

The Utility of Using Multiple Conceptual Models for the Design of Groundwater Remediation Systems

by

Philip Sheffield

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Civil Engineering

Waterloo, Ontario, Canada, 2014

© Philip Sheffield 2014

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

The design of pump and treat systems for groundwater remediation is often aided by numerical groundwater modelling. Model predictions are uncertain, with this uncertainty resulting from unknown parameter values, model structure and future system forcings. Researchers have begun to suggest that uncertainty in groundwater model predictions is largely dominated by structural/conceptual model uncertainty and that multiple conceptual models be developed in order to characterize this uncertainty. As regulatory bodies begin to endorse the more expensive multiple conceptual model approach, it is useful to assess whether a multiple model approach provides a significant improvement over a conventional single model approach for pump and treat system design, supplemented with a factor of safety. To investigate this question, a case study located in Tacoma, Washington which was provided by Conestoga-Rovers & Associates (CRA) was used.

Twelve conceptual models were developed to represent conceptual model uncertainty at the Tacoma, Washington site and a pump and treat system was optimally designed for each conceptual model. Each design was tested across all 12 conceptual models with no factor of safety applied, and a factor of safety of 1.5 and 2 applied. Adding a factor of safety of 1.5 decreased the risk of containment failure to 15 percent, compared to 21 percent with no factor of safety. Increasing the factor of safety from 1.5 to 2 further reduced the risk of containment failure to 9 percent, indicating that the application of a factor of safety reduces the risk of design failure at a cost directly proportional to the value of the factor of safety.

To provide a relatively independent estimate of a factor of safety approach a single “best” model developed by CRA was compared against the multiple model approach. With a factor of safety of 1.5 or greater, adequate capture was demonstrated across all 12 conceptual models. This demonstrated that in this case using the single “best” model developed by CRA with a factor of safety would have been a reasonable surrogate for a multiple model approach. This is of practical importance to engineers as it demonstrates that the a conventional single model approach may be sufficient. However, it is essential that the model used is a good model. Furthermore, a multiple model approach will likely be an excessive burden in cases such as pump and treat system design, where the cost of failure is low as the system can be adjusted during operation to respond to new data. This may not be the case for remedial systems with high capital costs such as permeable reactive barriers, which cannot be easily adjusted.

Acknowledgements

Throughout the course of this project I received assistance and guidance from many individuals who helped make this project possible. I would like to thank Steve Harris from Conestoga-Rovers & Associates, my supervisor James Craig and the rest of the research group.

I must also thank the Natural Sciences and Engineering Research Council of Canada and Conestoga-Rovers & Associates. Without their financial support this project would not have been possible.

Dedication

This is dedicated to my mother and my aunt Dianne who have supported me throughout this project, as well as my father who always tries to understand my research.

Table of Contents

List of Tables	ix
List of Figures	x
1 Introduction	1
1.1 Motivation	5
1.2 Objective	5
1.3 Outline of Thesis	6
2 Background	8
2.1 Groundwater Modelling	8
2.1.1 Overview	8
2.1.2 Contaminant Transport Modelling	11
2.1.3 Density-Dependent Groundwater Flow	12
2.1.4 Particle Tracking	13
2.2 Model Calibration	15
2.3 Model Uncertainty	19
2.3.1 Methods to Quantify Parameter Uncertainty	20
2.3.2 Methods to Quantify Conceptual Model Uncertainty	22
2.4 Pump-and-Treat System Design	27
2.5 Contribution in Relation to Previous Work	28

3	Methodology	30
3.1	Site Description	30
3.1.1	General Description	30
3.1.2	Site Contamination	31
3.1.3	Geology	33
3.1.4	Hydrology	46
3.2	Site Conceptual Models	48
3.2.1	Multiple Model Approach in the Context of Past Modelling Efforts	49
3.2.2	Model Construction	50
3.2.3	Stratigraphy	50
3.2.4	Recharge	53
3.2.5	Boundary Conditions	54
3.3	Calibration of Numerical Conceptual Models	60
3.3.1	Calibration Targets	60
3.3.2	Calibrated Method	61
3.3.3	Calibration Parameters	62
3.3.4	Calibration Results	64
3.4	Summary	68
4	Results and Discussion	70
4.1	Pump-and-Treat System Design	70
4.2	Optimal Pump-and-Treat Designs	72
4.3	Cross-Comparison of Remedial Designs	75
4.4	Factor-of-Safety Approach	78
4.5	Comparison of Models Developed in Practice to a Multiple Model Approach	80
4.6	Testing the Robustness of Remedial Designs	84
5	Conclusions and Recommendations	87

APPENDICES	91
A Scatter Plots of Simulated vs. Observed Heads	92
B Optimized Pumping Rates at Individual Wells	95
References	97

List of Tables

3.1	Initial Hydraulic Conductivity Values and Bounds Specified for Model Calibration	63
3.2	Boundary Condition Values and Bounds Used During Calibration	64
3.3	Residual Statistics	65
3.4	Mean Residual by Model Zone (ft)	67
3.5	Calibrated Parameter Values	68
4.1	Optimization Results, Ranked by Total Pumping Rate	73
4.2	Average Optimized Pumping Rates by Model Conceptualization	75
4.3	Particle Capture Results for Optimized Pumping Rates Applied in All Models	77
4.4	Cross-Conceptual Model Containment Performance with FOS of 1.5	79
4.5	Cross-Conceptual Model Containment Performance with FOS of 2	79
4.6	Optimized Pumping Rates for CRA Models	82
4.7	Capture Percentages for Optimized Pumping Rates from Models 1, 2, and 3	83
4.8	Capture Percentages for a Factor of Safety Applied to Model 1 Optimized Pumping Rates	84
B.1	Optimized pumping rates at individual wells (ft ³ /d)	96

List of Figures

3.1	Site Location (Modified from: USGS (2013))	31
3.2	Regional Plan View of Cross-Sections E-E' and F-F' (Modified from: Savoca et al. (2010))	36
3.3	Regional Cross-Sections E-E' and F-F' (Modified from: Savoca et al. (2010))	37
3.4	Regional Plan View of Cross-Section B-B' (Modified from: Robinson & Noble, Inc (1992))	40
3.5	Regional Cross-Section B-B' (Modified from: Robinson & Noble, Inc (1992))	41
3.6	Combined Regional Cross-Sections B-B' and E-E' (Modified from: Robinson & Noble, Inc (1992) and Savoca et al. (2010))	42
3.7	Site Under Predevelopment Conditions (Modified from: Hart Crowser (1975))	45
3.8	Hydrogeologic Cross-Section Typical of the Puget Sound Lowland (Source: Morgan and Jones (1996))	47
3.9	Plan View of Model Domain (Modified from: USGS (2013))	51
3.10	Model LR-S2-W Simulated vs. Observed Head	66
3.11	Model HR-S3-G Simulated vs. Observed Head	66
4.1	Percent Containment vs. Percent of Optimized Pumping Rate	86
A.1	Model LR-S1-G Simulated vs. Observed Head	92
A.2	Model LR-S1-W Simulated vs. Observed Head	92
A.3	Model HR-S1-G Simulated vs. Observed Head	93
A.4	Model HR-S1-W Simulated vs. Observed Head	93

A.5	Model LR-S2-G Simulated vs. Observed Head	93
A.6	Model HR-S3-W Simulated vs. Observed Head	93
A.7	Model HR-S2-G Simulated vs. Observed Head	94
A.8	Model HR-S2-W Simulated vs. Observed Head	94
A.9	Model LR-S3-G Simulated vs. Observed Head	94
A.10	Model LR-S3-W Simulated vs. Observed Head	94

Chapter 1

Introduction

In 1996, the percentage of population reliant on groundwater for municipal, domestic and rural use in Canada was 30 percent, ranging from 23 percent in British Columbia to 100 percent in Prince Edward Island ([François, 2006](#)). Over the last few decades, the vulnerability of groundwater aquifers to both natural and man-made contaminants has been repeatedly demonstrated. A famous example of man-made contaminants getting into an aquifer is the Love Canal in which 21,000 tons of buried chemicals led to a number of serious illnesses. It is estimated that there are at least 126,000 contaminated sites in the US alone ([National Research Council, 2013](#)), and in 2004, the U.S. Environmental Protection Agency (EPA) estimated that it would cost over \$209 billion to mitigate groundwater contamination hazards over the next 30 years, excluding the continued cost of operating sites already undergoing remediation ([National Research Council, 2013](#)). Given the importance of fresh drinking water and the significant investment required to clean up contaminated sites, it is very important to have good tools to characterize aquifers and to design and implement remedial solutions.

One of the most prevalent tools for supporting characterization and remediation are computational models. The computational model is typically a mathematical representa-

tion of a simplified physical system. These mathematical representations of groundwater systems can enhance understanding and allow for the testing of future scenarios including the implementation of proposed remedial systems. This ability to help understand how a system will respond to future changes is one of the main advantages of using computational models. The ability of a computational model to predict future scenarios depends on the assumptions made in its construction and how well the model reflects reality. With the correct assumptions computation models are useful tools aiding in the selection and design of remedial systems. Over the years, many different remediation techniques have been proposed and developed. One of the most widely used remediation techniques is a pump-and-treat system. Using this system, groundwater is pumped to the surface where it is treated and either discharged or pumped back into the ground. However, operation costs may be very high as it can easily take in excess of 30 years to sufficiently remediate an aquifer. For this reason, it is common to apply numerical models to properly design systems which require the selection of a number of design variables including the number of wells, and their respective locations and pumping rates. Pump-and-treat system design can include the use of field measurements, design equations and numerical and analytic groundwater models. The phased and integrated approach outlined by the U.S. EPA involves an initial characterization of the aquifer (usually accompanied by modelling) and design and implementation of a pump-and-treat system followed by continuous monitoring ([USEPA, 1997](#)). As more data are gathered, the pump-and-treat system is modified as necessary.

To numerically represent a groundwater flow system, the natural system must be simplified and broken down into model cells such that the governing flow and transport equations can be solved. This is typically done using finite difference, finite element or finite volume method. Although numerical groundwater flow models can be computationally intensive

and may require detailed site investigations for complex simulations, they can simulate a larger range of aquifer properties and conditions than analytical or data driven approaches. In addition, groundwater models can be used to test how a system will respond to future changes, such as the installation or modification of a remedial system.

In addition to being important for remedial design, computational groundwater models are also commonly used to assist engineers with other management decisions, such as delineation of capture zones or assessment of aquifer storage and recovery options. As with any other tool, it is important to recognize the assumptions and limitations of numerical groundwater models. Groundwater models only provide approximations of natural systems and cannot provide predictions with complete certainty, as it is infeasible to get a complete sample of the subsurface ([Bredehoeft, 2005](#)). Therefore, there is a degree of uncertainty related to predictions made using numerical groundwater models which must be taken into consideration when making management decisions, such as the design of a pump-and-treat system. Sources of uncertainty in a groundwater model are related to parameter indeterminacy, observation data accuracy and the conceptual model used ([Højberg and Refsgaard, 2005](#)). If a groundwater model is used in a predictive capacity, an additional source of uncertainty, referred to as scenario uncertainty, is introduced which results from an imperfect knowledge of future inputs or changes to the groundwater flow system.

The first and likely most important step in groundwater modelling is the development of the conceptual model. The conceptual model is a description of the groundwater flow system including system boundaries, geology, hydrology, geochemistry, and contaminant transport characteristics. A conceptual model can be represented by cross sections, figures, process diagrams, flow charts and/or qualitative descriptions ([Neuman and Wierenga, 2003](#)). It provides the basis for the numerical model which must be constructed such that it represents the structure and processes depicted by the conceptual model. Many researchers

have recently started to encourage the evaluation of alternative conceptual models as a means to avoid an underestimation of uncertainty in model predictions (Beven, 2002; Bredehoeft, 2003; Carrera and Neuman, 1986; Højberg and Refsgaard, 2005; Neuman, 2003; Neuman and Wierenga, 2003; Poeter and Anderson, 2005; Refsgaard et al., 2006; Rojas et al., 2008, 2010; Seifert et al., 2008, 2012; Ye et al., 2004, 2008, 2010). The main advantage of considering alternative conceptual models is that the effect of conceptual model uncertainty can be explicitly examined, thereby (in theory) increasing the robustness of model predictions (Refsgaard et al., 2007). However, this approach is limited by the fact that it cannot be known if the plausible model space was adequately sampled. Therefore, while the robustness of model predictions should increase, it is possible that important model structural features may have been overlooked, still resulting in an underestimation of model uncertainty.

The main disadvantage of a multiple conceptual model approach is the additional expense over a single conceptual model, in both person hours and computational burden. To construct a set of conceptual models, different interpretations of each hydrogeological component (i.e., stratigraphy, recharge and boundary conditions) are often combined (Højberg and Refsgaard, 2005; Nettasana, 2012; Neuman and Wierenga, 2003; Rojas et al., 2010). For example, three different stratigraphy interpretations, two recharge estimates and two possible boundary conditions result in 12 different conceptual models. Each conceptual model necessitates the construction and calibration of a corresponding groundwater model. Therefore, the cost of a multiple conceptual model approach can increase rapidly as addition hydrogeological interpretations are added.

1.1 Motivation

Recently some regulatory bodies have started to recognise the potential advantages of a multiple conceptual model approach and now endorse (or occasionally require) its use to address the impact of uncertainty at some sites. Although the use of multiple conceptual models may potentially provide a better understanding of site conditions and directly addresses the influence of conceptual model uncertainty, it is much more resource intensive than a traditional single model approach. While the concept of conceptual model uncertainty has received a lot of attention in the literature, few researchers have examined how conceptual model uncertainty impacts remedial design. In additions, no one has addressed the utility of multimodel approaches in contrast to conventional design methods. More specifically, how would the performance of a remedial system which was designed using multiple conceptual models compare to that designed using a single model approach combined with a factor-of-safety? As researchers and regulators have recently begun to recommend the use of multiple conceptual/numerical models to better identify the range of model uncertainty, it is critical to examine whether or not this more time- and resource-intensive approach is fundamentally more beneficial to the design process or whether the traditional approach using a single conceptual model with a factor-of-safety is sufficient.

1.2 Objective

The objective of this thesis is to assess the utility of a multiple conceptual model approach in contrast to a conventional factor-of-safety design method. A single case study will be used to evaluate whether or not a more expensive multiple conceptual model approach is a significant improvement over a conventional single model approach in the context of pump-

and-treat system optimization. The results will assist practicing engineers to determine where a more rigorous and expensive multiple-model approach would be most valuable. Attempts will be made to generalize results from this case study to remedial design as a practice, and some effort will be spent to identify pitfalls associated with the evaluation and comparison of multiple conceptual models.

1.3 Outline of Thesis

This thesis is organized as follows.

Chapter 2 provides a brief overview of groundwater modelling, including contaminant transport, density-dependent groundwater flow and particle tracking. Sources of uncertainty in groundwater models and methods to quantify that uncertainty are discussed. Within the literature review emphasis is placed on the quantification of conceptual model uncertainty in the context of a multiple model approach. A brief overview of model calibration and pump-and-treat system design is included.

Chapter 3 presents the methodology used to quantify conceptual uncertainty in this thesis. This method is demonstrated through the use of multiple groundwater models to design remedial systems for a real-world case study. Chapter 3 describes the development of multiple conceptual models for the case study and the construction and calibration of computation models corresponding to each conceptual model.

Chapter 4 presents the formulation of the pump-and-treat design problem, including objective functions, design variables and constraints. Pump-and-treat systems designed for each computation model were compared to illustrate the range of conceptual model uncertainty in pump-and-treat system designs. A factor-of-safety is applied to each pump-

and-treat system design which demonstrates how a factor-of-safety reduces the chance of design failure in the context of conceptual model uncertainty. To compare the multiple model approach to a conventional single model approach, the pump-and-treat system design from the multiple model approach are compared against a single model developed in practice and supplemented with a factor-of-safety.

Chapter 5 summarizes the main conclusions and recommendations from this work.

Chapter 2

Background

2.1 Groundwater Modelling

2.1.1 Overview

Groundwater models are often the only practical tool available to gain an understanding of a groundwater system's behaviour and its potential response to future stresses. While models are not a replacement for field and lab measurements, they provide useful insight into subsurface conditions where it is infeasible to collect sufficient observations or where conditions may change. Groundwater models are applied by regulators and practitioners to investigate a wide variety of subsurface problems including site remediation and the determination of sustainable well yields. One of the most important applications of groundwater models is the prediction of aquifer response to future stresses, such as changes in recharge rates resulting from development, the impact of increased groundwater extraction rates or the installation of groundwater remediation systems. The information obtained

from groundwater modelling helps practitioners make valuable predictions such as how much groundwater can be withdrawn for municipal use without dewatering private wells, or whether or not a given remedial design is likely to sufficiently contain a contaminant plume. To properly apply groundwater models, one must understand that models are best described as an approximation of reality and can only represent a physical system through simplifying assumptions. The applicability of a groundwater model depends on their assumptions and how well the model reflects reality. The successful application of a groundwater model to a given problem depends on proper understanding of the model assumptions, the uncertainty of model predictions, and potential limitations.

The governing equation used to model the three-dimensional saturated flow of groundwater of a constant density through porous media is described by the following partial differential equation ([Rushton and Redshaw, 1979](#)):

$$\frac{\partial}{\partial x} \left(K_{xx} \frac{\partial h}{\partial x} \right) + \frac{\partial}{\partial y} \left(K_{yy} \frac{\partial h}{\partial y} \right) + \frac{\partial}{\partial z} \left(K_{zz} \frac{\partial h}{\partial z} \right) + W = S_s \frac{\partial h}{\partial t} \quad (2.1)$$

Where K_{xx} , K_{yy} and K_{zz} are hydraulic conductivity values along the x, y and z coordinate axis, which are assumed to be parallel to the principal axis of hydraulic conductivity [L/T]; h is the hydraulic head [L]; W is the volumetric flow rate representing sink/source terms per unit volume of aquifer [1/T]; S_s is the specific storage of the porous media; and t is time [T].

The terms S_s , K_{xx} , K_{yy} and K_{zz} may be functions of space (e.g., $K_{yy} = K_{yy}(x, y, z)$ and $S_s = S_s(x, y, z)$) and the sink/source term W may be a function of space and time (e.g., $W = W(x, y, z, t)$) ([Harbaugh, 2005](#)). Therefore, Equation 2.1 can represent groundwater flow through homogeneous or heterogeneous and isotropic or anisotropic media, under the influence of any number of sink or source terms. Under steady-state conditions, $dh/dt = 0$,

and the right side of Equation 2.1 becomes zero. To solve the governing equation, boundary conditions are needed at the edges of the model domain and an initial condition must be specified at time $t = 0$. The three types of boundary conditions that may be implemented along system boundaries are; Type 1 (or Dirichlet) where the head h at the boundary is specified; Type 2 (or Neuman) where the gradient dh/dx at the boundary is specified and; Type 3 (or Cauchy) where some linear combination of the head h and the gradient dh/dx at the boundary are specified. By solving Equation 2.1 with boundary conditions at edges of the model domain and initial conditions one can determine the hydraulic head distribution as a function of space and time.

Solution methods for partial differential equations such as Equation 2.1 can be broken down into either analytical or numerical solution methods. For simpler problems, such as those with linear processes and simple geometry, analytic methods may be used to solve for $h(x, y, z, t)$ exactly. However, numerical methods are required for more general complex problems (Frind, 2003). Using numerical solutions, the system or model domain is discretized (broken down from a continuous model domain into discrete points) and the original differential equation is approximated at those points by a set of algebraic equations. The solution to the algebraic equations is the approximate solution to the original equation. The research herein uses the finite difference method, in which the model domain is broken into rectangular components (model cells).

The most commonly used groundwater model is the finite-difference groundwater model MODFLOW, which was originally developed by the U.S. Geologic Survey (McDonald and Harbaugh, 1984). MODFLOW uses a block-centered finite-difference approach in space and a backward-difference in time approach with which the hydraulic head h , is solved for at model cell centroids. MODFLOW can simulate confined or unconfined aquifer layers and includes many packages to simulate external stresses such as wells, recharge, rivers,

evapotranspiration and drains. To solve the system of algebraic equations, several solution methods are incorporated into MODFLOW. The Preconditioned Conjugate-Gradient method was selected for this research since it is an effective method for solving the large sets of matrix equations created by MODFLOW ([Hill, 2003](#)).

2.1.2 Contaminant Transport Modelling

In addition to naturally occurring substances, human activity has resulted in a large number of potentially harmful chemicals which have made their way into the environment. These harmful chemicals include pesticides and fertilizers which are sprayed on farmers fields, leachate from landfills, acid mine drainage and fuel leaking from storage tanks. Chemical compounds will migrate through the subsurface and can make their way into rivers or municipal wells, becoming a potential source of harm to humans and the environment. Once in the subsurface, chemical species move by advection and dispersion/diffusion and can transform into other chemical species or have their transport retarded by processes such as sorption. The partial differential equation used to describe the transient three-dimensional advective and dispersive transport of contaminant species in groundwater is as follows:

$$\frac{\partial(C)}{\partial t} = \frac{\partial}{\partial x_i} \left(D_{ij} \frac{\partial C}{\partial x_j} \right) - \frac{\partial}{\partial x_i} (v_i C) + W \quad (2.2)$$

Where C is the aqueous concentration [M/L]; t is time [T]; $x_{i,j}$ is the distance along the respective Cartesian coordinate axis [L]; v_i is the groundwater seepage velocity [L/T]; D_{ij} is the hydrodynamic dispersion coefficient tensor [L²/T]; and W is the sink/source term representing mass added or removed from the system [M/L³/T]. Equation 2.1 is typically solved to determine the groundwater seepage velocity v_i which is input into Equation 2.2. This is referred to as a coupled simulation.

One popular chemical transport model is MT3DMS ([Zheng and Wang, 1999](#)), which stands for Modular Transport 3-Dimensional Multi-Species. MT3DMS is the primary transport model used in conjunction with MODFLOW. As discussed in section [2.1.3](#) MT3DMS and MODFLOW were used.

2.1.3 Density-Dependent Groundwater Flow

Groundwater contains dissolved constituents such as dissolved minerals, salts, metals and man-made compounds. Typically they are present in relatively low concentrations such that they do not significantly affect the density of water. However, in some cases, such as in deep geologic formations or in coastal areas where fresh groundwater can come into contact with saltwater, the concentration of dissolved constituents within water can significantly affect its density. Even relatively small increases in density such as that when moving from fresh water (approximately 1,000 kg/m³) to salt water (approximately 1,025 kg/m³) have been shown to have a significant impact on groundwater flow patterns and rates ([Guo and Langevin, 2002](#)). Therefore, the effects of density-dependent flow must be considered in areas where dissolved constituents are present in sufficient concentrations to create density variations in the groundwater system. Since one of the assumptions of Equation [2.1](#) used in MODFLOW is that the groundwater is of constant density, a different governing equation must be applied to model density-dependent flow.

The governing equation for groundwater flow derived, from mass balance principles, is presented below ([Langevin et al., 2007](#)):

$$\nabla \left(\rho \frac{\mu_0}{\mu} K_0 \left(\nabla h_f + \frac{\rho - \rho_0}{\rho_0} \nabla_z \right) \right) = \rho S_{s,0} \frac{\partial h_f}{\partial t} + \theta \frac{\partial \rho}{\partial C} \frac{\partial C}{\partial t} - \rho_s q'_s \quad (2.3)$$

Where ρ_0 is the fluid density at the reference concentration and reference temperature

$[M/L^3]$; K_0 is the hydraulic conductivity tensor of the medium saturated with the reference fluid $[L/T]$; μ is the dynamic viscosity $[M/L/T]$; h_f is the hydraulic head measured in terms of the reference fluid at a specific concentration and temperature (typically fresh water is used as the reference fluid) $[L]$; $S_{s,0}$ is the specific storage $[1/L]$; t is time $[T]$; θ is porosity [dimensionless]; C is constituent concentration $[M/L^3]$; ρ_s is the fluid density of the sink/source $[M/L^3]$; and q'_s is the sink/source term per unit volume of aquifer $[1/T]$. For transient density-dependent flow problems Equation 2.3 must be solved in conjunction with Equation 2.2.

To model three dimensional, variable density, transient groundwater flow, SEAWAT was originally developed by Guo and Bennett (1998) to be compatible with MODFLOW-88 (McDonald and Harbaugh, 1988) and MT3D⁹⁶ (Zheng, 1996). The latest version of SEAWAT, version 4, is a coupled version of MODFLOW (Harbaugh et al., 2000) and MT3DMS (Zheng and Wang, 1999; Zheng, 2006) which uses the constant density groundwater flow process of MODFLOW to solve Equation 2.3. SEAWAT is an obvious choice to simulate density-dependent flow in this study as it works with the well established MODFLOW and MT3DMS codes and is supported by the U.S. Geological Survey.

2.1.4 Particle Tracking

In some cases, such as in the determination of well capture zones and assessment of hydraulic containment of a pump-and-treat system, it can be advantageous to simulate purely advective transport. Advective transport is simulated by releasing one or more hypothetical particles and tracking their movement based on groundwater velocities. This form of advective transport modelling is referred to as particle tracking. Particle tracking can be performed either forwards-in-time, determining where the particle will end up after a spec-

ified period of time, or backwards-in-time, determining where the particle originated from. Prior to particle tracking, groundwater velocities must be calculated. This is accomplished by the use of existing groundwater models such as MODFLOW. Once the groundwater velocity field is determined, particle tracking can be applied. The most popular particle tracking method is the semi-analytic Pollock Method (Pollock, 1988) which was designed to use the velocities generated from block-centred finite-difference ground-water flow models. Using the Pollock Method, the location where and time when the particle leaves the model cell can be calculated as follows:

$$x(t_2) = x_1 + \frac{1}{A_x} [v_x(t_1) \exp(A_x \nabla t) - v_{x_1}] \quad (2.4)$$

$$\nabla t_x = \frac{1}{A_x} \ln \left(\frac{v_{x_2}}{v_x(t_1)} \right) \quad (2.5)$$

Where $x(t_2)$ is the particles location along the x-axis at time $= t_2$ [L]; x_1 is the model cell wall location along the x-axis [L]; t_1 is the time at which the particle enters the model cell [T]; t_2 is the time at which the particle exits the cell [T]; A_x is a constant that correspond to the components of the velocity gradient within the cell ($A_x = (v_{x_2} - v_{x_1})/(x_2 - x_1)$) [1/T]; $\nabla t = t_2 - t_1$ [T]; v_{x_1} is the groundwater velocity at the model cell wall located at x_1 [L/T]; and v_{x_2} is the groundwater velocity at the model cell wall located at x_2 . The time and location for a particle to leave along the y- and z-axis are calculated in a similar manner.

To apply the Pollock Method one solves for the time it would take to reach each of the three possible exit faces, one along the x-, y- and z-axis using Equation 2.5. The smallest value for ∇t corresponds to the actual exit face. Then Equation 2.4 is applied to determine the location where the particle exits the selected cell face. The Pollock

Method is implemented in the MODPATH package ([Pollock, 1994](#)) which has been linked to MODFLOW. MODPATH and MODFLOW and have been used extensively for applications such as capture zone delineation and the evaluation of hydraulic containment of pump-and-treat systems.

2.2 Model Calibration

One of the fundamental issues when modelling groundwater systems is that it is impossible to get a complete sample of the subsurface. In most cases, less than a millionth of the site volume is sampled during field investigations ([Poeter and Mckenna, 1995](#)). Furthermore, when information is gathered regarding parameters such as hydraulic conductivity, the data typically only represents the aquifer in close proximity to the sample point and is therefore both qualitatively and quantitatively different than the information which is needed on the scale of a groundwater model ([Carrera et al., 2005](#)). Difficulties also arise pertaining to different types of parameters. Data are typically collected from borehole logs regarding aquifer transitivity, however data may be lacking on parameters such as boundary conditions and recharge ([Carrera et al., 2005](#)). Since field measurements are typically sparse and collected on a different scale than that which is required for groundwater modelling, inputting measured parameter values into the model will often result in a poor representation of the observed system response. To address this problem, inverse models are used to assist in parameter estimation.

Inverse modelling is the general framework or process by which observations of a system are used to gain information about the system which is not directly observed. In groundwater modelling this refers to the process of using observations of what is being modelled (i.e., hydraulic head) to gather information about the model (e.g., parameter values such

as transmissivity) (Carrera et al., 2005). The process of inverse modelling can be broken into two parts, model identification and parameter estimation. Model identification refers to the selection of an appropriate model structure, including governing equations, parameter zonation and boundary conditions. Parameter estimation, typically referred to as model calibration in groundwater modelling, is the process of assigning values to aquifer properties (e.g., hydraulic conductivity) (Carrera et al., 2005).

Calibration methods have been separated into two categories, consisting of either “direct” or “indirect” approaches (Yeh, 1986). The direct approach was originally based on substituting hydraulic heads into the groundwater flow equation to solve for transmissivities (treated as dependent variables). This approach has the advantage that it does not require repeated model simulations, however there were several drawbacks including numerical instabilities and the requirement that the hydraulic heads, boundary conditions and recharge be known (Carrera et al., 2005; Neuman, 1973). The indirect approach consists of adjusting parameter estimates until the model response sufficiently represents the real system. The indirect approach has become the industry standard for model calibration, since it does not require complete knowledge of hydraulic heads, boundary conditions and recharge, which are often highly uncertain due to sparse subsurface data.

To apply the indirect approach, an error measure (objective function) must be selected to quantify the degree of misfit between observed and model simulated values. The most commonly used objective function is the L_2 norm, however some authors have also applied other objective functions such as the L_1 norm (Xiang et al., 1992). To calibrate the model, the objective function is minimized by either manual or automatic methods (Anderson and Woessner, 1992). Traditionally, models were manually calibrated using a trial-and-error approach (Yeh, 1986) and this approach is still widely used today. The drawbacks of a manual trial-and-error approach are that it is often very time consuming and there is no

way of knowing whether or not an optimal solution has been found (Cooley, 1985). Many of the drawbacks of a manual calibrated are addressed using automated methods. Some advantages of automatic calibration method are that parameter statistics (i.e., standard deviations, sensitivities, correlation, etc) can be calculated, the calibration is more objective and likely to obtain optimal parameter values and the time savings allow the modeller to focus on other problems such as looking for improvements in the conceptual model (Cooley and Naff, 1992; Poeter and Hill, 1997; Sun, 1994).

Despite the advantages of automatic calibration, there are several potential pitfalls of which the modeller must be aware. The model calibration problem is often ill-posed. In many cases, a unique solution cannot be found, often a symptom of parameter insensitivity or correlation. This is most often the case when groundwater models are calibrated using only measurements of hydraulic head. While parameter insensitivity and non-uniqueness can be reduced by using addition types of field measurements such as flow rates at head-dependent boundaries (Hill, 1992), advective transport (Anderman et al., 1996) or concentration data (Harvey and Gorelick, 1995; Sun and Yeh, 1990) it is not feasible to collect this data in all cases. It should also be noted that hydraulic heads and flows are not a linear function of most parameters of interest (i.e., hydraulic conductivity), therefore the calibration problem is non-linear. This non-linearity can lead to automatic calibration methods getting stuck in local optimal solutions rather than finding the true or global optimum.

Most automatic calibration methods applied in groundwater use a form of non-linear least-squares regression. These methods allow the modeller to reduce the sum of squared difference between computer and observed values (typically hydraulic head) to improve the model fit and to quantify parameter uncertainty. Software packages commonly used in groundwater modelling which implement nonlinear least-square regression include UCODE

(Peoter et al., 2005; Peoter and Hill, 1998), MODFLOW-2000 (Hill et al., 2000) and PEST (Doherty, 2004). PEST is the model calibration and optimization software applied in this study.

The algorithm applied by PEST to solve the inverse problem is the Gauss-Marquardt-Levenberg algorithm. The Levenberg-Marquardt algorithm is a combination of two minimization algorithms, the steepest descent method and the Gauss-Newton method (Marquardt, 1963). It has been shown to efficiently solve nonlinear groundwater flow problems relative to other methods (Cooley, 1985). For nonlinear problems PEST works by iteratively moving towards an optimal solution. At the beginning of each iteration PEST calculates the Jacobian, containing the derivatives of all observations with respect to all parameters. PEST then solves for a better parameter set. By comparing the results of successive iterations PEST determines whether or not more iterations are necessary (Doherty, 2004).

PEST uses finite difference approximations to calculate the derivatives of observations with respect to parameters. PEST automatically switches from forward difference (more efficient) to central difference (more accurate) approximations based on user specified criteria to improve computational efficiency while maintaining accuracy. During each iteration, PEST records the sensitivity of each parameter with respect to the entire observation dataset (Doherty, 2004), which assists the modeller in identifying insensitive parameters which may degrade the performance of PEST and result in a nonunique solution. PEST also includes several features to handle highly parameterized inversion, including singular value decomposition and Tikhinov regularization.

2.3 Model Uncertainty

Groundwater models are an approximation of reality. They are based upon an imperfect and incomplete knowledge of the subsurface which has been characterized such that it can be modelled with available tools. This incomplete picture of the subsurface, resulting in part from sparse subsurface data, leads to uncertainty in model predictions (Bredehoeft, 2005). Other factors contributing to model uncertainty include variability in system properties, measurement and sampling errors and differences among sampling, simulation and actual scales of the system (Meyer et al., 2007). In literature, uncertainty has been classified in relation to its causes, or source. Commonly recognized sources of uncertainty include parameter uncertainty, scenario uncertainty, and conceptual model uncertainty (Neuman, 2003; Neuman and Wierenga, 2003; Refsgaard et al., 2006; Meyer et al., 2007; Rojas et al., 2008). Parameter uncertainty results from any incomplete or inadequate knowledge of model parameter values. This incomplete understanding can be attributed in part to measurement errors, scale discrepancies, time dependence and heterogeneity (Carrera et al., 2005). Conceptual model uncertainty is the result of choosing a poor model structure (i.e., poor governing equations, incorrect boundary conditions, forcing functions and/or parameter zonation). Often more than one conceptual model will fit the data equally well, therefore the wrong conceptual model may be selected, resulting in incorrect predictions (Bredehoeft, 2003). Scenario or future uncertainty results from not knowing the future conditions of the site when using the model in a predictive capacity. When using groundwater models to make predictions 10s, 100s or even 1000s of years in the future significant system changes could occur from geological events (e.g., earthquakes), climate events (i.e., flooding) or human activities (e.g., urbanization) (Meyer et al., 2007).

Traditionally, only a single conceptual model was constructed as a best approximation of

the system properties, based on available data and expert opinion (Neuman and Wierenga, 2003). Parameter uncertainty was then quantified within that single conceptual model. However, groundwater flow systems are often very complex, rendering them open to a wide range of interpretations and corresponding conceptualizations (Ye et al., 2010). More recently, authors have begun to suggest that uncertainty in groundwater model predictions is largely dominated by conceptual model uncertainty and that considering parameter uncertainty alone does not compensate for the effects of conceptual model uncertainty (Bredehoeft, 2003; Neuman, 2003; Neuman and Wierenga, 2003; Ye et al., 2004; Bredehoeft, 2005; Højberg and Refsgaard, 2005; Poeter and Anderson, 2005; Refsgaard et al., 2006; Seifert et al., 2008; Rojas et al., 2008, 2010). With increased attention given to conceptual model uncertainty it is important to ascertain whether or not a multiple model approach is a significant improvement over a traditional single model approach, using methods such as a factor of safety to account for uncertainty.

2.3.1 Methods to Quantify Parameter Uncertainty

Parameter uncertainty has traditionally been the main focus of uncertainty analysis in groundwater modelling and has received much attention in the literature (Carrera and Neuman, 1986; James and Oldenburg, 1997; Glasgow et al., 2003; Hill and Tiedeman, 2007). Typically a single deterministic model structure is assumed and the parameters are treated as being imperfectly known. Methods for handling parameter uncertainty include probabilistic approaches, sensitivity and regression analysis.

According to Anderson and Woessner (1992, p. 246), “the purpose of a sensitivity analysis is to quantify the uncertainty in the calibrated caused by uncertainty in the estimated parameter values”. Traditionally the calibrated model parameters are used as a reference

scenario in the sensitivity analysis. One parameter value is varied from the reference scenario at a time while holding all other parameter values constant. This is repeated for each parameter. The model results or objective function value for each varied parameter is compared to the base-case results to determine the sensitivity of the varied parameter. Parameter sensitivities can change as the parameter value increase or decrease, therefore it is important to test the parameter values over the entire plausible range. The major weaknesses of this method are (1) that the plausible range usually determined subjectively before model calibration and may include parameter values which do not adequately represent the system and (2) simultaneous changes in multiple parameters are not considered (Hill and Tiedeman, 2007).

Alternatively, probabilistic approaches can be used to quantify parameter uncertainty, the most common of which is a Monte Carlo analysis. A Monte Carlo simulation is conducted by (1) generating equally likely sets of randomized parameters; (2) computing the model outputs for each parameter set; and (3) analysing the model outputs. A statistically significant set of measured parameter data should be used to calculate parameter statistics. The data should also be of a scale compatible with the model used. However in cases where insufficient parameter data is available, the modeller must specify subjective probabilities which may introduce statistical bias (Neuman and Wierenga, 2003). Monte Carlo analyses are straightforward and can capture the full response of nonlinear systems, but they are computationally expensive (James and Oldenburg, 1997). Monte Carlo simulations are typically used as a benchmark against which to test other uncertainty propagation methods.

Other methods for propagating parameter uncertainty through a deterministic model include the first-order, second-moment method (Vecchia and Cooley, 1987; James and Oldenburg, 1997; Kunstmann et al., 2002; Glasgow et al., 2003), the stochastic response

surface method (Isukapalli et al., 1998), and stochastic moment methods (Dagan and Neuman, 1997; Zhang, 2001). Many of these methods have the advantage of being more computationally efficient than a Monte Carlo simulation, while still producing comparable results.

2.3.2 Methods to Quantify Conceptual Model Uncertainty

A conceptual model is both a quantitative and qualitative description of the system which is being modelled. According to Anderson and Woessner (1992, p. 28) “a conceptual model is a pictorial representation of the groundwater flow system, frequently in the form of a block diagram or a cross section”. The conceptual model is further described by Meyer and Gee (1999) as a working hypothesis that describes the main geological, hydrological and geochemical features and relationships within the site. The hypothesis or conceptual model can be tested by pairing it with a computational model (e.g., MODFLOW) and comparing the predictions made by the computational model with site measurements. For this study, the conventional definition that a conceptual model is a “mostly qualitative and often pictorial description of the groundwater system, including a delineation of the hydrogeological units, the system boundaries, input/outputs, and a description of the soils and sediments and their properties” (Meyer and Gee, 1999) is most appropriate.

The first step to quantifying conceptual model uncertainty is to construct a set of conceptual models. One method used to construct sets of conceptual models is to combine different interpretations for each hydrogeological component at the site. This includes different interpretations of stratigraphy, boundary conditions and recharge distributions (Højberg and Refsgaard 2005; Nettasana 2012; Neuman and Wierenga 2003; Rojas et al. 2010). For each conceptual model, a corresponding numerical groundwater flow model is

constructed and calibrated. This method can quickly become expensive as hydrogeological interpretations are added and combined; however, it ensures that a range of possible hydrogeological combinations and corresponding conceptual models are evaluated.

Once a set of alternative models is established, there are two ways to address conceptual model uncertainty. The first method is to evaluate each model and select the best one based on user specified criteria. This can be done through an informal comparison, through evaluations of model quality metrics such as calibration fitness, or through the evaluation of formal information criteria (Meyer et al., 2007). Commonly used information criteria include AIC (Akaike, 1973, 1974), AICc (Hurvich and Tsai, 1989), BIC (Schwarz, 1978) and KIC (Kashyap, 1982), all of which have received significant attention in the literature (Burnham, 2004; Singh et al., 2010; Engelhardt et al., 2012; Seifert et al., 2012; Foglia et al., 2013). In general, information criteria evaluate each model based on a combination of parsimony (typically represented by the number of model parameters) and model accuracy. The most parsimonious model which can represent observed data is deemed preferable. Each information criteria has its limitations. For example, AICc shouldn't be applied if the number of estimable parameters is less than 40 (Burnham, 2004) and the original BIC assumes that a true or quasi-true model is present in the set of alternatives (Poeter and Anderson, 2005) which is typically not the case as models are approximations of the natural or true system. It has been further argued by Ye et al. (2008) that BIC has been rederived by Cavanaugh and Neath (1999) in a way which does not require the assumption that a true or quasi-true model is present in the set of alternatives and that all published rigorous derivations of AIC and AICc require that the true model having generate the observational data be in the set of candidate models (Ye et al., 2008). In some cases information criteria have been shown to rank models differently when compared to each other (Ye et al., 2004; Singh et al., 2010) and there is no consensus regarding which criteria is best. By selecting

one “best” model from a set of alternatives, regardless of the selection method used, the uncertainty represented by the unselected models is omitted, potentially leading to an underestimation of uncertainty.

Results presented by [Duan et al. \(1992\)](#) highlight the difficulty of identifying the optimal parameter set and suggest that there may even be multiple optimal or near optimal parameter sets. This carries over to conceptual models as well. [Beven \(1993\)](#) refers to this problem as the equifinality concept. [Beven and Freer \(2001\)](#) argue that “given the current levels of understanding and measurement technologies, it may be endemic to mechanistic modelling of complex environmental systems that there are many different model structures and many different parameter sets within a chosen model structure that may be behavioral or acceptable in reproducing the observed behavior of that system.” The second strategy, referred to as model averaging, avoids the pitfall of rejecting plausible models by combining the results of all plausible models into a single set of predictions. Model averaging methods include Generalized Likelihood Uncertainty Estimation (GLUE) ([Beven and Binley, 1992](#)) and Maximum Likelihood Bayesian Model Averaging (MLBMA) ([Neuman, 2003](#)) or information criteria based model averaging.

The GLUE procedure, outlined in [Beven and Binley \(1992\)](#) is based on the assertion that all models are subject to error and it is only possible to assess the likelihood that a particular model is a sufficient simulator of the system. GLUE recognizes that different sets of calibrated parameters may be equivalent to each other. The GLUE procedure can be summarized as follows (1) define prior parameter distributions; (2) randomly sample parameter distributions to create equally likely parameter sets; (3) evaluate each parameter set and reject those which don't represent the system (e.g., have an objective function value above a specified threshold); (4) calculate the likelihood of each parameter set as a function of the difference between model simulated values and measured values; (5) normalize the

likelihoods to determine the probability of each model. Some disadvantages of GLUE include that parameter distributions can be highly subjective and that there is no guideline for the selection of an appropriate likelihood measurement. GLUE is also computationally expensive and could lead to an overestimation of model uncertainty as parameter values are not optimized (Nettasana, 2012).

According to Hoeting et al. (1999) “standard statistical practice ignores model uncertainty . . . leading to over-confident inferences and decisions . . . Bayesian Model Averaging (BMA) provides a coherent mechanism for accounting for this model uncertainty”. However, the implementation of BMA faces several difficulties including that the number of alternative models may be infeasible and that the specification of prior model probabilities is challenging (Hoeting et al., 1999). To address these problems and render the BMA approach outlined by Hoeting et al. (1999) computationally feasible, Neuman (2002, 2003) based the new MLBMA approach on a maximum likelihood approximation using KIC and the maximum likelihood parameter estimation method presented by Carrera and Neuman (1986). Using MLBMA, the posterior probability of model k can be calculated using Equation 2.6 shown below.

$$p(M_k|\mathbf{D}) = \frac{\exp(-0.5\nabla KIC_k)p(M_k)}{\sum_{k=1}^N \exp(-0.5\nabla KIC_l)p(M_l)} \quad (2.6)$$

$$\nabla KIC_k = KIC_k KIC_{min} \quad (2.7)$$

Where $p(M_k|\mathbf{D})$ is the posterior probability of model k given the set of data $\mathbf{D}$; KIC_k is the Kashyap information criterion (KIC) value for model k ; KIC_{min} is the minimum KIC value over all candidate models; N is the number of candidate models and; pM_k is the prior probability of model k .

The use of information criteria such as AIC is intended to allow the “selection of a parsimonious model that uses the smallest number of parameters needed to provide an adequate approximation of measured data (Engelhardt et al., 2012). Using information criteria the model weights w_j can be expressed as:

$$p(M_k|\mathbf{D}) = \frac{\exp(-0.5(IC_j IC_{min}))}{\sum_{j=1}^N \exp(-0.5(IC_j IC_{min}))} \quad (2.8)$$

Where IC_j is the information criterion value for model j ; IC_{min} is minimum information criterion value over all the candidate models and; N is the number of candidate models.

The formulation of information criteria based model weights is similar to that of MLBMA and may differ based on the information criteria used (e.g., AIC vs BIC/KIC which are typically used in MLBMA) and the prior model probabilities. Both information criteria based model weighting and MLBMA have been shown to place a high weight on a single candidate model, whereas GLUE places more uniform weights on the candidate models (Meyer et al., 2007; Ye et al., 2010; Engelhardt et al., 2012; Nettasana, 2012). A recent paper by Lu et al. (2013, in press) recognizes that evaluating model weights using AIC, AICc, BIC and KIC often leads to an “unrealistic situation where a high model weight (often close to 100%) is placed on a single candidate model. Lu et al. (2013) found that the problem was caused by using the covariance matrix of measurement errors to estimate the negative log likelihood function in model selection criteria and could be resolved by using the covariance matrix of total errors. As a result of these findings, more consideration should be given to the accuracy of information criteria model weighting and whether or not they are appropriate in a practical sense.

2.4 Pump-and-Treat System Design

Pump-and-treat systems are one of the most commonly applied methods for groundwater remediation (Ahlfeld et al., 1988; Wang and Zheng, 1997; USEPA, 2010). Pump-and-treat systems work by withdrawing contaminated groundwater from the subsurface through one or more pumping wells, then treating the contaminated water on the surface and either injecting the cleaned water back into the subsurface or sending it offsite for disposal. Pump-and-treat systems remediate the aquifer through containing the contaminant plume by changing groundwater flow paths and removing contaminant solute. Depending on type of contaminant and the extent of the source zone it can easily take decades to sufficiently remediate the site. Due to the lengthy operational time and high operational costs of some pump-and-treat systems, the application of numerical groundwater models to inform design decisions is recommended (Compernelle et al., 2013). The importance of optimization remedial design is highlighted by a USEPA (2005) study of 20 Superfund-financed pump-and-treat systems which identified cost savings in 17 of the 20 systems, on average translating to a potential 30% reduction in the annual operating cost.

Even with the assistance of numerical groundwater models, pump-and-treat optimization can be a challenging problem. The optimization problem is often nonlinear with multiple local optima and may include constraints or multiple competing objective functions. To solve this problem significant effort has been spent on the optimal design of pump-and-treat systems (Bear and Sun, 1998; Aly and Peralta, 1999; Russell and Rabideau, 2000; Ko et al., 2005; Matott et al., 2006; Sharief et al., 2012; Yang et al., 2013). Solution methods include global approximation, evolutionary algorithms, response surface approaches and both nonlinear and dynamic nonlinear programming (Mayer et al., 2002). The optimal design of a pump-and-treat system is often formulated as a cost minimization problem

subject to an environmental containment/remediation constraint with decision variables pertaining to the number of injection/extraction wells, the pumping rates of those well and the locations of those wells, making it a potentially complex and challenging problem to solve. In this study, only pumping rates are considered for optimization for the following two reasons: (1) the site has an existing pump-and-treat system for which pumping rates can be optimized (2) to focus on impact of conceptual model uncertainty on remedial design without clouding the arguments with a complicated pump-and-treat optimization problem.

2.5 Contribution in Relation to Previous Work

Authors have begun to suggest that uncertainty in groundwater model predictions is largely dominated by conceptual model uncertainty (Bredehoeft, 2003; Neuman, 2003; Neuman and Wierenga, 2003; Ye et al., 2004; Bredehoeft, 2005; Højberg and Refsgaard, 2005; Poeter and Anderson, 2005; Refsgaard et al., 2006; Seifert et al., 2008; Rojas et al., 2008, 2010). Many methods have been proposed to address conceptual model uncertainty including GLUE (Beven and Binley, 1992), the use of information criteria (Akaike, 1973, 1974; Hurvich and Tsai, 1989; Kashyap, 1982; Schwarz, 1978) for model selection and averaging and MLBMA (Neuman, 2002, 2003). While these method provide differing frameworks with which one can address conceptual model uncertainty in groundwater modelling, they have not been applied to some specific groundwater modelling applications including remedial design. Furthermore, the practicality of applying a multiple model model approach to remedial design in practice has not been compared to conventional single model approaches. This thesis provides a real-world comparison of a multiple model approach to a conventional single model approach supplemented by a factor-of-safety to examine whether a more

expensive multiple model approach is fundamentally more beneficial to the design process compared to a single model approach.

Chapter 3

Methodology

3.1 Site Description

3.1.1 General Description

The Tacoma, Washington site, herein referred to as the ‘Site’ is located on the peninsula situated between the Hylebos and Blair waterways in the Port of Tacoma as shown in Figure [3.1](#). This site was selected because it is sufficiently complex, allowing for the development of multiple plausible conceptual models describing the site. Since the development of the Port of Tacoma in the early 1900s, a number of industrial activities have taken place at the Site including bulk petroleum storage, production of chlorinated solvents, the construction and dismantling of ships and the dumping of waste material. These activities have resulted in elevated pH values (8.5-14) as well as high concentrations (in excess of 10,000 $\mu\text{g/L}$) of metals and volatile organic compounds in the subsurface. A pump-and-treat system is currently in operation at the Site to limit the migration of contaminants into the Blair and

Hylebos Waterways and into Commencement Bay to the North.



Figure 3.1: Site Location (Modified from: [USGS \(2013\)](#))

3.1.2 Site Contamination

Over the past century numerous industrial activities have taken place at and in the vicinity of the Site. Chemical manufacturing process by-products found at the Site include chlorine/caustic soda, sodium hypochlorite, trichloroethene/tetrachloroethene (TCE/PCE), ammonia, muriatic acid, calcium chloride, fish oils, aluminum chloride and sodium aluminate. The waste generated during these processes was managed on Site through the use of settling ponds, settling barges, landfills, disposal pits and waste piles. Other historic

activities include ship building, dismantling and maintenance, bulk petroleum storage, crude oil distillation, a Tetraethyl lead plant for blending lead with gasoline, and materials storage and dumping. These activities and chemical manufacturing processes have led to significant groundwater contamination at the Site.

Environmental investigations at the site have identified chlorinated volatile organic compounds, caustic soda (sodium hydroxide), salt (NaCl) and metals as the primary contaminants of concern. The release of chlorinated solvents (PCE and TCE) is associated with their on Site production and the management of production by-products with the use of settling ponds and a settling barge. PCE, TCE and their biodegradation product, vinyl chloride, are present in concentrations well in excess of drinking water guidelines. Salt and sodium hydroxide concentrations in groundwater are due to their use in the production of chlorine and chlorinated solvents. Salt was stored on Site in open piles which were sprayed with water to produce brine. The asphalt brine pads were sloped to direct the brine to sumps where it was pumped for processing. Due to high specific gravity of brine and caustic soda (approximately 1.2 and 1.3 to 1.5, respectively), the escape of brine and caustic soda to the subsurface created a high density plume, herein referred to as the anthropogenic density plume. The specific gravity of groundwater within the anthropogenic density plume can exceed 1.09 and has a significant impact on groundwater flow patterns in its vicinity. In addition to density impacts, the release of caustic soda is primarily responsible for elevated pH values, in excess of 13 in some areas at the Site. The elevated pH values also contribute to metal contamination at the Site. While metals migrated into the subsurface from waste material, the majority were released from in-situ soils as a result of geochemical conditions created by the release of other constituents. The high pH and ionic strength of compounds within the anthropogenic density plume mobilized metals naturally present in the sediments and inhibited their sorption to soil particles. Metal which

are found in elevated concentrations at the Site include arsenic, chromium, copper, lead, mercury, nickel, thallium and zinc.

3.1.3 Geology

Regional Geology

The Site lies within the Puget Sound Lowland. The Puget Sound Lowland is bounded on the north by the Fraser River, on the east by the drainage divide of the Cascade Mountain Range, on the west by the Strait of Juan de Fuca and the drainage divide of the Olympic Mountain Range and to the south by series of low hill that lie south of the town of Tenino Washington ([Jones, 1999](#)). The Puget Sound Lowland was formed by regional geologic processes including the convergence of the North American continental plate with the Pacific Plate and Juan de Fuca Plate. The continental plate convergence and subsequent subduction of the plates in the Cascadia subduction zone has resulting in the formation of the Olympic Mountains, Puget Sound Lowland and the Cascade Range.

At least four major glacial advances and several partial advances have occurred within the Puget Sound lowland ([Jones, 1999](#)). Glaciation has resulted in the deposit of approximately 3,000 feet of unconsolidated material in the Puget Sound Lowland and about 2,000 feet of deposits in the vicinity of the Site. The most recent glacial advance was that of the Vashon Stade of the Fraser Glaciation. Glacial deposits typically contain (in order of deposition) advanced outwash sand and gravels; glacial till (hard and poorly sorted mixture of clay, silt, sand and gravel) and ice-contact deposits, and recessional outwash sand and gravels ([Savoca et al., 2010](#)). The geological deposits by the Vashon Glacier include: Colvos Sands and glacial advance outwash, Vashon till, and recessional outwash and Stielcom gravels. These units comprise most upland deposits in the Puget Sound Lowland.

Within the Port of Tacoma, the glacial deposits of the Vashon Glacier are mostly not present and where present they are buried under a significant thickness of post-glacial fluvial and deltaic deposits. The primary geologic unit in the Port of Tacoma is the Puyallup River alluvial valley aquifer. It is very heterogeneous and consisting primarily of alluvial silt, sand and gravel which were deposited by mudflows, alluvial processes, tidal flat formation and the placement of artificial fill ([Savoca et al., 2010](#)). The Puyallup River alluvial valley aquifer is generally less than 100 feet thick, however it exceeds 250 feet in thickness as it nears Commencement Bay.

[Savoca et al. \(2010\)](#) developed hydrogeological cross-sections for the Chamber-Clover Creek Watershed and vicinity by grouping geologic units into hydrogeological groups based on lithology (grain size and sorting), relative stratigraphic position and hydrologic characteristics. This differed from previous studies in which all glacial deposits were classified as aquifers and all interglacial deposits were classified as confining units [Savoca et al. \(2010\)](#). As shown in Figure 3.2, cross-sections E-E' and F-F' terminate in close proximity to the Site. Cross-section E-E' ends just southwest of the Site and cross-section F-F' ends approximately 4 miles southeast of the Site. See Figure 3.3 for cross-sections E-E' and F-F'.

As shown in the cross-sections, the hydrogeologic units located beneath the Site from the ground surface down were classified by [Savoca et al. \(2010\)](#) as:

- AL (alluvial valley aquifer) - consists primarily of alluvial silt, sand and gravel deposits.
- C (aquifer) - consists of sand and gravel with minor lenses of silt, clay and till.
- D (confining unit) - consists of alluvial and lacustrine sand, silt and clay deposits and some deposits of volcanic ash.

- E (aquifer) - consists primarily of silt, sand and gravel, with discontinuous till and lacustrine deposits
- F (confining unit) - consists primarily of silt and clay with minor lenses of sand and gravel
- G (undifferentiated deposits) - consists primarily of stratified sand and gravel with discontinuous layers of till.
- Bedrock Unit - consists of sedimentary claystone, siltstone, bed of coal and volcanic rock.

The bluffs to the northwest and southeast of the Site have a similar geology to the Site at depth, however a shallow aquifer system is present rather than the alluvial valley aquifer. Ordered from the ground surface down, the shallow aquifer system consists of:

- A1 (Aquifer) - where present, consists of Vashon recessional outwash, stratified silt sand and gravel deposited by large meltwater streams.
- A2 (Confining Unit) - consists of Vashon till, ice-contact, moraine and fine-grained glaciolacustrine deposits.
- A3 (Aquifer) - consists of Vashon advance outwash, made up of well-sort sand or sand and gravel with lenses of silt and clay.
- B (Confining Unit) - consists of fine grained silts and sand deposited during the early Olympic interglacial period and glaciolacustrine clays deposited during the early Vashon period.



Figure 3.2: Regional Plan View of Cross-Sections E-E' and F-F' (Modified from: [Savoca et al. \(2010\)](#))

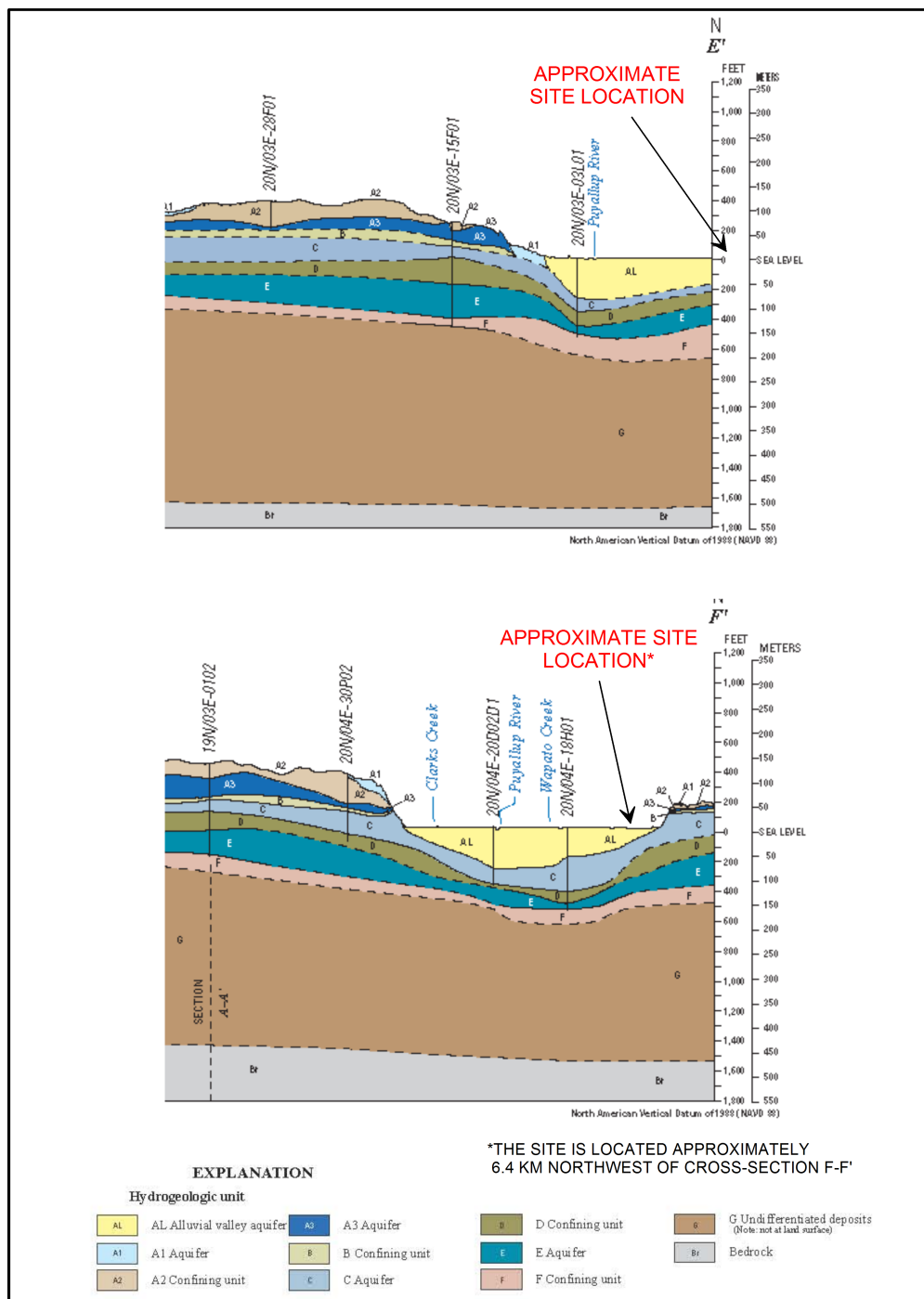


Figure 3.3: Regional Cross-Sections E-E' and F-F' (Modified from: Savoca et al. (2010))

Similar to the Port of Tacoma, the bluffs have a heterogeneous geologic composition and stratigraphy, consisting of sequence of glacial and interglacial deposits ([Johannessen and Maclellan, 2007](#)). A hydrogeological study of the Federal Way Area, including the bluffs to the northwest of the Site was conducted by [Robinson & Noble, Inc \(1992\)](#). Several hydrogeologic cross-sections were developed by [Robinson & Noble, Inc \(1992\)](#), one of which, cross-section B-B' terminates in close proximity to the Site. Cross-section B-B' and its location are shown in Figures [3.4](#) and [3.5](#), respectively. The hydrogeologic units beneath the bluffs, as interpreted by [Robinson & Noble, Inc \(1992\)](#), in order from the ground surface down are:

- Layer 1 - Consists of Vashon Till and Vashon recessional deposits which would correspond to layers A2 and A1 in [Savoca et al. \(2010\)](#), respectively.
- Layer 2 - Consists primarily of Vashon advance sediments. Contains distinct confined aquifers separated laterally by lower permeability zones.
- Layer 3 - Low permeability unit which includes Lawton Clay (Vashon Drift sequence), till, silt and clay. Thin zones of moderate permeability can occur
- Layer 4 - Intermediate aquifer systems which is predominantly low permeability with local pockets of higher permeability materials which make up aquifers. Generally coarse grained with silt or silt and clay matrixes.
- Layer 5 - Made up of low permeability sediments, mostly clay, silty clay, clayey silt, sandy silt and silty, fine sand. This layer includes isolated aquifers in upper portions, however most the layer is low permeability and it acts as a confining unit.
- Layer 6 - Deep aquifer systems, containing the Federal Way Deep Aquifer and the Puyallup Valley Deep Aquifer. The layer is very variable and consists of fine to

medium sand with silty zones in the Federal Way Deep Aquifer, some significant sand and gravel deposits associated with the Puyallup Valley Deep Aquifer and some silt and clay rich zones outside of the both aquifers.

The study area considered by [Robinson & Noble, Inc \(1992\)](#) is generally to the north of the study area used in [Savoca et al. \(2010\)](#) with some overlap. Cross-section B-B' ([Robinson & Noble, Inc, 1992](#)) starts in close proximity to the end of cross-section E-E' ([Savoca et al., 2010](#)). The two regional cross-sections are combined in Figure 3.6. Although both reports are in agreement about the order of geologic units in the bluffs on either side of the Port of Tacoma, there are discrepancies between both cross-sections about the thickness, location and order of geologic units directly beneath the site. Specifically, cross-section B-B' shows a shallow alluvial valley aquifer underlain by two confining units (Layers 3 and 5 ([Robinson & Noble, Inc, 1992](#))), whereas cross-section E-E' shows a much deeper alluvial valley aquifer underlain by an aquifer unit then a confining unit (layers C and D in [Savoca et al. \(2010\)](#)). While discrepancies in the location of contact zones may be expected in regional cross-sections based on sparse data, one would expect that the same zones to be present in the same order. In general, cross-section E-E' is more consistent with geologic conditions observed in Site borehole logs, as discussed in section 3.1.3.

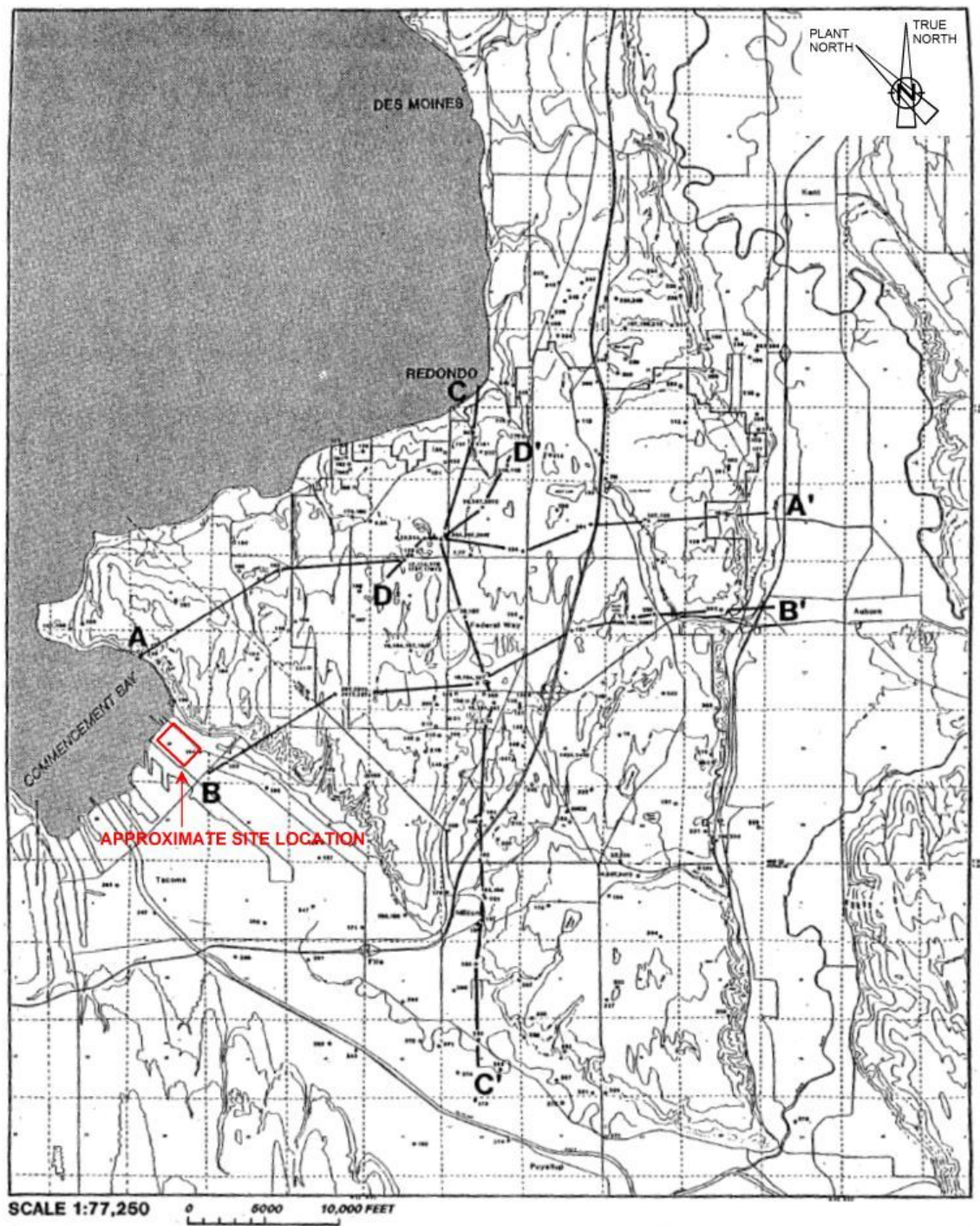


Figure 3.4: Regional Plan View of Cross-Section B-B' (Modified from: [Robinson & Noble, Inc \(1992\)](#))

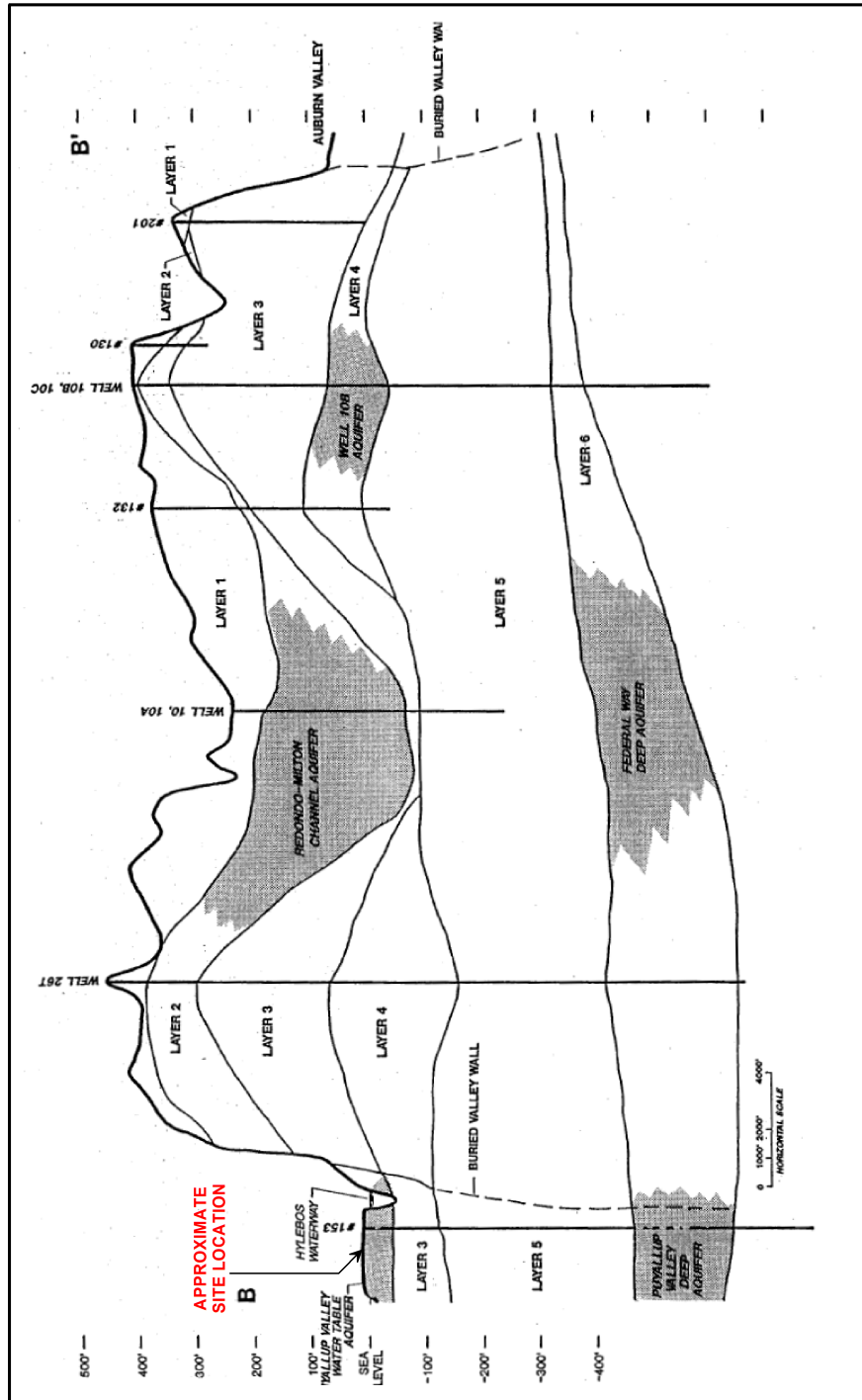


Figure 3.5: Regional Cross-Section B-B' (Modified from: [Robinson & Noble, Inc \(1992\)](#))

Site Geology

The Site sits on the Hylebos/Blair peninsula in the Puyallup River valley. To the north of the site bluffs slope downwards towards the Hylebos waterway. Beneath the Site are approximately 550 m (1,800 feet) of unconsolidated deposits underlain by bedrock that forms the basement confining unit ([Jones, 1999](#)). Site investigations are primarily focused on the upper 60 m (200 feet) or unconsolidated deposits. In general, the geologic conditions beneath the Site consist of fluvial deltaic deposits underlain by glacial till and overlain by fill material dredged from the waterways. This is consistent with the generalized geologic conditions in the vicinity of the Site described by [Savoca et al. \(2010\)](#).

Prior to 1920, the location of the Site consisted of undeveloped mudflats as shown in Figure 3.7. From 1920 to 1936, the Site was developed by dredging pre-existing streams and estuaries to create navigable waterways and placing fill on top of the existing mudflats to create usable peninsulas. The fill material consists primarily of silty sand largely derived from sediments dredged to create the waterways. The fill depth is approximately 20 feet and decreases in thickness as it moves inland until it reaches native soil ([Hart Crowser, 1975](#)). The mudflats which underlay the fill material likely consist of fine grained sediments including silts and organic matter deposited under pre-development conditions and may provide some hydraulic separation between the overlying fill and underlying deltaic deposits. The mudflats are on average approximately 3 m (10 feet) thick.

The deltaic deposits which underlay the mudflat and fill material are made up of later stage deposits from the Puyallup River. The deltaic deposits consist primarily of alluvial silt, sand and gravel ([Savoca et al., 2010](#); [Morgan and Jones, 1996](#)). This is consistent with Site borehole logs which identified clayey silt, sandy silt, silty sand, sand and gravel in the deltaic deposits. Sandy silt, silty sand and sand make up the bulk of the deltaic

deposits with some isolated pockets or lenses of gravel and clayey silt. The thickness of the deltaic deposit decreases from approximately 75 m (250 feet) as you move from the centre of the Puyallup River Valley towards the bluffs. Under the Site the deltaic deposits are approximately 45 m (150 feet) in depth.

The glacial material at the base of the deltaic deposits consists primarily of dense low permeability silty gravel. There are however, some discontinuous zones of more permeable silts, sands and gravels which also make up the upper portion of the glacial material. The surface of the glacial material slopes downwards from the bluffs just beyond the eastern shoreline of the Hylebos Waterway and extend across the Site. At the Site, the dense upper glacial material is found from an elevation of approximately -72 m (-170 feet) National Geodetic Vertical Datum (NGVD) to approximately -64 m (-210 feet) NGVD. Most boreholes advance below the upper glacial material identified more coarse soils consisting of medium sand to sand and gravel. This is consistent with the geological sequence presented by [Savoca et al. \(2010\)](#) which identified a confining unit (layer D) at approximately -200 feet NGVD, underlain by an aquifer (layer E) .

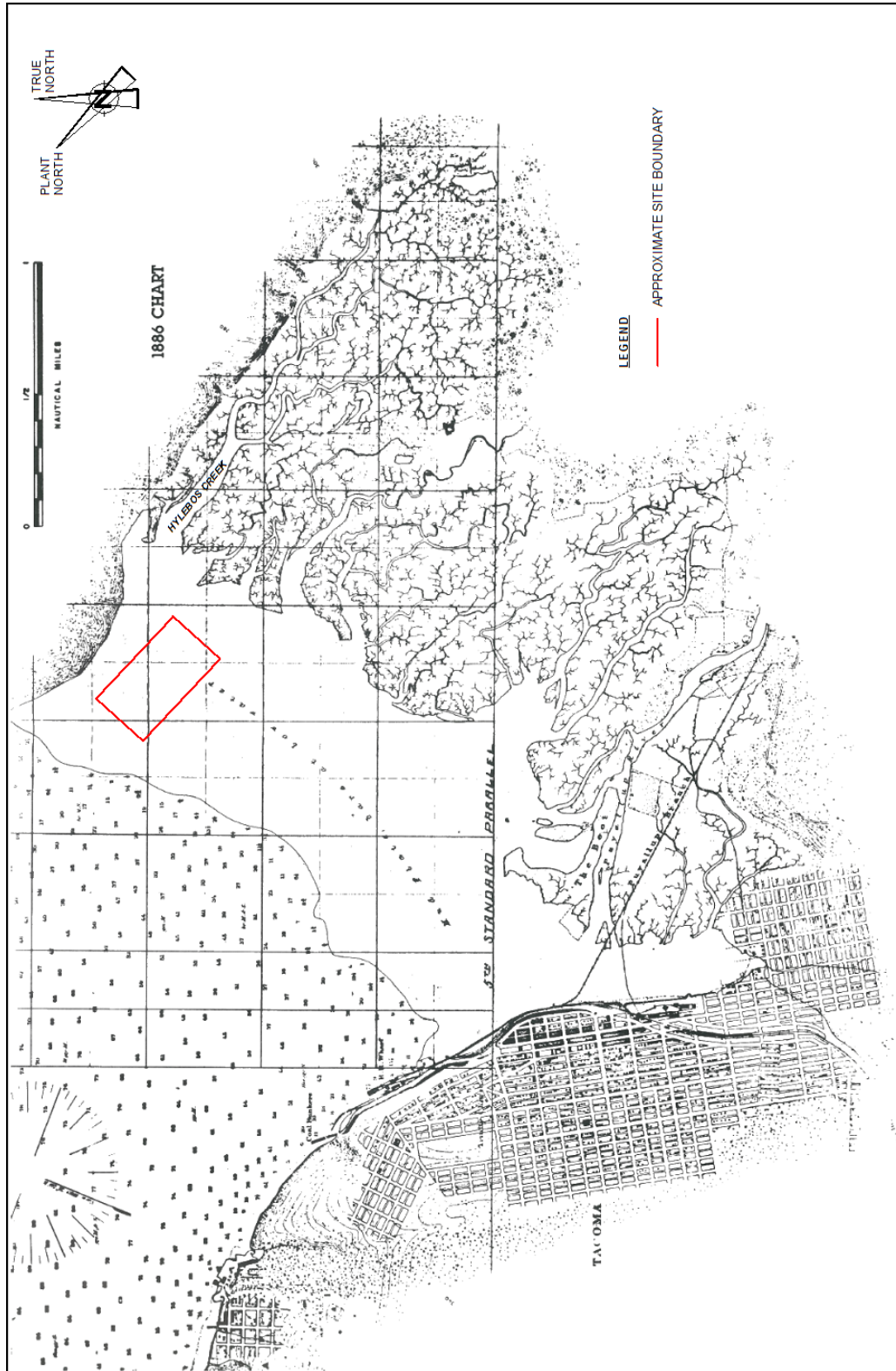


Figure 3.7: Site Under Predevelopment Conditions (Modified from: [Hart Crowser \(1975\)](#))

3.1.4 Hydrology

Regional Hydrology

The Site lies within the Puyallup River drainage basin. The major rivers in the basin are the Puyallup River and its two largest tributaries, the White and Carbon Rivers. The Puyallup River Watershed originates at Mount Rainer (approximately 4400 m (14,400 feet) above sea level) and the Puyallup River flows towards the sea where it empties into Commencement Bay ([Borden and Troost, 2001](#)).

The uplands on the north bank of the Hylebos Waterway include the Hylebos and Puget Sound Drainage basins which drain into Commencement Bay via Hylebos Creek. Together the Puyallup River and Hylebos Creek represent the major regional discharge features in the vicinity of the Site.

The regional groundwater flow is generally towards major streams which act as discharge boundaries for all local flow systems which typically exist in the upper few hundred feet of Quaternary sediments. Aquifers beneath the uppermost till units are generally confined except near their edges where they are truncated by bluffs and canyons formed by post-Pleistocene erosion ([Morgan and Jones, 1996](#)). Consequently, springs and seepage faces are observed along bluffs and canyons in the Puget Sound Lowland. An example near the Site would be the bluffs just north of the Hylebos Waterway, which exhibit leakage along the bluff face. Figure 3.8 shows a hypothetical hydrogeologic cross-section typical of the Puget Sound Lowland depicting recharge and discharge areas and generalized groundwater flow directions as presented in [Morgan and Jones \(1996\)](#).

In the Puget Sound Lowlands, precipitation mostly falls in the winter months. The quantity of precipitation varies significantly across the area, loosely correlated with ground surface elevation. The mean annual precipitation in the lowlands varies from 400 to 1350

mm. In the upland areas the annual mean precipitation ranges from 1150 to 1525 mm and in the mountains ranges from 1525 to more than 2540 mm (Jones, 1999).

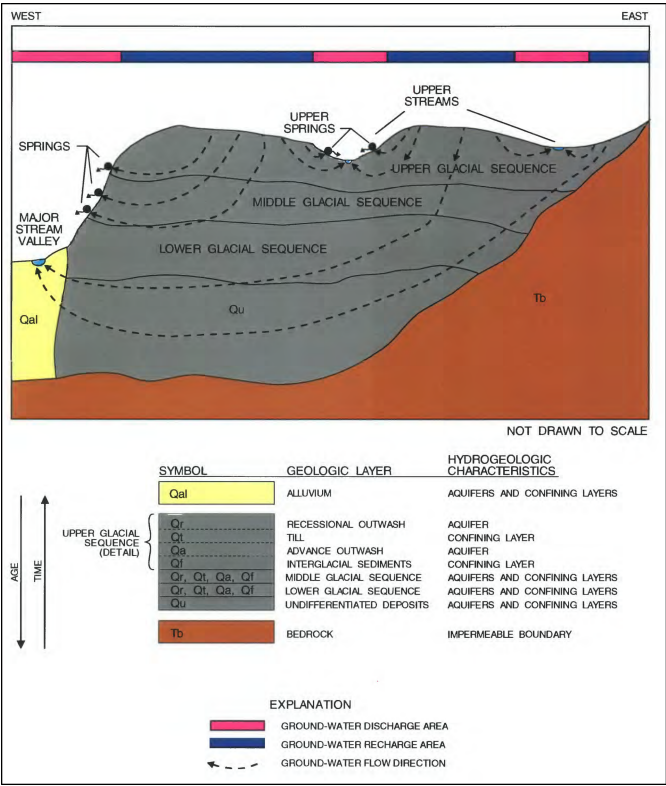


Figure 3.8: Hydrogeologic Cross-Section Typical of the Puget Sound Lowland (Source: Morgan and Jones (1996))

Site Hydrology

At the Site scale, the principal hydrologic features are the Hylebos and Blair Waterways and Commencement Bay to the north. The tidal fluctuations of these surface water bodies create transient flow conditions. Depending on the tidal stage groundwater flow may be directed towards the surface water bodies (i.e., during high tide) or flow may be directed

away from the surface water bodies (i.e., during low tide). Puget Sound typically experiences two high and two low tides every 24 hours and 50 minutes ([Lincoln, 2000](#)). The difference in water elevation between high and low tide at the Site is approximately 3 to 4 m (10 to 14 feet).

The surface waters adjacent to the Site are salt water bodies. The density of water in Commencement Bay and the Hylebos and Blair Waterways increases with depth, ranging in specific gravity from approximately 1.015 to 1.022. Density variations within the surface water bodies may be related to shallow groundwater discharges and surface runoff. Saltwater from Commencement Bay and the Hylebos and Blair Waterways migrates inland and mixes with fresh groundwater and the anthropogenic density plume. The differences in groundwater density affect groundwater flow at the Site.

The mean annual precipitation in the Tacoma area from 1971 to 2000 is 988 mm per year ([Savoca et al., 2010](#)).

3.2 Site Conceptual Models

Modelling a groundwater system is a task often involving many subjective decisions made by the modeller including the selection of boundary conditions, parameter zonation and model domain limits. These decisions can be complicated by sparse data, especially near the limits of the model domain where one may only have regional scale data. To completely address structural or conceptual model uncertainty with a multi-model approach, all plausible model structures should be evaluated, however this is neither practical nor feasible. It is more reasonable to use a subset of model structures which may sufficiently account for conceptual model uncertainty. In this study, a total of 12 conceptual models

were developed based on three different stratigraphic interpretations and two recharge and boundary condition interpretations.

3.2.1 Multiple Model Approach in the Context of Past Modelling Efforts

Site characterization and modelling of the Site is an ongoing progress. Conestoga-Rovers & Associates (CRA) provided several previous Site models in addition to observed data used in this study. While the models used in the multiple conceptual model approach differ from the CRA models, they share some similarities. Model domain limits between all models are set at approximately the same locations as described and rationalized in section 3.2.5 and similar model discretization was used. Shared boundary conditions between the multiple model approach presented in this study, and the models developed by CRA include the constant head boundary representing Commencement bay to the Northwest and the no flow boundary along the centre of the Blair-Sitcum peninsula constant inflow from the Southeast. Throughout the peer-reviewed model developed process CRA considered several plausible conceptual models including applying a no flow or general head boundary at the bottom of the model domain, a no flow boundary or constant inflow to represent conditions beneath the bluffs to the Northeast and a zoned stratigraphic model to represent the deltaic deposit below the site. While parameter values and implementation may differ, these conceptual models were carried forward in the multiple model approach. By incorporating conceptual models used in practice and using similar model domain limits and model discretization, the multiple conceptual model approach provides a reasonable approximation of an approach that could have been used in practice at this Site.

3.2.2 Model Construction

The groundwater flow models were constructed using Groundwater Vistas, a graphic user interface designed to work with several groundwater modelling packages, including SEAWAT and MODFLOW. The model domain is 2440 m (8,000 feet) by 3050 m (10,000 feet). The location of the model domain in relation to the Site and surrounding hydrogeologic features is shown in Figure 3.9. The model domain is 95 m (312) feet in vertical extent with the top set at 3.7 m (12 feet) NGVD (National Geodetic Vertical Datum) and the bottom set at -91 m (-300 feet) NGVD. Horizontal grid spacing is 6.1 m (20 feet) is applied within the Site, expanding to 61 m (200 feet) at the model domain limits. In the vertical direction, the first layer was made 5.2 m (17 feet) thick to ensure that the water table is within the first layer, avoiding instabilities associated with the drying and wetting of cells. Below the first layer, the grid spacing gradually increases from 0.76 m (2.5 feet) to 6.1 m (20 feet) at the bottom of the model domain.

3.2.3 Stratigraphy

Heterogeneity in soil properties may be treated in a variety of ways. It can range from a simple homogeneous model, where the same hydraulic conductivity is specified in every model cell, to complex models where a unique hydraulic conductivity is applied in each model cell. While there are no universally agreed upon guidelines to select the degree of model complexity, it's generally recognized that there is some trade-off between model fit and model complexity and that models should be only be as complex as required to provide a good representation of the modelled system (Hunt et al., 2007). To represent the range of potential model complexity, three different stratigraphy models are used, including a homogenous conductivity distribution, a zoned conductivity distribution and a pilot-point

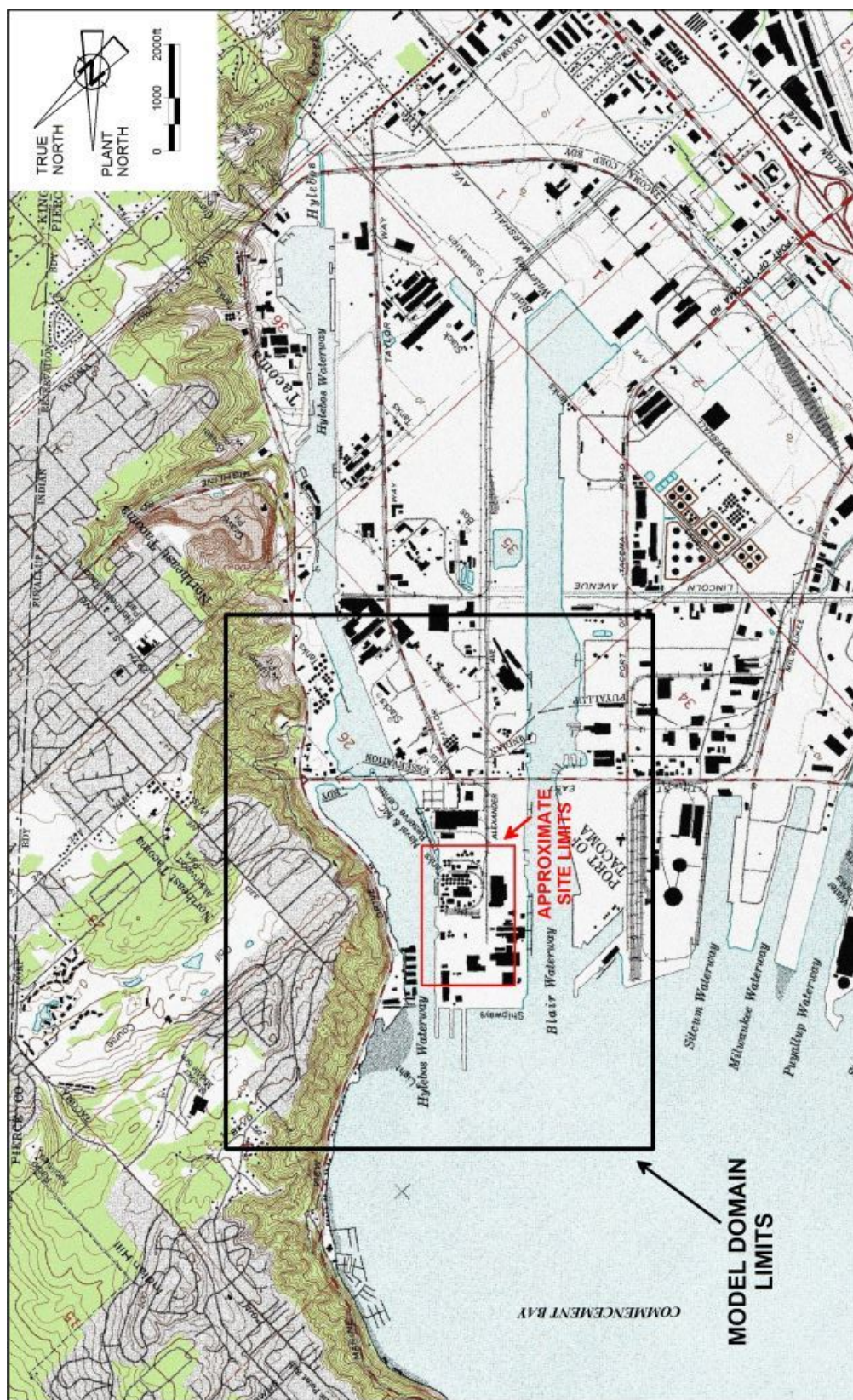


Figure 3.9: Plan View of Model Domain (Modified from: [USGS \(2013\)](#))

based conductivity distribution.

The simplest stratigraphy model applicable at the Site is a homogeneous anisotropic stratigraphic model. While the deltaic deposits below the Site are heterogeneous, the distribution of soil types is also highly variable with no distinct zones or uniform layers, therefore a homogeneous anisotropic strat model is likely appropriate for simulating the bulk response of the system. Furthermore, the dominant soil types are sands and silty sands, indicating that the bulk hydraulic properties of the subsurface beneath the Site will be similar to that observed in sands and silty sands.

The second plausible stratigraphy model was constructed by grouping model cells into zones of similar property values. Typically this is done by interpolating a layered aquifer system from borehole logs. However, there is no clear zonation or layering evident in the Site borehole logs. Therefore, kriging was used to create conductivity zones based on 9 observed soil types by interpolating soil type locations from borehole logs. Outside of the stratigraphic model, model cells were assigned conductivities based on the predominant soil type in that layer. Conductivity values were constrained to within expected values based on those presented in [Freeze and Cherry \(1979, p.29\)](#) as well as measured site values. In addition, parameter values were constrained relative to each other (i.e., sand should be more permeable than silt). Preliminary simulations indicated that the 9-zone stratigraphy model was overly constrained, with calibration runs providing either a worse fit than the homogeneous strat model or resulting in unrealistic parameter values. To address this problem, the Site stratigraphy model was simplified to four stratigraphy zones by grouping soil types with similar measured hydraulic conductivities to improve model calibration and achieve more representative estimated conductivity values.

While using zones of uniform conductivity has been the traditional approach in groundwater modelling, it has limitations. Parameter zones are only known approximately. Fur-

thermore, the geologic data or borehole logs from which parameter zones are estimated may not reflect the hydraulic properties of the soil (i.e., all sand does not have the same conductivity). To address these shortcomings and to provide a strongly contrasting conceptual model of system heterogeneity, a third approach based on pilot points and regularization was used (Doherty, 2003). This approach is applied by estimating the conductivity values at discrete points (pilot points) distributed throughout the model domain and creating a smooth conductivity distribution by interpolating conductivity values between points using kriging. However, using a large number of points can lead to the estimation of unrealistic parameter values during calibration or a high level of nonuniqueness making the problem numerically intractable (Doherty, 2003). To address this issue, regularization is used to stabilize the otherwise over parameterized inverse problem. Combining pilot points and regularization provides a method of estimating a heterogeneous conductivity distribution based on observed head values, without the limitations of predefined zones.

3.2.4 Recharge

The surface cover at the Site consists of approximately 65 percent impermeable material (concrete, asphalt and buildings) and 35 percent more permeable material (gravel and grass). Utilities at the Site include aging storm and sanitary sewers, and leakage from these systems is suspected to contribute to groundwater recharge. This adds to the uncertainty of recharge estimation as it is not feasible to get a direct measure of the contribution of sewer leakage to groundwater recharge. This uncertainty is compounded by correlation between groundwater recharge rates and hydraulic conductivity values. Using a homogeneous stratigraphic model, near zero (≤ 0.1 mm/yr) recharge rates and low conductivity values provided an equivalent match to observed head data as a model with higher recharge

rates and high conductivity values. Therefore, the observed head data was insufficient to determine a single appropriate recharge rate. To address this source of uncertainty in the conceptual model, both high and low recharge conditions were selected to encompass the range of plausible recharge values. This provided two conceptual models for recharge, a low recharge formulation corresponding to the presumption of mostly impermeable ground surface with no leakage from storm drains and a high recharge formulation corresponding to significant leakage from the storm drains source of uncertainty in the conceptual model.

3.2.5 Boundary Conditions

Boundary conditions are required at the model domain limits to uniquely solve the groundwater flow equation. For an adequate representation of the groundwater system, boundary conditions should ideally be linked to known conditions at the model domain limits. Boundary conditions should also be specified a significant distance away from the area of interest, to avoid boundary artifacts (undue influence of boundary conditions in the Site vicinity). To complicate the task of assigning boundary conditions, data at the model domain limits is often significantly more sparse than that in the area of interest. To help assess the uncertainty at model domain limits, multiple types of boundary conditions are considered where supported by available data. The boundary conditions assigned at the model boundaries are outlined in the following sections.

Top of model domain

The top of the model domain is set at 3.65 m (12 feet) NGVD which corresponds to the approximate average ground surface elevation of the Site. Consequently, the top of the model domain transects the bluffs to the north east of the Site. Constant recharge is

specified at the top of the model where surface water is not present. Two recharge zones are used, one specified beneath the bluffs to represent leakage and the other specified at the ground surface to represent groundwater recharge from rainfall and storm sewer leakage as discussed in section [3.2.4](#).

Northwest Boundary

The northwest boundary is set in Commencement Bay, approximately 915 m (3,000 feet) from the Site. Constant head cells are specified where Commencement Bay contacts the sea floor, based on observed head values in Commencement Bay.

Southwest Boundary

The southwest boundary is set approximate 760 m (2,500 feet) from the Site, corresponding to the centre of the peninsula running between the Blair and Sitcum Waterways. The centre of the Blair/Sitcum Water also corresponds approximately to the centre of the Puyallup River Valley Aquifer. Groundwater flow along the Blair/Sitcum peninsula is primarily southeast to northwest, corresponding to regional flow along the Puyallup River Valley Aquifer, discharging to Commencement Bay. At shallow depths, local flow patterns are along the peninsula, discharging to Commencement Bay, and outwards discharging to the Blair and Sitcum Waterways. Therefore, this boundary can realistically be simulated by a no-flow boundary condition. Furthermore, it is located a sufficient distance from the Site such that the boundary condition assigned at this location will not unduly influence simulated groundwater flow patterns at the Site.

Northeast Boundary

The northeast boundary is set approximately 1220 m (4,000 feet) from the Site, such that the boundary conditions at this location do not unduly influence the simulated groundwater flow conditions at the Site. With the exception of regional borehole logs and studies there is little data indicating the conditions beneath the bluffs. The studies by [Savoca et al. \(2010\)](#) and [Robinson & Noble, Inc \(1992\)](#) indicate that there is a stratified aquifer system beneath the bluffs. The hypothetical hydrogeologic cross section presented [Morgan and Jones \(1996\)](#) and shown in Figure 3.8, indicates that any regional aquifers would likely discharge to Puyallup River Valley Aquifer. However, there is no conclusive evidence to confirm that the regional aquifers are hydraulically connected to the deltaic deposits beneath the site which form part of the Puyallup River Valley Aquifer. The [Robinson & Noble, Inc \(1992\)](#) study shows that the regional aquifers are truncated by a buried valley wall and may not be hydraulically connected to the Site, whereas the study by [Savoca et al. \(2010\)](#) indicates that the aquifers may be connected to the Site. To address this uncertainty, two conceptual models are used. In one case constant inflow is specified along the northeast boundary, and in the second case a no-flow boundary condition is specified along the northeast boundary.

Southeast Boundary

The southeast boundary is located approximately 1220 m (4,000 feet) beyond the Site to provide sufficient distance such that the boundary conditions assigned at this location do not unduly influence the simulated groundwater flow patterns at the Site. A constant inflow condition is assigned along the southeast boundary to simulate regional groundwater flow along the Puyallup River Valley Aquifer.

Bottom of Model Domain

The bottom of the model domain is set at -91 m (-300 feet) NGVD, corresponding roughly to the deepest depth investigated during Site characterization. This allows the model to simulate leakage through the low permeability glacial material located from approximately an average of -52 m (-170 feet) NGVD to -67 m (-220 feet) NGVD. This also provides sufficient separation between the anthropogenic density plume which extends to approximately -46 m (-150 feet) NGVD and bottom of the model domain. Only a handful of boreholes were extended below -49 m (-160 feet) NGVD, including two boreholes advanced to -88 m (-290 feet) NGVD and three boreholes advanced to -70 m (-230 feet) NGVD, each of which identified coarse deposits in the toward the bottom of the borehole. As discussed in section [3.1.3](#) this likely corresponds to an aquifer, referred to as layer E, as identified by [Savoca et al. \(2010\)](#). On average, the thickness of layer E ranges from a thin veneer to approximately 37 m (120 feet). The thickness of coarse deposits range from approximately 1524 m (-50-80 feet) at the bottom of the deepest boreholes, however the underlying confining unit interpreted by [Savoca et al. \(2010\)](#) is not encountered in those boreholes. Therefore, two possible boundary conditions are proposed for the bottom of the model domain. The first is a no-flow boundary condition which would correspond to either the continuous low permeability unit suspected to be present below the model, or primarily horizontal flow along layer E towards Commencement Bay. The second is a general head boundary condition which would correspond to leakage from lower layers and the net upwards hydraulic gradient identified in the deepest observation well. Both boundary conditions are considered to address conceptual model uncertainty brought about by limited knowledge of subsurface conditions at depths below -61 m (-200 feet) NGVD.

Initial Density Distribution

Density values were assigned using a density distribution provided by CRA. Where observed data was available density values were assigned based on that data. Beyond the limits of observed data a transient saltwater intrusion simulation was used to determine appropriate density values. To simulate saltwater intrusion the concentration of salt along Commencement Bay and the Hylebos and Blair Waterways were assigned based on measurements taken from the waterways. A transient simulation was used to simulate saltwater intrusion from Commencement Bay and the Hylebos and Blair Waterways inland until salt concentrations within the model reached equilibrium. The density distribution from the transient saltwater intrusion simulation was merged with observed density values from Site measurements to assign density values throughout the model domain. During model calibration, the density distribution was held constant as the density values correspond to the observed head data used for model calibration. The density distribution was also held constant during pump-and-treat optimization simulations, such that the optimized pump-and-treat systems would simulate containment under current Site conditions.

Forming Numerical Computational Models

To construct a set of conceptual models for the Site, the different interpretations for stratigraphy and boundary conditions were combined, similar to the approach used by [Højberg and Refsgaard \(2005\)](#); [Nettasana \(2012\)](#); [Neuman and Wierenga \(2003\)](#) and [Rojas et al. \(2010\)](#). Combining three stratigraphy interpretations with two possibilities for the recharge, northeast boundary conditions and the domain bottom boundary condition resulted in 16 possible conceptual models. This was later reduced to 12 conceptual models as a general head boundary combined with inflow on the northeast boundary provided a

nonunique solution. When combined with the general head boundary, the value of the northeastern inflow condition could not be effectively determine through model calibration as the objective function became insensitive to changes in the inflow condition. This is supported by parameter sensitivities calculated by PEST which showed a two order of magnitude drop in the sensitivity of the objective function to the northeast inflow boundary dropped when the general head boundary was added at the bottom of the model domain. In the models, the northeast inflow boundary serves a similar purpose to a general head boundary, to maintain groundwater heads at depth beneath the Site. However, when combined, water flows downwards and exists the model domain through the general head boundary, rather than flowing along the bottom of the model domain and upwards beneath the Site. This mechanism prevents the northeast inflow boundary from having the intended effect. Furthermore, the size of the general head boundary and its proximity to the Site dominates the potential impact of the northeast inflow condition. In addition, the solution obtained using a combination of the general head and northeast inflow boundary would likely be similar to that of using the general head boundary alone, defeating the purpose of a multiple conceptual model approach, to have different solutions which account for conceptual model uncertainty. Therefore, to obtain a unique solution and use sufficiently different conceptual models, the two boundary conditions were considered separately. The amended configuration contained of 12 different conceptual models which consisted of three strat interpretations, a low recharge condition, and either a general head boundary specified at the bottom of the model domain or an inflow condition specified along the northeastern boundary.

3.3 Calibration of Numerical Conceptual Models

3.3.1 Calibration Targets

Two site-wide 72 hour monitoring events were conducted in May 2006, referred to as Event 1 and Event 2. Event 1 was conducted under pumping conditions with the Site pump-and-treat system active, while Event 2 was conducted under non-pumping conditions. The pumping wells had been shut down for maintenance prior to the commencement of Event 1 and had not been allowed sufficient time to reach steady pumping conditions before Event 1 monitoring. Therefore, Event 1 data corresponds to neither steady-state pumping nor non-pumping conditions and is consequently excluded from this analysis.

Over the 72 hour Events, pressures were recorded at five minute intervals using pressure transducers attached to data loggers. To facilitate data interpretation and modelling, the observed hydraulic responses were averaged using an approach presented in [Serfes \(1991\)](#), which was developed to filter out tidal fluctuations in observed data. Although, data was recorded in five minute intervals, only hourly measurement were required for the [Serfes \(1991\)](#) averaging approach. The [Serfes \(1991\)](#) averaging procedure requires 71 hourly readings and calculates the average at hour 36 using three consecutive averaging steps. The first step is to take the moving 24 hour average over the 71 hour data set, resulting in 48 consecutive 24-hour averages. The second step is to take the 24 hour moving average of the 48 consecutive 24 hour averages taken in the first step. To calculate the average pressure at hour 36 of the 71 hour period, the arithmetic average is taken of the 25 averaged values calculated in the second step.

Pressure data obtained for the Site is recorded in term of fresh water equivalent heads (FEH), where FEH is defined as the level of fresh water in a well required to balance

the pressure existing at a given point (Luszczynski, 1961). FEHs are useful as they are used to determine the hydraulic gradient in the horizontal, however the latest version of SEAWAT.v4 no longer reads and writes head in terms of FEH. Instead point water heads are used. Point water heads are defined as the level to a well would have to be filled with water of the type observed at a point would have to rise in order to balance the pressure existing at that point (Luszczynski, 1961). To convert FEH to point-water heads, Equation 3.1 was used. In total, 166 averaged point water heads are used as calibration targets.

$$H_{ip} = Z_i + \frac{\rho_f}{\rho_i}(H_{if} - Z_i) \quad (3.1)$$

Where H_{ip} is the point water head at point i [L]; Z_i is the elevation of point i [L]; ρ_f is the density of freshwater [M/L³]; ρ_i is the density of water at point i [M/L³] and; H_{if} is the FEH at point i [L].

3.3.2 Calibrated Method

The models were calibrated by minimizing the sum of squared residuals (SSR) between observed¹ and modelled values by both manually adjusting parameters and automatically adjusting parameters using PEST. Parameters used in model calibration are outlined in section 3.3.3 below. Due to high correlation between the influence of hydraulic conductivity and recharge in the homogeneous model, both parameters had to be calibrated using an iterative process, since simultaneous calibration of both parameters resulted in unreasonable parameter values. PEST would select near zero conductivities and recharge rates as optimal parameters, however similar SSR values could be obtained by reasonable conductivity values and recharge rates. Therefore, either the recharge or conductivity had

¹Observed data is filtered with Serfes (1991) approach.

to be fixed in each calibration run to obtain an equally good match to the observed data with reasonable parameter values. Each model calibration consisted of numerous PEST runs in combination with manual changes to parameter values and PEST variables. Model parameters were adjusted within a reasonable range until simulated heads provided an acceptable match to observed values.

3.3.3 Calibration Parameters

The parameters adjusted during model calibration include vertical and horizontal conductivities, recharge values, inflow along the southeastern boundary, and boundary conditions at the bottom and northeast model boundaries. The most sensitive parameters in model calibration are the Site recharge and the hydraulic conductivity values. The bounds and initial guesses for hydraulic conductivity were set based on observed Site values and the ranges presented for appropriately matched soil types included in [Freeze and Cherry \(1979, p.29\)](#). For the pilot point approach and the homogeneous stratigraphy model, the same bounds and initial values were used. For the zoned stratigraphy model, initial hydraulic conductivity values were set at their measured values. Bounds were only used in the zoned stratigraphy model when required to constrain less permeable soils (i.e., silts) from being assigned high hydraulic conductivities than more permeable soils (i.e., sands) and vice versa. As discussed in section [3.2.3](#), the zones were created by grouping soil types with similar measured hydraulic conductivities. Initial hydraulic conductivity values and bounds are presented in Table [3.1](#). Site recharge was set based on the recommended values of 0-430 mm per year presented by [Robinson & Noble, Inc \(1992\)](#). Initial estimates of 25 mm per year and 380 mm per year were used for the low and high recharge conditions, respectively. The bounds for the general head boundary assigned at the bottom of the

model domain were set from 0 to 3 m (0 to 10 feet) based with an initial point water head of 1.3 m (4.25 feet), corresponding to the point water head observed at the deepest observation well (approximately -70 m or -230 feet NGVD).

Table 3.1: Initial Hydraulic Conductivity Values and Bounds Specified for Model Calibration

Strat Zone	Description	Initial Value cm/s	Lower Bound cm/s	Upper Bound cm/s
Homogeneous	One homogeneous anisotropic zone assigned throughout model domain.	5.30E-03	1.00E-06	7.06E-02
Pilot Points	Conductivity values calibrated at discrete points which are kriged to create a smooth conductivity distribution.	5.30E-03	1.00E-06	7.06E-02
0	Includes Clayey Silt, Silty Clay, Clay and Sandy Clay.	3.98E-05	-	-
1	Includes Sandy Silt and Silt.	2.04E-04	-	-
2	Includes Silty Sand and Silty Gravels.	8.10E-04	-	-
3	Includes Sand, Sand and Gravel and Gravel.	2.01E-02	-	-

For the least sensitive boundary conditions, the recharge applied beneath the bluffs and the inflow applied to the southeast and northeast boundaries, observations were not available to inform parameter bounds. Therefore, the bounds were set subjectively based on experience with previous models developed for the Site and are presented in Table 3.2

Table 3.2: Boundary Condition Values and Bounds Used During Calibration

Boundary Condition	Initial Value	Lower Bound	Upper Bound
Northeastern Inflow	10,000 ft ³ /d	0 ft ³ /d	30,000 ft ³ /d
Southwestern Inflow	10,000 ft ³ /d	0 ft ³ /d	20,000 ft ³ /d
Bluffs Recharge	380 mm/yr	0 mm/yr	2300 mm/yr

3.3.4 Calibration Results

A properly calibrated model should (1) provide a reasonable match to observations in that lack of fit is small and weighted residuals are randomly distributed relative to observed values and (2) have optimized parameter values which fall within reasonable bounds ([Hill and Tiedeman, 2007](#)). Although the sum of squared errors was used as the objective function for calibration, model fit can be measured using any number of residual statistics. One measure that is used is the Normalized Root Mean Squared Error (NRMSE) as presented in Equation 3.2. NRMSE is a measure of the average model error per calibration target.

$$NRMSE = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (h_s - h_m)_i^2}}{\max(h_m) - \min(h_m)} \quad (3.2)$$

Where n is the number of observations [dimensionless]; h_s is the simulated head at point i [L] and; h_m is the measured head value at point i [L]

The British Columbia Ministry of the Environment states that a NRMSE under 10 percent is considered good in many models, and less than 5 percent is very good with respect to average residual fit ([Robertson GeoConsultants Inc. and SRK Consulting Inc., 2012](#)). A similar goodness of fit goal proposed by [Environmental Simulations \(2011\)](#) is that the residual standard deviation divided by the range (referred to as normalized standard deviation) in observations should be less than 10 percent. For comparison, final calibration statistics are presented in Table 3.3. Each model is identified by: (1) the recharge condition

specified, either low recharge (LR) or high recharge (HR), (2) the conductivity model used, either homogeneous (S1), zoned (S2) or pilot points (S3) and (3) whether a general head boundary was applied at the bottom of the model domain (G) or inflow was applied along the north east model boundary (W). As shown in Table 3.3, the normalized standard deviation and NRMSE are between 4.0 and 6.5 percent indicating a good to very good fit with respect average to residual values. The model calibrations are also consistent with the expectation that model fit improves with an increasing number of parameters as is illustrated by the lower SSE values of models with more complex conductivity fields, characterized by an increase in the number of parameters.

Table 3.3: Residual Statistics

Model	SSE	Mean Residual (ft)	Absolute Residual Mean	Residual Standard Deviation	Normalized Standard Deviation	NRMSE
LR-S1-G	140.72	-0.08	0.68	0.92	0.061	6.1%
LR-S1-W	140.11	-0.03	0.68	0.92	0.061	6.1%
HR-S1-G	139.43	-0.09	0.69	0.91	0.060	6.1%
HR-S1-W	139.56	-0.03	0.68	0.92	0.061	6.1%
LR-S2-G	135.72	0.00	0.67	0.90	0.060	6.0%
LR-S2-W	137.49	-0.02	0.66	0.91	0.060	6.0%
HR-S2-G	135.77	-0.02	0.65	0.90	0.060	6.0%
HR-S2-W	136.85	-0.04	0.66	0.91	0.060	6.0%
LR-S3-G	78.22	-0.04	0.47	0.69	0.045	4.5%
LR-S3-W	77.68	0.00	0.47	0.68	0.045	4.5%
HR-S3-G	65.99	-0.02	0.43	0.63	0.042	4.2%
HR-S3-W	74.08	0.01	0.48	0.67	0.044	4.4%

Scatter plots of simulated vs. observed heads for models LR-S2-W and HR-S3-G are presented in Figures 3.10 and 3.11, respectively. Scatter plots for the remaining models are shown in Appendix A. Residuals presented in the scatter plots are colour coded by grouping targets based on their approximate depths. The scatter plots illustrate that the

model calibrations are reasonable with most residuals falling within 10 percent of the head range (1.5 feet) of the exact match line. In all models, the overall distribution of residuals around the exact match line is approximately random, showing little bias, which is also indicative of a reasonable calibration.

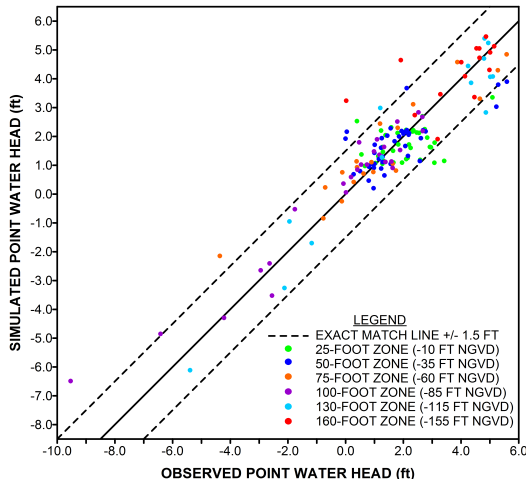


Figure 3.10: Model LR-S2-W Simulated vs. Observed Head

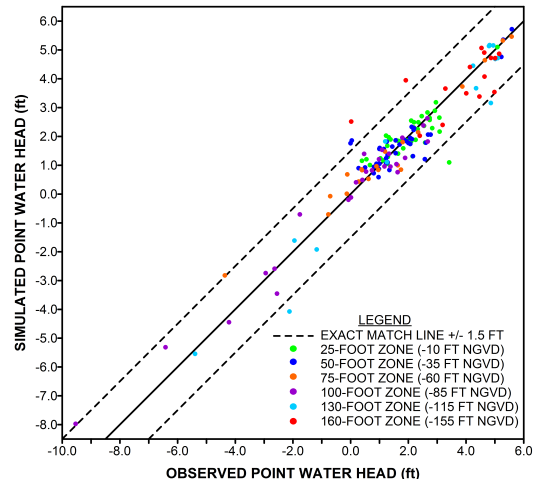


Figure 3.11: Model HR-S3-G Simulated vs. Observed Head

Visual inspection of the calibration targets by depth zone indicates that the calibration of some zones may be slightly positively or negatively biased. For example, the 25 foot zone in Figure 3.10 appears to show a minor positive bias (observed head is greater than simulated head). Table 3.4 breaks the residual means down by depth. Table 3.4 confirms the small positive bias observed in the scatter plots. While it is observed that most zones exhibit a small positive or negative bias, it is not correlated with depth (i.e., positive bias can be observed both in the 25 foot zone and the 130 foot zone). Given that the Site is hydraulically constrained by the Hylebos and Blair Waterway the calibration involves tradeoffs in model fit between different zones. Any effort to decrease the heads at the bottom of the model domain may reduce the negative bias in the 160 foot zone, but would

increase the positive bias in the 130 foot zone above it. Similarly increasing the heads in the 25 foot zone would reduce the positive bias in the 25 foot zone, but it would likely result in a more negative bias in the 50 and 75 foot zones since the increasing heads in the 25 foot zone would also increase heads in the 50 and 75 foot zones. Therefore, given the tradeoffs in positive and negative bias between layers and the limiting factors of Site hydrology the least biased option was selected.

Table 3.4: Mean Residual by Model Zone (ft)

Model	25-ft Zone	50-ft Zone	75 ft Zone	100 ft Zone	130 ft Zone	160 ft Zone
LR-S1-G	0.20	-0.08	-0.3	-0.42	0.36	-0.09
LR-S1-W	0.28	0.02	-0.16	-0.3	0.2	-0.37
HR-S1-G	0.02	-0.17	-0.33	-0.34	0.49	0.22
HR-S1-W	0.27	0.02	-0.15	-0.3	0.23	-0.34
LR-S2-G	0.31	0.03	-0.2	-0.28	0.31	-0.15
LR-S2-W	0.32	0.04	-0.2	-0.3	0.22	-0.36
HR-S2-G	0.34	0.05	-0.25	-0.3	0.17	-0.26
HR-S2-W	0.32	0.04	-0.24	-0.32	0.15	-0.35
LR-S3-G	0.2	-0.09	-0.17	-0.26	0.33	-0.14
LR-S3-W	0.2	-0.01	-0.18	-0.19	0.32	-0.08
HR-S3-G	0.03	-0.03	-0.12	-0.1	0.28	-0.04
HR-S3-W	0.14	-0.04	-0.06	-0.08	0.31	-0.12

Another measure of the quality of model calibration is whether or not the optimized parameter values are reasonable and/or whether calibrated parameter values are pushed to bounds. In this study, parameter values were generally constrained within a reasonable range based on available studies or field data. Optimized Parameter values are presented in Table 3.5. To provide grounds for comparison between all models, cell-averaged conductivity values are reported. All horizontal conductivity values fall between the 10th percentile (0.29 ft/d) and 90th percentile (47.34 ft/day) of the measured horizontal conductivity for sand and silty sand, the two most common soil types encountered beneath the Site. All

vertical conductivity values also compare well with the 10th percentile (0.01 ft/d) and 90th percentile (3.69 ft/day) of the measured vertical conductivity values for sand and silty sand. All optimized parameters are within the bounds specified in section 3.3.3.

Table 3.5: Calibrated Parameter Values

Model	kx	kz	ani	Site	Bluffs	Southern	Northeast	GHB
	ft/d	ft/d		RCH mm/yr	RCH mm/yr	Inflow ft ³ /d	Inflow ft ³ /d	ft
LR-S1-G	4.88	0.07	70.11	48	728	11,110.01	-	1.10
LR-S1-W	1.88	0.07	26.37	28	175	2,817.37	3,120.07	-
HR-S1-G	15.61	0.18	85.37	318	2259	10,111.50	-	0.19
HR-S1-W	19.99	0.72	27.83	325	1946	7,602.07	9,147.01	-
LR-S2-G	6.05	0.26	23.64	39	390	0.40	-	0.75
LR-S2-W	5.28	0.18	29.18	44	195	0.54	9,341.25	-
HR-S2-G	41.81	1.15	36.32	277	1773	0.39	-	1.23
HR-S2-W	41.64	1.74	23.87	286	1642	10.97	5.46	-
LR-S3-G	14.46	2.82	5.14	60	2199	10,040.23	-	1.12
LR-S3-W	7.02	0.39	17.93	38	130	5,943.37	6,608.21	-
HR-S3-G	37.04	1.02	36.27	320	2278	782.33	-	0.45
HR-S3-W	31.25	4.46	7.00	316	2086	8,306.03	7,135.66	-

3.4 Summary

All 12 models used in the multiple model approach met and exceeded the goals set out for model calibration. The Normalized Root Mean Squared Error for all the models ranged from 4 to 6.5 percent, below the goal of 10 percent specified by the British Columbia Ministry of the Environment to indicate good model calibration. Some negative and positive biases were observed in individual model zones, however given the constrained nature of the Site the bias in one zone could not be improved without increasing the bias in another zone. Overall, model mean residuals were close to zero, ranging from 0.01 to -0.09 ft, indicating

little overall bias. In addition, all calibrated parameter values for each model fell within the specified ranges, which were supported by field data where available. The sum squared error (SSE) of all 12 models ranges from 66 to 141 ft². Models with a pilot point based stratigraphy conceptualization had the lowest SSE since 49 parameter were used in calibration, compared to 7 parameter in models with homogeneous stratigraphy. It's expected that the models with a larger number of parameter perform better in calibration as the additional parameters make the models more flexible. Therefore, the calibration results of the simple homogeneous models are comparable to those which used pilot points, when the number of parameters is taken into consideration. These calibration results demonstrate that all 12 models are comparable and adequately represent the observed data. This is significant because it shows that the models can be treated equally and be used for remedial design in the context of a multiple model approach.

Chapter 4

Results and Discussion

4.1 Pump-and-Treat System Design

The objective of pump-and-treat system design is a system that minimizes capital and maintenance costs while achieving adequate containment of the contaminant. Pump-and-treat system design variables can include the number of pumping and injection wells, and the locations and pumping rates of those wells, all of which must be constrained to reasonable values. Due to the size of the design space, automatic optimization is often useful. In this study, only pumping rates are considered for optimization for the following reasons: (1) the site has an existing pump-and-treat system for which pumping rates can be optimized (2) to focus on the impact of conceptual model uncertainty on remedial design without clouding arguments with a complicated pump-and-treat optimization problem.

The pump-and-treat system was designed by applying an optimization algorithm to determine the minimum pumping rates at existing well locations required to achieve contaminant at the Site. The contaminant plume is defined as the anthropogenic density

plume which contains the majority of highly concentrated contaminants of concern and containment is feasible without changing well locations or adding additional wells. Containment is evaluated using particle tracking to determine what percent of the plume is hydraulically contained. Theoretical particles are released from each cell within the anthropogenic density plume and tracked through the model domain using MODPATH. Particles which exit the model domain through the pump-and-treat system are contained while those which reach other model boundaries are not.

The optimization problem was formulated to minimize the total pumping rate summed across all groundwater extraction wells at the Site while meeting the containment constraint. There are a total of 21 extraction wells (extracting water from the subsurface) and 13 injection wells (injecting water to the subsurface). Injection well pumping rates were fixed at their current pumping rates, since the groundwater chemistry at the Site causes precipitate to build up around the injection wells, reducing flow rates and making the injection system inefficient to run; larger flow rates are impractical. The optimization problem was implemented as a general constrained optimization problem using the Ostrich software package (Matott, 2006). The optimization objective function is shown in Equation 4.1. Extraction well pumping rates were constrained between 0 and 10,000 ft³/d. The upper bound was set at 10,000 ft³/d because the preprocessor used to assign pumping rates to model cells based on cell transmissivity could not assign a cell specific pumping rates in excess of 10,000 ft³/d. In addition, the upper bound is sufficient to achieve containment at the Site. Containment of the anthropogenic density plume was controlled using the penalty function P , in the optimization objective function. If containment, measured as the percentage of particles captured by the pump-and-treat system, was greater or equal to 99 percent $P = 1$, otherwise $P = 10^7$.

$$F(Q, P) = P \sum_{i=1}^N |Q_i| \quad (4.1)$$

Where P is the penalty function which accounts for any constraint violations [L^3/T]; N is the number of pumping wells [dimensionless] and; Q is the pumping rate of well i [L^3/T].

To optimize the pumping rates at existing well locations, the dynamically dimensioned search algorithm (DDS), which has been shown to be an effective optimizer for computationally expensive models (Tolson and Shoemaker, 2007), was used. The optimization process for each model consisted of 24 optimization runs, each with different random seeds, and consisting of 500 model evaluations. The best solution from the set of 24 runs was used as the starting point for final DDS run consisting of 1000 model evaluations to determine to verify that the DDS algorithm had reached an optimized solution. The default perturbation values of 0.2 was used in DDS.

4.2 Optimal Pump-and-Treat Designs

The optimized total pumping rates calculated by DDS for each model are shown in Table 4.1. Optimized pumping rates broken down by well are shown in Appendix B. In all cases optimized pumping rates corresponded to a particle capture percentage between 99.000 and 99.005 percent, indicating that DDS successfully converged to a solution without violating the capture constraint.

As shown in Table 4.1, optimized pumping rates between all models range from approximately 7,000 to 131,000 ft^3/d . The optimized pumping rates have a mean and standard deviation of 41,328 ft^3/d and 34,428 ft^3/d , respectively. This high variation in optimized pumping rates shows that models which perform similarly in calibration to head observa-

Table 4.1: Optimization Results, Ranked by Total Pumping Rate

Model	Total Pumping Rate (ft ³ /d)
LR-S1-W	7,295
LR-S2-W	7,942
LR-S2-G	13,351
LR-S1-G	13,447
HR-S3-G	23,242
LR-S3-G	37,926
LR-S3-W	39,470
HR-S1-G	45,673
HR-S2-W	49,865
HR-S2-G	51,347
HR-S1-W	67,208
HR-S3-W	130,812

tions can produce very dissimilar results when used in a predictive capacity, supporting the findings presented by [Højberg and Refsgaard \(2005\)](#). The large range in pumping rates, differing by a factor of 18 from the lowest to highest, further supports that the 12 models characterize a large range of uncertainty at the Site.

As one would expect, models with lower recharge and hydraulic conductivity generally have lower optimized pumping rates than those with higher recharge rates and hydraulic conductivity. For comparison, the six models with the lowest optimized pumping rates have an average recharge and hydraulic conductivity of 90 mm/yr and 11.60 ft/d, compared to an average recharge and hydraulic conductivity of 260 mm/yr and 23.22 ft/d for the six models with the highest optimized pumping rates. These results are expected since higher recharge and hydraulic conductivity will increase flow through the system, thereby increasing the total pumping required to achieve containment at the Site as this additional recharge water must also be removed. In addition, the link between recharge and pumping rates shown in [Table 4.1](#) indicates that if this exercise were used to suggest areas for additional data

collection, better characterization of recharge could significantly reduce conceptual model uncertainty.

To compare the effects of different conceptual model choices on the optimized pumping rates, the average optimized pumping rate for each conceptual model choice is shown in Table 4.2. The choice of recharge condition showed the largest difference, with the average pumping rates for LR and HR separated by approximately a factor of three. In comparison, the average optimized pumping rates for the W and G conceptualizations are separated by a factor of 1.7 and the S3 conceptualization is approximately a factor of 1.9 above S1 and S2. These results are, in part, dominated by the high pumping rate for conceptual model HR-S3-W, which is almost 100 percent greater than the second highest optimized pumping rate. A third column showing the average pumping rates for each conceptualization excluding model HR-S3-W is included in Table 4.2 to facilitate the comparison between each conceptualization. The average optimized pumping rates from removing HR-S3-W further support that the choice of the conceptual model for recharge has the greatest impact on optimized pumping rates. This holds true when HR or LR is combined with the S1 and S2 conceptualizations, with the LR conceptualization combined with S1 and S2 models having consistently lower optimized pumping rates than the HR conceptualization S1 and S2 models. However, in the case of the S3 conceptualization HR models have the highest and lowest optimized pumping rates, with the LR models falling in the middle. It is possible that the more complex S3 conceptualization leads to a more variable stratigraphy resulting in preferential flow paths or zones of low conductivity. This could explain how the optimized pumping rates can be less dependent on the recharge conceptualizations when combined with a pilot point stratigraphy model.

Table 4.2: Average Optimized Pumping Rates by Model Conceptualization		
Conceptualization	Average Pumping ft ³ /d	Average Pumping Excluding HR-S3-W ft ³ /d
W	50,432	34,356
G	30,831	30,831
LR	19,905	19,905
HR	61,358	47,467
S1	33,406	33,406
S2	30,626	30,626
S3	57,863	33,546

4.3 Cross-Comparison of Remedial Designs

While all 12 models used together cover a range of conceptual model uncertainty, if any individual model was used without the characterization of this uncertainty, under- or over-design of the system is very plausible. To examine the robustness of a given model/pump-and-treat design to conceptual model uncertainty, each of the 12 pumping system designs were applied across all of the models. This is equivalent to testing the designs in 12 different plausible realities. Results of this comparison are shown in Table 4.3. The total pumping rate and the model for which it was optimized are shown in the left most columns, while the models in which the pumping rates are tested are shown across the top row of Table 4.3. Where each system design was tested in the model for which it was optimized, as shown along the diagonal, the capture percentage is 99 percent, corresponding to the capture constraint imposed during optimization. The seven highest optimized pumping rates (those above 37,000 ft³/d) had above 90 percent capture in all comparisons, leading to an effective over-design for the majority of conceptual models. As would be expected, the lower optimized pumping rates for model LR-S1-W, LR-S2-W, LR-S2-G and LR-S1-

G, showed poor performance when tested in less forgiving conceptual models where high optimized pumping rates were needed for full containment (e.g., HR-S1-W and HR-S3-W). To distinguish between clear design failures and reasonable performance, a threshold for containment failure was specified at 85 percent capture. Using this definition of failure, 21 percent of pumping system designs failed to achieve the target of 85 percent capture when tested in all 12 models (shown by bold highlighted text in Table 4.3). The average capture for comparisons which failed to reach the threshold was 61 percent, while those above the threshold had an average capture of 98 percent. This comparison is useful because it identifies the sensitivity of design choices to conceptual model uncertainty. For this case study it shows that the selection of a single non-conservative model would lead to a high probability of containment failure.

Table 4.3: Particle Capture Results for Optimized Pumping Rates Applied in All Models

Model	Q (ft ³ /d)	LR-S1-W	LR-S2-W	LR-S2-G	LR-S1-G	HR-S3-G	LR-S3-G	LR-S3-W	HR-S1-G	HR-S2-W	HR-S2-G	HR-S1-W	HR-S3-W
LR-S1-W	7,295	99.00	97.59	90.43	89.67	62.45	51.50	90.71	57.94	50.93	52.49	31.02	40.08
LR-S2-W	7,942	98.89	99.00	94.59	92.82	54.07	49.35	87.36	53.75	49.60	51.74	29.76	38.88
LR-S2-G	13,351	99.77	99.76	99.00	98.67	78.90	69.39	96.56	76.82	69.55	71.38	58.39	61.82
LR-S1-G	13,447	99.80	99.83	98.45	99.00	82.86	75.93	96.35	77.83	67.68	69.37	63.62	61.41
HR-S3-G	23,242	100.00	100.00	99.93	99.86	99.00	95.53	97.89	93.62	87.81	88.14	84.85	83.74
LR-S3-G	37,926	100.00	100.00	100.00	100.00	99.58	99.00	98.75	98.27	96.39	96.56	90.74	89.92
LR-S3-W	39,470	100.00	100.00	100.00	100.00	99.70	98.97	99.00	98.92	95.90	96.13	90.46	95.89
HR-S1-G	45,673	100.00	100.00	100.00	100.00	99.69	99.44	99.04	99.00	97.33	97.09	91.28	92.85
HR-S2-W	49,865	100.00	100.00	100.00	100.00	99.70	99.62	99.28	99.20	99.00	98.99	98.23	96.01
HR-S2-G	51,347	100.00	100.00	100.00	100.00	99.67	99.56	99.27	99.12	99.03	99.00	98.33	93.00
HR-S1-W	67,208	100.00	100.00	100.00	100.00	99.78	99.09	99.53	99.89	99.70	99.56	99.00	96.57
HR-S3-W	130,812	100.00	100.00	100.00	100.00	99.90	99.75	100.00	100.00	99.93	99.86	99.73	99.00

4.4 Factor-of-Safety Approach

Typically, only a single "best" model is used to aid in remedial design. An evaluation of parameter uncertainty and/or the application of a factor-of-safety (FOS) to resultant designs would be used to account for uncertainty. However, it is recognized that conceptual model uncertainty is often the main source of uncertainty in model prediction and the evaluation of parameter uncertainty alone does not sufficiently account for all sources of uncertainty (Neuman and Wierenga, 2003; Højberg and Refsgaard, 2005; Poeter and Anderson, 2005). Presumably, the FOS approach implicitly addresses this, with a much smaller effort than a multiple conceptual model framework. Here, the factor-of-safety approach is directly compared against and evaluated using the multiple model approach. Theoretically, an appropriately selected factor-of-safety could account for model uncertainty. However, one runs the risk of over-designing the remedial system and there is no concrete methodology which can be used to select an appropriate factor of safety. Although there is no universally accepted factor-of-safety recommended for remedial design, the (USEPA, 1997, 2005, 2008) provides some guidance, citing a factor-of-safety ranging from 1.5 to 2 for applications of plume containment, injection well design and pump and treat system design. Therefore, a factor-of-safety of 1.5 and 2 were selected for comparison with the multiple model approach.

To apply the factor-of-safety, the total flow of each optimized pump and treat system was increased by uniformly multiplying the optimized pumping rate at each well by the given factor-of-safety. Tables 4.4 and 4.5 present the cross-conceptual model comparison of containment performance for the optimized pumping rates for each conceptual model implemented each other conceptual model for a FOS of 1.5 and 2, respectively.

Using factor-of-safety of 1.5 resulted in an average capture rate of 94 percent, compared to 90 percent with no factor-of-safety applied to the optimized pumping rates. Fifteen

Table 4.4: Cross-Conceptual Model Containment Performance with FOS of 1.5

Model	Q (ft ³ /d)	LR-S1-W	LR-S2-W	LR-S2-G	LR-S1-G	HR-S3-G	LR-S3-G	LR-S3-W	HR-S1-G	HR-S2-W	HR-S2-G	HR-S1-W	HR-S3-W
LR-S1-W	10,943	99.58	99.45	96.65	96.34	74.98	65.69	95.39	72.33	63.55	64.84	53.74	53.24
LR-S2-W	11,913	99.55	99.53	98.19	97.22	66.07	59.95	95.93	71.22	65.28	66.85	51.74	52.23
LR-S2-G	20,026	99.99	99.98	99.86	99.72	90.91	85.16	97.56	88.51	80.06	81.11	78.29	73.29
LR-S1-G	20,171	100.00	99.99	99.89	99.77	97.83	89.51	97.44	88.96	80.32	81.70	79.40	73.88
HR-S3-G	34,864	100.00	100.00	100.00	100.00	99.63	98.80	98.82	98.28	96.62	96.78	91.30	92.18
LR-S3-G	56,888	100.00	100.00	100.00	100.00	99.75	99.68	99.40	99.62	99.34	99.05	98.05	97.10
LR-S3-W	59,206	100.00	100.00	100.00	100.00	99.80	99.69	99.53	99.83	99.37	99.08	97.75	97.76
HR-S1-G	68,509	100.00	100.00	100.00	100.00	99.79	99.71	99.53	99.88	99.71	99.46	98.31	97.78
HR-S2-W	74,797	100.00	100.00	100.00	100.00	99.80	99.71	99.64	99.89	99.75	99.69	99.01	97.98
HR-S2-G	77,020	100.00	100.00	100.00	100.00	99.78	99.72	99.62	99.91	99.76	99.70	99.04	98.01
HR-S1-W	100,811	100.00	100.00	100.00	100.00	99.87	99.77	99.90	100.00	99.70	99.56	99.56	98.47
HR-S3-W	196,219	100.00	100.00	100.00	100.00	99.94	99.75	100.00	100.00	99.93	99.86	99.98	99.35

Table 4.5: Cross-Conceptual Model Containment Performance with FOS of 2

Model	Q (ft ³ /d)	LR-S1-W	LR-S2-W	LR-S2-G	LR-S1-G	HR-S3-G	LR-S3-G	LR-S3-W	HR-S1-G	HR-S2-W	HR-S2-G	HR-S1-W	HR-S3-W
LR-S1-W	14,590	99.86	99.85	98.76	99.22	85.21	77.04	96.56	79.75	69.51	71.10	66.67	63.08
LR-S2-W	15,884	99.85	99.82	99.25	98.78	75.45	67.05	96.81	79.83	73.38	74.94	67.59	60.52
LR-S2-G	26,702	100.00	100.00	99.97	99.91	98.85	91.34	98.16	95.82	88.68	89.03	87.00	79.03
LR-S1-G	26,895	100.00	100.00	99.99	99.94	99.13	92.76	98.11	93.58	89.20	89.53	85.45	81.08
HR-S3-G	46,485	100.00	100.00	100.00	100.00	99.73	99.59	99.29	99.35	98.73	98.69	97.39	96.57
LR-S3-G	75,851	100.00	100.00	100.00	100.00	99.81	99.72	99.64	99.92	99.81	99.73	99.10	97.96
LR-S3-W	78,941	100.00	100.00	100.00	100.00	99.87	99.71	99.83	99.93	99.91	99.84	99.23	98.18
HR-S1-G	91,346	100.00	100.00	100.00	100.00	99.85	99.73	99.80	99.96	99.71	99.46	99.38	98.18
HR-S2-W	99,729	100.00	100.00	100.00	100.00	99.87	99.74	99.94	99.97	99.75	99.82	99.43	98.41
HR-S2-G	102,693	100.00	100.00	100.00	100.00	99.85	99.74	99.91	99.97	99.76	99.70	99.46	98.39
HR-S1-W	134,415	100.00	100.00	100.00	100.00	99.91	99.78	100.00	100.00	99.70	99.56	99.82	98.90
HR-S3-W	261,625	100.00	100.00	100.00	100.00	99.94	99.75	100.00	100.00	99.93	99.86	100.00	99.51

percent of pumping rates failed to achieve the containment thresholds, while the capture percentage of those which failed increased to 67 percent from 61 percent. Using a factor of safety of 2, the average capture was 96 percent and the number of comparisons under the 85 percent capture threshold decreased further to 9 percent. Assuming all model are equally likely, use of any of the single models developed here with a factor of safety would result in a 9 to 15 percent chance of failure, as compared to a 21 percent chance of failure with no FOS. This clearly illustrates that the FOS reduces the likelihood of failure, at a cost directly proportional to the value selected for the FOS. However, even a conservative FOS of 2.0 does not protect entirely against conceptual model uncertainty. While Tables 4.4 and 4.4 show that non-conservative models such LR-S1-W and LR-S2-W which likely underestimate flow through the Site, have high chance of failure (greater than 50 percent) despite the a FOS of 2.0, it should be recognized that the models used in the multiple model approach were designed to encompass the range, rather than be used individually without explicitly considering conceptual model uncertainty. This does, however, support the argument that the use of single non-conservative could lead to high chance of failure despite the use of a FOS. To better examine the use of a FOS approach compared to a multiple model approach, Section 4.5 presents a comparison of a FOS approach using models developed in practice against the multiple model approach.

4.5 Comparison of Models Developed in Practice to a Multiple Model Approach

Site characterization and groundwater modelling of the Site is an ongoing process. Several models have been developed of the site by engineering practitioners. Three groundwa-

ter models developed by Conestoga-Rover & Associates (CRA) were used for comparison against the multiple model approach. The purpose of this comparison is to provide a relatively independent estimate of a FOS approach with an actual model used by practitioners. The first model, referred to as Model 1, was developed in 2010 to provide a predictive tool to aid in the comparison and evaluation of remedial alternatives for the Site. It was developed as a single "best" approximation of Site. The next two models, Models 2 and 3 were developed in 2011 to bracket the upper range of flow through the Site, rather than provide a single "best" estimate of Site conditions. Model 2 represents what was believed to be a plausible upper range of flow through the Site. Model 3 was intentionally designed to represent unrealistically high flow rates at the Site, with the specified recharge rate at the Site over double the annual average precipitation, which would lead to an ultra conservative pump-and-treat system design. While Models 2 and 3 may not be applicable with a FOS approach as they are already designed to be conservative, comparing them to the multiple model approach provides an indication of the value of expert judgment in model design.

Optimized flow rates were developed for Models 1, 2 and 3 using the same method used in the multiple model approach and are presented in Table 4.6. The optimized pumping rate for Model 1 falls between the fifth and sixth lowest pumping rates when compared to the 12 models used in the multiple model approach. Model 2, which was designed to bracket the upper range of flow expected at the Site corresponds well to the highest pumping rate from the multiple model approach, with a very large optimized pumping rate of 120,753 ft³/d compared to 130,812 ft³/d for model HR-S3-W. The third model, Model 3 had an optimized pumping rate of 434,770 ft³/d, well in excess of any model developed for the multiple model approach. This was expected given the high recharge rate specified in Model 3, which was intended to overpredict flow through the Site. The

optimized pumping rates for the three models developed in practice suggest that they may perform as was intended by those developing the models when compared to the multiple model approach; however, capture percentages must be examined for a more complete comparison.

Table 4.6: Optimized Pumping Rates for CRA Models

	Pumping Rate (ft ³ /d)
Model 1	30,008
Model 2	120,753
Model 3	434,770

Table 4.7 shows the optimized pumping rates for Models 1, 2 and 3 tested in all 12 models developed for the multiple model approach. The optimized pumping rates for Model 3 perform very well when tested in all 12 Models used in the multiple model approach, achieving above 99.5 percent capture in each model. This is expected since Model 3 was designed to predict very high flow through the Site, leading to a very conservative estimate of the required pumping for containment. Model 2, which was designed to predict the upper plausible range of flow through the Site achieved good containment in all models, ranging from 98.5 to 100 percent containment. Model 2, designed to be reasonably conservative, does a good job of capturing the upper end of flow predicted at the Site as evident by the high particle capture percentages across all models. Model 1, which was designed as a single “best” model to approximate Site conditions shows the lowest capture of the three. However, the capture percentage only drops below the 85 percent threshold for a single conceptual model, HR-S3-W. Table 4.7 illustrates that in this case study, expert judgment was used effectively to predict an upper range of flow through the Site which is comparable to that predicted by the multiple model approach. This is evident in the similar pumping rates of Model 2 and model HR-S3-W. This is further supported by the high capture percentages obtained by model 2 across all conceptual models used in the

multiple model approach.

Model 1 was designed as a single “best” approximation of Site conditions, recognizing that there will be uncertainty associated with any design predictions. Applying a factor of safety to remedial designs is one method to address this uncertainty. Table 4.8 shows particle capture percentages for Model 1 with a factor-of-safety of 1.5 and 2 applied to optimized pumping rates. This evaluation was not carried out with Models 2 and 3 as they were intended to represent the upper range of flow through the Site, and therefore would likely be considered conservative in practice without the addition of a FOS. In both cases, with a FOS of 1.5 and 2, optimized pumping rates for Model 1 achieved above 85 percent containment across all conceptual models, indicating that reasonable containment is achieved in all cases. This indicates that applying Model 1 with a FOS of 1.5 to 2.0, is a reasonable and far less costly surrogate for a multiple model approach.

Table 4.7: Capture Percentages for Optimized Pumping Rates from Models 1, 2, and 3

Model	Model 1	Model 2	Model 3
LR-S1-W	99.92	100.00	100.00
LR-S2-W	100.00	100.00	100.00
LR-S2-G	96.49	100.00	100.00
LR-S1-G	90.08	99.67	100.00
HR-S3-G	99.98	100.00	100.00
LR-S3-G	100.00	100.00	100.00
LR-S3-W	90.61	99.65	99.81
HR-S1-G	90.63	99.84	99.87
HR-S2-W	88.81	99.79	99.86
HR-S2-G	98.31	99.98	100.00
HR-S1-W	98.47	99.89	99.98
HR-S3-W	78.79	98.53	99.58
Average	94.34	99.78	99.93

Table 4.8: Capture Percentages for a Factor of Safety Applied to Model 1 Optimized Pumping Rates

Model	FOS = 1	FOS = 1.5	FOS = 2
LR-S1-W	99.92	100.00	100.00
LR-S2-W	100.00	100.00	100.00
LR-S2-G	96.49	98.55	99.56
LR-S1-G	90.08	97.86	98.63
HR-S3-G	99.98	100.00	100.00
LR-S3-G	100.00	100.00	100.00
LR-S3-W	90.61	98.05	99.27
HR-S1-G	90.63	98.27	99.41
HR-S2-W	88.81	94.97	97.76
HR-S2-G	98.31	99.06	99.41
HR-S1-W	98.47	99.35	99.70
HR-S3-W	78.79	85.28	91.82
Average	94.34	97.62	98.80

4.6 Testing the Robustness of Remedial Designs

Robustness is the ability of a system to overcome change or adverse conditions. The most common changes which would occur during the operation of a pump-and-treat system includes silting, or fouling of wells which would reduce system pumping rates and overall containment. However, system robustness may also consider the adversity of uncertainty; a robust design performs well despite an unknown system, a less robust design will be highly sensitive to the unknown. To test the robustness of each optimized pump-and-treat system developed from the multiple model approach, the well flow rates were incrementally reduced and particle capture was re-evaluated. To the author's best knowledge, this method of testing the robustness of pump-and-treat designs has not been presented in the literature. Figure 4.1 shows the percentage of particles captured plotted against the percent of the optimized pumping rate for each of the 12 models. A truly robust though

potentially oversized system would provide high containment even with a reduction in flow rate, whereas a non-robust system would be characterized by failure with relatively small reductions in pumping rate. As shown in Figure 4.1, there is considerable variation in the performance of the optimized remedial system as their pumping rate is reduced. For example, at 15 percent of their respective optimized pumping rates, there is close to a 50 percent difference in capture between model LR-S3-W and model HR-S1-W. Of all 12 models, LR-S3-W may be considered the most robust with respect to a decrease in total pumping rates since it displays the lower decrease in capture percent as the pumping rate is decreased. This is also evident in particle capture percentages when the optimized pumping rates from other models are tested in LR-S3-W, all of which achieved capture greater than 85 percent in model LR-S3-W. Comparatively, less robust models with lower optimized pumping rates (HR-S3-G and LR-S3-G) exhibited lower capture than LR-S3-W when the four lower optimized pumping rates were tested in those models. However, this trend is not observed in HR-S3-W, which would be the next most robust model based on Figure 4.1. This is likely due to the high flow through HR-S3-W which resulted a large optimized pumping rate for that model.

Figure 4.1 also illustrates that a large increase in pumping rate, on the order of 40 to 50 percent of the optimized pumping rate is required to achieve the final 10 percent of simulated capture (i.e., to move from 89 to 99 percent capture). This can be expected since the pumping rates were optimized at existing well locations. Significant increases of pumping rates are required to capture particles that are not located in close proximity to the pumping wells. Strategically adding additional wells at locations where capture is poor could improve the efficiency of the pump-and-treat system. However, adding additional wells will increase capital costs and may reduce the robustness to decreases in pumping rates. Furthermore, installing additional wells does not prevent over-design of the pump-

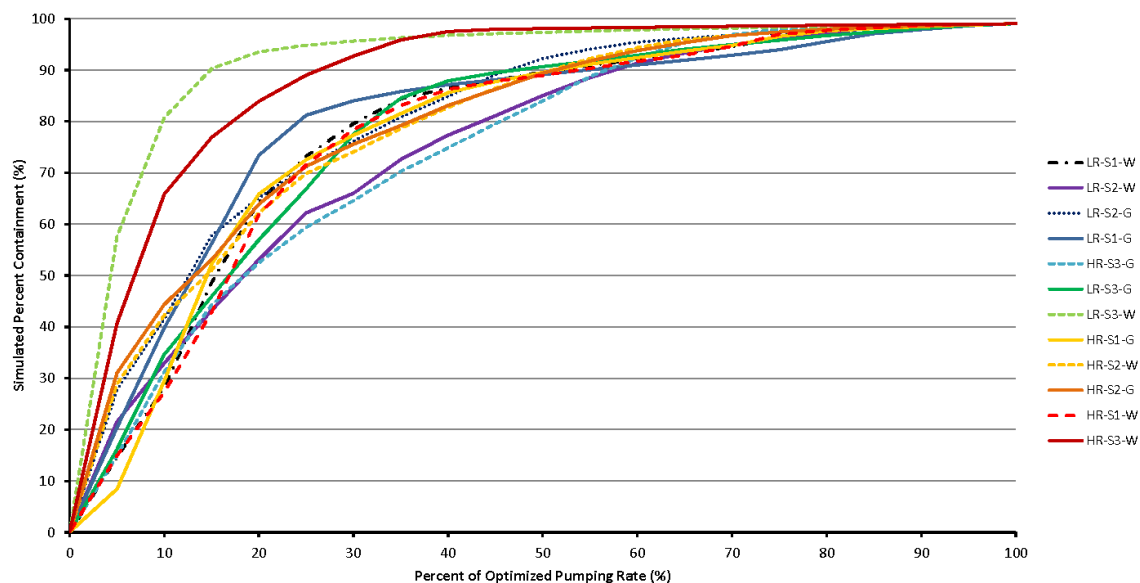


Figure 4.1: Percent Containment vs. Percent of Optimized Pumping Rate

and-treat system since predictions may be made using overly conservative models. For example, if actual Site conditions correspond to the middle or lower range of the models using in the multiple model approach and the highest optimized pumping rate is applied as a conservative design, pumping rates could be over 10 times in excess of what may be required as shown by the difference in total pumping between Models LR-S1-W and HR-S3-W in Table 4.1.

Chapter 5

Conclusions and Recommendations

The utility of using multiple conceptual models for the design of groundwater remediation systems was investigated through the use of a case study. A set of 12 conceptual models were constructed to represent the range of conceptual model uncertainty at the Site, and a pump and treat system was designed/optimized for each model. An evaluation of optimized pumping rates indicated that the choice of recharge condition had the largest effect on pumping rates. This shows that the construction and evaluation of multiple conceptual models could be used to identify areas for additional data collection (in this case for recharge) which could most effectively reduce conceptual model uncertainty.

To examine the robustness of individual pump and treat designs to conceptual model uncertainty each pump and treat system was compared across all 12 conceptual models, with no factor-of-safety, a factor-of-safety of 1.5, and a factor-of-safety of 2. This comparison illustrated that there was considerable variation in containment between the pump and treat systems designed for each conceptual model. Nonconservative models with lower pumping rates did not achieve adequate containment across all conceptual models, even

with the consideration of a relatively conservative factor-of-safety of 2. However, it was demonstrated that a factor-of-safety reduces the chance of design failure at a cost directly proportional to the value of the factor-of-safety. Alternatively, conservative models with high pumping rates achieved adequate containment across all conceptual models even without the addition of a factor-of-safety, indicating that the use of a conservative model could lead to over design of the remedial system with respect to most conceptual models. Overall, the comparison of pump and treat designs within the multiple model approach showed that a factor of safety does not protect entirely against conceptual model uncertainty if a nonconservative base model is used, and that a multiple model approach does not necessarily address over design of remedial systems. Recognizing this, a factor-of-safety approach applied to an expertly developed model will in most cases be preferable to a multi-model approach, if for no other reason than the cost of the latter does not provide sufficiently large insurance against the impacts of uncertainty for consideration in practice.

The models used in the multiple model approach were designed to encompass the range of conceptual model uncertainty at the Site and not for any one model to be applied individually to make design decisions. Therefore, to obtain a relatively independent comparison of a factor-of-safety approach, optimized pumping rates from three models developed in practice by CRA were compared against all 12 computational models used in the multiple model approach. Supporting the value of expert judgment, the single “best” model developed by CRA demonstrated above 85 percent capture in all 12 computational models with a factor-of-safety of either 1.5 or 2 applied to its optimized pumping rates. The second model designed by CRA, to bracket the plausible upper range of flow through the Site, compared well to the upper range predicted by the multiple model approach, achieving above 98.5 percent capture in all cases with no factor-of-safety applied. In the context of this case study, this shows that the standard engineering approach of applying a factor-of-

safety can be a reasonable surrogate for a multiple model approach. This is of practical importance to consultants as it supports the standard approach, however it is critical (and unsurprising) that the model used is a good model.

Finally, the robustness of the pump-and-treat systems designed using the multiple model approach was examined by incrementally reducing the total system pumping rate and recalculating particle capture. The results showed considerable variation in the performance of the optimized pump-and-treat systems as the pumping rate was reduced. For example, model LR-S3-W maintained above to 90 percent capture at 20 percent of its optimized pumping rate, while several models showed closer to 50 percent capture at 20 percent of their optimized pumping rates. Additionally, the results showed that approximately 40 to 50 percent of the optimized pumping rate was required to achieve the last 10 percent of particle capture in each model. This suggests that adding additional wells in areas of poor capture may be able to improve the pump-and-treat system efficiency by reducing the total pumping rate required for containment. However, it is also possible that this might reduce the robustness of the system. Since robust designs are preferred, this test could be applied in practice to assess the robustness of pump and treat systems to change.

While a multiple model approach may capture a greater range of uncertainty than a single model, there is no definitive methodology with which to average predictions made by the ensemble of models such that a design is both robust and efficient. Prior to testing the pump-and-treat system in the field, its performance cannot be known with certainty. Since most systems can be corrected during operation, a multiple model approach is likely an excessive burden for remediation options which are adjustable and have low capital costs relative operation and maintenance costs (i.e., the cost of failure for a pump-and-treat system is low as the system can be corrected, but this may not be true for high capital

cost systems such as permeable reactive barriers). This is consistent with the phased and integrated approach outlined by the [USEPA \(1997\)](#), which involves initial characterization of the aquifer and design and implementation of a pump and treat system, likely aided by groundwater modelling, followed by continuous monitoring of the system's performance. As more data is gathered, the system is modified as necessary. The process of implementing a system, monitoring its performance, and modifying it as needed can reduce groundwater modelling costs, address system uncertainty and reduce the need for overly conservative designs. It likewise, in this case, reduces the justification for using a costly, imprecise, and relatively subjective multi-model approach for characterizing uncertainty.

APPENDICES

Appendix A

Scatter Plots of Simulated vs. Observed Heads

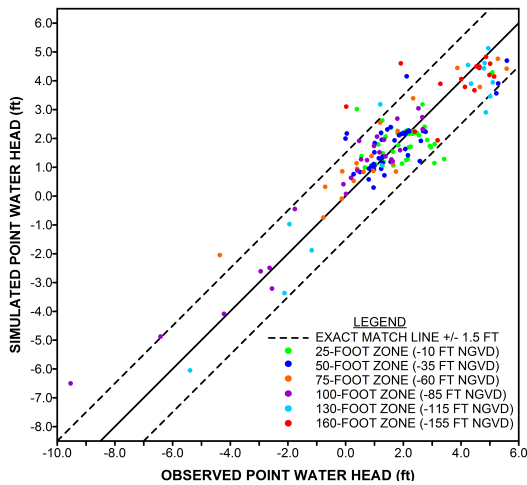


Figure A.1: Model LR-S1-G Simulated vs. Observed Head

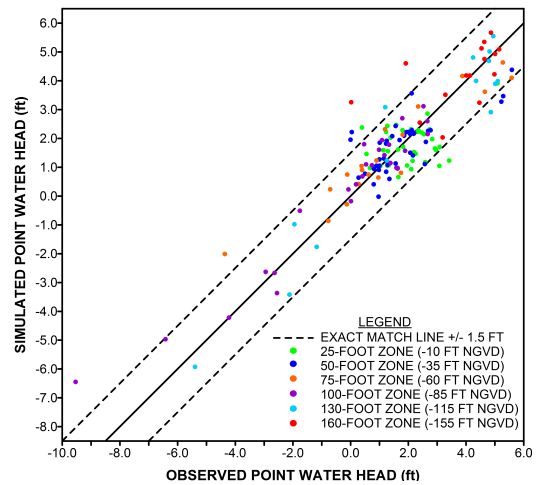


Figure A.2: Model LR-S1-W Simulated vs. Observed Head

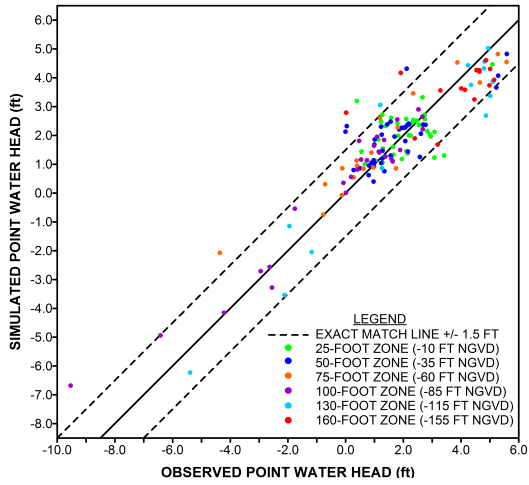


Figure A.3: Model HR-S1-G Simulated vs. Observed Head

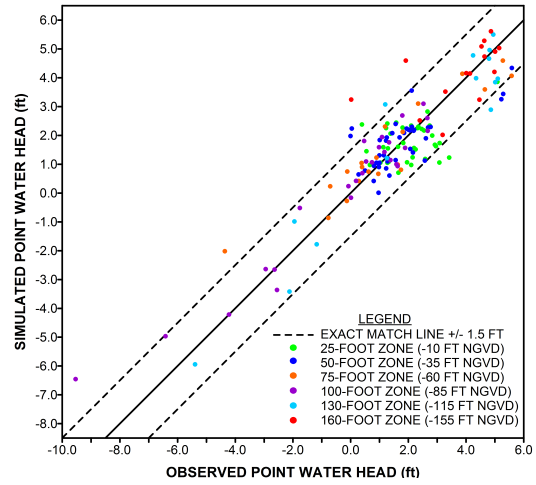


Figure A.4: Model HR-S1-W Simulated vs. Observed Head

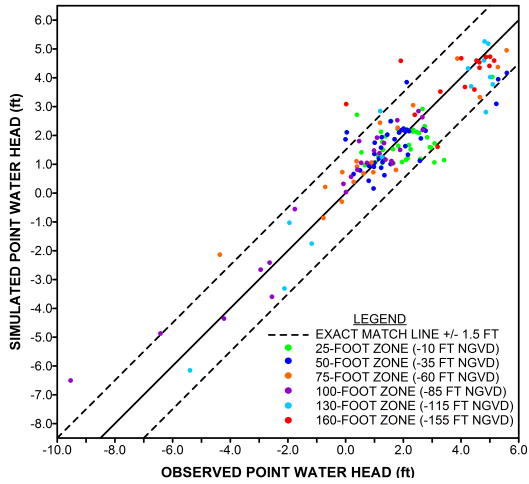


Figure A.5: Model LR-S2-G Simulated vs. Observed Head

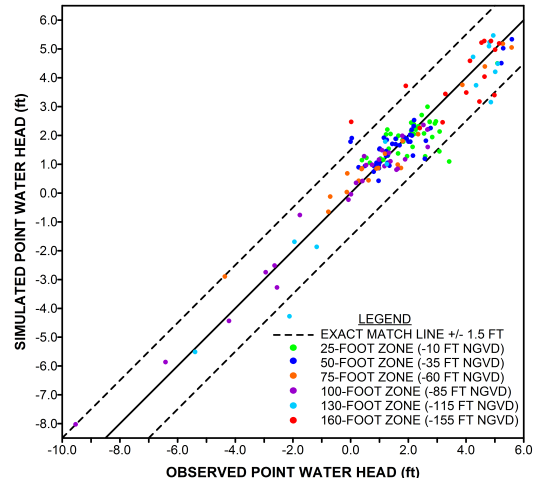


Figure A.6: Model HR-S3-W Simulated vs. Observed Head

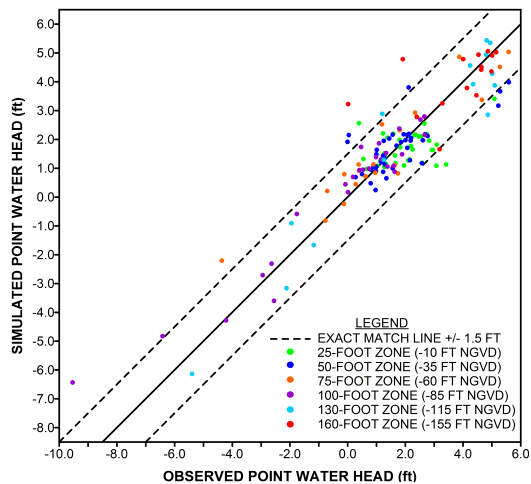


Figure A.7: Model HR-S2-G Simulated vs. Observed Head

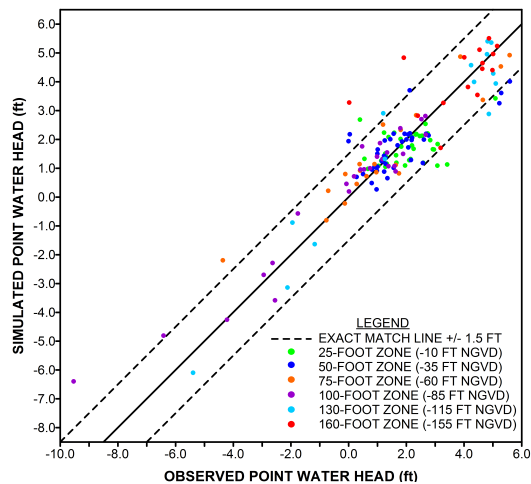


Figure A.8: Model HR-S2-W Simulated vs. Observed Head

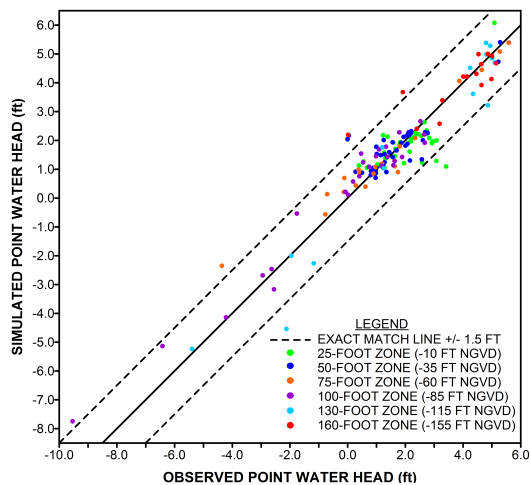


Figure A.9: Model LR-S3-G Simulated vs. Observed Head

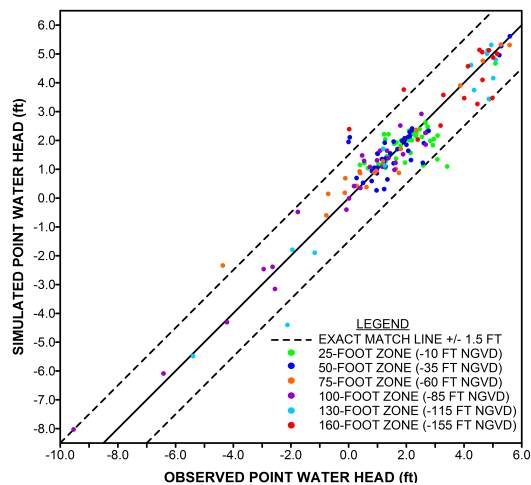


Figure A.10: Model LR-S3-W Simulated vs. Observed Head

Appendix B

Optimized Pumping Rates at Individual Wells

Table B.1: Optimized pumping rates at individual wells (ft³/d)

Well Name	LR-S1-W	LR-S2-W	LR-S2-G	LR-S1-G	HR-S3-G	LR-S3-G	LR-S3-W	HR-S1-G	HR-S2-W	HR-S2-G	HR-S1-W	HR-S3-W
Ext1	215	115	263	63	101	138	2,414	357	412	96	3	9,362
Ext2	294	56	447	655	3,768	3,858	7,677	3,757	6,756	3,434	9,102	9,321
Ext3	44	376	282	741	145	508	531	3,538	764	3,360	0	9,501
Ext4	93	474	168	973	161	3,916	1,186	3,761	2,088	1,811	863	3,546
Ext5	906	328	346	339	462	1,823	2,391	3,265	445	3,125	1,143	8,458
Ext6	333	307	335	1,352	239	3,226	1,805	3,583	3,416	94	7	1,896
Ext7	336	47	14	409	466	9	612	2,172	218	1,492	14	124
Ext8	248	393	3	156	348	1,179	494	319	1,367	153	189	1,940
Ext9	28	50	399	74	248	113	81	394	258	13	14	2,013
Ext10	264	125	374	422	202	1,333	5	913	224	389	77	4,261
Ext11	11	327	489	163	2	1,214	399	2,709	2,278	2,046	658	9,606
Ext12	1,699	709	50	196	1,141	1,300	66	3,103	1,981	2,629	136	6,748
Ext13	379	102	1,337	430	3,388	578	687	2,426	4,315	3,575	6,364	8,192
Ext14	419	2,130	411	1,367	1,089	3,496	7,488	1,299	3,815	3,876	5,888	7,872
Ext15	178	134	1,127	1,238	2,215	3,732	2,129	2,502	118	3,534	13	3,334
Ext16	70	605	53	38	2,754	2,445	1,816	567	5,811	3,962	9,203	7,987
Ext17	8	75	2,385	167	1,519	739	420	337	2,338	3,993	762	9,647
Ext18	165	639	263	131	280	612	149	32	6,262	3,616	4	1,805
Ext19	32	314	254	763	31	3,575	34	3,942	1,303	3,178	9,964	7,135
Ext20	61	377	2,552	1,034	1,108	530	3,730	2,852	2,041	3,677	9,939	8,199
Ext21	1,513	261	1,800	2,736	3,575	3,601	5,354	3,845	3,655	3,292	9,993	9,865
Total	7,295	7,942	13,351	13,447	23,242	37,926	39,470	45,673	49,865	51,347	64,336	130,812

References

- Ahlfeld, D., J. Mulvey, G. Pinder, and E. Wood (1988). Contaminated groundwater remediation design using simulation, optimization, and sensitivity theory 1. model development. *Water Resources Research* 24(3), 431–441.
- Akaike, H. (1973). *Information theory and an extension of the maximum likelihood principle*, in: *Breaking through Statistics, vol. 1, Foundation and Basic Theory*. Springer-Verlag, New York, USA, 610–624.
- Akaike, H. (1974). New look at the statistical model identification. *IEEE Transactions on Automatic Control* AC-19(6), 716–723.
- Aly, A. and R. Peralta (1999). Optimal design of aquifer cleanup systems under uncertainty using a neural network and a genetic algorithm. *Water Resources Research* 35(8), 2523–2532.
- Anderman, E., M. Hill, and E. Poeter (1996). Two-dimensional advective transport in ground-water flow parameter estimation. *Ground Water* 34(6), 1001–1009.
- Anderson, M. and W. Woessner (1992). *Applied Groundwater Modeling Simulation of Flow and Advective Transport*. Academic Press.
- Bear, J. and Y. Sun (1998). Optimization of pump-treat-inject (pti) design for the remediation of a contaminated aquifer: Multi-stage design with chance constraints. *Journal of Contaminant Hydrology* 29(3), 225–244.
- Beven, K. (1993). Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources* 16(1), 41–51.
- Beven, K. (2002). Towards an alternative blueprint for a physically based digitally simulated hydrologic response modelling system. *Hydrological Processes* 16(2), 189–206.

- Beven, K. and A. Binley (1992). The future of distributed models: model calibration and uncertainty prediction. *Hydrological Processes* 6(3), 279–298.
- Beven, K. and J. Freer (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the glue methodology. *Journal of Hydrology* 249(1-4), 11–29.
- Borden, R. and K. Troost (2001). Late Pleistocene Stratigraphy in the South-Central Puget Lowland, Pierce County, Washington. Technical report, Washington State Department of Natural Resources.
- Bredehoeft, J. (2003). From models to performance assessment: The conceptualization problem. *Ground Water* 41(5), 571–577.
- Bredehoeft, J. (2005). The conceptualization model problem - surprise. *Hydrogeology Journal* 13(1), 37–46.
- Burnham, K.P. and Anderson, D. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods and Research* 33(2), 261–304.
- Carrera, J., A. Alcolea, A. Medina, J. Hidalgo, and L. Slooten (2005). Inverse problem in hydrogeology. *Hydrogeology Journal* 13(1), 206–222.
- Carrera, J. and S. Neuman (1986). Estimation of aquifer parameters under transient and steady state conditions: I. maximum likelihood method incorporating prior information. *Water Resources Research* 22(2), 199–210.
- Cavanaugh, J. and A. Neath (1999). Generalizing the derivation of the schwarz information criterion. *Communications in Statistics - Theory and Methods* 28(1), 49–66.
- Compernelle, T., S. Van Passel, and L. Lebbe (2013). The value of groundwater modeling to support a pump and treat design. *Groundwater Monitoring and Remediation* 33(3), 111–118.
- Cooley, R. (1985). A comparison of several methods of solving nonlinear regression ground-water flow problems. *Water Resources Research* 21(10), 1525–1538.
- Cooley, R. and R. Naff (1992). Regression modeling of ground-water flow. Technical report, U.S. Geologic Survey Techniques in Water Resources Investigations. Chapter B4, Book 3, 232 pp.

- Dagan, G. and S. Neuman (1997). *Subsurface Flow and Transport: A Stochastic Approach*. Cabridge University Press, Cambridge, United Kingdom.
- Doherty, J. (2003). Ground water model calibration using pilot points and regularization. *Ground Water* 41(2), 170–177.
- Doherty, J. (2004). *PEST Model-Independent Parameter Estimation User Manual: 5th Edition* (5th ed.). Watermark Numerical Computing.
- Duan, Q., S. Sorooshian, and V. Gupta (1992). Effective and efficient global optimization for conceptual rainfall- runoff models. *Water Resources Research* 28(4), 1015–1031.
- Engelhardt, I., J. G. De Aguinaga, H. Mikat, C. Schüth, O. Lenz, and R. Liedl (2012). Complexity versus simplicity: an example of groundwater model ranking with the akaike information criterion. *Hydrology and Earth System Sciences Discussions* 9(8), 9687–9714.
- Environmental Simulations, I. E. (2011). Guide to using groundwater vistas. *Environmental Simulations. Inc.*, 336.
- Foglia, L., S. Mehl, M. Hill, and P. Burlando (2013). Evaluating model structure adequacy: The case of the maggia valley groundwater system, southern switzerland. *Water Resources Research* 49(1), 260–282.
- François, C. (2006). Freshwater management in canada: IV. groundwater. *Parliamentary Information and Research Service, Library of the Parliament*.
- Freeze, R. and J. Cherry (1979). *Groundwater*. Prentice Hall, Inc.
- Frind, E. (2003). *Fundamentals of Groundwater Modelling*. Department of Earth Sciences University of Waterloo.
- Glasgow, H., M. Fortney, J. Lee, A. Graettinger, and H. Reeves (2003). MODFLOW 2000 head uncertainty, a first-order second moment method. *Ground Water* 41(3), 342–350.
- Guo, W. and G. Bennett (1998). Simulation of saline/fresh water flows. *Proceddings of MODFLOW '98, Colorado School of Mines, Golden, Colorado* 1, 267–274.
- Guo, W. and C. Langevin (2002). Users guide to SEAWAT: A computer program for simulation of three-dimensional variable-density ground-water flow. *Techniques of Water-Resources Investiations of the United States Geological Survey*.

- Harbaugh, A. W. (2005). MODFLOW-2005, the U.S. Geological Survey modular ground-water model the ground-water flow process. Technical report, U.S. Geological Survey.
- Harbaugh, A. W., E. Banta, M. Hill, and M. McDonald (2000). MODFLOW2000, the us geological survey modular ground-water model-user guide to modularization concepts and the ground-water flow process. Technical Report Open-File Rep 00-92, US Geological Survey.
- Hart Crowser, H. C. (1975). Geology of the port of tacoma. Technical report. Prepared in cooperation with the Port of Tacoma, Washington. Hart Crowser, Seattle, Washington.
- Harvey, C. and S. Gorelick (1995). Mapping hydraulic conductivity: sequential conditioning with measurements of solute arrival time, hydraulic head, and local conductivity. *Water Resources Research* 31(7), 1615–1626.
- Hill, M. (1992). A computer program (modflowp) for estimating parameters of a transient, three-dimensional, groundwater flow model using nonlinear regression. Technical report, U.S. Geological Survey Open-File Report 91-484. 358 pp.
- Hill, M. (2003). Preconditioned conjugate-gradient 2 (PCG2) a computer program for solving ground-water flow equations. Technical report, U.S. Geological Survey.
- Hill, M., E. Banta, A. W. Harbaugh, and E. Anderman (2000). MODFLOW-2000, the u.s. geological survey modular ground-water model user guide to the observation, sensitivity, and parameter-estimation processes and three post-processing programs. Technical report, U.S. Geological Survey.
- Hill, M. and C. Tiedeman (2007). *Effective Groundwater Model Calibration*. John Wiley & Sons, Inc, Hoboken, New Jersey.
- Hoeting, J., D. Madigan, A. Raftery, and C. Volinsky (1999). Bayesian model averaging: A tutorial. *Statistical Science* 14(4), 382–417.
- Højberg, A. and J. Refsgaard (2005). *Model uncertainty - Parameter uncertainty versus conceptual models*, Volume 52 of *Water Science and Technology*.
- Hunt, R., J. Doherty, and M. Tonkin (2007). Are models too simple? arguments for increased parameterization. *Ground Water* 45(3), 254–262.
- Hurvich, C. and C. Tsai (1989). Regression and time series model selection in small samples. *Biometrika* 76(2), 297–307.

- Isukapalli, S., A. Roy, and P. Georgopoulos (1998). Stochastic response surface methods (srsms) for uncertainty propagation: Application to environmental and biological systems. *Risk Analysis* 18(3), 351–363.
- James, A. and C. Oldenburg (1997). Linear and monte carlo uncertainty analysis for subsurface contaminant transport simulation. *Water Resources Research* 33(11), 2495–2508.
- Johannessen, J. and A. MacLennan (2007). Beaches and bluffs of puget sound and the northern straits. Technical report, Coastal Geologic Services, Inc. Technical Report 2007-04, 27 p.
- Jones, M. (1999). Geologic framework for the puget sound aquifer system, washington and british columbia. Technical report, U.S. Geologic Survey.
- Kashyap, R. (1982). Optimal choice of ar and ma parts in augoregressive moving average models. *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-4*(2), 99–104.
- Ko, N.-Y., K.-K. Lee, and Y. Hyun (2005). Optimal groundwater remediation design of a pump and treat system considering clean-up time. *Geosciences Journal* 9(1), 23–31.
- Kunstmann, H., W. Kinzelbach, and T. Siegfried (2002). Conditional first-order second-moment method and its application to the quantification of uncertainty in groundwater modeling. *Water Resources Research* 38(4), 61–615.
- Langevin, C., D. Thorne, A. Jr., Dausman, M. Sukop, and W. Guo (2007). Seawat version4: A computer program for simulation of multi-species solute and heat transport: U.s. geological survey techniques and methods book 6, chapter a22, 39 p. Technical report, U.S. Geologic Survey.
- Lincoln, J. (2000). The puget sound model summary. Retrieved from: <http://exhibits.pacsci.org>.
- Lu, D., M. Ye, P. Meyer, G. Curtis, X. d. Shi, X.-F. Niu, and S. Yabusaki (2013). Effects of error covariance structure on estimation of model averaging weights and predictive performance. *Water Resources Research* 49.
- Luszczynski, N. (1961). Head and flow of ground water of variable density. *Journal of Geophysical Research* v. 66, no. 12, 4247–4256.

- Marquardt, D. (1963). An algorithm for least squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics* 11(2), 431–441.
- Matott, L. (2006). Ostrich: an optimization software tool; documentation and user’s guide, version 1.8. Buffalo, NY: University of Buffalo, Department of Civil, Structural and Environmental Engineering.
- Matott, L., A. Rabideau, and J. Craig (2006). Pump-and-treat optimization using analytic element method flow models. *Advances in Water Resources* 29(5), 760–775.
- Mayer, A., C. Kelley, and C. Miller (2002). Optimal design for problems involving flow and transport phenomena in saturated subsurface systems. *Advances in Water Resources* 25(8-12), 1233–1256.
- McDonald, M. and A. Harbaugh (1988). A modular three-dimensional finite-difference ground-water flow model. *Book 6: Modeling techniques, Section A. Ground-water*, 78.
- McDonald, M. and A. W. Harbaugh (1984). A modular three-dimensional finite-different groundwater flow model. Technical report, U.S. Geologic Survey.
- Meyer, P. and G. Gee (1999). Information on hydrologic conceptual models, parameters, uncertainty analysis, and data sources for dose assessments at decommissioning sites. *NUREG/CR-6656, U.S. Nuclear Regulatory Commission, Washington, D.C.*
- Meyer, P., M. Ye, M. Rockhold, S. Neuman, and K. Cantrell (2007). Combined estimation of hydrogeologic conceptual model parameter and scenario uncertainty with application to uranium transport at the hanford site 300 area.
- Morgan, D. and J. Jones (1996). Numerical model analysis of the effects of ground-water withdrawals on discharge to streams and springs in small basins typical of the puget sound lowland, washington. Technical report, U.S. Geological Survey. Open-File Report 95-470.
- National Research Council (2013). Alternatives for managing the nation’s complex contaminated groundwater sites. *The National Academies Press*.
- Nettasana, T. (2012). *Conceptual Model Uncertainty in the Management of the Chi River Basin, Thailand*. Ph. D. thesis, University of Waterloo.
- Neuman, S. (1973). Calibration of distributed paramter groundwater flow models viewed as a multiple-objective decision process under uncertinainty. *Water Resources Research* 9(4), 1006–1021.

- Neuman, S. (2002). Accounting for conceptual model uncertainty via maximum likelihood model averaging. In *Proceedings of 4th International Conference of Calibration and Reliability in Groundwater Modelling (ModelCARE 2002)*. edited by K. Kovar and Z. Hrkal, pp. 529–534, Charles Univ., Prague, Czech Republic.
- Neuman, S. (2003). Maximum likelihood bayesian averaging of uncertain model predictions. *Stochastic Environmental Research and Risk Assessment* 17(5), 291–305.
- Neuman, S. and P. Wierenga (2003). A comprehensive strategy of hydrogeologic modeling and uncertainty analysis for nuclear facilities and sites. Technical report, U.S. Nucl. Regul. Comm.
- Peoter, E. and M. Hill (1998). Documentation of ucode, a computer code for universal inverse modeling. Technical report, U.S. Geological Survey Water-Resources Investigations Report 98-4080.
- Peoter, E., M. Hill, E. Banta, S. Mehl, and S. Christensen (2005). Ucode 2005 and six other computer codes for universal sensitivity analysis, calibration, and uncertainty evaluation. Technical report, U.S. Geological Survey Techniques and Methods 6-A11, 283 pp.
- Poeter, E. and D. Anderson (2005). Multimodel ranking and inference in ground water modeling. *Ground Water* 43(4), 597–605.
- Poeter, E. and M. Hill (1997). Inverse models: A necessary next step in ground-water modeling. *Ground Water* 35(2), 250–260.
- Poeter, E. and S. McKenna (1995). Reducing uncertainty associated with ground-water flow and transport predictions. *Ground Water* 33(5), 899–904.
- Pollock, D. (1988). Semianalytical computation of path lines for finite-difference models. *Ground Water* 26(6), 743–750.
- Pollock, D. (1994, September). Users guide for MODPATH/MODPATH-PLOT, version 3: A particle tracking post-processing package for MODFLOW, the U. S. Geological Survey finite-difference ground-water flow model. Open-file report 94-464, U.S. Geologic Survey.
- Refsgaard, J., J. van der Sluijs, J. Brown, and P. van der Keur (2006). A framework for dealing with uncertainty due to model structure error. *Advances in Water Resources* 29(11), 1586–1597.

- Robertson GeoConsultants Inc. and SRK Consulting Inc. (2012). Guidelines for ground-water modelling to assess impacts of proposed natural resource development activities. Technical report, British Columbia Ministry of Environment.
- Robinson & Noble, Inc (1992). Hydrogeologic analysis of the federal way area, Washington. Technical report, Robinson & Noble, Inc.
- Rojas, R., L. Feyen, and A. Dassargues (2008). Conceptual model uncertainty in ground-water modeling: Combining generalized likelihood uncertainty estimation and Bayesian model averaging. *Water Resources Research* 44(12).
- Rojas, R., S. Kahunde, L. Peeters, O. Batelaan, L. Feyen, and A. Dassargues (2010). Application of a multimodel approach to account for conceptual model and scenario uncertainties in groundwater modelling. *Journal of Hydrology* 394(3-4), 416–435.
- Rushton, K. and S. Redshaw (1979). *Seepage and groundwater flow: numerical analysis by analog and digital methods*.
- Russell, K. and A. Rabideau (2000). Decision analysis for pump-and-treat design. *Ground Water Monitoring & Remediation* 20(3), 159–168.
- Savoca, M., W. Welch, K. Johnson, R. Lane, B. Clothier, and E. Fassler (2010). Hydrogeologic framework, groundwater movement, and water budget in the Chambers-Clover Creek watershed and vicinity, Pierce County, Washington. Technical report.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics* 6(2), pp. 461–464.
- Seifert, D., T. Sonnenborg, J. Refsgaard, A. Højberg, and L. Troldborg (2012). Assessment of hydrological model predictive ability given multiple conceptual geological models. *Water Resources Research* 48.
- Seifert, D., T. Sonnenborg, P. Scharling, and K. Hinsby (2008). Use of alternative conceptual models to assess the impact of a buried valley on groundwater vulnerability. *Hydrogeology Journal* 16(4), 659–674.
- Serfes, M. E. (1991). Determining the mean hydraulic gradient of ground water affected by tidal fluctuations. *Ground Water* 29(4), 549–555.
- Sharief, S. b., T. Eldho, A. Rastogi, and V. Gurunadha Rao (2012). Optimal groundwater remediation by pump and treat using fem- and ega-based simulation-optimization model. *Journal of Hazardous, Toxic, and Radioactive Waste* 16(2), 106–117.

- Singh, A., S. Mishra, and G. Ruskauff (2010). Model averaging techniques for quantifying conceptual model uncertainty. *Ground Water* 48(5), 701–715.
- Sun, N. (1994). *Inverse Problems in Groundwater Modeling*. Springer.
- Sun, N.-Z. and W.-G. Yeh (1990). Coupled inverse problems in groundwater modeling 1. sensitivity analysis and parameter identification. *Water Resources Research* 26(10), 2507–2525.
- Tolson, B. c. and C. d. Shoemaker (2007). Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resources Research* 43(1).
- USEPA (1997). *Design Guidelines for Conventional Pump-and-Treat Systems*. U.S. EPA.
- USEPA (2005). *Cost-Effective Design of Pump and Treat Systems*. U.S. EPA.
- USEPA (2008). A systematic approach for evaluation of capture zones at pump and treat systems. Technical report.
- USEPA (2010). Superfund remedy report. 13th. ed. washington, dc: Environmental protection agency office of solid waste and emergency response. Technical report.
- USGS (2013, November). US topo and historical topographic map collection. Retrieved From: <http://geonames.usgs.gov/pls/topomaps/>.
- Vecchia, A. and R. Cooley (1987). Simultaneous confidence and prediction intervals for nonlinear regression models with application to a groundwater flow model. *Water Resources Research* 23(7), 1237–1250.
- Wang, M. and C. Zheng (1997). Optimal remediation policy selection under general conditions. *Ground Water* 35(5), 757–764.
- Xiang, Y., J. Sykes, and N. Thomson (1992). A composite l1 parameter estimator for model fitting in groundwater flow and solute transport simulation. *Water Resources Research* 29(6), 1661–1673.
- Yang, Y., J. Wu, X. Sun, J. Wu, and C. Zheng (2013). A niched pareto tabu search for multi-objective optimal design of groundwater remediation systems. *Journal of Hydrology* 490, 56–73.
- Ye, M., P. Meyer, and S. Neuman (2008). On model selection criteria in multimodel analysis. *Water Resources Research* 44(3).

- Ye, M., S. Neuman, and P. Meyer (2004). Maximum likelihood bayesian averaging of spatial variability models in unsaturated fractured tuff. *Water Resources Research* 40(5), W051131–W0511317.
- Ye, M., K. Pohlmann, J. Chapman, G. Pohll, and D. Reeves (2010). A model-averaging method for assessing groundwater conceptual model uncertainty. *Ground Water* 48(5), 716–728.
- Yeh, W.-G. (1986). Review of parameter identification procedures in groundwater hydrology: the inverse problem. *Water Resources Research* 22(2), 95–108.
- Zhang, D. (2001). *Stochastic Methods for Flow in Porous Media*. Academic press.
- Zheng, C. (1996). MT3D-A modular three-dimensional transport model for simulation of advection, dispersion and chemical reactions of contaminants in groundwater systems. *S.S. Papadopoulos & Associates, Inc., Bethesda, Md.*
- Zheng, C. (2006). MT3DMS v5.2 supplemental user’s guide. Technical report, Technical report to the U.S. Army Engineer Research and Development Centre, Department of Geolocial Sciences, University of Alabama, 24p.
- Zheng, C. and P. Wang (1999). MT3DMS: A modular three-dimensional multispecies transport model for simulation of advection, dispersion, and chemical reactions of contaminants in groundwater systems; documentation and users guide. Technical report, Contract Report SERDP-99-1, U.S. Army Engineer Research and Development Center, Vicksburg, MS.