologies to pricing the risk capital for the guarantee period at hand with the intent of demonstrating our approach for incorporating the worth of hydrogeological data.

4.3.1 The Actuarial Premium Calculation Principle

Within the actuarial literature, one of the non-life insurance premium calculation principles involves charging a premium based on the expected value of the claim enhanced by a safety loading term. In the context of the guarantee period, this principle is described as:

$$V(H) = \sum_{i=1}^{n_{times}} \sum_{j=1}^{n_{houses}} \{E^{\mathbb{P}}[H_{i,j}] + A(H_{i,j})\} \tag{4.18}$$

where: $V(H)$ denotes the risk capital arising from the guarantee period for houses 1, 3, 4 and 6; $E^{\mathbb{P}}$ is the expectation of claim $H_{i,j}$ under the physical measure $\mathbb{P}$; and $A(H_{i,j})$ is the safety loading term.

The expectation of the claim $E^{\mathbb{P}}[H]$ can be calculated as follows:

$$E^{\mathbb{P}}[H] = \sum_{i=1}^{n_{times}} \sum_{j=1}^{n_{houses}} E^{\mathbb{P}}[H_{i,j}]$$

$$= \sum_{i=1}^{n_{times}} \sum_{j=1}^{n_{houses}} E^{\mathbb{P}}[C_{failure_i} \exp\{-\rho t_i\} \mathbf{1}_{c_{IA_{i,j}}>c_{regu}}]$$

$$\tag{4.19}$$

$$= \sum_{i=1}^{n_{times}} \sum_{j=1}^{n_{houses}} E^{\mathbb{P}}[C_{failure_i} \exp\{-\rho t_i\}] \; E^{\mathbb{P}}[\mathbf{1}_{c_{IA_{i,j}}>c_{regu}}]$$

$$= \sum_{i=1}^{n_{times}} \sum_{j=1}^{n_{houses}} \eta \, S_0 \exp(\alpha_S t_i) \exp(-\rho t_i) \, P_{i,j}$$

which is simply the present t_0 value of the probability of failure times the cost of failure. In Chapter 3, we explore the worth of data to reduce prediction uncertainty on the probability of exceedence term $P_{i,j}$ as discussed above in Section 4.2.1. For the problem at hand, we note in Chapter 3 that $P_{i,j} = 0 \; \forall \, i,j$ which may not be true in general. Furthermore, in Figure 3.9 the root mean square error

between the predicted and actual probability of exceedence diminishes as the number of permeability and concentration measurements increases. Therefore, this same progression of data will minimize $E^{\mathbb{P}}[\,H\,]$ for this particular guarantee period.

The risk loading term $A(H_{i,j})$ is evaluated using the standard deviation principle (Møller, 2001), and is adapted here for the guarantee period as:

$$A(H_{i,j}) = a_{(ac)_{i,j}} \sqrt{\text{Var}[H_{i,j}]} \tag{4.20}$$

where $a_{(ac)_{i,j}}$ is a data informed scalar for the safety loading factor under the actuarial principle (denoted here by the subscript "$(ac)_{i,j}$") in the time interval $(t_{i-1}, t_i]$ for the jth house. We begin by calculating $\text{Var}[H_{i,j}]$ as a random sum of independent random variables (Goodman, 1962) as:

$$
\begin{aligned}
\text{Var}[H_{i,j}] &= \text{Var}\left[\mathcal{C}_{failure_i} \exp\{-\rho t_i\}\, \mathbf{1}_{c_{IA_{i,j}} > c_{regu}}\right] \\
&= \left(E\left[\mathcal{C}_{failure_i} \exp\{-\rho t_i\}\right]\right)^2 \text{Var}\left[\mathbf{1}_{c_{IA_{i,j}} > c_{regu}}\right] \\
&\quad + \left(E\left[\mathbf{1}_{c_{IA_{i,j}} > c_{regu}}\right]\right)^2 \text{Var}\left[\mathcal{C}_{failure_i} \exp\{-\rho t_i\}\right] \\
&\quad + \text{Var}\left[\mathcal{C}_{failure_i} \exp\{-\rho t_i\}\right] \text{Var}\left[\mathbf{1}_{c_{IA_{i,j}} > c_{regu}}\right] \\
&= \exp\{-2\rho t_i\}\left\{\text{Var}[\mathcal{C}_{failure_i}]P_{i,j} + E[\mathcal{C}_{failure_i}]^2 P_{i,j}(1 - P_{i,j})\right\}.
\end{aligned} \tag{4.21}
$$

While $\text{Var}[H_{i,j}]$ is informed by the probability of failure in an identical manner to $E^{\mathbb{P}}[\,H\,]$, intuitively we would expect that it would also have some dependence on the first two metrics used to quantify predication uncertainty: first, the ability to correctly estimate the actual indoor air concentrations at unmonitored locations (i.e. within houses 1, 3, 4 and 6) at any point in time; second, the ability to use available soil gas concentration data to reduce the standard deviation in indoor air

concentration at unmonitored locations. Therefore, we introduce these terms into $a_{(ac)_{i,j}}$ which is a scalar for the safety loading factor in the time interval $(t_{i-1}, t_i]$ for the jth house as:

$$a_{(ac)_{i,j}} = \min \left[a_{(ac)h} \min \left(b_{(ac)1} \omega_{1_{i,j}} + b_{(ac)2} \omega_{2_{i,j}}, 1.0 \right), 1.0 \right] \tag{4.22}$$

with;

$$\omega_{1_{i,j}} = \left| \frac{\sigma_{IA_{i,j}}^{+}}{\mu_{IA_{i,j}}^{+}} \right|$$

$$\omega_{2_{i,j}} = \left| \frac{\hat{c}_{IA_{i,j}}^{+} - \hat{c}_{IA,reality_{i,j}}}{\hat{c}_{IA,reality_{i,j}}} \right| . \tag{4.23}$$

Equation (4.22) entirely represents hydrogeological uncertainty as scaled by the parameter $a_{(ac)h}$. Parameterization of the scalars $a_{(ac)_{i,j}}$, $a_{(ac)h}$, $b_{(ac)1}$, and $b_{(ac)2}$ is discussed later in Section 4.3.3. Equation (4.23) indicates that $\omega_{1_{i,j}}$ is the expression of the coefficient of variation in an absolute form, which is a measure of variability associated with uncertainty and is dimensionless and independent of scale. In statistics, it is a normalized measure of dispersion of a probability distribution and is known as the relative standard deviation. Zhang et al. (2005) used the coefficient of variation as a criterion to select target sampling locations. The parameter $\omega_{2_{i,j}}$ is a measure of relative accuracy and it is defined as a ratio of the absolute error of a measurement to the accepted value of the measurement. In combination, $\omega_{1_{i,j}}$ and $\omega_{2_{i,j}}$ address the first two prediction uncertainty metrics, and are illustrated in Figure 4.2 for each house during the guarantee period using the progression of permeability and soil gas concentration sampling strategies itemized as 1K2C, 1K4C, 7K4C and 7K7C.

Finally, we complete this section with a closed-form expression of the risk capital during the guarantee period as:

$$V(H) =$$

$$\sum_{i=1}^{n_{times}} \sum_{j=1}^{n_{houses}} \left(\underbrace{\eta S_0 \exp\{\alpha_S t_i\} \exp\{-\rho t_i\} P_{i,j}}_{E^{\mathbb{P}}[H_{i,j}]} \right. \tag{4.24}$$

$$\left. + \underbrace{a_{(ac)_{i,j}} \sqrt{\exp\{-\rho t_i\} \left[\eta^2 S_0^2 \exp\{2\alpha_S t_i\} [\exp\{\sigma_S^2 t_i\} - 1] P_{i,j} + \eta^2 S_0^2 \exp\{2\alpha_S t_i\} P_{j,i}(1 - P_{i,j})\right]}}_{A(H_{i,j})} \right) .$$

For projects with a non-zero $P_{i,j}$ in reality, Equation (4.24) shows that the as progressively more hydrogeological data are collected then $\omega_{1_{i,j}}$ and $\omega_{2_{i,j}}$ in $a_{(ac)_{i,j}}$ are driven to "zero" leaving only the financial risk in the housing market to surcharge the expected cost of failure via the project discount rate ρ as shown in Equation (4.14).

4.3.2 The Martingale Premium Calculation Principle in a Financial Market

In the previous section, the premium for a risky asset is defined as the expectation of the total claim amount to be paid in a given time interval surcharged by a safety loading factor. In the financial markets, observed prices for insurance premiums on risky assets are never equal to the mathematical expectations of the underlying assets under the $\mathbb{P}$ measure because they do not take into account the risk averseness of investors. Sondermann (1991) introduced a method for evaluating the insurance premium in an arbitrage free market by taking a risk-neutral probability distribution under the $\mathbb{Q}$ measure. Delbaen and Haezendonck (1989) applied this principle to a process with a compound Poisson probability distribution. We will adopt this principle for the risk capital valuation of our risk-based brownfields redevelopment project under the risk-neutral $\mathbb{Q}$ measure. For clarity and completeness of notation, we review method 5 of Yu et al. (2012) with application to pricing the risk capital for this guarantee period.

To begin with, we no longer focus on the probability of failure of each individual house, but rather expand it to the number of occurrences of failure within a residential area using a compound Poisson distribution. Suppose that:

$${N(t_i), t_i \geq 0} \sim Poisson(\lambda_{i,j}) \tag{4.25}$$

which is a counting process using $N(t_i)$ a random variable in time interval $(t_{i-1}, t_i]$ with a Poisson distribution. The probability of obtaining $k_{i,j}$ occurrences of failure in the time interval $(t_{i-1}, t_i]$ for the jth house (which can alternatively be expressed as the number of houses $k_{i,j}$ being affected by the TCE gas concentration in exceedence of the regulatory limit) can be formulated as:

$$P(N(t_i) = k_{i,j}) = f(k_{i,j}, \lambda_{i,j}\Delta t_i) = \frac{(\lambda_{i,j}\Delta t_i)^{k_{i,j}}}{k_{i,j}!} e^{-\lambda_{i,j}\Delta t_i} \tag{4.26}$$

where: $\lambda_{i,j}$ is the rate of the Poisson process and is equal to the expected number of occurrences during the given time interval $(t_{i-1}, t_i]$ for the jth house,

$$\lambda_{i,j} = \frac{n_{c_{IA_{i,j}} > c_{regu}}}{\Delta t_i} \ ; \tag{4.27}$$

and, $n_{c_{IA_{i,j}} > c_{regu}}$ is the number of realizations where the indoor air concentration of TCE exceeds the regulatory limit over the time period $(t_{i-1}, t_i]$. The choice of $\lambda_{i,j}$ is made to ensure that

$$\frac{\lambda_{i,j}\Delta t_i}{n_{realizations}} = P_{i,j} \ . \tag{4.28}$$

The expected value and the variance of variable $k_{i,j}$ exhibit as:

$$E[k_{i,j}] = \text{Var}[k_{i,j}] = \lambda_{i,j}\Delta t_i \quad . \tag{4.29}$$

The present value of cost of failure within the time interval $(t_{i-1}, t_i]$ for the jth house can be derived from Equation (4.16) as;

$$H_{i,j}(t_0) = \exp\{-\rho t_i\} \sum_{z=1}^{k_{i,j}} (C_{failure_i})_j^z \tag{4.30}$$

and implies that the cost of failure for any single house within the development is the same. The expected cost of failure can then be evaluated under the physical $\mathbb{P}$ measure as;

$$E^{\mathbb{P}}[H_{i,j}] = \exp\{-\rho t_i\} \, E^{\mathbb{P}}\left[\sum_{z=1}^{k_{i,j}} (C_{failure_i})_j^z\right] = \exp\{-\rho t_i\} \, E^{\mathbb{P}}[k_{i,j}] \, E^{\mathbb{P}}[(C_{failure_i})_j^1] \tag{4.31}$$

which now involves the loss on a "unit" house $(C_{failure_i})_j^1$ during the time interval $(t_{i-1}, t_i]$. In terms of the problem at hand, the unit house implies that each of houses 1, 3, 4 and 6 has the same value and that we do not attempt to distinguish between them in terms of their contribution to the "cost" of failure.

The probability distribution defining the cost of failure under the $\mathbb{P}$ measure can be converted into a probability distribution under the $\mathbb{Q}$ measure, which is also a compound Poisson process. Furthermore, these distributions are progressively equivalent. Under $\mathbb{Q}$, the price process becomes a Martingale. The probability under $\mathbb{Q}$ tends to give more weight to less favourable events in a risk-averse environment. The expected cost of failure under the $\mathbb{Q}$ measure can be defined as;

$$\mathrm{E}^{\mathbb{Q}}[H_{i,j}] = \exp(-r_{\mathcal{N}}t_i)\,\mathrm{E}^{\mathbb{Q}}\left[\sum_{z=1}^{k_{i,j}}(\mathcal{C}_{failure_i})_j^z\right] = \exp(-r_{\mathcal{N}}t_i)\,\mathrm{E}^{\mathbb{Q}}[k_{i,j}]\,\mathrm{E}^{\mathbb{Q}}[(\mathcal{C}_{failure_i})_j^1] \qquad (4.32)$$

where the discount rate ρ in Equation (4.31) is replace with the risk-free rate $r_{\mathcal{N}}$. Having established the notation for $\mathrm{E}^{\mathbb{Q}}[H_{i,j}]$ for the guarantee period at hand, we are now in a position to summarize how Equations (28-33) of Yu et al. (2012) contribute to pricing the risk capital. At issue is the need to relate $\mathrm{E}^{\mathbb{Q}}[k_{i,j}]\,\mathrm{E}^{\mathbb{Q}}[(\mathcal{C}_{failure_i})_j^1]$ in Equation (4.32) to their same expectations under the physical $\mathbb{P}$ measure which can then be evaluated using the available hydrogeological data. The transformation is accomplished using Proposition 2.1 of Delbaen and Haezendonck (1989) combined with Equation (33) of Yu et al. (2012). The premise of this transformation is that the $\mathbb{Q}$ measure achieves its risk loading via an interest rate surcharge in excess of α_s, which is the expected appreciation rate of the brownfields project arising from the sale of the houses. This interest rate surcharge varies with time interval $(t_{i-1}, t_i]$ as well as for houses $j \in \{1,3,4,6\}$ and is expressed as;

$$a_{(f)i,j} = \underbrace{\min\left[a_{(f)h}\,\min\left(b_{(f)1}\omega_{1_{i,j}} + b_{(f)2}\omega_{2_{i,j}}, 1.0\right), 1.0\right]}_{hydrogeological\ risk} - \underbrace{\hat{q}_S\sigma_S}_{\substack{financial\\risk}}. \qquad (4.33)$$

Finally, the total value of the risk capital can be expressed as;

$$\begin{aligned}
V(H) &= \sum_{i=1}^{n_{times}}\sum_{j=1}^{n_{houses}} \exp\{-r_{\mathcal{N}}t_i\}\,\mathrm{E}^{\mathbb{P}}[k_{j,i}]\,\mathrm{E}^{\mathbb{P}}[\mathcal{C}_{failure_i}]\exp\{a_{(f)i,j}t_i\} \\
&= \sum_{i=1}^{n_{times}}\sum_{j=1}^{n_{houses}} S_0 \exp\{(\alpha_S + a_{(f)i,j} - r_{\mathcal{N}})t_i\}\lambda_{i,j}\Delta t_i.
\end{aligned} \qquad (4.34)$$

The estimation of parameters $a_{(f)h}$, $b_{(f)1}$ and $b_{(f)2}$ will be discussed in the next section. Once again we note that for projects with a non-zero $P_{i,j}$ in reality, Equations (4.33-4.34) show that as progressively more hydrogeological data are collected then $\omega_{1_{i,j}}$ and $\omega_{2_{i,j}}$ in $a_{(f)_{i,j}}$ are driven to "zero" leaving only the financial risk in the housing market $\hat{q}_S \sigma_S$ to surcharge the risk capital. This financial risk term can be combined with $r_{\mathcal{N}}$ to yield the project discount rate ρ as shown in Equation (4.14).

4.3.3 Optimization and Parameter Estimation

The optimization and parameterization problem that embodies this work is best expressed by Equation (4.2), and involves our attempt to find the least cost strategy for the guarantee period. The premise here is that we can assess the financially sustainable market rate for companies specializing in this type of activity by observing the interest rate spread that shares in these companies trade at in excess of appreciation in the US housing market α_S. Insurance policies for the guarantee period are likely to be quoted in an over-the-counter market (as with insurance policies in general), and cannot be directly observed. In Chapter 3, we alluded to the idea that 7K4C and 7K7C appeared to be optimal sampling strategies in that they provided the "best" estimates of the three statistical metrics quantifying the worth of the hydrogeological data with a reasonable (i.e. not excessive as with 80K) amount of data. In this section, we build upon this idea and demonstrate that the interest rate spread can be used to price the worth of the hydrogeological data in terms of minimizing the risk capital for the guarantee period.

We begin by recasting $PV(\mathcal{C}_{IV})$ in Equation (4.2) as the value of an objective function Obj, and then $PV(\mathcal{C}_{failure})$ using $V(H)$ from either Equation (4.24) or (4.34). We now state the objective function as;

$$Obj = V(H) + \mathrm{E}[\mathcal{C}_{data}] \quad \text{subject to} \quad 0 \leq Obj \leq S_0\, n_{house}\,(1-\eta) \tag{4.35}$$

where $n_{houses} \in \{1,3,4,6\}$, $S_0 = \$200{,}000$, and $\eta = 0.2$. Therefore, the expected value of the risk capital cannot exceed \$640,000 otherwise the developer should never attempt to resell any house

should it become impacted by indoor air quality issues. The following optimization and parameterization effort is built upon the idea that the minimum value of the objective function is located within the vicinity of the 7K4C and 7K7C sampling strategies. Clearly, the objective function seeks to balance the worth of the hydrogeological data in terms of minimizing $V(H)$, and the expense of ever increasing site characterization costs valued by $\mathrm{E}[\mathcal{C}_{data}]$.

The first step in the procedure is to parameterize four unknown values of $b_{(ac)1}$, $b_{(f)1}$, $b_{(ac)2}$ and $b_{(f)2}$ using the 7K4C sampling case strategy. The idea is to balance the contributions of the first two statistical metrics that quantify prediction uncertainty, measured by $\omega_{1_{i,j}}^{7K4C}$ and $\omega_{2_{i,j}}^{7K4C}$, towards the risk capital. We achieve this balance by simultaneously solving the following fours equations for the above four unknowns;

$$b_{(ac)1} \frac{1}{n_{houses}} \sum_{j=1}^{n_{houses}} \left(\frac{1}{n_{times}} \sum_{i=1}^{n_{times}} \omega_{1_{i,j}}^{7K4C} \right) = b_{(ac)2} \frac{1}{n_{houses}} \sum_{j=1}^{n_{houses}} \left(\frac{1}{n_{times}} \sum_{i=1}^{n_{times}} \omega_{2_{i,j}}^{7K4C} \right)$$

$$b_{(f)1} \frac{1}{n_{houses}} \sum_{j=1}^{n_{houses}} \left(\frac{1}{n_{times}} \sum_{i=1}^{n_{times}} \omega_{1_{i,j}}^{7K4C} \right) = b_{(f)2} \frac{1}{n_{houses}} \sum_{j=1}^{n_{houses}} \left(\frac{1}{n_{times}} \sum_{i=1}^{n_{times}} \omega_{2_{i,j}}^{7K4C} \right) \qquad (4.36)$$

$$b_{(ac)1} + b_{(ac)2} = 1$$

$$b_{(f)1} + b_{(f)2} = 1$$

where $n_{houses} \in \{1,3,4,6\}$. Values of $\omega_{1_{i,j}}^{7K4C}$ and $\omega_{2_{i,j}}^{7K4C}$ are provided in Figure 4.2. As a resulting of solving Equation (4.36), we found that the weighting coefficients $b_{(ac)1}$ and $b_{(f)1}$ to be approximately 0.6, and $b_{(ac)2}$, $b_{(f)2}$ to be approximately 0.4. These values are listed on Table 4.1. We repeated the same procedure for the 7K7C sampling strategy (i.e. by using $\omega_{1_{i,j}}^{7K7C}$ and $\omega_{2_{i,j}}^{7K7C}$) and found similar results.

The second step in the procedure is to parameterize $a_{(f)h}$. We begin by observing a long-term average interest rate surcharge $\bar{a}_{(f)}$ (in excess of α_S) in the stock of companies that specialize in trading the risk associated with the guarantee period. We anticipate that this might be difficult given

that it is more likely that companies will be involved in all activities associated with the brownfields development as expressed by Equation (4.1). To proceed, we assume $0 \leq \bar{a}_{(f)} \leq 0.2$ per annum with a base case value of $\bar{a}_{(f)} = 0.1$ per annum given the anecdotal evidence that brownfields project rarely receive funding given the multitude of other "safer" investment opportunities in the financial markets. We now use $\bar{a}_{(f)}$ to parameterize $a_{(f)h}$ for the project at hand assuming the optimal solution lies somewhere between the 7K4C and 7K7C sampling strategies. We formulate this as;

$$\chi^{7K4C}$$

$$= \frac{1}{n_{houses}} \sum_{j=1}^{n_{houses}} \left(\frac{1}{n_{times}} \sum_{i=1}^{n_{times}} \left(\min\left[a_{(f)h} \ \min\left(b_{(f)1} \ \omega_{1_{i,j}}^{7K4C} + b_{(f)2} \ \omega_{2_{i,j}}^{7K4C}, 1.0 \right), 1.0 \right] \right. \right.$$

$$\left. \left. - \hat{q}_S \sigma_S \right) \right)$$

$$\chi^{7K7C} \tag{4.37}$$

$$= \frac{1}{n_{houses}} \sum_{j=1}^{n_{houses}} \left(\frac{1}{n_{times}} \sum_{i=1}^{n_{times}} \left(\min\left[a_{(f)h} \ \min\left(b_{(f)1} \ \omega_{1_{i,j}}^{7K7C} + b_{(f)2} \ \omega_{2_{i,j}}^{7K7C}, 1.0 \right), 1.0 \right] \right. \right.$$

$$\left. \left. - \hat{q}_S \sigma_S \right) \right)$$

$$\frac{\chi^{7K4C} + \chi^{7K7C}}{2} = \bar{a}_{(f)}$$

where $n_{houses} \in \{1,3,4,6\}$. The value of $a_{(f)h}$ on Table 4.1 is based on $\bar{a}_{(f)} = 0.1$ per annum.

The third step in the procedure is to parameterize $a_{(ac)h}$ and simultaneously equate the actuarial and financial methodologies for calculating the risk premium as expressed by Equations (4.24) and (4.34), respectively. We begin by recasting the objective function expressed by Equation (4.35) as;

$$Obj_{(ac)}^{7K4C} + Obj_{(ac)}^{7K7C} = Obj_{(f)}^{7K4C} + Obj_{(f)}^{7K7C} \qquad (4.38)$$

under the assumption that the minimum value lies somewhere between the 7K4C and 7K7C sampling strategies. The value of $a_{(ac)h}$ is found by minimizing Equation (4.38), with a value provided on Table 4.1 when $\bar{a}_{(f)} = 0.1$ per annum.

4.4 Results

The optimal cost of the risk capital is such, that when added to the cost of the hydrogeological data, the least cost for guarantee period is achieved. This concept is directly expressed by Equation (4.35). In Section 4.3.3, we conducted a joint parameter estimation and optimization exercise with the understanding that the least cost for the guarantee period lies somewhere between the 7K4C and 7K7C sampling strategies. This notion was informed by Sections 3.2.3.1 and 3.2.3.2 in which we observed that beyond the 7K4C sampling strategy, there was not a statistically significant reduction in the prediction uncertainty of the first two metrics at the 5% level of significance. In this section, we now price all components of the risk capital and data collection in Equation (4.35). Specifically, this involves evaluating the risk capital using both the actuarial (see Section 4.3.1) and financial (see Section 4.3.2) methodologies, for all data collection strategies. The worth of data becomes apparent by comparing and contrasting the expected cost of the guarantee period for the alternative data collection strategies.

We begin the discussion on the optimal cost of the risk capital by focusing on the financial methodology. In Figure 4.3, we provide values of $a_{(f)_{i,j}}$ in time interval $(t_{i-1}, t_i]$ for the jth house when $\bar{a}_{(f)} = 0.1$ per annum. Values of $a_{(f)_{i,j}}$ are calculated using Equation (4.33) with parameters $b_{(ac)1}, b_{(f)1}, \ b_{(ac)2}, b_{(f)2}$ and $a_{(f)h}$ estimated from the first two steps outlined in Section 4.3.3. The time-varying nature of $a_{(f)_{i,j}}$ denotes the transient risk during the guarantee period, which mimics the shape of $\omega_{1_{i,j}}$ and $\omega_{2_{i,j}}$ for sampling strategies 7K4C and 7K7C as shown in Figure 4.2. Specifically, the interest rate surcharge $a_{(f)_{i,j}}$ quantifying hydrogeological (and financial) risk is minimal at early time before any of the Monte Carlo permeability realization plumes arrive. Then, as the leading edge of the groundwater plume reaches the houses, the risk is greatest because the Kalman filter is least effective at constraining the prediction uncertainty as measured by the first two metrics. This is

largely a consequence of the very low arrival-time soil gas and indoor air concentrations inflating the value of $\sigma_{SG_{i,j}}{}^+$ and $\sigma_{IA_{i,j}}{}^+$ of Equation (3.9). Later, $a_{(f)_{i,j}}$ declines and asymptotically reaches $\bar{a}_{(f)}$. Finally, we note that the contribution of the hydrogeological risk to $a_{(f)_{i,j}}$ far exceeds that of the financial risk when $\bar{a}_{(f)} = 0.1$ per annum.

On Table 4.3, we provide values of the objective function (see Equation (4.35)) with the risk capital $V(H)$ from Equation (4.34), and the probability of failure $P_{i,j}$ defined by Equation (4.3) and as shown in Figure 3.8. The value of $a_{(f)h}$ arising from the use of the 7K4C and 7K7C sampling strategies, as discussed in the second step of the parameter estimation and optimization procedure outlined in Section 4.3.3, was applied to calculate $V(H)$ for all remaining sampling strategies. Surprisingly, the minimum value of the objective function occurred for 1K4C. However, in Section 3.3 we recognized that while the 1K4C provided accurate estimates of $P_{i,j}$, it performed poorly based on the first two performance metrics measuring prediction uncertainty. Therefore, we instead accept 7K4C as the correct minimum. The minimum remained near 7K4C when we adjusted $\bar{a}_{(f)}$ over the interval $0 \leq \bar{a}_{(f)} \leq 0.2$ per annum, and we provide further discussion on this issue later in this section.

Professional engineers, geoscientists, and actuaries involved in managing the guarantee period of the brownfields development project are more likely to envision the time-varying cost of the risk capital $V(H)$ by decomposing it into a contribution from a first and second moment analogous to the actuarial methodology in Equation (4.24). In step 3 of Section 4.3.3, we chose a value of $a_{(ac)h}$ in Equation (4.33) so that the costs of the risk capital between the financial and actuarial methodologies, and for the average of the 7K4C and 7K7C sampling strategies, were identical. This can be observed on Tables 4.3 and 4.4. Note that the cost of the objective function using either the actuarial or financial methodologies to evaluate $V(H)$ exhibits the same trends for the various sampling strategies. Note that 1K4C still remains the minimum. However, we use our judgement to denote 7K4C as the correct minimum.

In the first column of Figure 4.4, we show values of various components of the objective function calculated using the probability of exceedence from Equation (4.3). These components include; $E^{\mathbb{P}}[H]$ which is the expected cost of the risk capital (i.e. first moment), $A[H]$ which is the risk loading term arising from the standard deviation in the risk capital (i.e. second moment), $E[C_{data}]$ which is the expected cost of data, and Obj which is the cost of the objective function from Equation (4.35). The value of the risk capital in Equation (4.35) is; $V(H) = E^{\mathbb{P}}[H] + A[H]$. The solid lines

indicate the cost of each component while the dashed lines indicate the percentage by which they contribute to the total cost of the objective function. The expected costs of failure for both 7K4C and 7K7C are nearly identical, and almost "zero" in keeping with reality for the problem at hand in which the indoor air concentration never exceeds the regulatory limit for houses 1, 3, 4 and 6. The risk loading term for 7K4C is slightly larger than for 7K7C due to the worth of the extra soil gas concentration data in 7K7C for reducing prediction uncertainty. However, these extra soil gas concentration measurements cause the cost of data for 7K7C to exceed that of 7K4C. The optimal sampling strategy 7K4C achieves a least cost by balancing the worth of the cost of hydrogeological data against its value in reducing the risk loading term arising from prediction uncertainty. In contrast, 7K7C places too much emphasis on data collection relative to its actual worth in reducing the cost of the risk loading term. The optimal balance achieved by 7K4C indicates that the distribution of costs for the guarantee period should be: $E^{\mathbb{P}}[H] = 10.09\%$, $A[H] = 10.53\%$, and $E[\mathcal{C}_{data}] = 79.38\%$ for a total cost of \$91883.66. We reiterate that this breakdown of costs was established under the assumption that financial risk $\bar{a}_{(f)} = 0.1$ per annum.

Sampling strategies 1K and 80K provide extreme opposite insights into the merits of Equation (4.24) for evaluating the risk capital. For instance, 1K exhibits erratic behaviour with the cost of $E^{\mathbb{P}}[H]$ and $A[H]$ for 4C being less than 7C. This is irrational and is a consequence of Equation (4.24) being strongly dependent on the probability of failure $P_{i,j}$. This is a due to our earlier assumption in Equation (4.16) that the total claim or risk H is the product of the probability and cost of failure. Therefore, when we have insufficient data as with 1K4C, we may erroneously underestimate $P_{i,j}$ (as the third performance metric quantifying prediction uncertainty) yielding a low value for the total claim or risk and ultimately the risk capital. Given that the cost of data is also at a minimum, the objective function is minimized implying an optimal management strategy. Our attempts to alleviate this problem by introducing the first and second performance metrics for prediction uncertainty (i.e. $\omega_{1_{i,j}}$ and $\omega_{2_{i,j}}$) into $a_{(f)_{i,j}}$ and $a_{(ac)_{i,j}}$ were not entirely successful. This could create a problem for optimization algorithms that do not depend on human judgement. At the opposite end of the spectrum, 80K exhibits perfectly rational behaviour in that the extensive site characterization effort yields a cost for $E^{\mathbb{P}}[H]$ and $A[H]$ which are effectively "zero" in keeping with reality for the problem at hand. However, $E[\mathcal{C}_{data}]$ is exorbitantly large yielding a high total cost and hence sub-optimal management strategy.

In the second and third columns of Figure 4.4, once again we show values of various components of the objective function except using the probability of exceedence $P_{i,j}$ derived by fitting the log-

normal and beta distributions to $c_{IA_{i,j}}{}^+$. This adjustment to $P_{i,j}$ has a direct impact on both $\mathrm{E}^{\mathbb{P}}[\,H\,]$ and $A[\,H\,]$ as shown by Equation (4.24). Focusing specifically on the 7K sampling strategies, we see on Table 4.4 and Figure 4.4 that the beta distribution yields the lowest total cost of the objective function. This is a consequence of the fact that it yields the lowest estimates of $P_{i,j}$ as discussed in Section 3.3. Similarly, the log-normal distribution yields the greatest total cost given higher estimates of $P_{i,j}$. Of particular interest is the observation that the use of the log-normal distribution to calculate $P_{i,j}$ appears to place the most value on $\mathrm{E}[\mathcal{C}_{data}]$ as a percentage of the total costs relative to using Equation (4.3) to calculate $P_{i,j}$. It also inflates the percentage of contribution of $\mathrm{E}^{\mathbb{P}}[\,H\,]$ and $A[\,H\,]$ to total costs. In contrast, the beta distribution has the opposite effect relative to using Equation (4.3) to calculate $P_{i,j}$, although the impact is only slight. This should be of concern to professional engineers, geoscientists and actuaries who will jointly share the wealth created by the brownfields project. As we progressively remove permeability data and transition down to the 5K, 3K and finally 1K sampling strategies, the above observations become less relevant. We simply reiterate our previous statement from Section 3.3 that there is little value in using either the log-normal or beta distributions to extrapolate the probability distribution for extreme events when $\mu_{IA_{i,j}}{}^+$ and $\sigma_{IA_{i,j}}{}^+$ are poorly constrained by insufficient data.

The remaining question that we wish to examine is the sensitivity of the total cost of the guarantee period to the long-term average interest rate surcharge $\bar{a}_{(f)}$ (in excess of α_S) on the stock of companies that specialize in trading the risk associated with the guarantee period. The motivation for this sensitive analysis follows from the fact that we are unaware of the availability of market data to estimate $\bar{a}_{(f)}$ and expect that it may instead come from public disclosure of over-the-counter financial products. We surmise that a successful brownfields project will be one whose progress is least sensitive to potential market fluctuations in $\bar{a}_{(f)}$, as well as being managed at the least cost. We conduct this sensitivity analysis by adjusting $\bar{a}_{(f)}$ over the interval $0 \leq \bar{a}_{(f)} \leq 0.2$ per annum, and re-establishing steps 2 and 3 of the parameter estimation and optimization processes discussed above in Section 4.3.3 for each increment of $\bar{a}_{(f)}$. Each incremental value of $\bar{a}_{(f)}$ yields a new estimate of $a_{(ac)h}$ and ultimately the cost of the safety loading term $A[\,H\,]$. The cost of $\mathrm{E}^{\mathbb{P}}[\,H\,]$ and $\mathrm{E}[\mathcal{C}_{data}]$ in the objective function remain unaffected.

Figure 4.5 provides the cost of $A[\,H\,]$ and the objective function with $0 \leq \bar{a}_{(f)} \leq 0.2$ per annum, for sampling strategies 1K2C, 1K4C, 7K4C and 7K7C. Once again, 1K4C appears optimal although it should be discarded based on the first two performance metrics quantifying prediction uncertainty.

113

For sampling strategies 7K4C and 7K7C, the cost of the safety loading term $A[\,H\,]$ increases monotonically with $\bar{a}_{(f)}$. This occurs because $A[\,H\,]$ is proportional to $a_{(ac)h}$, and $a_{(ac)h}$ increases with $\bar{a}_{(f)}$. Because $A[\,H\,]$ for sampling strategy 7K4C is greater than that for 7K7C, the spread between their safety loading terms also increases with $\bar{a}_{(f)}$. In other words, as the market places a greater premium on risk during the guarantee period by increasing the average interest rate surcharge $\bar{a}_{(f)}$, the contribution of the safety loading term to the total cost of the guarantee period also increases. At some point, the optimal least cost management strategy becomes one that places greater worth on the value of hydrogeological data $\mathrm{E}[\mathcal{C}_{data}]$ as a means to reduce prediction uncertainty and hence the cost of the safety loading term $A[\,H\,]$. This transition whereby 7K7C becomes the optimal least cost management strategy is shown in Figure 4.5 as $\bar{a}_{(f)}$ increases beyond 0.17 per annum when using Equation (4.3) to calculate $P_{i,j}$. It is reassuring to note that the total cost of the objective function for both 7K4C and 7K7C does not increase appreciably with $\bar{a}_{(f)}$ because both sampling strategies already place a significant emphasis on the worth of data as a means of reducing prediction uncertainty. Finally, we note that the above observations are insensitive to whether Equation (4.3), the log-normal, or beta distributions are used to calculate $P_{i,j}$.

4.5 Conclusions

In this work, we develop the methodology to price the risk in a brownfields redevelopment project using the actuarial premium calculation principle and financial principle respectively. The possessed risk mainly results from the hydrogeological and the financial uncertainties. The results indicate that with more hydrogeological measurements through the site characterization process as well as during the monitoring period, part of the loss reserves which is a measure of the level of risk aversion will be reduced. In another word, when the insurers or developers are more certain about the occurrence of failure in the future, the less they need to put aside to compensate against the uncertainties. On the other hand, more data could become excessive to further investigate the project and not be worth it due to the high cost of data. The optimal cost of the project occurs when we obtain sufficient amount of data. In this study, 7K4C gives the least cost and comparable to 7K7C.

The safety loading factors $a_{(ac)_{i,j}}$, $a_{(f)_{i,j}}$ accounted for the insurers' preference of risk aversion and incorporated into both actuarial and financial methodologies are dependent on the first two metrics which are used to quantify predication uncertainty: first, the ability to correctly estimate the actual indoor air concentrations at unmonitored locations at any point in time; second, the ability to use

available soil gas concentration data to reduce the standard deviation in indoor air concentration at unmonitored locations. The principle of the optimization and parameterization is to balance the contributions of the two statistical metrics and also to equate the actuarial and financial methodologies for calculating the risk premium for the 7K4C and 7K7C sampling strategies. Using the financial methodology, we can see that the contribution of the hydrogeological risk to $a_{(f)i,j}$ is far beyond that of the financial risk when the average interest rate surcharge $\bar{a}_{(f)} = 0.1$ per annum. Also, the risk capital and the total cost of this project vary when adjusting the average $\bar{a}_{(f)}$ for the 7K4C and 7K7C cases between 0 and 0.2. We can conclude that when the site information is not sufficient enough to infer the potential risk of the indoor air in the residential buildings exceeding the regulatory limit, with the change of $\bar{a}_{(f)}$, the risk capital and the total cost could change significantly. On the contrary, with sufficient amount of data prior to the construction and during the monitoring period, when the $\bar{a}_{(f)}$ changes within a practical range, the risk capital and the total cost would be expected to stay within an accepted level (7K case). The reason for that is the safety loading factor is only a small fraction of the total risk capital with sufficient datasets. Also, the minimum value of the objective function becomes 7K7C which is the optimal least cost management strategy as $\bar{a}_{(f)}$ increases beyond 0.17 per annum.

The sensitivity of the methods of defining the probability of failure for the calculation of the total cost, by either using Equation (4.3), fitting the log-normal or beta distributions to $c_{IA_{i,j}}{}^{+}$, is discussed. Log-normal distribution or beta distribution can either overestimate or underestimate the risk capital depending on the characteristics of the population. The log-normal distribution can be used for heavy-tailed population to increase the probability of failure which will be in favor of determining the loss reserve in a conservative manner. However, if the data is enough, we will need less money to ensure the uncertainty, and the safety loading factor will contribute less to the risk capital. Since the safety loading takes a small portion of the risk capital compared with the expected value of losses, the risk capital and the corresponding total cost in this case will not differentiate dramatically no matter which method is applied to define the probability of failure.

This chapter shows the importance of site characterization and sampling for a project, and the way to quantify the risk in an economic way. It has to be noted that the methodology presented here can be extended and applied, and other data can be incorporated into the framework for different risk-based engineering projects.
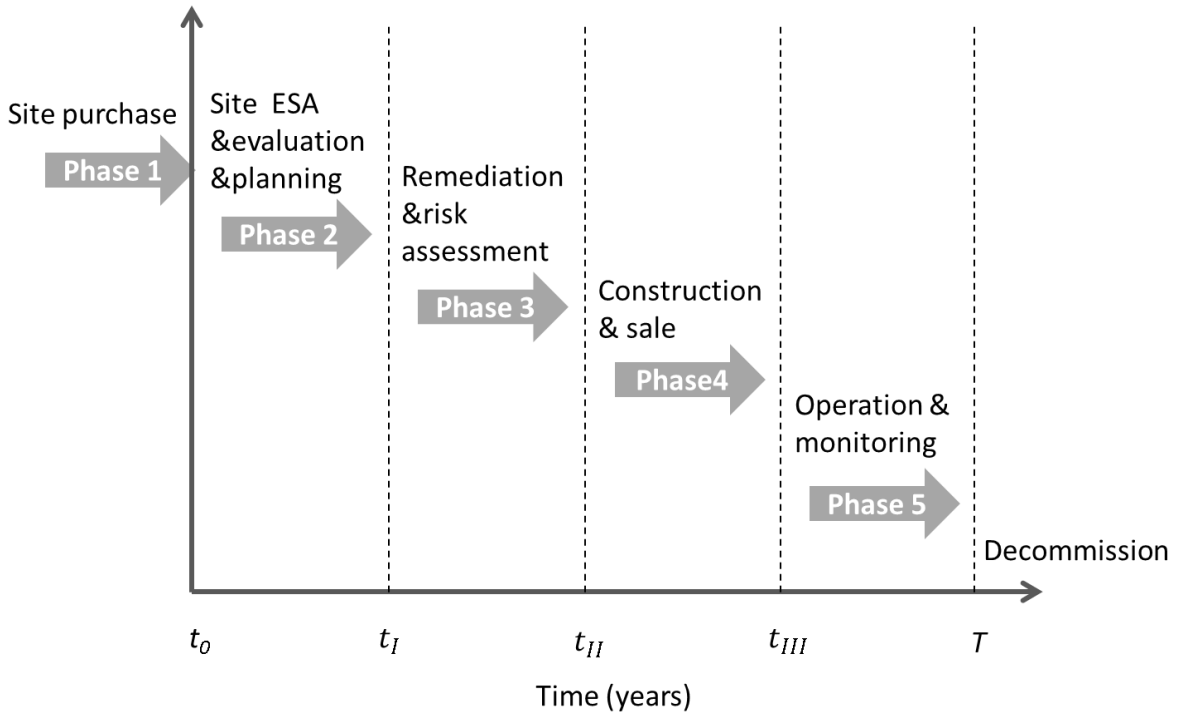
Figure 4.1: The five stages of a brownfields redevelopment project.
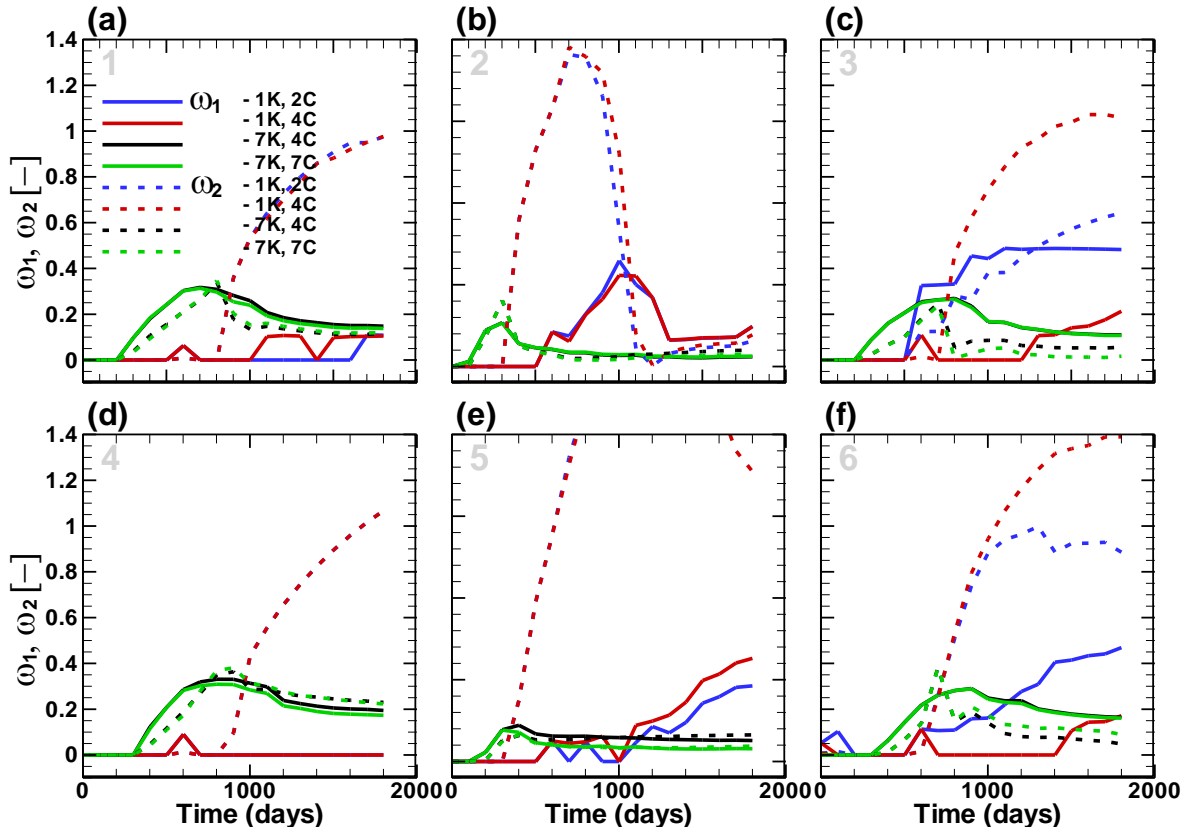
Figure 4.2: Transient behaviour of ω_1 and ω_2 with time for sampling strategies 1K2C, 1K4C, 7K4C and 7K7C in (a) house 1, (b) house 2, (c) house 3, (d) house 4, (e) house 5, and (f) house 6.
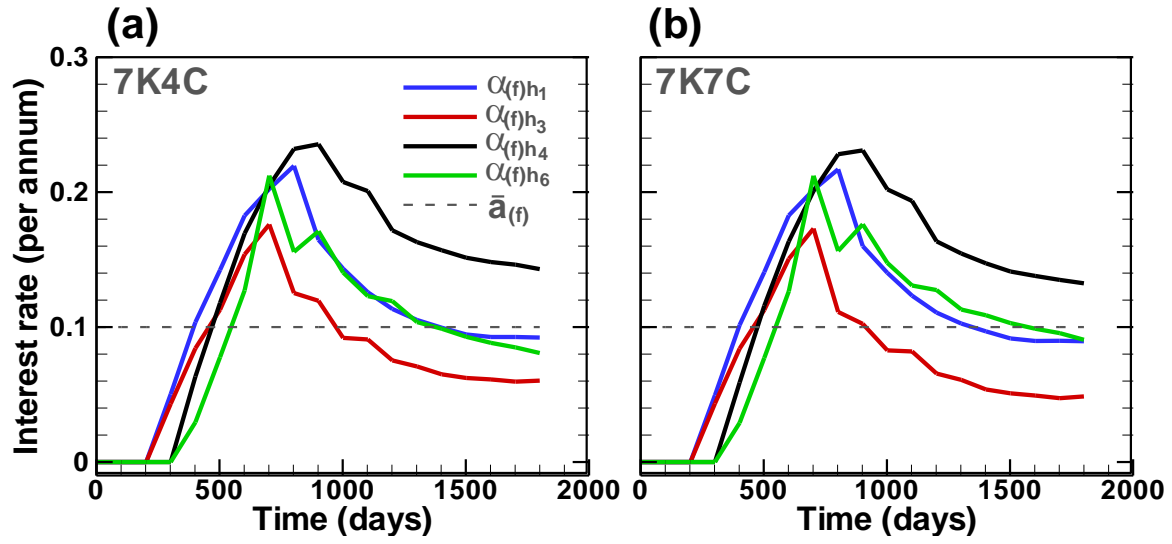
Figure 4.3: Transient behaviour of $a_{(f)_{i,j}}$ for houses 1, 3, 4 and 6 using sampling strategies (a) 7K4C and (b) 7K7C.
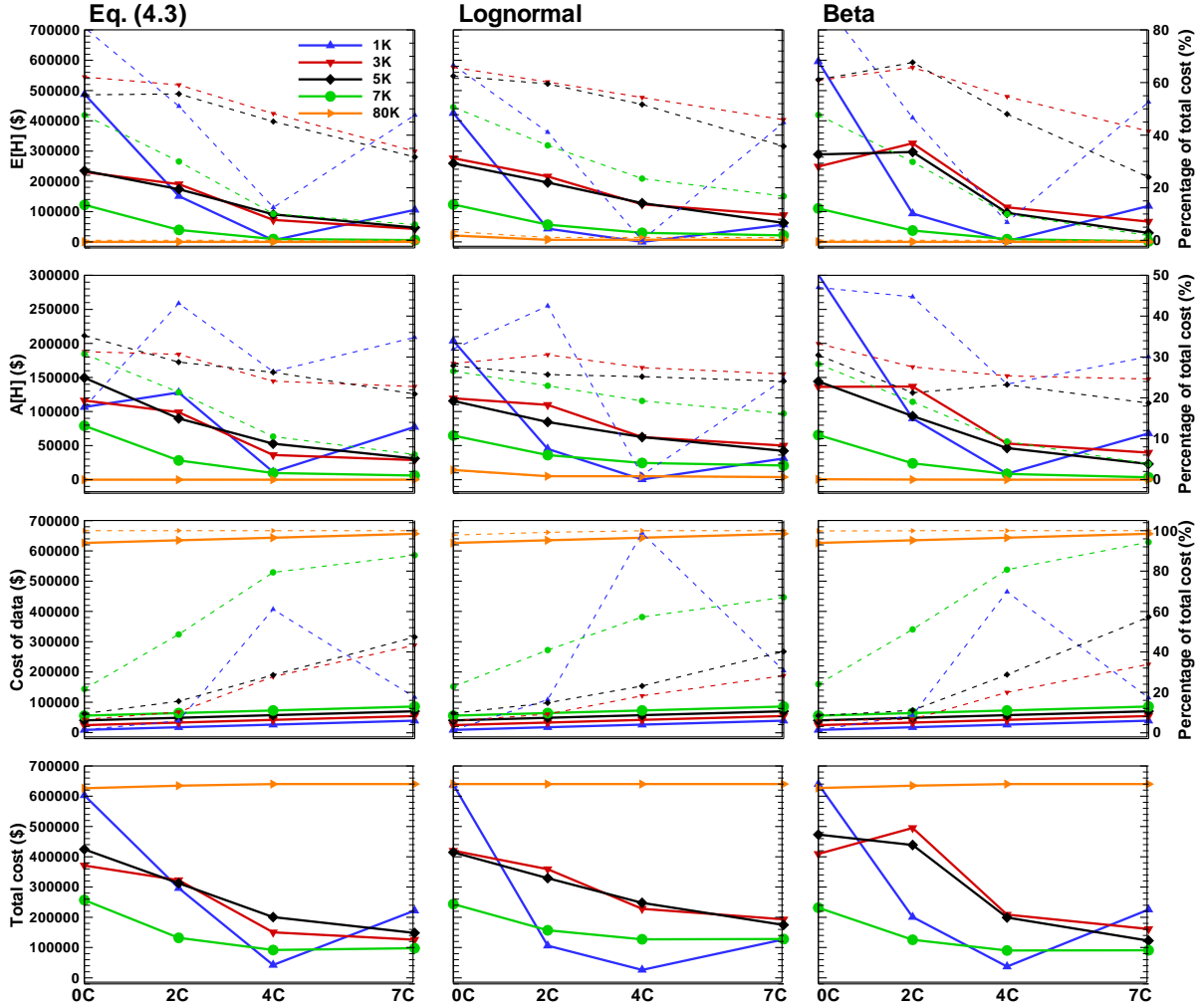
Figure 4.4: Transient variability as a function of sampling strategy for; $E[H]$ which is the expected cost of failure, $A[H]$ which is the safety loading term, $E[\mathcal{C}_{data}]$ which is the expected cost of data, and Obj which is the total cost of the objective function (see Equation (4.35)). The first column involves $P_{i,j}$ estimated using Equation (4.3), the second column involves $P_{i,j}$ estimated using a lognormal distribution, and the third column involves $P_{i,j}$ estimated using a beta distribution.
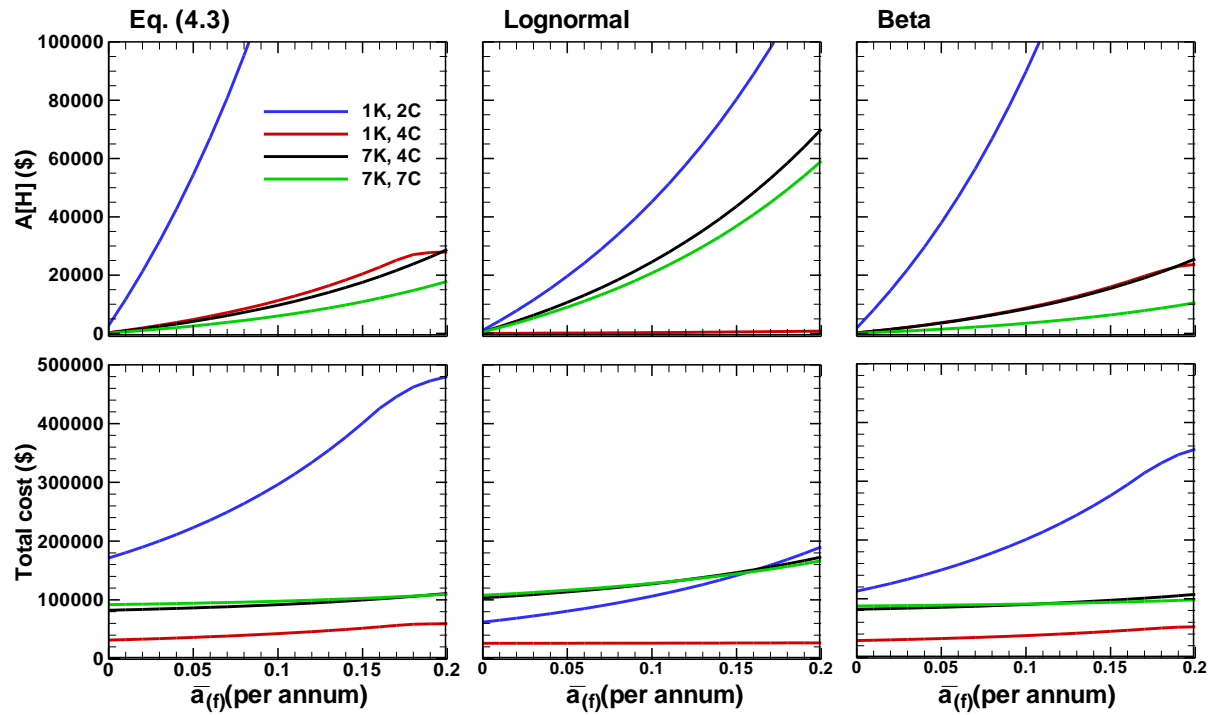
Figure 4.5: Sensitivity in the cost of; $A[H]$ which is the safety loading term, and Obj which is the total cost of the objective function (see Equation (4.35)), to $\bar{a}_{(f)}$ over the interval $0 \le \bar{a}_{(f)} \le 0.2$ per annum. The first column involves $P_{i,j}$ estimated using Equation (4.3), the second column involves $P_{i,j}$ estimated using a log-normal distribution, and the third column involves $P_{i,j}$ estimated using a beta distribution.

Table 4.1: Risk capital parameter values

Parameter		Value
General		
t_0	start date for brownfields project	March 2008
$\Delta t_{monitor}$	indoor air monitoring interval	100 days
T	guarantee period for brownfields project	2000 days
η	house value reduction after resold	0.2
US national housing index parameters for S		
S_0	unit price of house at time t_0	\$200,000
α_S	annual rate of appreciation	0.0456 per annum
σ_S	volatility	0.0375
$\hat{q}_S$	market price of risk	0.0316
Annual rate of inflation $r_{\jmath}$		
$\kappa_{\jmath}$	speed of adjustment	5.83
$\theta_{\jmath}$	reversion level	3.12 per annum
$\sigma_{\jmath}$	Volatility	6.19
$\hat{q}_{\jmath}$	market price of risk	-0.1 to 0.1
Annual discount rate ρ		
ρ	discount rate $\rho = r_{\mathcal{N}} + \hat{q}_s \sigma_S$	0.0612
Nominal annual interest rate $r_{\mathcal{N}}$		
$\kappa_{\mathcal{N}}$	speed of adjustment	0.2
$\theta_{\mathcal{N}}$	reversion level	0.06 per annum
$\sigma_{\mathcal{N}}$	Volatility	0.07
$\hat{q}_{\mathcal{N}}$	market price of risk	-0.1
S&P 500 parameters for M		
α_M	annual rate of appreciation	0.0662 per annum
σ_M	volatility	0.1537
$E[\mu_M]$	expected return on the market portfolio M	0.12 per annum
β_S	correlation between μ_S and μ_M	0.0197
Safety loading parameters		
$b_{(ac)1}, b_{(f)1}$	weighting parameter for coefficient of variation	0.6
$b_{(ac)2}, b_{(f)2}$	weighting parameter for relative accuracy	0.4
$a_{(f)h}$	hydrogeological coefficient in financial method	0.7210
$a_{(ac)h}$	hydrogeological coefficient in actuarial method	0.7916

note: modified from Table 4.2. of Yu (2009) and based on $\bar{a}_{(f)} = 0.1$ per annum.

Table 4.2: Estimated sampling costs for soil gas concentration and permeability data

Cost of sampling concentration (\$)		Cost of permeability data (\$)	
Type of method	Buried collector	Type of method	Auger drilling
Equipment		Drill	430.71
Mobilization		Equipment	100
Operation		Mobilization	400
Per sample	250	Sampling per borehole	7800
Total	$250 n_C$	Total	$930.71 + 7800 n_k$

note: n_C and n_K are the number of concentration and boreholes for permeability data.

Table 4.3: Values of objective function for different sampling strategies using the financial premium calculation principle

Obj	0C	2C	4C	7C
1K – Eqn. (3)	$640,000.00	$640,000.00	$51,602.64	$521,271.29
3K – Eqn. (3)	$493,275.59	$454,147.37	$170,055.57	$129,264.35
5K – Eqn. (3)	$631,132.02	$524,726.16	$261,609.59	$151,430.53
7K – Eqn. (3)	$326,670.09	$149,458.83	$91,740.90	$97,637.68
80K – Eqn. (3)	$626,298.71	$634,893.85	$640,000.00	$640,000.00

note: the safety loading coefficients are calculate based on $\bar{a}_{(f)} = 0.1$ per annum.

Table 4.4: Values of objective function for different sampling strategies using the actuarial premium calculation principle

Obj	0C	2C	4C	7C
1K – Eqn. (3)	$603,020.27	$296,437.90	$42,618.83	$222,052.10
3K – Eqn. (3)	$371,076.73	$322,743.19	$150,245.09	$126,015.93
5K – Eqn. (3)	$424,568.91	$312,474.46	$200,695.79	$148,339.10
7K – Eqn. (3)	$257,024.86	$132,073.46	$91,883.66	$97,766.69
7K – lognormal	$243,745.38	$157,082.49	$127,352.46	$128,136.10
7K – beta	$231,657.80	$125,717.41	$90,354.70	$91,032.11
80K – Eqn. (3)	$626,298.71	$634,943.26	$640,000.00	$640,000.00

note: the safety loading coefficients are calculate based on $\bar{a}_{(f)} = 0.1$ per annum.

Chapter 5 Summary and Recommendations

5.1 Summary of Accomplishments

The main goal of this research is to develop a framework for optimal design of site characterization of a risk-based brownfield redevelopment project under uncertainty. To accomplish this goal, three major elements are presented and comprise the core of this thesis. First, the numerical model *CompFlow Bio* was applied to analyze the hydrogeological process governing the fate and transport of TCE (a volatile chlorinated solvent) from groundwater to indoor air. It provided an increased understanding of the impact of various influential factors (including recharge, a clay layer with and without recharge, a thicker vadose zone, and heterogeneity in the subsurface permeability structure of a stratigraphically continuous aquifer) on the distribution of the indoor air concentrations within multiple residential houses located directly above and adjacent to the groundwater plume. Due to limiting factors such as cost, time, and accuracy involved in site characterization and monitoring during redevelopment, an integrated approach was developed as the second element of this thesis to reduce uncertainty and improve accuracy with effective information. To reduce prediction uncertainty, this approach took into account the worth of permeability data and soil gas concentration data and was measured statistically using three metrics. A series of permeability data was conditioned and used to reconstruct the geological aquifer based on the kriging technique in order to build the spatially correlated error covariance matrix using the numerical model to simulate indoor air concentration in space and time. Different amounts of soil gas concentration data were also used through the Kalman filter to update estimates of the indoor air concentration obtained from the hydrogeological model. For the third element, prediction results, after data assimilation was applied, were used to investigate the cost of hydrogeological and financial risks in different sampling strategies under the two risk capital valuation approaches defined and developed in this study. Eventually, an optimal sampling strategy was selected based on the least cost criterion.

Specific accomplishments are described below.

- The heterogeneous and statistically isotropic Borden-like permeability field realization and the problem geometry, reminiscent of the Rivett (1995) field experiment in the Borden aquifer, were presented for the numerical model. The three-dimensional geometry was motivated by the need to accommodate multiple houses, with some laterally offset from the direction of groundwater plume advection.

- The relative importance of various influential parameters was observed in Chapter 2. The combination of inserting a clay layer and recharge created an impervious barrier for TCE beneath the foundation slab effectively negating the flux of TCE across the foundation slab. Heterogeneity in the subsurface permeability structure was the next parameter of importance. It implied that characterizing the site stratigraphy plays an important role when assessing the fate and transport of TCE and detailed soil core permeability measurements to characterize the heterogeneity within each stratigraphic unit is necessary.

- Simulation results using the numerical model *CompFlow Bio* in Chapter 2 indicated that houses which were laterally offset from the groundwater plume were less affected by vapour intrusion due to limited transverse horizontal flux of TCE within the groundwater plume, in agreement with the ASTM (2008) guidance. The J&E model generated an exclusion zone that would be significantly further away from the groundwater plume than what either *CompFlow Bio* or ASTM (2008) would predict.

- By comparing J&E-calculated attenuation coefficients ($\alpha_{J\&E}$) against those computed using *CompFlow Bio* (α_{CFB}) based on 50 alterative permeability realizations, the result showed that the J&E model is conservative and the degree to which the J&E model is conservative is unrealistic from a practical engineering design perspective. When site characterization budgets are limited, the J&E model is an excellent choice to assess the fate and transport of TCE in a manner that is both economical and conservative.

- Chapter 3 presented a methodology for reducing the uncertainty in predicting indoor air concentration when field data (permeability and concentration data) became available. The overall performance of the prediction was measured by three metrics: the ability to correctly estimate the actual indoor air concentration; the ability to use available soil gas concentration data to reduce the standard deviation of the indoor air concentration; and the ability to accurately forecast the probability of indoor air concentrations exceeding the regulatory limit. Based on the RMSE and the statistical test results with respect to the three metrics, we concluded that 7K4C produced the optimal estimation which balanced the type and quantity of sampling data.

- Permeability data were extracted from borehole logging prior to the redevelopment, therefore the geological information was imbedded in the model before assimilating soil

124

gas sampling data through the Kalman filter to reduce the model prediction uncertainty. The results in Chapter 3 showed that the value of permeability data was worth more than the value of concentration data under this framework by comparing different measures of uncertainty reduction, which implied that the site characterization of a geological aquifer is the most important step in a risk-based project.

- The uncertainty reduction was significantly improved after incorporating sampling data, even if a limited number of samplings were taken due to the availability of samplings in residential dwellings and high sampling costs. It is very important for the decision maker to choose the appropriate number of samplings, location and time of samplings, and category of samplings (permeability versus concentration). We obtained better estimations of indoor air concentration and probability of exceedence when more data were acquired.

- The third metric was presented by three different evaluations of the probability of exceedence (Equation (3.19) and fitting the tail of the indoor air concentration distribution to a log-normal and beta distribution). It showed that the log-normal distribution provided the most conservative estimates for prediction uncertainty as characterized by the probability of exceedence and the beta distribution and Equation (3.19) provided intermediate and least conservative estimates. This may help regulators to make an appropriate strategy for managing uncertainty based on the characteristics of indoor air concentration population.

- The use of Monte Carlo simulations and the Kalman filter requires a great number of groundwater flow and solute transport simulations per optimal sampling design, which can be computationally expensive. Therefore, the efficiency of this methodology depends on the type of application and the computation speed of the model. Judgment from experienced professionals for the initial sampling design is needed.

- Chapter 4 proposed an optimal least cost solution for designing a sampling network and managing the risks from hydrogeological and financial perspectives. It showed that drilling 7 boreholes to extract permeability data and taking soil gas samplings in 4 locations or 7 locations alternatively gave the minimum total cost.

- Two risk capital valuation approaches were presented based on two different principles, actuarial and financial. With the increasing amount of data, the value of the safety loading term for compensating against the uncertainties became smaller, however, the cost of samplings increased. A trade-off between an increasing sampling budget and a decreasing risk capital was observed.

- The hydrogeological uncertainty was represented in the risk capital valuation by the coefficient of variation, the relative accuracy, and the probability of exceedence summarized in the results obtained by the Kalman filter. The financial uncertainty mainly resulted from the market price of risk for a contract that depended linearly on the underlying stochastic US national housing price index.

- The risk capital was composed of the expected value of risk capital, the safety loading term, and the cost of data. By fitting the indoor air concentration of TCE into a log-normal distribution to define the probability of failure, a more conservative estimation of loss reserve was obtained.

- The safety loading term only contributed to a small portion of the risk capital compared with the expected value of losses. The risk capital and the corresponding total cost would not differentiate dramatically from varying the average interest rate surcharge which represents the market price of risk with a greater volatility and they were not sensitive to the method used to define the probability of failure.

- This framework can also be applied for different engineering projects.

5.2 Recommendations for Future Research

This research has provided some insights into the risk capital valuation for a risk-based engineering project and helped understand the process affecting exposure pathways of volatile organic compounds (VOCs) originating from a NAPL source zone located below the water table into indoor air, but some questions remain and others have arisen from this study.

126

The effects of more hydrogeological factors and factors with respect to structure of residential houses on the vapour intrusion pathways should be further investigated. In the modeling part of this thesis, the impacts of the influential parameters, recharge, barometric pressure, inserting a clay layer with and without recharge, extending the thickness of the vadose zone, heterogeneity in the subsurface permeability structure of the stratigraphically continuous aquifer, the aperture of the foundation slab, and the negative pressure in the basement on the indoor air concentration were examined. However, there are some other parameters which may also play important roles in soil gas transport process. First, non-equilibrium dissolution models used to show the persistent dissolution behaviour of DNAPLs is crucial for governing the mass transfer in the source zone. Different formulation and parameter values of the dissolution rate should be evaluated. Second, in previous studies, the thickness of the capillary fringe has been shown to be a dominant factor in the vertical mass flux across the capillary fringe since that DNAPLs located below the water table must diffuse across the capillary fringe before moving upwards in the gas phase. It would be helpful if the importance of the thickness of the capillary fringe was fully investigated.

The problem with using the three-dimensional *CompFlow Bio* model is the high computational cost. It took a couple of weeks to finish one standard simulation for this application. Therefore, there is a need to enhance the computational efficiency without simplifying the model. Parallel computing is a form of high-performance computation that can be considered to mitigate this problem. Building a Jacobian matrix and the iterative solver are the two core components contributing to the execution time and should be the main focus for improvement.

This study was intended to provide some insights and practical guidance for redeveloping a brownfield, but the conceptual model was highly stylized and hypothetical. It will be valuable to use real world data to verify our strategy. This thesis has shown that the geological information of an aquifer is essential for determining the temporal and spatial distribution of contaminated vapour, so more characteristic sites with different lithology parameters should be taken into account to evaluate this impact. Also, the scale of this study was relatively small. There were only six houses set up in the problem geometry. The negative pressure in the basement of each house causes the soil gas to be pulled out of the subsurface into the basement and deflects the soil gas plume in the vadose zone. If there are more houses with the same structure built in the geometry, the plume may spread more laterally and the delineation of the exclusion zone may change. At the same time, the problem of boundary effect could be alleviated. A larger scale model is necessary with up-scaling techniques introduced to capture the effects of fine scale heterogeneity.

The thesis used an ordinary kriging technique to produce smooth images based on the spatial correlation generated from the sampled data. In doing so, short scale variability was poorly reproduced while it underestimated extremes. Compared with the classical kriging method, a conditional transition probability/Markov chain approach (TP/MC) can generate a high-resolution aquifer analogue, improve consideration of spatial cross-correlations, and facilitate the integration of geologic interpretation of facies architecture. In order to better capture the structurally complex geology, this approach can be applied to reconstruct aquifer heterogeneity in a future study.

In the cost-benefit-risk analysis, the cost of acquiring permeability and soil gas sampling data was roughly estimated. The cost of labour and reporting results were not taken into consideration, which could be a significant part of the total cost. The hypothesis in the risk capital valuation was that the cost of demolishing a house, remediating the site and rebuilding the house was neglected, so the risk of this cost appreciated over time was overlooked. If the liability period for developers is long enough, the inflation rate could be an important factor for cost analysis in order to reserve enough capital to avoid insolvency. In future studies, all of these factors should be taken into account.

Appendices

Appendix A Model Parameters

In order to better understand the numerical model and parameter estimations in Chapter 2, a great number of materials from Section 2 of Yu et al. (2009) are directly used as supporting documentation, including; (1) formulation, (2) the parameterization of all porous media and chemical properties governing flow and transport, (3) a methodology for incorporating the foundation of the building that is below grade into the model and simulating the interaction between the aquifer and the building, and (4) a methodology for calculating the indoor air TCE concentration based on the interaction between the aquifer and building.

Based on Equations (2.1) – (2.3), there exist the following constraints among the primary variables:

$$S_q + S_n + S_g = 1$$

$$P_g = P_n - \hat{\alpha}P_{cgn}(S_g) + (1 - \hat{\alpha})[P_{cgq}(S_g) - P_{cnq}(S_q = 1)] \tag{A.1}$$

$$P_n = P_q - \hat{\alpha}P_{cnq}(S_q) + (1 - \hat{\alpha})P_{cnq}(S_q = 1)$$

where

$$\hat{\alpha} = \min(1, S_n/S_n^*) \tag{A.2}$$

and P_{cgn} , P_{cgq} , and P_{cnq} are experimentally derived capillary pressure curves $[kPa]$. S_n^* is a blending parameter used to ensure that the capillary pressure has the correct form as the non-aqueous phase saturation goes to zero. The capillary pressure curves are scaled using a modified form of the function proposed by Leverett (1941):

$$P^D_{cl_1 l_2} = \frac{P_{cl_1 l_2}}{\sigma_{l_1 l_2}} \left(\frac{\mathbf{K}}{\phi}\right)^{\tilde{\alpha}}$$

(A.3)

where $P^D_{cl_1 l_2}$ and $P_{cl_1 l_2}$ are the dimensionless and dimensional capillary pressures, respectively, between any two phases, $l_1 l_2 = \{gn, gq, nq\}$, $\sigma_{l_1 l_2}$ is the surface/interfacial tension between the two phases, and $\tilde{\alpha}$ is a parameter related to the pore size distribution of the porous media.

Stone's second model (Stone, 1973) is used in *CompFlow Bio* to estimate k_{rn} and is normalized using the method introduced by Aziz and Settari (1979). Therefore, the relative permeability of the aqueous, gas and non-aqueous phases for the three-phase system is given by:

$$k_{rq} = \mathrm{F}(S_q)$$

$$k_{rg} = \mathrm{F}(S_g)$$

$$k_{rn} = k_{r(nq_c)}\left[\left(\frac{k_{r(nq)}}{k_{r(nq_c)}} + k_{rq}\right)\left(\frac{k_{r(ng)}}{k_{r(nq_c)}} + k_{rg}\right) - \left(k_{rq} + k_{rg}\right)\right]$$

(A.4)

where $k_{r(nq)}$ and $k_{r(ng)}$ are the non-aqueous phase relative permeabilities measured in the two-phase non-aqueous/aqueous and non-aqueous/gas systems at the aqueous and gas phase saturations in the full three-phase system, respectively.

Partitioning of components between phases is assumed to be at equilibrium. In this context, the following constraint applies to any phase l which is present:

$$\sum_p X_{pl} = 1 \quad .$$

(A.5)

Component p partitions between the gas and non-aqueous phases according to:

$$X_{pg} = Z_{pgn}X_{pn} \text{ where } Z_{pgn} = exp\{a_{pgn}\}/P_g \qquad (A.6)$$

while between the gas and aqueous phases, the relationship is:

$$X_{pg} = Z_{pgq}X_{pq} \text{ where } Z_{pgq} = exp\{a_{pgq}\}/P_g$$

$$X_{wg} = Z_{wgq}X_{wq} \text{ where } Z_{wgq} = a_{wgq}/P_g$$

$$(A.7)$$

where both Z_{pgn} and Z_{pgq} are dependent on the systems pressure P_g.

Although Equation (2.1) is given in the most general sense, some simplifying assumptions are made that restrict the composition of various phases. Specifically, the non-aqueous phase consists of only component $p = \{c_{TCE}\}$ which implies $X_{wn} = 0$ and $X_{an} = 0$, the aqueous phase consists of components $p = \{w, c_{TCE}\}$ which implies $X_{aq} = 0$, and the gas phase consists of all available components $p = \{w, a, c_{TCE}\}$. The rationale for this restricted system is provided in Forsyth (1993).

Phase appearance and disappearance rules are used to develop meaningful tests to detect which subset of phases exists at a node given the moles of component p present. These rules follow from the set of primary variables, which in general consist of P_l, X_{pl}, and S_l and are aligned with Equations (2.1) to (2.3) . Following the formulation introduced by Forsyth (1993), we also assume the gas phase is always present to at least some minimal saturation $S_g^{min} = 10^{-3}$ to alleviate numerical issues associated with the non-condensable air component.

All relevant physical and chemical properties of the porous media and components are itemized on Table A.1 – A.4.

Interaction between the soil gas and the indoor air is limited to occur across the foundation slab only. To facilitate this in the numerical model, we assumed there was a 2.2 m length of crack per unit

surface area of foundation slab, where the genesis of the cracks is due to foundation settlement. The aperture of the crack for the base scenario is $2b = 100\ \mu m$ resulting in a permeability of $\mathbf{K}_{slab_C} = 8.33 \times 10^{-10}\ m^2$, while the foundation slab matrix is comprised of concrete and is essentially impermeable which is approximated by assigning it a permeability of $\mathbf{K}_{slab_M} = 1 \times 10^{-20}\ m^2$, with the resulting bulk permeability of the foundation slab being $\mathbf{K}_{slab_B} = 1.83 \times 10^{-13}\ m^2$. These properties were calculated using the following series of relationships:

$$\phi_{slab_C} = \frac{V_C}{V_{slab}}$$

$$\phi_{slab_M} = 1 - \phi_{slab_C}$$

$$\mathbf{K}_{slab_C} = \frac{(2b)^2}{12}$$

$$\mathbf{K}_{slab_B} = \mathbf{K}_{slab_C}\,\phi_{slab_C} + \mathbf{K}_{slab_M}\,\phi_{slab_M}$$

(A.8)

where V_C and V_{slab} are the volume $[m^3]$ of the crack and entire foundation slab, respectively. Advective flux of the gas phase is forced to occur across the foundation slab only by reducing the gas phase pressure along the top row of the basement interior nodes so that for the base scenario they are $\Delta P_g^{basement} = 10$ Pa below that of the ambient atmospheric pressure. The interior nodes of the basement are assigned a permeability of $\mathbf{K}_{basement} = 1 \times 10^{-9}\ m^2$ to ensure that the foundation slab is the limiting factor allowing gas phase flow to enter the basement. The porosity of the nodes representing the interior volume of the basement was reduced to $\phi_{basement} = 0.001$ to minimize the mixing volume within the basement. These properties are summarized on Table A.5.

Once the two metrics described above are computed using *CompFlow Bio*, the indoor air concentration of TCE is computed as a post-processing step. This methodology proceeds by first computing the concentration of TCE in the soil gas $\mathbb{C}_{c_{TCE}g}$ that is entering the basement at current *CompFlow Bio* simulation time $n + 1$ as:

$$\mathbb{C}^{n+1}_{c_{TCE}g} = \frac{Q^{n+1}_{c_{TCE}}}{Q^{n+1}_a M^*_a} \times 10^6 \; [ppmV] \tag{A.9}$$

where $Q^{n+1}_{c_{TCE}} \; [moles/day]$ and $Q^{n+1}_a \; [m^3/day]$ are the flow rate of TCE and air into the basement at timestep $n+1$. Next, we assume two possibilities concerning the ventilated condition of the house, namely; without air exchange, and alternatively with air exchange. Without air exchange, the high efficiency furnace does not introduce any outside air into the house, the air within the house is perfectly circulated via the ventilation system attached to the furnace, and that any soil gas entering $Q^{n+1}_a|_1$ (or exiting $Q^{n+1}_a|_2$) the indoor air of the house across the foundation slab is compensated by an equal flow of indoor air out of $Q^{n+1}_a|_3$ (or atmospheric air into $Q^{n+1}_a|_4$) the house through leaks around the windows and doors. The indoor air concentration of TCE, $\mathbb{C}^{n+1}_{c_{TCE}IA}$, at current *CompFlow Bio* time step $n+1$ is computed as:

$$\frac{\mathbb{C}^{n+1}_{c_{TCE}IA} - \mathbb{C}^{n}_{c_{TCE}IA}}{\Delta t} V_H = \underbrace{Q^{n+1}_a|_1 \, \mathbb{C}^{n+1}_{c_{TCE}g}}_{F_1 \left\{ \substack{\text{into house} \\ \text{across slab}} \right.} - \underbrace{Q^{n+1}_a|_2 \, \mathbb{C}^{n+1}_{c_{TCE}IA}}_{F_2 \left\{ \substack{\text{out of house} \\ \text{across slab}} \right.}$$

$$\qquad\qquad - \underbrace{Q^{n+1}_a|_3 \, \mathbb{C}^{n+1}_{c_{TCE}IA}}_{F_3 \left\{ \substack{\text{out of house} \\ \text{around doors} \\ \text{and windows}} \right.} + \underbrace{Q^{n+1}_a|_4 \, \mathbb{C}^{n+1}_{c_{TCE}ATM}}_{F_4 \left\{ \substack{\text{into house} \\ \text{from atmosphere}} \right.} \tag{A.10}$$

where $Q^{n+1}_a|_1 + Q^{n+1}_a|_4 = Q^{n+1}_a|_2 + Q^{n+1}_a|_3$ and the concentration of TCE in the atmospheric air $\mathbb{C}^{n+1}_{c_{TCE}ATM} = 0$ ppmV for the base scenario. The fluxes of TCE, F_1, F_2, F_3 and F_4 $[ppmV \; TCE \times m^3 air/day]$, are depicted on Figure A.1. The volume of the house V_H is calculated under the assumption that it consists of a basement and main floor, and each floor including supporting beams is 3 m tall, resulting in a total indoor air volume of $V_H = 900 \; m^3$. Note that timestep $\Delta t = t^{n+1} - t^n$, where n denotes the prior *CompFlow Bio* simulation time. With air exchange, $Q^{n+1}_a|_4 = V_H \times A_{ex}$ where $A_{ex} = 0.5/hr$ is the number of air volumes flowing into the house per unit time for the base scenario model. Note that $Q^{n+1}_a|_1$ is enforced using a penalty sink term for Q_a in Equation (2.1) and is implemented as:

$$Q_a^{n+1}|_{1,i} \frac{M_a^*}{V_i} = B \, X_{ag|i} \, M_{g|i} \frac{k_{rg|i}}{\mu_{g|i}} \left(P_g^{basement} - P_{g|i}^{n+1} \right) \tag{A.11}$$

where $P_g^{basement} = P^{ref} - \Delta P_g^{basement}$, node i is basement node in which the sink term is placed, V_i is the volume of node i, and $B = 10^5$ in order to ensure a grid-converged air flow rate $Q_a^{n+1}|_{1,i}$ for the simulated pressure difference.
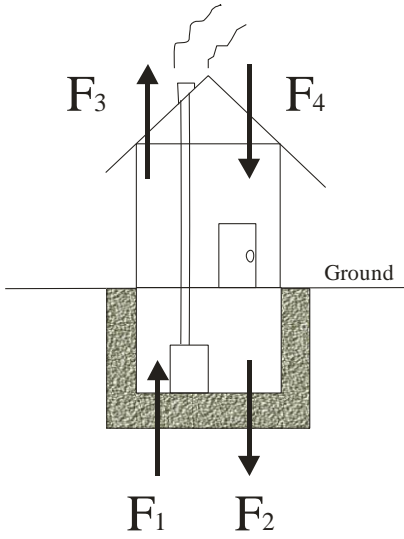
Figure A.1: Contaminant flux into and out of house, denoted by F_1, F_2, F_3 and F_4 $\left[\text{ppmV TCE} \times \text{m}^3\text{air/day}\right]$, as calculated using Equation (A.10).

Table A.1: Capillary pressure and relative permeability aqueous/non-aqueous phase table

S_q [−]	k_{rq} [−]	$k_{r(nq)}$ [−]	P_{cnq} [kPa]
0.2	0.0	0.68	9.0
0.3	0.04	0.55	5.4
0.4	0.10	0.43	3.9
0.5	0.18	0.31	3.3
0.6	0.30	0.20	3.0
0.7	0.44	0.12	2.7
0.8	0.60	0.05	2.4
0.9	0.80	0.0	1.53
1.0	1.0	0.0	0.0

note: cited from Table 2. of Yu et al. (2009)

Table A.2: Capillary pressure and relative permeability liquid/gas phase table

$S_q + S_n$ [−]	k_{rg} [−]	$k_{r(ng)}$ [−]	P_{cng} [kPa]	P_{cgq} [kPa] [a]	P_{cgq} [kPa] [b]
0.2	0.64	0.0	9.0	6.6	11.6
0.32	0.46	0.0	3.0	4.5	9.5
0.4	0.36	0.0009	2.4	3.9	8.9
0.5	0.25	0.045	2.1	3.6	8.6
0.6	0.16	0.116	1.8	3.3	8.3
0.7	0.09	0.21	1.5	3.0	8.0
0.8	0.04	0.34	1.2	2.0	7.0
0.9	0.01	0.49	0.9	1.0	6.0
0.95	0.0	0.58	0.5	0.5	5.5
1.0	0.0	0.68	0.0	0.0	5.0

[a] base scenario capillary fringe
[b] scenario with thicker capillary fringe
note: cited from Table 3. of Yu et al. (2009)

Table A.3: Equilibrium partitioning coefficients (at $10^{\circ}C$)

Property	Value
water: $p = \{w\}$	
$\quad Z_{wgn}$ $[-]$	—[a]
$\quad a_{wgq}$ $[-]$	1.07
air: $p = \{a\}$	
$\quad Z_{agn}$ $[-]$	—[b]
$\quad Z_{agq}$ $[-]$	—[b]
TCE: $p = \{c_{TCE}\}$	
$\quad a_{c_{TCE}gn}$ $[-]$	1.96
$\quad a_{c_{TCE}gq}$ $[-]$	10.52

[a] water is not allowed to partition from the gas into the non-aqueous phases
[b] air is a non-condensable component and is not allowed to partition from the gas phase into either the aqueous or non-aqueous phases
note: cited from Table 4. of Yu et al. (2009)

Table A.4: Component and phase property data (at 10°C)

Property	Value
Compressibilities	
$\hat{C}_q$ [kPa^{-1}]	3.0×10^{-6}
$\hat{C}_n$ [kPa^{-1}]	4.3×10^{-7}
$\hat{C}_r$ [kPa^{-1}]	1.0×10^{-7}
Standard component densities	
M_w^* [mole/m^3]	5.5×10^4
M_a^* [mole/m^3]	41.1
$M_{c_{TCE}}^*$ [mole/m^3]	1.1×10^4
Molecular weights	
ω_w [kg/mole]	18.02×10^{-3}
ω_a [kg/mole]	28.97×10^{-3}
$\omega_{c_{TCE}}$ [kg/mole]	131.4×10^{-3}
Reference pressure and temperature	
P^{ref} [kPa]	100.0
T^{ref} [$^{\circ}$K]	283.0
Capillary pressure blending parameter	
S_n^* [$-$]	0.1
Viscosities	
μ_q [kPa $\cdot$ day]	2.44×10^{-11}
μ_g [kPa $\cdot$ day]	1.62×10^{-13}
$\mu_{c_{TCE}}$ [kPa $\cdot$ day]	9.75×10^{-12}
Molecular diffusion coefficient	
d_q [m^2/day]	3.8×10^{-5}
d_n [m^2/day]	3.8×10^{-5}
d_g [m^2/day]	0.394 [a]

Molar density

$$M_l = \frac{1 + \hat{C}_l\left(P_l - P^{\text{ref}}\right)}{\sum_p \max\left(0, X_{pl}\right)/M_p^*} \quad l = \{q,n\}, \qquad M_g = \frac{P_g}{RT}$$

Mass density

$$\rho_l = \sum_p X_{pl}\omega_p$$

note: cited from Table 5. of Yu et al. (2009)

Table A.5: Foundation slab and basement properties

Property	Value
Foundation slab	
aperture of crack in slab: $2b$ [μm]	100
permeability of slab matrix (concrete): $\mathbf{K}_{slab_M}$ [m^2]	1×10^{-20}
permeability of crack in slab: $\mathbf{K}_{slab_C}$ [m^2]	8.33×10^{-10}
bulk permeability: $\mathbf{K}_{slab_B}$ [m^2]	1.83×10^{-13}
bulk porosity: ϕ_{slab_B} [$-$]	0.00022
Basement (walls and interior)	
vertical permeability (i.e. interior): $\mathbf{K}_{basement}$ [m^2]	1×10^{-9}
porosity: $\phi_{basement}$ [$-$]	0.001

Appendix B Kalman Filter Theory

Filtering can be referred as an estimator of the state vector at the current state time based upon all past measurements. The Kalman filter addresses the general problem of trying to estimate the state vector of a discrete-time controlled process that is governed by the linear stochastic equation. The estimate minimizes the mean of the squared error and is essentially a Bayesian approach. The application of the Kalman filter is presented in Section 3.2.2.1. The discrete Kalman filter (Welch and Bishop, 2006) is derived as follows.

The state equation can be expressed as

$$c_n = A c_{n-1} + B u_{n-1} + w_{n-1} \tag{B.1}$$

with a noise-corrupted measurement z_n

$$z_n = H_n c_n + v_n \tag{B.2}$$

where c_n, c_{n-1} are the vectors of discrete estimates of the state variable at time t_n and t_{n-1} respectively; u_{n-1} is a known control input to the system; the random variables w_{n-1} and v_n represent the process and measurement noise and are assumed to be independent, white and with normal probability distributions

$$p(w) \sim N(0, Q)$$
$$\tag{B.3}$$
$$p(v) \sim N(0, R) \, ,$$

where the process noise covariance Q and measurement noise covariance R matrics might change with each time step or measurement, however here they are assumed to be constant. The matrix A is the known state transition matrix and it allows calculation of the state vector at time t_n given complete knowledge of the state vector at time t_{n-1} in the absence of either a driving function or

process noise. The matrix B relates the optional control input to the state vector c_n. The matrix H_n relates the state to the measurement z_n at time t_n and is also called the measurement matrix.

An estimate, $\hat{c}$, is the computed value of a quantity (which has a different definition from the one in Chapter 3), c, based upon a set of measurements, z. $\hat{c}_n^-$ is defined as a priori state estimate at time t_n and $\hat{c}_n^+$ is a posteriori state estimate at time t_n given measurement z_n. And a priori and a posteriori estimate errors can be defined as

$$e_n^- = c_n - \hat{c}_n^- \, ,$$

$$e_n^+ = c_n - \hat{c}_n^+ \, .$$

(B.4)

The a priori estimate error covariance is then

$$P_n^- = \mathrm{E}[e_n^- e_n^{-T}] \, ,$$

(B.5)

and the a posteriori estimate error covariance is

$$P_n^+ = \mathrm{E}[e_n^+ e_n^{+T}] \, .$$

(B.6)

In deriving the equation for the Kalman filter, a posteriori estimate $\hat{c}_n^+$ is sought as a linear combination of a priori estimate $\hat{c}_n^-$ and a weighted difference between an actual measurement z_n and a measurement prediction $H_n \hat{c}_n^-$ as shown in the equation below,

$$\hat{c}_n^+ = \hat{c}_n^- + K_n(z_n - H_n \hat{c}_n^-)$$

(B.7)

and

$$P_n^+ = P_n^- - K_n H_n P_n^- \quad . \tag{B.8}$$

The difference $(z_n - H_n \hat{c}_n^-)$ is called the measurement innovation, or the residual. The residual reflects the discrepancy between the predicted measurement $H_n \hat{c}_n^-$ and the actual measurement z_n. A residual of zero means that the two are in complete agreement. The matrix K_n is called the Kalman gain that minimizes the a posteriori error covariance at time t_n. The minimization can be accomplished by first rewriting e_n^+ with Equation (B.7), inserting that into Equation (B.6), and then solving K by setting the error covariance equal to zero. As a result, the Kalman gain K can be given by

$$K_n = P_n^- H_n^{\ T} (H_n P_n^- H_n^{\ T} + R_n)^{-1} \quad . \tag{B.9}$$

As the measurement error covariance R approaches zero in Equation (B.9), the Kalman gain K has more weights on the residual, specifically,

$$\lim_{R_n \to 0} K_n = H_n^{\ -1} \quad . \tag{B.10}$$

On the other hand, as the a priori estimate error covariance P_n^- approaches zero, the Kalman gain K has less weights on the residual, specifically,

$$\lim_{P_n^- \to 0} K_n = 0 \quad . \tag{B.11}$$

In summary, the discrete Kalman filter estimates a process by predicting the process state at each time step and obtaining updates of estimation after incorporating noisy measurements recursively. The time update equations are responsible for projecting forward the current state and error covariance estimates to obtain a priori estimate for the next time step. This is so-called predictor-corrector algorithm, and the specific equations for the time and measurement updates are presented as follows.

The time update equations:

$$\hat{c}_n^- = A\hat{c}_{n-1} + Bu_{n-1}$$

(B.12)

$$P_n^- = AP_{n-1}A^T + Q_{n-1}$$

The measurement update equations:

$$K_n = P_n^- H_n{}^T (H_n P_n^- H_n{}^T + R_n)^{-1}$$

$$\hat{c}_n^+ = \hat{c}_n^- + K_n(z_n - H_n\hat{c}_n^-)$$

(B.13)

$$P_n^+ = (I - K_n H_n)P_n^- \quad .$$

After each time and measurement update, the whole process is repeated with the previous posteriori estimates used to project or predict the new priori estimates.

Here, an example of the use of Kalman filter is presented (see Pinder, 2010) for demonstration. Consider the case of four a priori (model derived) values of concentration as shown in Figure B.1, where $c_1 = 4.5$, $c_2 = 4.0$, $c_3 = 4.5$ and $c_4 = 5.0$ with covariance matrix

$$[P] = \begin{bmatrix} 0.67 & -0.12 & -0.36 & -0.16 \\ -0.12 & 0.67 & 0.19 & -0.34 \\ -0.36 & 0.19 & 1.5 & 0.30 \\ -0.16 & -0.34 & 0.30 & 0.67 \end{bmatrix} \quad .$$

Let us now see how we can accommodate two new measurements, namely $z_1 = c_2 = 5$ and $z_2 = c_4 = 6$ so as to obtain an updated estimate of these values in accordance with the following steps.

1. First we define the sampling matrix H,

$$H = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} .$$

2. Then we have

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ 5 \\ c_3 \\ 6 \end{bmatrix} = \begin{bmatrix} 5 \\ 6 \end{bmatrix} .$$

3. Assume the values of the measurement noise are given by

$$v = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

with correlation

$$R = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} .$$

4. The Kalman filter gain is given by substituting the above parameters into Equation (B.9)

$$K = \begin{bmatrix} 0.67 & -0.12 & -0.36 & -0.16 \\ -0.12 & 0.67 & 0.19 & -0.34 \\ -0.36 & 0.19 & 1.5 & 0.30 \\ -0.16 & -0.34 & 0.30 & 0.67 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}$$
$$\times \left(\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0.67 & -0.12 & -0.36 & -0.16 \\ -0.12 & 0.67 & 0.19 & -0.34 \\ -0.36 & 0.19 & 1.5 & 0.30 \\ -0.16 & -0.34 & 0.30 & 0.67 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right)^{-1}$$

which yields,

$$K = \begin{bmatrix} -0.12 & -0.16 \\ 0.67 & -0.34 \\ 0.19 & 0.30 \\ -0.34 & 0.67 \end{bmatrix} \times \left(\begin{bmatrix} -0.12 & 0.67 & 0.19 & -0.34 \\ -0.16 & -0.34 & 0.30 & 0.67 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right)^{-1}$$

$$= \begin{bmatrix} -0.12 & -0.16 \\ 0.67 & -0.34 \\ 0.19 & 0.30 \\ -0.34 & 0.67 \end{bmatrix} \times \left(\begin{bmatrix} 0.67 & -0.34 \\ -0.34 & 0.67 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right)^{-1}$$

$$= \begin{bmatrix} -0.12 & -0.16 \\ 0.67 & -0.34 \\ 0.19 & 0.30 \\ -0.34 & 0.67 \end{bmatrix} \times \left(\begin{bmatrix} 1.67 & -0.34 \\ -0.34 & 1.67 \end{bmatrix} \right)^{-1}$$

$$= \begin{bmatrix} -0.12 & -0.16 \\ 0.67 & -0.34 \\ 0.19 & 0.30 \\ -0.34 & 0.67 \end{bmatrix} \times \frac{1}{(1.67^2 - 0.34^2)} \begin{bmatrix} 1.67 & -0.34 \\ -0.34 & 1.67 \end{bmatrix}$$

$$= \begin{bmatrix} -0.09 & -0.12 \\ 0.38 & -0.13 \\ 0.16 & 0.21 \\ -0.13 & 0.37 \end{bmatrix}.$$

5. The update of the concentration is given by Equation (B.7) based on

$$\hat{c}^- = \begin{bmatrix} 4.5 \\ 4.0 \\ 4.5 \\ 5.0 \end{bmatrix}$$

thus the updated values of the concentration are

$$\begin{bmatrix} \hat{c}_1^+ \\ \hat{c}_2^+ \\ \hat{c}_3^+ \\ \hat{c}_4^+ \end{bmatrix} = \begin{bmatrix} 4.5 \\ 4.0 \\ 4.5 \\ 5.0 \end{bmatrix} + \begin{bmatrix} -0.09 & -0.12 \\ 0.38 & -0.13 \\ 0.16 & 0.21 \\ -0.13 & 0.37 \end{bmatrix} \left(\begin{bmatrix} 5 \\ 6 \end{bmatrix} - \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 4.5 \\ 4.0 \\ 4.5 \\ 5.0 \end{bmatrix} \right)$$

$$= \begin{bmatrix} 4.5 \\ 4.0 \\ 4.5 \\ 5.0 \end{bmatrix} + \begin{bmatrix} -0.09 & -0.12 \\ 0.38 & -0.13 \\ 0.16 & 0.21 \\ -0.13 & 0.37 \end{bmatrix} \left(\begin{bmatrix} 5 \\ 6 \end{bmatrix} - \begin{bmatrix} 4 \\ 5 \end{bmatrix} \right) = \begin{bmatrix} 4.5 \\ 4.0 \\ 4.5 \\ 5.0 \end{bmatrix} + \begin{bmatrix} -0.09 & -0.12 \\ 0.38 & -0.13 \\ 0.16 & 0.21 \\ -0.13 & 0.37 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 4.5 \\ 4.0 \\ 4.5 \\ 5.0 \end{bmatrix} + \begin{bmatrix} -0.21 \\ 0.25 \\ 0.37 \\ 0.24 \end{bmatrix} = \begin{bmatrix} 4.29 \\ 4.25 \\ 4.87 \\ 5.24 \end{bmatrix} .$$

6. Next we update the covariance matrix using Equation (B.8), which in our case is

$$P^+ = \left(\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} -0.09 & -0.12 \\ 0.38 & -0.13 \\ 0.16 & 0.21 \\ -0.13 & 0.37 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \right) \begin{bmatrix} 0.67 & -0.12 & -0.36 & -0.16 \\ -0.12 & 0.67 & 0.19 & -0.34 \\ -0.36 & 0.19 & 1.5 & 0.30 \\ -0.16 & -0.34 & 0.30 & 0.67 \end{bmatrix}$$

$$= \left(\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \right.$$

$$\left. - \begin{bmatrix} 0 & -0.09 & 0 & -0.12 \\ 0 & 0.38 & 0 & -0.13 \\ 0 & 0.16 & 0 & 0.21 \\ 0 & -0.13 & 0 & 0.37 \end{bmatrix} \right) \begin{bmatrix} 0.67 & -0.12 & -0.36 & -0.16 \\ -0.12 & 0.67 & 0.19 & -0.34 \\ -0.36 & 0.19 & 1.5 & 0.30 \\ -0.16 & -0.34 & 0.30 & 0.67 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & -0.09 & 0 & -0.12 \\ 0 & 0.62 & 0 & 0.13 \\ 0 & -0.16 & 1 & -0.21 \\ 0 & 0.13 & 0 & 0.63 \end{bmatrix} \begin{bmatrix} 0.67 & -0.12 & -0.36 & -0.16 \\ -0.12 & 0.67 & 0.19 & -0.34 \\ -0.36 & 0.19 & 1.5 & 0.30 \\ -0.16 & -0.34 & 0.30 & 0.67 \end{bmatrix}$$

$$= \begin{bmatrix} 0.68 & -0.019 & -0.38 & -0.27 \\ -0.10 & 0.37 & 0.16 & -0.12 \\ -0.31 & 0.15 & 1.40 & 0.21 \\ -0.12 & -0.13 & 0.21 & 0.38 \end{bmatrix} .$$
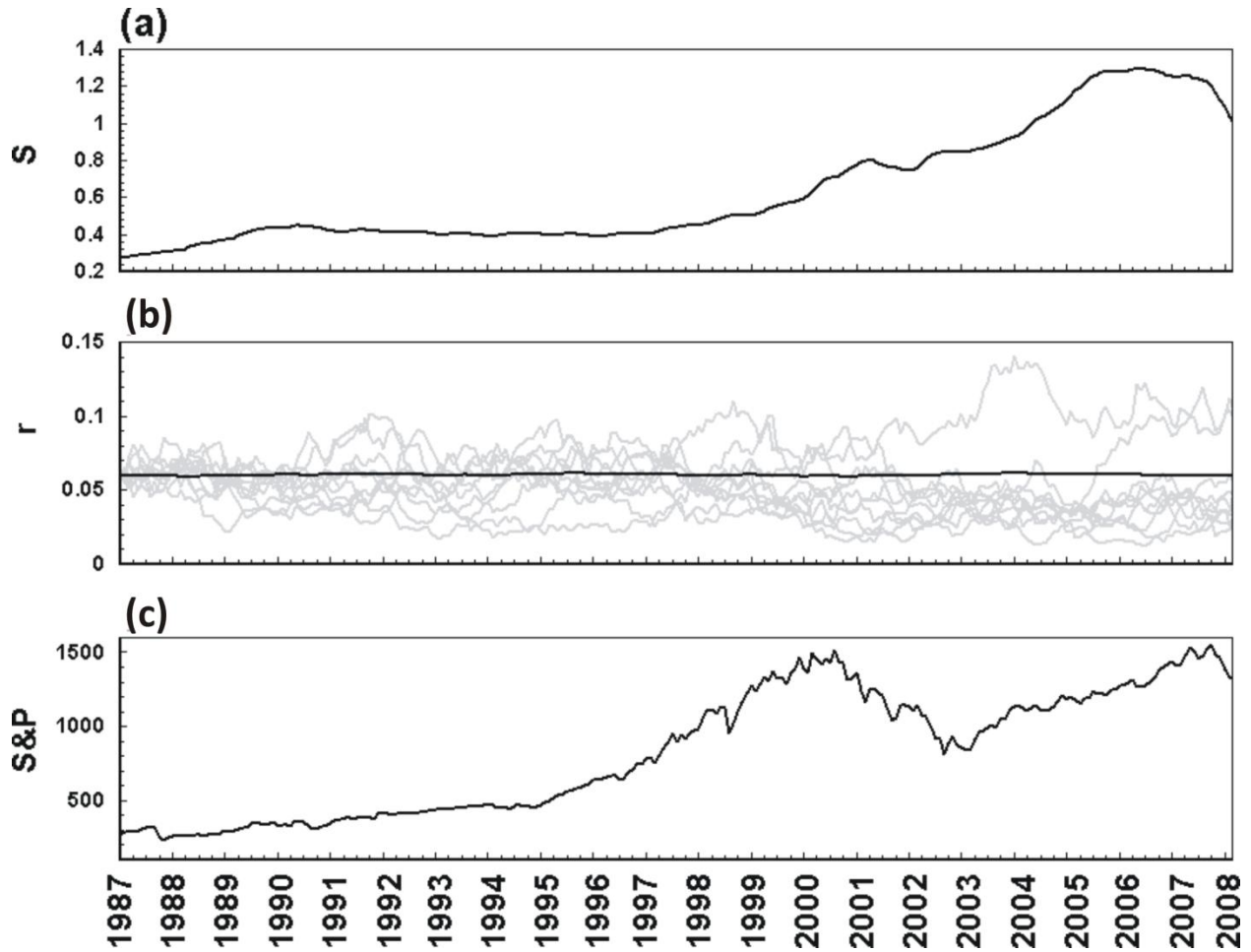
If one now proceeds to take another sample, the entire state variable and covariance sequence will be repeated.

Figure B.1: Spatial correlation of model derived values of concentration.

Appendix C Supporting Figure For Chapter 4

Figure C.1: (a) the US national home price index, S_i normalized by the index of March 2008 (source: http://www2.standardandpoors.com), (b) the nominal annual interest rate inferred from risk-free US Treasury securities, r_N; Monte Carlo simulations of short term interest rate using parameters for the CIR model in Table 4.1. Interest rate at the first quarter in 1987 was assumed to 0.06 and Δt of $1/4$ year was used for the Monte Carlo simulation. The black solid line represents the mean of 1000 realizations at each time, and (c) the general performance of the US market using the S&P 500 as a proxy. The closing price on the last trading day of the month was used (source: http://finance.yahoo.com/q?s=%5EGSPC). Cited from Figure 1 of Yu (2009).

Bibliography

Abbaspour, K. C., Van Genuchten, M. T., Schulin, R. and Schläppi, E. (1997). A sequential uncertainty domain inverse procedure for estimating subsurface flow and transport parameters. *Water Resources Research, 33*(8), 1879-1892.

Abbaspour, K. C., Schulin, R., Van Genuchten, M. T. and Schläppi, E. (1998). Procedures for uncertainty analyses applied to a landfill leachate plume. *Ground Water, 36*(6), 874-883.

AboutRemediation[TM] (2012). Brownfields redevelopment toolbox[TM]. Available at http://www.aboutremediation.com/Toolbox/Introduction.asp (accessed April 2012).

Abreu, L. D. V. and Johnson, P. C. (2005). Effect of vapour source – building separation and building construction on soil vapour intrusion as studied with a three-dimensional numerical model. *Environmental Science and Technology, 39*, 4550-4561.

Abreu, L. D. V. and Johnson, P. C. (2006). Simulating the effect of aerobic biodegradation on soil vapour intrusion into buildings: Influence of degradation rate, source concentration, and depth. *Environmental Science and Technology, 40*, 2304-2315.

Aller, L., Bennett, T. W., Hackett, G., Petty, R. J., Lehr, J. H., Sedoris, H., Nielsen, D. M. and Denne, J. E. (1989). Handbook of Suggested Practices for the Design and Installation of Ground-Water Monitoring Wells: Technology Support Centre, Environmental Monitoring Systems Laboratory, EPA/600/4-89/034, United States Environmental Protection Agency.

Analytic Sciences Corporation, technical Staff, Gelb, A. (1974). Applied optimal estimation. *Cambridge, Massachusetts: M.I.T. Press.*

Andricevic, R. (1990). Cost-effective network design for groundwater flow monitoring. *Stochastic Hydrology and Hydraulics, 4*(1), 27-41.

Andricevic, R. (1993). Coupled withdrawal and sampling designs for groundwater supply models. *Water Resources Research, 29*(1), 5-16.

Armstrong, M. (1998). Basic linear geostatistics. Springer, Berlin 146.

ASHRAE (the American Society of Heating, Refrigerating, and Air Condition Engineers) (2009). Ventilation and infiltration. ASHRAE Handbook - Fundamentals, Chapter 16.

ASTM (American Society for Testing and Materials Standard) (2008). Standard specification for biodiesel fuel (B100) blend stock for distillate fuels, in: Annual Book of ASTM Standards, ASTM Press, West Conshohocken, Method D6751.

Back, P. -E., Rosén, L. and Norberg, T. (2007). Value of information analysis in remedial investigations. *Ambio, 36*(6), 486-493.

Balatbat, M. C. A., Lin, C.-Y. and Carmichael, D.G. (2010). Comparative performance of publicly listed construction companies: Australian evidence. Construction Management and Economics, *28*, 919-932.

Barnes, D. L. and McWhorter, D. B. (2000). Design of soil vapor extraction systems under conditions of uncertainty. *Ground Water, 38*(2), 209-217.

Barnes, R. J. (1991). The variogram sill and the sample variance. *Mathematical Geology, 23*(4), 673-678.

Béland-Pelletier, C., Fraser, M., Barker, J. and Ptak, T. (2011). Estimating contaminant mass discharge: A field comparison of the multilevel point measurement and the integral pumping investigation approaches and their uncertainties. *Journal of Contaminant Hydrology*, *122*(1-4), 63-75.

Bianchi, M., Zheng, C., Wilson, C., Tick, G. R., Liu, G. and Gorelick, S. M. (2011). Spatial connectivity in a highly heterogeneous aquifer: From cores to preferential flow paths. *Water Resources Research, 47*(5).

Blum, P. and Annable, M. D. (2008). Partial source zone removal. *Journal of Contaminant Hydrology*, *102*(1-2), 1-2.

Bockelmann, A., Zamfirescu, D., Ptak, T., Grathwohl, P. and Teutsch, G. (2003). Quantification of mass fluxes and natural attenuation rates at an industrial site with a limited monitoring network: A case study. *Journal of Contaminant Hydrology*, *60*(1-2), 97-121.

Bozkurt, O., Pennell, K. G. and Suuberg, E. M. (2009). Simulation of the Vapor intrusion process for nonhomogeneous soils using a three-dimensional numerical model. *Ground Water Monitoring and Remediation*, *29*(1), 92-104.

Bredehoeft, J. D. (2003). From models to performance assessment: The conceptualization problem. *Ground Water, 41*(5), 571-577.

Bredehoeft, J. D. (2005). The conceptualization model problem - surprise. *Hydrogeology Journal, 13*(1), 37-46.

Brooks, M. C., Wood, A. L., Annable, M. D., Hatfield, K., Cho, J. and Holbert, C., et al. (2008). Changes in contaminant mass discharge from DNAPL source mass depletion: Evaluation at two field sites. *Journal of Contaminant Hydrology, 102*(1-2), 140-153.

Cai, Z., Wilson, R. D., Cardiff, M. A. and Kitanidis, P. K. (2011). Increasing confidence in mass discharge estimates using geostatistical methods. *Ground Water, 49*(2), 197-208.

Celia, M. A. and Binning, P. (1992). A mass conservative numerical solution for two-phase flow in porous media with application to unsaturated flow. *Water Resources Research*, *28*(10), 2819–2828.

Christensen, C.V. (2001). Implied loss distributions for catastrophe insurance derivatives. In: Working paper series no. 80, Centre for analytical finance. University of Aarhus.

Conant, B. H., Gillham, R. W. and Mendoza, C. A. (1996). Vapor transport of trichloroethylene in the unsaturated zone: Field and numerical modeling investigations. *Water Resources Research, 32*(1), 9-22.

Cox, J., Ingersoll, J. and Ross, S. (1985). The term structure of interest rates. *Econometrica, 53*, 385-407.

Culver, T. B., Shoemarker, C. A. and Lion, L. W. (1991). Impact of vapour sorption on the subsurface transport of volatile organic compounds: a numerical model and analysis. *Water Resources Research, 27*(9), 2259–2270.

Dagan, G. (1984). Solute transport in heterogeneous porous formations. *Journal of Fluid Mechanics, 145*, 151-177.

151

Delbaen, F. and Haezendonck, J. (1989). A martingale approach to premium calculation principles in an arbitrage free market. *Insurance Mathematics and Economics, 8*(4), 269-277.

Deutsch, C. V. and Journel, A. G. (1992). GSLIB: Geostatistical Software Library and User's Guide. *Oxford University Press, Oxford.*

DeVaull, G., Ettinger, R. and Gustafson, J. (2002). Chemical vapor intrusion from soil or groundwater to indoor air: Significance of unsaturated zone biodegradation of aromatic hydrocarbons. *Soil and Sediment Contamination, 11*(4), 625-641.

Devlin, J. F. and Barker, J. F. (1996). Field investigation of nutrient pulse mixing in an in situ biostimulation experiment. *Water Resources Research, 32*(9), 2869-2877.

Dokou, Z. and Pinder, G. F. (2011). Extension and field application of an integrated DNAPL source identification algorithm that utilizes stochastic modeling and a kalman filter. *Journal of Hydrology, 398*(3-4), 277-291.

Duan, J.-.C. and Simonato, J.-G. (1999). Estimating and testing exponential e affine term structure models by Kalman filter. *Review of Quantitative Finance and Accounting, 13*, 111-135.

Eggleston, J. R., Rojstaczer, S. A. and Peirce, J. J. (1996). Identification of hydraulic conductivity structure in sand and gravel aquifers: Cape Cod data set. *Water Resources Research, 32*(5), 1209-1222.

Ek, C. W., Anisko, S. A. and Gregg, G. O. (1990). Air leakage tests of manufactured housing in the Northwest United States, in: Air change rate and airtightness in buildings, STP 1067, 152-164. M.H. Sherman, ed. American Society for Testing and Materials, West Conshohocken, PA.

Essaid, H. I., and Hess, K. M. (1993). Monte Carlo simulations of multiphase flow incorporating spatial variability of hydraulic properties. *Groundwater, 31*(1), 123–143.

Evensen, G. (2003). The ensemble kalman filter: theoretical formulation and practical implementation. *Ocean Dynamics*, *53*(4), 343-367.

Ezzedine, S. and Rubin, Y. (1997). Analysis of the Cape Cod tracer data. *Water Resources Research, 33*(1), 1-11.

Feehley, C. E., Zheng, C. and Molz, F. J. (2000). A dual-domain mass transfer approach for modeling solute transport in heterogeneous aquifers: Application to the macrodispersion experiment (MADE) site. *Water Resources Research, 36*(9), 2501-2515.

Fitzpatrick, N. A. and Fitzgerald, J. J. (2002). An evaluation of vapor intrusion into buildings through a study of field data. *Soil and Sediment Contamination, 11*(4), 603-623.

Forsyth, P. A. (1993). A positivity preserving method for simulation of steam injection for NAPL site remediation. *Advances in Water Resources, 16*(2), 351-370.

Forsyth, P. A., Unger, A. J. A. and Sudicky, E. A. (1998). Nonlinear iteration methods for nonequilibrium multiphase subsurface flow. *Advances in Water Resources, 21*(6), 433-449.

Freeze, R., James, B., Massmann, J., Sperling, T. and Smith, L. (1992). Hydrogeological decision analysis: 4. the concept of data worth and its use in the development of site investigation strategies. *Ground Water, 30*(4), 574-588.

Freeze, R. A., Massmann, H., Smith, L., Sperling, T. and James, B. (1990). Hydrogeological decision analysis: 1. A framework. *Ground Water, 28*(5), 738-766.

Gao, H., Wang, J. and Zhao, P. (1996). The updated kriging variance and optimal sample design. *Mathematical Geology, 28*(3), 295-313.

Garbesl, K., Sextro, R. G., Fisk, W. J., Modera, M. P. and Revzan, K. L. (1993). Soil-Gas entry into an experimental basement: model measurement comparisons and seasonal effects. *Environmental Science and Technology, 27*, 466-473.

Goodman, L.A. (1962). The variance of the product of k random variables. *Journal of the American Statistical Association, 57*(297), 54-60.

Graham, W. D. and McLaughlin, D. B. (1989). Stochastic analysis of nonstationary subsurface solute transport. 2. conditional moments. *Water Resources Research, 25*(11), 2331-2355.

Graham, W. D. and McLaughlin, D. B. (1991). A stochastic model of solute transport in groundwater: Application to the Borden, Ontario, tracer test. *Water Resources Research, 27*(6), 1345-1359.

Guadagnini, L., Guadagnini, A. and Tartakovsky, D. M. (2004). Probabilistic reconstruction of geologic facies. *Journal of Hydrology, 294*(1-3), 57-67.

Hardle, W. and Lopez Cabrera, B. (2007). Calibrating CAT bonds to Mexican earthquakes. In: Proceedings of the 101st EAAE (European Association of Agricultural Economists) Seminar 'Management of Climate Risks in Agriculture', Berlin Germany. Available at http://ageconsearch.umn.edu/bitstream/9265/1/sp07ha01.pdf (accessed November 2012).

Hayek, C. and Ghanem, R. (2002). Impact of uncertainty in catastrophe losses on insurance derivatives. 15th ASCE Engineering Mechanics Conference, 2-5 June 2002. Columbia University, New York, USA.

Heine, G. W. (1986). A controlled study of some two-dimensional interpolation methods. *COGS Computer Contributions 2*(2), 60–72.

Heinz, J. (2001). Sedimentary geology of glacial and periglacial gravel bodies (SW-Germany): Dynamic stratigraphy and aquifer sedimentology. Ph.D. dissertation, Institute for Geology and Paleontology, University of Tübingen.

Heinz, J., Kleineidam, S., Teutsch, G. and Aigner, T. (2003). Heterogeneity patterns of quaternary glaciofluvial gravel bodies (SW-Germany): Application to hydrogeology. *Sedimentary Geology, 158*(1-2), 1-23.

Herrera, G. S. and Pinder, G. F. (2005). Space-time optimization of groundwater quality sampling networks. *Water Resources Research, 41*(12), 1-15.

Hers, I., Zapf-Gilje, R., Johnson, P. C. and Li, L. (2003). Evaluation of the Johnson and Ettinger model for prediction of indoor air quality. *Ground Water Monitoring and Remediation, 23*(1), 62-76.

Hess, K. M., Davis, J. A., Kent, D. B. and Coston, J. A. (2002). Multispecies reactive tracer test in an aquifer with spatially variable chemical conditions, Cape Cod, Massachusetts: Dispersive transport of bromide and nickel. *Water Resources Research, 38*(8), 361-3617.

Hess, K.M., Wolf, S.H. and Celia, M.A. (1992). Large scale natural gradient tracer test in sand and gravel, Cape Cod, Massachusetts: 3. Hydraulic conductivity and calculated macrodispersivities. *Water Resources Research, 28*(8), 2011–2027.

Ho, Y. C. and Lee, R. C. K. (1964). A Bayesian approach to problems in stochastic estimation and control. *IEEE Transactions on Automatic Control, 9*(4), 333–339.

Hyndman, D. W., Day-Lewis, F. D. and Singha, K. (2007). Subsurface hydrology: data integration for properties and processes. American Geophysical Union, Washington, DC.

Insley, M. C. and Wirjanto, T. S. (2010). Contrasting two approaches in real options valuation: contingent claims versus dynamic programming. *Journal of Forest Economics, 16*(2), 157-176.

Interstate Technology & Regulatory Council (ITRC) (2007a). Vapor Intrusion Pathway: A Practical Guideline. VI-1. Washington, D.C.: Interstate Technology & Regulatory Council, Vapor Intrusion Team. Available at http://www.itrcweb.org/Documents/VI-1.pdf (accessed February 2012).

Interstate Technology & Regulatory Council (ITRC) (2007b). Vapor Intrusion Pathway: Investigative Approaches for Typical Scenarios. VI-1A. Washington, D.C.: Interstate Technology & Regulatory Council, Vapor Intrusion Team. Available at http://www.itrcweb.org/Documents/VI-1A.pdf (accessed February 2012).

James, B. R. and Freeze, R. A. (1993). The worth of data in predicting aquitard continuity in hydrogeological design. *Water Resources Research, 29*(7), 2049-2065.

James, B. R. and Gorelick, S. M. (1994). When enough is enough: The worth of monitoring data in aquifer remediation design. *Water Resources Research, 30*(12), 3499-3514.

James, B. R., Gwo, J. -P. and Toran, L. (1996a). Risk-cost decision framework for aquifer remediation design. *Journal of Water Resources Planning and Management, 122*(6), 414-420.

James, B. R., Huff, D. D., Trabalka, J. R., Ketelle, R. H. and Rightmire, C. T. (1996b). Allocation of environmental remediation funds using economic risk-cost-benefit analysis: A case study. *Ground Water Monitoring and Remediation, 16*(4), 95-105.

Johnson, P. C. and Ettinger, R. A. (1991). Heuristic model for predicting the intrusion rate of contaminant vapors into buildings. *Environmental Science and Technology, 25*, 1445-1452.

Johnson, P. C., Luo, E., Dahlen, P. and Holton, C. (2011). Temporal changes in VI behavior: considerations for pathway assessment. Indoor air vapour intrusion database (IAVI), 07 - Approaches to sampling frequency and data collection, and impacts on conclusions, Category: 00 – EPA/AEHS, San Diego – workshop: Addressing regulatory challenges in vapour intrusion.

Johnston, J. E. and Gibson, J. M. (2011). Probabilistic approach to estimating indoor air concentrations of chlorinated volatile organic compounds from contaminated groundwater: A case study in San Antonio, Texas. *Environmental Science and Technology, 45*(3), 1007-1013.

Journel, A. G. and Huijbregts, C. J. (1992). Mining Geostatistics. *Academic Press, New York.*

Khadam, I. and Kaluarachchi, J. J. (2003). Applicability of risk-based management and the need for risk-based economic decision analysis at hazardous waste contaminated sites. *Environment International, 29*(4), 503-519.

King, M. W. G., Barker, J. F., Devlin, J. F. and Butler, B. J. (1999). Migration and natural fate of a coal tar creosote plume. 2. Mass balance and biodegradation indicators. *Journal of Contaminant Hydrology, 39*(3-4), 281-307.

Korte, N., Wagner, S. and Nyquist, J. (1992). Choosing an appropriate soil-gas survey method. *Environmental Monitoring and Assessment, 21*(1), 27-35.

Kram, M. L., Keller, A. A., Rossabi, J. and Everett, L. G. (2002). DNAPL characterization methods and approaches, part 2: Cost comparisons. *Ground Water Monitoring and Remediation, 22*(1), 46-61.

Laird, C. L., Biegler, L. T., van Bloemen Waanders, B. G. and Bartlett, R. A. (2005). Contamination source determination for water networks. *Journal of Water Resources Planning and Management*, *131*(2), 125-134.

Lam, N. S. -N. (1983). Spatial interpolation methods: A review. *American Cartographer, 10*(2), 129-149.

Laryea, S. and Hughes, W. (2008). How contractors price risk in bids: Theory and practice. *Construction Management and Economics, 26*(9), 911-924.

Leblanc, D. R., Garabedian, S. P., Hess, K. M., Gelhar, L. W., Quadri, R. D., Stollenwerk, K. G. and Wood, W. W. (1991). Large-scale natural gradient tracer test in sand and gravel, Cape Cod, Massachusetts, 1. Experimental design and observed tracer movement. *Water Resources Research, 27*(5), 895-910.

Lesage, S. and Jackson, R. E. (1992). Groundwater contamination and analysis at hazardous waste sites. CRC Press.

Leverett, M. C. (1941). Capillary behaviour in porous solids. Transactions of the American Institute of Mining. *Metallurgical and Petroleum Engineers, 142*, 152–169.

Loaiciga, H. A. (1989). An optimization approach for groundwater quality monitoring network design. *Water Resources Research, 25*(8), 1771-1782.

Longino, B. (2005). DNAPL source zone characterization and remediation: an ongoing challenge. Groundwater Resources Association of California (GRAC). Available at http://www.grac.org/dnaplsummary.asp (accessed February 2012).

Luo, H., Dahlen, P. and Johnson, P. C. (2010). Hydrocarbon and oxygen transport in the vicinity of a building overlying a NAPL source zone. Paper presented at the Air and Waste Management Association - Vapor Intrusion 2010, 1, 155-185.

Mackay, D. M., De Sieyes, N. R., Einarson, M. D., Feris, K. P., Pappas, A. A., Wood, I. A., et al. (2006). Impact of ethanol on the natural attenuation of benzene, toluene, and o-xylene in a normally sulfate-reducing aquifer. *Environmental Science and Technology, 40*(19), 6123-6130.

Mackay, D. M., Freyberg, D. L., Roberts, P. V. and Cherry, J. A. (1986). Natural gradient experiment on solute transport in a sand aquifer: 1. Approach and overview of plume movement. *Water Resources Research, 22*(13), 2017-2029.

Maji, R., Sudicky, E. A., Panday, S. and Teutsch, G. (2006). Transition probability/Markov chain analyses of DNAPL source zones and plumes. *Ground Water, 44*(6), 853-863.

Marrin, D. L. and Thompson, G. M. (1987). Gaseous behavior of TCE overlying a contaminated aquifer. *Ground Water, 25*(1 1), 21-27.

Marryott, R. A. (1996). Optimal ground-water remediation design using multiple control technologies. *Ground Water, 34*(3), 425-433.

Massey, F. J. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association, 46*(253), 68–78.

Massmann, J. and Farrier, D. F. (1992). Effects of atmospheric pressure on gas transport in vadose zone. *Water Resources Research, 28*(3), 777-791.

Massmann, J. and Freeze, R. A. (1987a). Groundwater contamination from waste management sites: The interaction between risk-based engineering design and regulatory policy. I. methodology. *Water Resources Research, 23*(2), 351-367.

Massmann, J. and Freeze, R. A. (1987b). Groundwater contamination from waste management sites: The interaction between risk-based engineering design and regulatory policy. II. Results. *Water Resources Research, 23*(2), 368-380.

Massmann, J., Freeze, R. A., Smith, L., Sperling, T. and James, B. (1991). Hydrogeological decision analysis. 2. Applications to ground-water contamination. *Ground Water, 29*(4), 536-548.

McCarthy, L. (2001). Brownfield redevelopment: A resource guide for Toledo and other Ohio governments, developers, and communities: A report for the urban affairs centre. Available at http://uac.utoledo.edu/Publications/brownfield-redevelopment.pdf (accessed April 2012).

McCarthy, K. A. and Johnson, R. L. (1993). Transport of volatile organic compounds across the capillary fringe. *Water Resources Research, 29*(6), 1675-1683.

McGrath, W. A. and Pinder, G. F. (2003). Search strategy for groundwater contaminant plume delineation. *Water Resources Research, 39*(10), SBH141-SBH1412.

Mendoza, C. A. and Frind, E. O. (1990a). Advective-dispersive transport of dense organic vapors in the unsaturated zone 1. Model development. *Water Resources Research, 26*(3), 379–387.

Mendoza, C. A. and Frind, E. O. (1990b). Advective-dispersive transport of dense organic vapors in the unsaturated zone 2. Sensitivity analysis. *Water Resources Research, 26*(3), 388–398.

Meyer, P. B., Chilton, K. M. and KY, L. (1997). Environmental insurance for brownfields redevelopment: A feasibility study. Available at http://www.huduser.org/Publications/pdf/envins.pdf (accessed April 2012).

Miller, L. H. (1956). Table of percentage points of Kolmogorov statistics. *Journal of the American Statistical Association, 51*(273), 111–121.

Millington, R. J. (1959). Gas diffusion in porous media. Science, 130: 100-102, Washington, DC.

Mills, W. B., Liu, S., Rigby, M. C. and Brenner, D. (2007). Time-variable simulation of soil vapor intrusion into a building with a combined crawl space and basement. *Environmental Science and Technology, 41*(14), 4993-5001.

Moller, T. (2002). On valuation and risk management at the interface of insurance and finance. *British Actuarial Journal, 8*(4), 787-827.

Møller, T. (2001). On transformations of actuarial valuation principles. *Insurance: Mathematics and Economics, 28*(3), 281-303.

Mueller, G. R. (2005). Brownfields capital - unlocking value in environmental redevelopment. *Journal of Real Estate Portfolio Management, 11*(1), 81-92.

National Round Table on the Environment and the Economy (NRTEE) Canada, 2003. Cleaning up the Past, Building the Future: A National Brownfield Redevelopment Strategy for Canada. Available at http://nrtee-trnee.ca/wp-content/uploads/2011/06/brownfield-redevelopment-strategy-eng.pdf (accessed February 2012).

Neufville, R. and King, D. (1991). Risk and need-for-work premiums in contactor bidding. ASCE Journal of Construction Engineering and Management, *117*(4), 659-673.

Neuman, S. P. (2003). Maximum likelihood Bayesian averaging of uncertain model predictions. *Stochastic Environmental Research and Risk Assessment, 17*(5), 291-305.

Neuman, S. P., Riva, M. and Guadagnini, A. (2008). On the geostatistical characterization of hierarchical media. *Water Resources Research, 44*(2), doi: 10.1029/2007WR006228.

New York State Department of Health (NYSDOH) (2005). Fact sheet: February 2005 "Trichloroethene (TCE) in indoor and outdoor air". Available at http://www.health.state.ny.us/environmental/investigations/soil_gas/svi_guidance/docs/fs_tce.pdf.

Norberg, T. and Rosén, L. (2006). Calculating the optimal number of contaminant samples by means of data worth analysis. *Environmetrics, 17*(7), 705-719.

Office of Environmental Health Hazard Assessment – California Environmental Protection Agency (OEHHA-CEPA) (2007). Occupational health hazard risk assessment project for California: Identification of chemicals of concern, possible risk assessment methods, and examples of health

protective occupational air concentrations. Reproductive and Cancer Hazard Assessment Branch, December. Available at http://www.dhs.ca.gov/ohb/HESIS/risksummary.pdf.

Olea, R. A. (1974). Optimal contour mapping using universal kriging. *Journal of Geophysical Research, 79*(5), 695-702.

Ontario Ministry of Municipal Affairs and Housing (OMMAH) (2007). A Practical Guide to Brownfields Redevelopment in Ontario. Available at http://www.mah.gov.on.ca/AssetFactory.aspx?did=4995 (accessed February 2012).

Ontario Ministry of the Environment (OME) (1997). Guideline for use at contaminated sites in Ontario. Ministry of Environment and Energy.

Pantazidou, M. and Liu, K. (2008). DNAPL distribution in the source zone: Effect of soil structure and uncertainty reduction with increased sampling density. *Journal of Contaminant Hydrology, 96*(1-4), 169-186.

Parker, J. C. (2003). Modeling volatile chemical transport, biodecay, and emission to indoor air. *Ground Water Monitoring and Remediation, 23*(1), 107-120.

Pelmulder, S. D., Yeh, W. W. -. and Kastenberg, W. E. (1996). Regional scale framework for modeling water resources and health risk problems. *Water Resources Research, 32*(6), 1851-1861.

Pinder, G. (2010). Kalman filter, lecture 3. Advanced Course in Optimization of Subsurface Source Finding, Monitoring Well Network, and Remedial Design: HydroGeoLogic, Inc., Reston, Virginia.

Pratt, S. P. and Grabowski, R. J. (2008). Cost of capital: Applications and examples (3rd edition). In *Cost of capital basics* (pp. 3–65). Hoboken, New Jersey: Wiley.

Provoost, J., Bosman, A., Reijnders, L., Bronders, J., Touchant, K. and Swartjes, F. (2010). Vapour intrusion from the vadose zone-seven algorithms compared. *Journal of Soils and Sediments, 10*(3), 473-483.

Provoost, J., Reijnders, L., Swartjes, F., Bronders, J., Seuntjens, P. and Lijzen, J. (2009). Accuracy of seven vapour intrusion algorithms for VOC in groundwater. *Journal of Soils and Sediments, 9*(1), 62-73.

Reddy, K. R., Adams, J. A. and Richardson, C. (1999). Potential technologies for remediation of brownfields. *Practice Periodical of Hazardous, Toxic, and Radioactive Waste Management*, *3*(2), 19821.

Refsgaard, J. C., van der Sluijs, J. P., Brown, J. and van der Keur, P. (2006). A framework for dealing with uncertainty due to model structure error. *Advances in Water Resources, 29*(11), 1586-1597.

Reichard, E. G. and Evans, J. S. (1989). Assessing the value of hydrogeologic information for risk-based remedial action decisions. *Water Resources Research, 25*(7), 1451-1460.

Riley, W. J., Robinson, A. L., Gadgil, A. J. and Nazaroff, W. W. (1999). Effects of variable wind speed and direction on radon transport from soil into buildings: Model development and exploratory results. *Atmospheric Environment, 33*(14), 2157-2168.

Rivett, M. O. (1995). Soil-gas signatures from volatile chlorinated solvents: Borden field experiments. *Ground Water, 33*(1), 84-98.

Rivett, M. O. and Cherry, J. A. (1991). The effectiveness of soil gas surveys in delineation of groundwater contamination: controlled experiments at the Borden field site. Proceedings of the Conference on Petroleum Hydrocarbons and Organic Chemicals in Groundwater. National Water Well Association, Houston, Texas, pp. 107–124. Nov. 20–22.

Rivett, M. O., Feenstra, S. and Cherry, J. A. (2001). A controlled field experiment on groundwater contamination by a multicomponent DNAPL: Creation of the emplaced-source and overview of dissolved plume development. *Journal of Contaminant Hydrology, 49*(1-2), 111-149.

Robin, M. J. L., Gutjahr, A. L., Sudicky, E. A. and Wilson, J. L. (1993). Cross-correlated random field generation with the direct Fourier transform method. *Water Resources Research, 29*(7), 2385-2397.

Robinson, N. I. and Turczynowicz, L. (2005). One- and three-dimensional soil transportation models for volatiles migrating from soils to house interiors. *Transport in Porous Media, 59*(3), 301-323.

Rogerson, P. A., Delmelle, E., Batta, R., Akella, M., Blatt, A. and Wilson, G. (2004). Optimal sampling design for variables with varying spatial importance. *Geographical Analysis, 36*(2), 177-194.

Royle, A. G., Clausen, F. L. and Frederiksen, P. (1981). Practical universal kriging and automatic contouring. *Geo-Processing, 1*(4), 377-394.

Sale, T. C., Zimbron, J. A. and Dandy, D. S. (2008). Effects of reduced contaminant loading on downgradient water quality in an idealized two-layer granular porous media. *Journal of Contaminant Hydrology, 102*(1-2), 72-85.

Schirmer, M. and Barker, J. F. (1998). A study of long-term MTBE attenuation in the Borden aquifer, Ontario, Canada. *Ground Water Monitoring and Remediation, 18*(2), 113-122.

Schreuder, W. A. (2006). Uncertainty Approach to the Johnson and Ettinger Vapor Intrusion Model. Principia Mathematica, Lakewood, CO.

Schwartz, E. (1994). Review of investment under uncertainty. *The Journal of Finance, 49*(5), 1924-1928.

Schwede, R. L., Cirpka, O. A., Nowak, W. and Neuweiler, I. (2008). Impact of sampling volume on the probability density function of steady state concentration. *Water Resources Research, 44*, W12433, doi: 10.1029/2007WR006668.

Siegel, J. and Schwartz, J. D. (2004). The long term returns on the original S&P 500 firms. The Wharton School, University of Pennsylvania. Available at http://finance.wharton.upenn.edu/~rlwctr/papers/0429.pdf (accessed November 2012).

Slough, K. J., Sudicky, E. A. and Forsyth, P. A. (1999). Grid refinement for modeling multiphase flow in discretely fractured porous media. *Advances in Water Resources, 23*(3), 261-269.

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

Smedes, H. W., Spycher, N. and Leonard Allen, R. (1993). Case history of one of the few successful superfund remediation sites: A site at Salinas, California, USA. *Engineering Geology, 34*(3-4), 189-203.

Solomon, D. K., Poreda, R. J., Schiff, S. L. and Cherry, J. A. (1992). Tritium and Helium 3 as groundwater age tracers in the Borden aquifer. *Water Resource Research, 28* (3), 741–755.

Sondermann, D. (1991). Reinsurance in arbitrage-free markets. *Insurance Mathematics and Economics, 10*(3), 191-202.

Stone, H. L. (1973). Estimation of three-phase relative permeability and residual oil data. *Journal of Canadian Petroleum Technology 12*, 53–61.

Sudicky, E. A. (1986). Natural gradient experiment on solute transport in a sand aquifer: spatial variability of hydraulic conductivity and its role in the dispersion process. *Water Resources Research, 22*(13), 2069-2082.

Thomson, N. R., Sykes, J. F. and Van Vliet, D. (1997). A numerical investigation into factors affecting gas and aqueous phase plumes in the subsurface. *Journal of Contaminant Hydrology, 28*, 39-70.

Torresetti, R., Brigo, D. and Pallavicini, A. (2007). Risk neutral versus objective loss distribution and CDO tranches valuation. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=900784 (accessed November 2012).

Turcke, M. A. and Kueper, B. H. (1996). Geostatistical analysis of the Borden aquifer hydraulic conductivity field. *Journal of Hydrology, 178*(1-4), 223-240.

UK Environment Agency (2003). Brownfield Land Redevelopment: Position Statement. May 2003. Available at http://www.environment-Agency.gov.uk/static/documents/Research/brownfield_land_908146.pdf (accessed February 2012).

Unger, A. J. A., Forsyth, P. A. and Sudicky, E. A. (1996). Variable spatial and temporal weighting schemes for use in multi-phase compositional problems. *Advances in Water Resources*, *19*(1), 1-27.

Unger, A. J. A., Younis, R., Rehan, R., Yu, S., Budimir, F., Nazir, A., Knight, M. and Darrall, R. (2011). Forecasting the unit cost of watermain and sanitary sewer capital works. Construction Management and Economics (submitted).

United States Air Force (USAF) (2006). Guide for the Assessment of the Vapor Intrusion Pathway, prepared by Douglas N. Cox, Mitretek Systems, Inc. for Air Force Institute for Operational Health, Risk Analysis Directorate, Risk Assessment Division.

United States Environmental Protection Agency (US EPA) (1999). A sustainable brownfields model framework. January 1999. Publication Number: EPA-500-R-99-001.

United States Environmental Protection Agency (US EPA) (1998, August). Environmental technology verification program verification statement. EPA-VS-SCM-22, Office of research and development, Washington. Available at http://www.epa.gov/etv/pubs/01_vs_quadrel.pdf.

Van Groenigen, J. W. (2000). The influence of variogram parameters on optimal sampling schemes for mapping by kriging. *Geoderma, 97*(3-4), 223-236.

Wade VanLandingham, H., The stormstown group and Meyer, P. B. (2002). Public strategies for cost-effective community brownfield redevelopment. Southeast Regional Environmental Finance Center, EPA Region 4, University of Louisville. Available at http://cepm.louisville.edu/Pubs_WPapers/practiceguides/PG1.pdf.

Ward, C. H., Cherry, J. A. and Scalf, M. R. (1997). Subsurface restoration. CRC Press.

Wagner, B. J. (1999). Evaluating data worth for ground-water management under uncertainty. *Journal of Water Resources Planning and Management, 125*(5), 281-288.

Welch, G. and Bishop, G. (2006). An introduction to the Kalman filter. University of North Carolina at Chapel Hill. Available at http://www.cs.unc.edu/~welch/media/pdf/kalman_intro.pdf.

Wernstedt, K., Meyer, P. B. and Yount, K. R. (2003). Insuring redevelopment at contaminated urban properties. *Public Works Management & Policy, 8*(2), 85-98.

Wong, H. S. and Yeh, W. W. -G. (2002). Uncertainty analysis in contaminated aquifer management. *Journal of Water Resources Planning and Management, 128*(1), 33-45.

Woodbury, A. D. and Sudicky, E. A. (1991). The Geostatistical Characteristics of the Borden Aquifer. *Water Resources Research*, 27 (4), 533-546.

Yao, Y., Shen, R., Pennell, K. G. and Suuberg, E. M. (2011). Comparison of the Johnson-Ettinger vapor intrusion screening model predictions with full three-dimensional model results. *Environmental Science and Technology, 45*(6), 2227-2235.

Yeh, W. W. -G. (1986). Review of parameter identification procedures in groundwater hydrology: The inverse problem. *Water Resources Research, 22*(2), 95-108.

Yount, K. R. and Meyer, P. B. (2005). Environmental Insurance Products available for Brownfields redevelopment, 2005. Northern Kentucky University.

Yu, S., Unger, A. J. A. and Parker, B. (2009). Simulating the fate and transport of TCE from groundwater to indoor air. *Journal of Contaminant Hydrology, 107*(3-4), 140-161.

Yu, S., Unger, A. J. A., Parker, B. and Kim, T. (2012). Allocating risk capital for a brownfields redevelopment project under hydrogeological and financial uncertainty. *Journal of Environmental Management*, *100*(0), 96-108.

Zhang, Y., Pinder, G. F. and Herrera, G. S. (2005). Least cost design of groundwater quality monitoring networks. *Water Resources Research, 41*(8), 1-12.

Zheng, C. and Jiao, J. J. (1998). Numerical simulation of tracer tests in heterogeneous aquifer. *Journal of Environmental Engineering, 124*(6), 510-516.

Zhou, Y., Te Stroet, C. B. M. and Van Geer, F. C. (1991). Using kalman filtering to improve and quantify the uncertainty of numerical groundwater simulations, 2. Application to monitoring network design. *Water Resources Research, 27*(8), 1995-2006.

Zou, S. and Parr, A. (1995). Optimal estimation of two dimensional contaminant transport. *Ground Water, 33*(2), 319-325.