2011

Adaptive Kalman Filtering Scheme For The Simulation Of Benzene In Subsurface Environment

Linkel Kwabena Boateng
North Carolina Agricultural and Technical State University

ADAPTIVE KALMAN FILTERING SCHEME FOR THE SIMULATION OF BENZENE IN SUBSURFACE ENVIRONMENT

by

Linkel Kwabena Boateng

A thesis submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of
MASTER OF SCIENCE

Department: Civil, Architectural and Environmental Engineering
Major: Civil Engineering
Major Professor: Dr. Shoou-Yuh Chang

North Carolina A&T State University
Greensboro, North Carolina
2011

School of Graduate Studies
North Carolina Agricultural and Technical State University

This is to certify that the Master's Thesis of

Linkel Kwabena Boateng

has met the thesis requirements of
North Carolina Agricultural and Technical State University

Greensboro, North Carolina
2011

Approved by:

Dr. Shoou-Yuh Chang Major Professor	Dr. Stephanie Luster-Teasley Committee Member
Dr. Manoj K. Jha Committee Member	Dr. Sameer Hamoush Department Chairperson

Dr. Sanjiv Sarin
Dean of Graduate Studies

DEDICATION

This work is dedicated to my uncle, Dr. Robert A. Baffour, whose support and continual inspiration has created a stirring atmosphere for my academic excellence.

BIOGRAPHICAL SKETCH

Linkel Kwabena Boateng was born on January, 12 1982 in New Edubiase, Ghana. He graduated with a Bachelor of Science degree in Chemical Engineering from Kwame Nkrumah University of Science and Technology, Ghana in 2007. In 2009, he enrolled in the master's program in Civil and Environmental Engineering at North Carolina Agricultural and Technical State University. His research under the supervision of Dr. Shoou-Yuh Chang focuses on the use of stochastic data assimilation schemes in improving predictions of subsurface contaminant transport models.

Linkel has also worked on various aspects of Solid Waste Technology and Management and currently has a certificate in Waste Management from the Interdisciplinary Waste Management Institute at North Carolina A&T State University. Mr. Boateng has received various recognitions for academic excellence including the W.L. Kennedy Scholars' Award for 4.0 GPA and a certificate of outstanding academic performance from the International Students' and Scholars Office. In 2009, he became a member of the American Academy of Environmental Engineers and subsequently received the Six Sigma Green Belt Certification in 2010. As a member of the Phi Kappa Phi Honors Society, Linkel is committed to pursuing higher education and hopes to attain a Ph.D. someday. He is a candidate for the Master of Science degree in Civil Engineering.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude and appreciation to my advisor, Dr. Shoou-Yuh Chang, for his insightful suggestions and professional guidance throughout the course of this study. I would also like to thank other members of my thesis committee, Dr. Stephanie Luster-Teasley and Dr. Manoj K. Jha for their review and invaluable comments on the thesis. Furthermore, I would like to extend special thanks to my family, friends and colleague graduate students for their support and encouragement which has made this endeavor a possibility.

This work was sponsored by the Department of Energy's Samuel Massie Chair of Excellence Program under grant number DF-FG01-94EW11425. The views and conclusions expressed herein are exclusively those of the author and do not necessarily represent the official policies or endorsement of the funding agency.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURES **PAGE**

LIST OF SYMBOLS

AKF	Adaptive Kalman Filter
FTCS	Forward Time Central Space
ICS-GC	Innovation Covariance Scaling and Gain Correction
KF	Kalman Filter
$\mathbf{A}$	State Transition Matrix (STM)
b	Aquifer thickness
C	Concentration of benzene in solute phase
C_0	Initial benzene concentration
$\mathbf{C}_t$	Predicted innovation covariance
$\overline{\mathbf{C}}_t$	Estimated true innovation covariance
$\mathbf{d}_t$	Innovation sequence at time step t
D_x, D_y	Dispersion coefficients in x and y directions respectively
$E\{\cdot\}$	Expectation function
$\mathbf{H}$	Observation data pattern matrix
$\mathbf{I}$	Identity matrix
$\mathbf{K}_{t+1}$	Optimal Kalman gain matrix at time step t+1
$\overline{\mathbf{K}}_{t+1}$	Corrected optimal Kalman gain matrix at time step t+1
k	First-order benzene decay rate
M	Moving average window size
M_0	Initial benzene mass input

m	Dimension of observation vector
n	Number of nodes in model domain
$\mathbf{o}_t$	Observation noise vector at time t
$\mathbf{p}_t$	Model system error or process noise vector at time step t
$\mathbf{P}_{t+1}$	Optimal estimate error covariance matrix at time step t+1
$\overline{\mathbf{P}}_{t+1}$	True optimal estimate error covariance matrix at time step t+1
$\mathbf{Q}_t$	Model system error covariance matrix at time step t
$\mathbf{R}_t$	Observation error covariance matrix at time step t
R	Retardation factor
RMSE (t)	Root-Mean-Squared-Error at time step t
Δt	Time step
V	Linear pore liquid velocity
x,y	Cartesian coordinates
$\mathbf{x}_{t+1}$	Vector of benzene concentration at time step t+1
$\mathbf{x}_t^T$	True optimal benzene concentration at time step t
Δx, Δy	Space interval along x and y directions respectively
$\mathbf{z}_t$	Observations at time step t
α_t	Adaptive factor at time step t
λ_t	Forgetting factor at time step t
δ_{tj}	Dirac delta function
η	Porosity of porous medium
Γ	Boundary of studied area

ABSTRACT

Boateng, Linkel Kwabena. ADAPTIVE KALMAN FILTERING SCHEME FOR THE SIMULATION OF BENZENE IN SUBSURFACE ENVIRONMENT. **(Advisor: Shoou-Yuh Chang)**, North Carolina Agricultural and Technical State University.

Environmental legislation in several states has become more stringent on the clean up procedures for benzene and other toxic chemicals since the enactment of the Comprehensive Environmental Response, Compensation, and Liability Act (Superfund). In order to comply with the Superfund requirements for hazardous pollutants, accurate information about the nature of contaminants is required to carry out risk assessment and effective site remediation. The use of subsurface contaminant transport models, coupled with stochastic data assimilation schemes, can provide accurate predictions of contaminant transport to enhance the reliability of risk assessment in the area of environmental remediation.

In this study, a two-dimensional deterministic model was used to simulate the advective and diffusive transport of benzene in the subsurface. A robust Adaptive Kalman Filter (AKF) has been constructed as a stochastic data assimilation scheme to improve the prediction of the benzene contaminant plume. The AKF has been proposed to improve the performance of the conventional Kalman Filter (KF) by reducing the uncertainties associated with the process and observation noise statistics. The impact of the adaptive filter on the KF performance was examined by comparing model predictions with a simulated true field which was created by introducing some random Gaussian noise into an observation model. The simulation results indicated an improvement in

filter performance after the implementation of the adaptive Kalman filter scheme. Although the Kalman filter was successful in reducing the prediction error of the deterministic model from 5.0 mg/L to 1.1 mg/L at the end of the simulation period, the introduction of the AKF scheme further improved the prediction accuracy of the KF by about 18%. In all, the AKF scheme successfully improved the prediction accuracy of the deterministic model by about 82%. Furthermore, the results of sensitivity test suggest that for the AKF under consideration, using a window size of five can give a much improved accuracy and stability.

Keywords: Benzene; subsurface environment; deterministic model; stochastic process; random Gaussian noise; data assimilation; Adaptive Kalman Filter (AKF).

CHAPTER 1

Introduction

1.1 Background Information

Benzene and other petroleum-derived contaminants in the subsurface pose considerable risk to groundwater quality due to their toxic and carcinogenic nature. Many petroleum-contaminated sites are typically polluted by: benzene, toluene, ethylbenzene and xylene (BTEX) which constitute the major hazardous components in gasoline. Benzene has been classified as the most hazardous component of gasoline and is therefore of most concern to regulators and water purveyors (Dean 1985). The principal sources of these petroleum-derived contaminants, commonly referred to as Non-aqueous Phase Liquids (NAPLs) include accidental spills of petroleum products, leachate from landfills and leaking underground storage tanks (Gary and Rixey 1999; Tam and Beyer 2002). According to the United States Environmental Protection Agency (USEPA), about 15-20% of the over 1.8 million underground storage tanks containing petroleum products and other hazardous chemicals are either leaking or are expected to leak in the future (Chen et al. 1996). Nationally, over 488,000 confirmed releases of petroleum-based fuels from leaking underground storage tanks have been recorded (USEPA 2009).

These tanks release gasoline and other chemicals into the subsurface layer which migrate downwards through the vadose zone to encounter groundwater aquifers. At the groundwater table, the light NAPLs spread laterally along the capillary fringe while the dense NAPLs may displace water and continue its migration under pressure and gravity

forces (Qin et. al 2009). The soluble components of NAPLs dissolve to form a solute plume which is transported over long distances by the movement of groundwater (Mackay et al. 1985).The transport of benzene and other NAPLs in a porous media is generally controlled by advection, hydrodynamic dispersion and sorption. In the subsurface, the migration process tends to exhibit a retardation effect due to sorption onto the surfaces of solid particles which results in the reduction of contaminant concentration. Generally, the effect of retardation on the transport process greatly depends on the properties of the NAPL and the soil structure; for unsaturated soils, the retardation coefficients are usually higher than those of saturated soils (Maraqa et al. 1999).

Natural attenuation of aqueous benzene serves as an essential remediation process which plays an important role in minimizing the spread of groundwater plumes. A number of factors including dispersion, sorption, decay, volatilization and the oxygen level in a contaminated aquifer can significantly affect the attenuation rate of benzene. Among these mechanisms, intrinsic biodegradation has been identified as the major process that accounts for the reduction of benzene concentration in the subsurface environment (Kao et al. 2010). Moreover, the use of fuel oxygenates such as *tert*-butyl alcohol (TBA) and ethanol can notably affect the transport and attenuation rate of benzene and the other BTEX compounds in the subsurface. Various studies have shown that using ethanol as a fuel oxygenate can inhibit benzene biodegradation rate which tends to increase the plume life span and the risk of groundwater contamination (Fu and Luthy 1986; Herman and Powers 1996; Corseuil et al. 1998; Da Silva and Alvarez 2002; Deeb et al. 2002).

Over the last decades, several cases of benzene contamination in groundwater have been detected at various contaminated sites. In the United States alone, there are about 100,000 different sites for which benzene contamination have been reported (Chang and Latif 2007). At certain Superfund restricted sites, alarming benzene concentrations have been detected in groundwater which poses considerable risk to potable water resources. The effect of long term exposure to benzene has been linked to cancer of the lungs, kidney and liver. In some cases, respiratory effects have also been reported in humans after acute exposure to benzene vapors (Avis and Hutton 1993). As a result, the USEPA has set the Maximum Contaminant Level (MCL) of benzene at 5µg/L to protect consumers of public water systems from the effects of long term chronic exposure to benzene (USEPA 2010). Groundwater accounts for an estimated 20% of the total water usage in the United States and approximately 53% of the population rely on groundwater as a source of drinking water (NGWA and USGS 2003). Therefore, the migration of benzene and other NAPLs in groundwater poses significant threat to public health and the environment.

Although the transport of NAPLs may vary from site to site, the basic principles governing the transport process is the same; hence, understanding contaminant migration process is key for the implementation of effective remediation strategies. Several studies have been conducted to investigate the fate and transport of benzene in the subsurface environment (Smith and Jaffe 1994; Deeb et al. 2002; Choi et al. 2005; Zhang et al. 2005; Batlle-Aguilar et al. 2009). Most of these studies employ mathematical models to simulate the processes that underlie the transport of benzene in subsurface environment.

In recent time, due to the complex and heterogeneous nature of the subsurface environment, simulation models have been combined with optimization techniques to address the complexities and the variability in the transport process. This combined system can help identify the best management and remediation strategy for petroleum-contaminated sites (Qin et al. 2009).

The implementation of remediation systems for contaminated sites requires the use of suitable models which can accurately predict the nature of the contaminant plumes. Subsurface contaminant transport models can enhance the effectiveness of site remediation processes by providing accurate information about the shapes and sizes of these plumes. However, predictions by traditional numerical models usually result in unavoidable deviation from the true field due to inherent uncertainties in model parameters, boundary and initial conditions, numerical schemes as well as inaccurate representation of physical processes (Johnson et al. 2006). In order to minimize this deviation, stochastic data assimilation schemes have been introduced in contaminant transport modeling to guide the prediction process and to account for the variability in the subsurface flow. Such a scheme has the capacity to integrate observation information into deterministic models to provide a more accurate representation of the transport process.

The Kalman Filter (KF) data assimilation tool has been widely applied in the areas of subsurface contaminant transport modeling and water resources (Van Geer 1982; Zou and Par 1995; McLaughlin 2002; Chang and Jin 2005). For the last three decades, the application of the KF approach has received considerable attention in areas of energy and environmental concern. The KF scheme can be used in conjunction with traditional

models to predict contaminant migration in cases where the construction of observation wells are very expensive or infeasible. Unlike traditional numerical models, the KF adopts a probabilistic approach to project the state of a dynamic system with minimum deviation. In its implementation, the KF utilizes a process model and observation information obtained from field measurements or a given criteria. The KF operation is a Bayesian recursive process consisting of a series of prediction and update equations.

1.2 Problem Statement

Although the KF scheme can significantly improve the prediction accuracy of traditional numerical models, its effectives is usually limited by the uncertainties in the process and observation noise structures. Actually, the effectiveness of the KF depends significantly on how accurately these noise structures are defined. However, the specific values of these errors cannot be determined and have to be initially estimated before filter operation. This assumption however may not be accurate and could result in inaccurate definition of the process and observation noise statistics which may depend on the process dynamics and hydro-geological factors. In general, the innovation of the KF is affected by unaccounted errors such as inaccurate model dynamics and unknown sources which can eventually lead to increased predicted error covariance. These errors may also affect the optimal Kalman gain and the overall effectiveness of the KF scheme.

Therefore, an Adaptive Kalman Filter (AKF) scheme is constructed by adjusting the system error covariance and the optimal Kalman gain in order to reduce the effect of errors in the process and observation noise statistics during filter operation.

1.3 Objectives

In this study, an Adaptive Kalman Filter (AKF) is constructed as a stochastic data assimilation scheme to improve the prediction accuracy of a hydrodynamic transport model for benzene. The adaptive filtering scheme has been introduced to optimize the performance of the conventional KF by reducing the effect of errors in the process and observation noise covariances. Furthermore, the AKF scheme is expected to improve the stability and convergence performance of the filter. The adaptive filtering technique is based on the Innovation Covariance Scaling and Gain Correction (ICS-GC) approach which manipulates the system covariance and the optimal Kalman gain to reduce the effect of errors in the process and observation noise statistics (Fathabadi et al. 2009).

Specifically, the objectives of this study are:

- To improve the predictions of benzene contaminant plume by the use of the Kalman Filter (KF) data assimilation scheme.
- To examine the impact of an adaptive filter on the performance of the KF by constructing an Adaptive Kalman Filter (AKF) scheme.
- To examine the stability and convergence performance of the AKF scheme.

CHAPTER 2

Literature Review

2.1 Contaminants in the Subsurface

Point and nonpoint source pollution have resulted in severe groundwater contamination through the release of large quantities of toxic and hazardous chemicals into the environment. Point source pollution may be associated with a point location such as a toxic-waste spill site, underground storage tank, municipal and hazardous waste dumpsite and septic tanks. Nonpoint source pollution on the other hand results from agricultural activities such as the application of fertilizers and pesticides, urban and industrial runoff, channelization and mining activities. In addition to these forms of pollution, various natural phenomena such as volcano eruption have also contributed to the release of toxic and radioactive chemicals into the environment. Contaminated sites such as Brownfield's are typically polluted by polycyclic aromatic hydrocarbons (PAHs), Volatile Organic Contaminants (VOC), BTEX and heavy metals such as lead, arsenic and cadmium which constitutes the major contaminants in groundwater (Fischer et al. 2007).

In addition to BTEX, methyl *tert*-butyl ether (MTBE) is one of the most severe groundwater contaminants found in gasoline hydrocarbon plumes. MTBE has been used as a fuel oxygenate for over two decades to increase the octane rating of gasoline and to fulfill oxygen content requirements of the Clean Air Act. However, recent legislation has called for the removal of MTBE from gasoline due to reported cases of groundwater contamination by MTBE in several states (Deeb et al. 2002). A significant portion of

contaminated soil and groundwater sites contain NAPLs such as dry cleaning fluids, aviation fuel and gasoline which result mostly from human activities. Most NAPLs such as the chlorinated hydrocarbons can penetrate deep below the water table and thus become difficult to identify when investigating sites for contamination. As a result, NAPL contamination can affect aquifers for several years and this makes it difficult to restore a NAPL-contaminated aquifer to drinking-water quality.

Landfills receiving municipal and industrial waste have become one of the major causes of groundwater pollution and soil contamination. Approximately 150 million tons of municipal solid waste and 240 million tons of industrial solid waste are deposited in landfills annually (Smith and Jaffe 1994). Industrial waste deposited in landfills may contain significant amount of hazardous chemicals such as acids, trichloroethylene (TCE), pesticides and radioactive agents which can contaminate the underlying groundwater. In recent times, growing concern about public health and degradation of potable water resources from existing unregulated landfills have resulted in the enforcement of appropriate regulatory measures at such landfills. Although the enforcement of control measures has regulated the operations of these dumpsites, several cases of groundwater contamination have been reported as a result of leachate migration from various landfills ((Mor et al. 2006; Singh et al. 2009).

The transport of dissolved contaminants in the subsurface is generally controlled by advection of contaminants and molecular diffusion, also known as hydrodynamic dispersion. From a macroscopic viewpoint, Darcy's law provides an accurate description of the flow of ground water in hydrogeologic environments. It expresses the flux of water

through a porous medium due to the advection of contaminants. The hydraulic head decreases during fluid flow and the direction of flow moves from a higher to a lower hydraulic gradient (Freeze and Cherry 1979). Similarly, the transport of pollutants due to diffusion in a porous medium can explicitly be expressed by Fick's first law.

In order to design cost-effective remediation systems for contaminated sites, a detailed understanding of the flow and transport of contaminants is required. Over the past few decades, various efforts have been made to understand the fate and transport of pollutants in the subsurface. Mathematical models have been recognized as the most effective tool for explaining how a plume of contaminant evolves. Subsurface contaminant transport models can describe the spatial and temporal distribution of contaminant as well as their movement to potential receptor points (Tam and Beyer 2002). However, a number of factors including the difficulty in determining the field coefficient of dispersion and velocity of water as well as the challenges involved in the quantification of chemical reaction terms have limited the routine use of such models. Regional size models which neglect the effects of dispersion have also had limited success due to the scarcity and poor quality of field data.

Furthermore, subsurface contaminant transport models suffer from prediction errors due to inherent uncertainties in model parameters, space and time limits of numerical schemes and inaccurate representation of physical processes which describe the transport phenomena. The use of subsurface transport models, coupled with stochastic data assimilation schemes such as the Kalman and Particle filtering can help improve model predictions by integrating observation information into the dynamic system.

2.2 Benzene Contamination in Groundwater

Subsurface contamination by benzene and other petroleum hydrocarbons is a pervasive environmental problem which can affect aquifers for several years and render groundwater unsafe for drinking purposes. Groundwater pollution by petroleum hydrocarbons has been reported extensively in several states including Louisiana, North Carolina, Texas and Delaware. Most petroleum contaminated sites may contain significant quantities of benzene, toluene, ethylbenzene, xylene (BTEX) and polycyclic aromatic hydrocarbons. Among these, the BTEX compounds are of major concern due to their toxic nature. Gasoline contains approximately 10-20% of BTEX which persists in a pure liquid phase within pore spaces at many gasoline spill sites (Kao et al. 2010). When NAPLs become entrapped in the subsurface, the soluble components of residual BTEX dissolve at a slower rate to form a contaminated plume of groundwater which can be transported over long distances from the spill zone.

Benzene contamination in groundwater results from both natural and anthropogenic activities although the human-induced sources are undoubtedly the most severe. Several cases of benzene soil and groundwater contamination have been reported over the past decades. In the United States, over 100,000 sites have been identified as benzene contaminated zones. At certain sites, reported benzene concentrations have exceeded state and federal levels for groundwater quality, posing significant threats to human health and jeopardizing future opportunities for economic development. According to the USEPA, about 2 million pounds of benzene were released into the environment between 1987 and 1992. In the state of Texas, approximately, 1,436 pounds

of benzene were released into the environment through petroleum refining and related activities between 1987 and 1992 (USEPA, 2009). In 2005, an explosion of a petrochemical plant in China polluted the water supply of over 9 million people in the city of Harbin (Fu et al. 2008). More recently, an estimated 206 million gallons of crude oil was released into the Gulf of Mexico during a lethal explosion at the deepwater horizon oil platform operated by British Petroleum (Kerr et. al 2010). This resulted in the destruction of several aquatic lives along with the creation of dead zones which perpetrate aquatic mortality. The development of remediation systems for these contaminated sites, coupled with the reimbursement of affected individuals and businesses have run into several billions of dollars over the years.

Remediation of NAPL-contaminated aquifers remains a challenging problem in the United States due to the difficulty involved in identifying certain NAPLs during risk assessment and site investigations. As a result, NAPL contamination may persist for several years thereby increasing the risk of groundwater contamination. The Comprehensive Environmental Response, Compensation, and Liability Act of 1990 was enacted to regulate clean-up procedures of hazardous waste disposal sites. The law was designed to clean-up abandoned and uncontrolled hazardous waste sites around the country through the support of an established Superfund Trust Fund (Reisch 2003). A number of remediation strategies have been developed over the years to clean up NAPL-contaminated aquifers and other Superfund sites. However, enhanced attenuation processes such as pump-and-treat and In-situ Chemical Oxidation (ISCO) may require extensive risk assessment and planning. The urgency of site remediation has motivated

the identification of efficient and cost-effective techniques for implementing remedial actions. Stochastic data assimilation schemes can give a much improved predictions of the contaminant transport process. The use of such schemes over the years has provided accurate information about contaminant transport to enhance the reliability of risk assessment and emergency response procedures.

2.3 Modeling Fate and Transport of Petroleum Hydrocarbons

Several models have been developed to simulate the fate and transport of petroleum hydrocarbons in the environment. A number of studies have adopted a more advanced reactive transport modeling that incorporates geochemical processes to provide a comprehensive understanding of the contaminant dispersion at contaminated sites. Some of the most recent developments have been summarized below.

Delshad et al. (1996) developed a three-dimensional multi-component multiphase compositional finite difference simulator for analyzing surfactant-enhanced remediation of NAPL-contaminated aquifers. A phase behavior model was constructed to account for observed micellar and micro-emulsion behavior pertinent to SEAR. The other surfactant related properties including adsorption, interfacial tension, capillary pressure, capillary number and micro-emulsion viscosity were modeled based on the phase behavior model.

A one-dimensional multi-component model for the biodegradation of BTEX in groundwater was presented by Prommer et al. (1998). The model was solved numerically by an operator-splitting method to couple the advective-dispersive transport with a geochemical equilibrium package PHREEQC and a biodegradation module. The total

aqueous concentrations for each component were determined by solving the transport equations. An application to two illustrative contamination scenarios demonstrated the model's ability to handle the complex interactions in groundwater.

The effect of ethanol on benzene plume length was investigated by Deeb et al. (2002) by the use of a microbial and modeling approach. Fuel oxygenates such as *tert-butyl* alcohol (TBA) and ethanol has been recommended to reduce emissions and to comply with the Clean Air Act Amendments (1990) requirements. However, the results of this study have indicated that using ethanol as a fuel oxygenate can inhibit BTEX biodegradation in the subsurface environment and increase its plume length.

In another study, Kim et al. (2006) investigated the fate of benzene in subsurface environment by focusing on the role of sorption and biodegradation in the attenuation of aqueous benzene under dynamic flow conditions. The results indicated that aqueous benzene could be attenuated by irreversible sorption and biodegradation during transport through a porous medium. Furthermore, it was concluded that, the attenuation of benzene in the subsurface was closely related to the organic carbon content and the oxygen level in contaminated aquifers.

A three dimensional numerical model was developed by Guo and Wang (2008) to simulate the spilled oil in coastal areas. The model was based on a hybrid particle tracking and Eulerian-Lagrangian approach which simulated the processes of advection and diffusion while other weathering processes like evaporation, dissolution and emulsification were simultaneously considered. The model predicted the horizontal movement of surface oil slick, the vertical distribution of oil particles and the mass

balance of spilled oil. In their model validation, they simulated an accidental oil release using the proposed approach; the simulation results indicated a reasonable level of accuracy after comparison with field observations.

Fu et al. (2008) simulated the concentration of benzene and nitrobenzene in the Songhua River using a one dimensional advection-dispersion-decay equation which incorporated a first order volatilization mechanism. An explicit finite-difference approach was adopted in generating the numerical results. The simulation was carried out at various monitoring stations along the river and the results were compared with the observations from the monitoring stations. Simulated benzene and nitrobenzene concentrations closely matched the observed values.

The use of purely deterministic models in studying the transport of NAPLs usually results in imprecision and unavoidable deviation from the true field. In order to minimize the errors associated with deterministic models, stochastic and Bayesian probabilistic based models must be introduced to account for the variability of subsurface transport processes.

2.4 Environmental Modeling of Stochastic Processes

A simulation model can be considered stochastic if it incorporates random components into the modeling process (Law and Kelton 2000). Such a scheme includes an input that introduces random variables into deterministic dynamics to simulate complicated real-world transport systems with uncertain sources. The random noise scheme accounts for the heterogeneity that exists in natural subsurface flow systems as a

result of inherent randomness of hydro-geological factors (Jin 2009). Stochastic data assimilation models dealing with variability have received widespread application in atmospheric dispersion modeling, weather foresting and subsurface contaminant transport modeling (Zou and Parr 1995; Li and Graham 1999; Chang and Jin 2005). Over the last decades, such models have provided a scheme to assimilate observations into traditional numerical models to obtain a more accurate representation of dynamic systems.

The Kalman filter is a statistical filtering data assimilation scheme based on the theory of Bayesian inference. The concept of Bayesian inference provides a mechanism for integrating observation information into a prior probability distribution for a state in order to obtain an improved posterior probability distribution. In general, the Bayesian filtering technique involves the prediction and update stages; the prediction stage estimates the state probability density function (pdf) based on the system model and the update process modifies the predicted pdf using the latest observation information.

In the operation of the KF, the contaminant state distribution and its corresponding noise fields are assumed to be Gaussian. This premise is based on the fact that the Gaussian distribution is widely used in statistics and also most statistical tests are based on the assumption of normality. The KF data assimilation scheme is particularly effective when both system and observation models are linear (Welch and Bishop, 1995). In cases where the models are non-linear, the implementation of other extensions of the KF (Extended KF, Ensemble KF, and Unscented KF) is recommended. Numerous applications of the KF and its extensions have been demonstrated in various areas of environmental process modeling (Van Gerr 1982; Zou and Parr 1995; Franssen et al.

2008; Jin 2009; Chang and Latif 2010).

Zou and Par (1995) demonstrated the effectiveness of the Kalman filtering data assimilation scheme in predicting optimal concentrations for two-dimensional plumes in groundwater. The state-space technique was employed to obtain a distribution of contaminant concentrations using the process modeling data and measurement data. The results suggested that, the KF scheme produces significant improvement in the prediction of plume concentration distributions.

Li and Graham (1999) combined numerical schemes with Fourier transform algorithm to solve stochastic transport model in heterogeneous aquifers. In their study, stochastic scheme of groundwater flow and transport equations were constructed for subsurface system variables such as hydraulic properties and concentration which were treated as a random field.

Chang and Jin (2005) investigated the performance of the KF data assimilation scheme with a regional noise structure. The KF scheme was constructed based on a two-dimensional advection-dispersion hydrodynamic model for a radioactive contaminant. The effectiveness of the KF scheme was examined based on a standard error parameter. The results indicated that, by applying the correct regional noise structure, the KF data assimilation scheme can reduce the prediction error from 25 to 10 ppm, indicating a 60% improvement in prediction accuracy.

Chang and Li (2006) implemented a particle filter (PF) data assimilation scheme to simulate the fate and transport of chlorobenzene leachate in a landfill. The discrete advection-dispersion transport equation was used as a deterministic model. The model

was solved using the Forward Time Central Space (FTCS) numerical scheme. A random Gaussian error was introduced into the deterministic model to simulate the observations and the particle filter recursive process was then applied in the state space model. At each time step, random samples representing the density of state vector were propagated through the state equations and updated by observation data assimilation. The results indicated that the PF can give a more accurate estimate of the contaminant concentration profile than the pure deterministic model.

Nerger et al. (2007) compared the performance of the Ensemble Kalman Filtering (EnKF) and the Singular Evolution Interpolated Kalman (SEIK) Filter within a parallel data assimilation framework in the Finite Element Ocean Model (FEOM) of the North Atlantic. Several data assimilation experiments were implemented to assess the effectiveness of the two schemes. The filter algorithms were applied with a model configuration of FEOM to assimilate the sea surface height in twin experiments. The results indicated that, SEIK can provide more accurate estimates that the EnKF scheme.

In another study, Franssen et al. (2008) employed the Ensemble Kalman filter (EnKF) to assimilate the hydraulic head data from 90 locations in a groundwater flow model for a two-year period. A deterministic model based on a three dimensional finite element groundwater flow model was developed. A statistical model for the uncertain hydraulic conductivities was constructed on the basis of a geostatistical analysis of the data, geostatistical up scaling techniques and results from the inverse modeling. The results suggested that the EnKF gives a considerable improvement in terms of absolute hydraulic head errors, as compared with the deterministically calibrated model.

More recently, He et al. (2011) presented a generalized uncertainty estimation method (GUEM) to study the fate and transport of petroleum-hydrocarbons in a subsurface environment. GUEM employs the Metropolis-Hastings sampling algorithm to replace the conventional Monte Carlo algorithm which has the advantage of reducing the computational cost for obtaining realizations of uncertain parameters. The scheme was applied to both a hypothetical homogenous and a real-world heterogonous site for demonstration studies. Simulation results indicated that benzene concentrations still exceeded the regulated limits after three years of natural attenuation and remediation actions. Therefore, groundwater would not be suitable for drinking purposes after this period unless enhanced remediation actions are taken to improve groundwater quality.

The effectiveness of conventional numerical models is usually limited by uncertainties in model parameters, initialization data and observation of system behavior. Although, the implementation of the KF data assimilation scheme can help address this discrepancy, its performance may also suffer from errors that are associated with the process and observation noise statistics. The uncertainty of the process and observation noise statistics has a significant impact on the KF performance (Grewal and Weil 2001; Weidong et al.2006). In practice, the values of the noise statistics are usually fixed during the KF operation since a precise evaluation of the statistical structures is not feasible. This assumption however, may result in deviation of prediction results since the statistical structures may depend on process dynamics and hydro-geological factors.

A number of adaptive filtering techniques have been developed to improve the performance of the conventional KF. The covariance scaling approach improves filter

stability and convergence performance by adjusting the state covariance matrix using a scalar factor. The scalar factor can be estimated based on an empirical approach or a criteria derived from filter innovations (Hu et al. 2003; Yang and Xu 2003; Yang and Gao 2006). In the multi-model adaptive estimation, an ensemble of Kalman filters with different stochastic models is simultaneously operated to produce an output which is the weighted sum of each individual filter's output. The weighting factor for this approach can be estimated using the residual probability function (Hide et al. 2004b).

Although adaptive filtering has had limited application in contaminant transport modeling, it has received widespread application in Geographical Positioning Systems (GPS) integration, vehicle navigation, signal processing and process control. Hu et al. (2003) demonstrated the effectives of an adaptive Kalman filter in vehicle navigation. Hajiyev (2007) successfully implemented an adaptive Kalman filter with gain correction for an integrated navigation system, consisting of inertial and radar altimeters. A new covariance-based adaptive process noise scaling method was developed and tested by Weidong et al. (2006). The proposed adaptive mechanism was effective in tuning the process noise covariance to its optimal magnitude, resulting in a significant improvement in filtering performance. Recently, Fathabadi et al. (2009) presented an adaptive Kalman filter based on covariance scaling and gain correction for state estimation in plants with different communication delays in their sensors. The effectiveness of the proposed scheme was demonstrated through simulation studies performed on a continuous stirred tank reactor (CSTR) benchmark problem. The adaptive filter indicated better performance than Kalman filtering based on each single sensor's information.

Due to the difficulty involved in determining the exact statistical structures of the process and observation noise, the effectiveness of the conventional KF is usually limited in most dynamic applications. The effect of unaccounted errors resulting from inaccurate noise statistics can significantly degrade the KF performance. In this study, an Innovation Covariance Scaling and Gain Correction (ICS-GC) adaptive mechanisms are simultaneously implemented in order to reduce the effect of errors in both system and observation noise statistics. The ICS-GC scheme is constructed to improve the prediction accuracy of a subsurface transport model for benzene.

CHAPTER 3

Methodology

3.1 Deterministic Model Formulation

The conceptual model for the transport of benzene in a subsurface layer can be described by the traditional advection-dispersion equation. In a saturated porous media, the migration process is primarily controlled by molecular diffusion in response to chemical concentration gradient and advection of contaminants due to hydraulic gradients. In this model, sorption is considered as an important process in the migration of benzene, which is classified as a non-conservative contaminant with a first-order decay mechanism. Therefore the advection-dispersion equation for a two-dimensional transport in the x-y plane can be expressed as

$$\frac{\partial C}{\partial t} = \frac{D_x}{R}\left(\frac{\partial^2 C}{\partial x^2}\right) + \frac{D_y}{R}\left(\frac{\partial^2 C}{\partial y^2}\right) - \frac{V}{R}\left(\frac{\partial C}{\partial x}\right) - \frac{k}{R}C \tag{3.1}$$

where C is the concentration of benzene in the solute phase (mg/L); D_x and D_y are the dispersion coefficients in x and y directions, respectively (m^2/d); x, y are the Cartesian coordinates (m); V is linear velocity in the x-direction (m/d); R is the retardation factor (dimensionless), k is the first-order benzene decay rate (1/day) and t is time (day).

In the model implementation, a constant boundary condition was assumed and the initial conditions were specified based on an instantaneous point source at (x_0, y_0) as:

$$C(x,y,t)\big|_{t=0} = C_0(x_0, y_0) \tag{3.2a}$$

$$C(x,y,t)_{t=\Gamma} = C_I \tag{3.2b}$$

where (x_0, y_0) is the initial pulse input coordinate and C_I is constant concentration at the aquifer boundary, Γ which is chosen as a square boundary in this case. The aquifer is assumed to be square in shape for demonstration purposes.

In order to express the deterministic model in a more convenient form for the implementation of the data assimilation schemes, an explicit finite-difference scheme, Forward Time Central Space (FTCS) is used to develop the state-space form of the above partial differential equation. The deterministic model is discretized by substituting the appropriate finite difference approximations for the Partial Differential Equation (PDE) in equation (3.1). The resulting iteration scheme after discretizing the concentration in the model domain is expressed as

$$\begin{aligned} C(i, j, t+1) \approx\ & b_1 C(i-1, j, t) + b_2 C(i, j, t) + b_3 C(i+1, j, t) \\ & + b_4 C(i, j-1, t) + b_5 C(i, j+1, t) \end{aligned} \tag{3.3}$$

where the coefficients b_1, b_2, b_3,b_4,b_5 are functions of the hydrodynamic flow in the porous media:

$$b_1 = \frac{D_x \Delta t}{R \Delta x^2} + \frac{V \Delta t}{2R \Delta x} \tag{3.4a}$$

$$b_2 = 1 - \frac{2D_x \Delta t}{R \Delta x^2} - \frac{2D_y \Delta t}{R \Delta y^2} \tag{3.4b}$$

$$b_3 = \frac{D_x \Delta t}{R \Delta x^2} - \frac{V \Delta t}{2R \Delta x} \tag{3.4c}$$

$$b_4 = b_5 = \frac{D_y \Delta t}{R \Delta y^2} \tag{3.4d}$$

In order to avoid divergence of the numerical solution, the convergence and stability criterion for the FTCS scheme is checked by satisfying the following conditions:

$$\Delta x < 2D_x / V \tag{3.5a}$$

$$\Delta t < (2\Delta y D_x / (\Delta x^2 R)) \tag{3.5b}$$

where Δx and Δt are the spatial and temporal step sizes, respectively. Expressing the above numerical scheme in the general state-space form gives

$$\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} \tag{3.6}$$

where $\mathbf{x}_t$ and $\mathbf{x}_{t-1}$ are the subsurface state variables at time t and t-1 respectively, **A** is the State Transition Matrix (STM) which incorporates the model parameters and advances the current state to the next time step.

A purely deterministic model is totally controlled by pre-specified algorithms which do not involve any random conditions (Law and Kelton 2000). The use of such models in environmental prediction may lead to imprecision and unavoidable deviation from the true field. Stochastic data assimilation schemes can simulate real-world transport processes with uncertain sources and inaccurate transport parameters by introducing a random noise term in the deterministic dynamics (Saad 2007). Such a scheme accounts for the heterogeneity of the subsurface by incorporating randomness into the system to provide an accurate representation of the transport process.

Considering a linear stochastic process

$$\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{p}_t,\ t = 0,1,2,3... \tag{3.7}$$

where $\mathbf{p}_t$ is the model system error which is the difference between the optimal estimate of the true state and the model prediction. The observation, $\mathbf{z}_t$ in this study is defined as

$$\mathbf{z}_t = \mathbf{H}\mathbf{x}_t^T + \mathbf{o}_t \tag{3.8}$$

where **H** is the measurement sensitivity matrix which reflects the observations in the field. **H** is constructed as an $n \times n$ identity matrix with n being the number of nodes in the model domain which is 400 in this case; $\mathbf{x}_t^T$ is the true optimal estimate of the state. The model error, $\mathbf{p}_t$ and the observation error, $\mathbf{o}_t$ are assumed to have zero means and covariance matrix $\mathbf{Q}_t$ and $\mathbf{R}_t$ respectively. In this study, $\mathbf{p}_t$ and $\mathbf{o}_t$ are drawn from a normal distribution based on the assumption of normality which holds for most linear dynamic systems. Furthermore, $\mathbf{p}_t$ and $\mathbf{o}_t$ are uncorrelated random Gaussian noise distributions with the following statistical structures:

$$E\{\mathbf{p_t}\} = E\{\mathbf{o_t}\} = 0 \tag{3.9a}$$

$$E\left\{\mathbf{p_t p_j^T}\right\} = \mathbf{Q_t}\delta_{tj} \tag{3.9b}$$

$$E\left\{\mathbf{o_t o_j^T}\right\} = \mathbf{R_t}\delta_{tj} \tag{3.9c}$$

$$\delta_{tj} = 0 \ \ if \ \ t \neq j \tag{3.9d}$$

$$\delta_{tj} = 1 \ if \ \ t = j \tag{3.9e}$$

where $\{\cdot\}$ denotes expectation function, $\mathbf{Q}_t$ and $\mathbf{R}_t$ are symmetric and positive semi-definite matrices. In order to avoid the formation of badly scaled matrix, the system covariance error may be constructed in a unique way to avoid the issue of singularity.

3.2 Implementation of Conventional Kalman Filter (KF)

The KF is a very effective data assimilation tool in subsurface contaminant transport due to its ability to deal with dynamic and stochastic processes. It falls in the class of estimation problems where the process and observation models are approximated

by linear Gaussian state space models (Welch and Bishop 1995). The process equation simulates the dynamics of the transport phenomenon whereas the observation model relates the observation information to the state variables (Zou and Par 1995). The conventional KF utilizes these two models to estimate the state of a dynamic system through a predictor-corrector recursive process which minimizes the estimated error covariance. The prediction process estimates the current state and the error covariance to obtain the before-adjustment estimates for the next step while the correction process serves as a feedback medium which introduces new observation data to estimate the new state (Welch and Bishop 1995). The KF as a data assimilation scheme combines observation data and model dynamics to give an improved prediction of the state at discrete time and spatial points. As a stochastic approach, it incorporates random conditions into the filtering process to simulate real world situations.

The KF algorithm is a recursive Bayesian process involving the following prediction and update stages (Kalman 1960):

Prediction Stage:

$$\mathbf{x}_t(-) = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{p}_t \tag{3.10}$$

$$\mathbf{P}_t(-) = \mathbf{A}\mathbf{P}_{t-1}\mathbf{A}^T + \mathbf{Q}_t \tag{3.11}$$

Update Stage:

$$\mathbf{K}_t = \mathbf{P}_t(-)\mathbf{H}^T(\mathbf{H}\mathbf{P}_t(-)\mathbf{H}^T + \mathbf{R}_t)^{-1} \tag{3.12}$$

$$\mathbf{x}_t(+) = \mathbf{x}_t(-) + \mathbf{K}_t[\mathbf{z}_t - \mathbf{H}\mathbf{x}_t(-)] \tag{3.13}$$

$$\mathbf{P}_t(+) = (I - \mathbf{K}_t\mathbf{H})\mathbf{P}_t(-) \tag{3.14}$$

where, $\mathbf{x}_t(+)$ is the estimated state after the KF adjustment with covariance $\mathbf{P}_t(+)$; $\mathbf{x}_t(-)$ is the estimated state before the Kalman filter adjustment with covariance $\mathbf{P}_t(-)$; $\mathbf{K}_t$ is the optimal Kalman gain matrix which minimizes the optimal estimate error covariance matrix, $\mathbf{P}_t$; $()^T$ and $()^{-1}$ denote matrix transpose and inverse respectively and $\mathbf{I}$ is an identity matrix. It must be noted that $\mathbf{P}_{t-1}(-)$ and $\mathbf{x}_{t-1}(-)$ are initially specified to start the KF data assimilation process.

3.3 Data Assimilation with Adaptive Kalman Filter (AKF)

One of the key limitations of the conventional KF is how to assign suitable statistical properties to both the process and observation models (Weidong et al. 2006). The KF assumes complete knowledge of the process and observation noise covariance matrices, $\mathbf{Q}_t$ and $\mathbf{R}_t$. However, in practice these noise statistics may not be precisely known and have to be initially estimated. The conventional way of determining $\mathbf{Q}_t$ and $\mathbf{R}_t$ requires a priori knowledge of the noise statistics which can be achieved through intensive empirical analysis (Weidong et al. 2006). The process and observation noise covariances influence the weight assignment that the filter applies between the existing process information and the latest observation. Therefore inaccurate estimation of these noise statistics may have a significant effect on the optimality of the KF which can result in filter divergence in severe cases (Hide et al. 2004). Furthermore, subsurface contaminant transport is highly heterogeneous and randomized due to uncertainties in the hydraulic parameters and initial conditions, hence assigning constant noise statistics for such applications may not be a plausible assumption in many practical situations.

Therefore, it is expected that, the implementation of an adaptive filter will help to reduce the effect of unaccounted errors in the process and observation noise covariances. The adaptive filtering technique which is based on the Innovation Covariance Scaling and Gain Correction (ICS-GC) addresses the effect of unaccounted errors in the process and observation models. The covariance scaling scheme compensates for the unaccounted errors in the process model by increasing the magnitude of the system error covariance using an adaptive factor (Kim and Lee 2006).The KF assigns more weight to the estimate with the least uncertainty; hence increasing the variance of the predicted state will results in assigning more weight to the observation data. The gain correction process decreases the magnitude of the optimal Kalman gain in order to reduce the weight assigned to the observations thereby reducing the effect of unaccounted errors in the observation model (Fathabadi et al. 2009). In the implementation of the ICS-GC scheme, a scalar factor greater than one is required to manipulate the error covariance and the optimal Kalman gain in order to create an effective adaptive system.

For the KF described above, the innovation sequence, $\mathbf{d}_t$, is defined as

$$\mathbf{d}_t = \mathbf{z}_t - \mathbf{H}\mathbf{x}_t(-) \tag{3.15}$$

The above innovation represents the predicted residuals which is the difference between the observation and the predicted state at a given time. For any linear dynamic system, the predicted covariance of the above innovation sequence, $\mathbf{C}_t$, can be expressed as

$$\mathbf{C}_t = E[\mathbf{d}_t \mathbf{d}_t^T] = \mathbf{H}\mathbf{P}_t(-)\mathbf{H}^T + \mathbf{R}_t \tag{3.16}$$

The KF estimates the above innovation covariance based on inaccurate initial process and observation noise covariance. In order to ensure optimal filter performance, a “true

innovation covariance", $\overline{\mathbf{C}}_t$, is approximated as the average of the predicted covariance within a fixed window, M (Fathabadi et al. 2009).

$$\overline{\mathbf{C}}_t = \frac{1}{\mathrm{M}-1}\sum_{\mathrm{i}=t-M+1}^{\mathrm{t}} \mathbf{d}_\mathrm{i}\mathbf{d}_\mathrm{i}^\mathrm{T} \tag{3.17}$$

where, M is the moving window size. The moving average technique was used in estimating the "true innovation covariance". Using a window size of five, the "true innovation covariance" for each time step was approximated as the average of the previous four innovation covariances. In this case, the adaptive filtering process starts at the fifth time step of the simulation. The predicted covariance and the "true covariance" of the innovation sequence can be related as

$$\overline{\mathbf{C}}_\mathrm{t} = \alpha_\mathrm{t}\mathbf{C}_t \tag{3.18}$$

$$\alpha_t = \max\left\{1, \frac{1}{m} tr\,(\overline{\mathbf{C}}_t\mathbf{C}_t^{-1})\right\} \tag{3.19}$$

where, α_t is the adaptive factor and $\alpha_t > 1$; *m* is the dimension of the observation vector and tr (·) denotes the trace of a matrix which is the sum of the elements in the main diagonal. The estimated adaptive factor is used to increase the magnitude of the error covariance, $\mathbf{P}_t(-)$. In this case, the AKF assigns less weight to the system model thereby reducing the effect of inaccuracies in the initial process noise covariance. The error covariance estimation equation for the AKF can then be expressed as (Hu et al. 2003; Fathabadi et al. 2009).

$$\overline{\mathbf{P}}_\mathrm{t}(-) = \alpha_\mathrm{t}[\mathbf{A}\mathbf{P}_{\mathrm{t}\text{-}1}\mathbf{A}^\mathrm{T} + \mathbf{Q}_\mathrm{t}] \tag{3.20}$$

Thus, equation (3.20) is used to estimate the covariance error during the AKF operation.

As filtering proceeds, the adaptive scheme adjusts the initially specified $\mathbf{P}_{t-1}(-)$ to a "true system covariance" by the use of the adaptive factor. It should be noted however that for an optimal filter, the predicted innovation covariance and the estimated innovation covariance are consistently equal at each time step in which case the adaptive factor becomes stabilized at unity. In practice, $\alpha_t \leq 1$ indicates the filtering is in a steady state process. However, when $\alpha_t > 1$, the filter may be in an unstable state or in a state of divergence which serves as a threshold for the implementation of the AKF scheme (Hu et al. 2003; Weidong et al. 2006).

For the second case where the observation model is assumed to have unaccounted errors, a scalar variable is estimated to adjust the observation covariance matrix, $\mathbf{R}_t$. This is done to indirectly decrease the magnitude of the optimal Kalman gain so as to reduce the weight assigned to the observation data. In this case, "the true innovation covariance" is estimated as (Hajiyev 2007)

$$\overline{\mathbf{C}}_t = \mathbf{H}\overline{\mathbf{P}}_t(-)\mathbf{H}^T + \lambda_t \mathbf{R}_t \tag{3.21}$$

where λ_t is called a forgetting factor and $\lambda_t > 1$. For an optimum AKF, the predicted innovation covariance and the estimated innovation covariance must be consistently equal at each time step;

$$E[\mathbf{d}_t \mathbf{d}_t^T] = \mathbf{H}\overline{\mathbf{P}}_t(-)\mathbf{H}^T + \lambda_t \mathbf{R}_t \tag{3.22}$$

Taking the trace of both sides and re-arranging equation (3.22) gives the expression for the forgetting factor

$$\lambda_t = \frac{tr(\mathbf{d}_t \mathbf{d}_t^T) - tr(\mathbf{H}\overline{\mathbf{P}}_t(-)\mathbf{H}^T)}{tr(\mathbf{R}_t)} \tag{3.23}$$

The corrected optimal Kalman gain expression for the AKF is given as (Hajiyev 2007)

$$\overline{\mathbf{K}}_t = \overline{\mathbf{P}}_t(-)\mathbf{H}^T(\mathbf{H}\overline{\mathbf{P}}_t(-)\mathbf{H}^T + \lambda_t\mathbf{R}_t)^{-1} \tag{3.24}$$

A discrete-time AKF based on the ICS-GC scheme can be described by the following coupled equations:

$$\mathbf{x}_t(-) = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{p}_t \tag{3.25}$$

$$\overline{\mathbf{P}}_t(-) = \alpha_t[\mathbf{A}\mathbf{P}_{t-1}\mathbf{A}^T + \mathbf{Q}_t] \tag{3.26}$$

$$\overline{\mathbf{K}}_t = \overline{\mathbf{P}}_t(-)\mathbf{H}^T(\mathbf{H}\overline{\mathbf{P}}_t(-)\mathbf{H}^T + \lambda_t\mathbf{R}_t)^{-1} \tag{3.27}$$

$$\mathbf{x}_t(+) = \mathbf{x}_t(-) + \overline{\mathbf{K}}_t[\mathbf{z}_t - \mathbf{H}\mathbf{x}_t(-)] \tag{3.28}$$

$$\overline{\mathbf{P}}_t(+) = (\mathbf{I} - \overline{\mathbf{K}}_t\mathbf{H})\overline{\mathbf{P}}_t(-) \tag{3.29}$$

where α_t and λ_t are as defined in equations (3.19) and (3.23) respectively. The above equations represent the adaptive system which has been created by the introduction of a true system covariance and a corrected optimal Kalman gain. Figure 3.1 shows the schematic representation of the ICS-GC adaptive filter scheme.

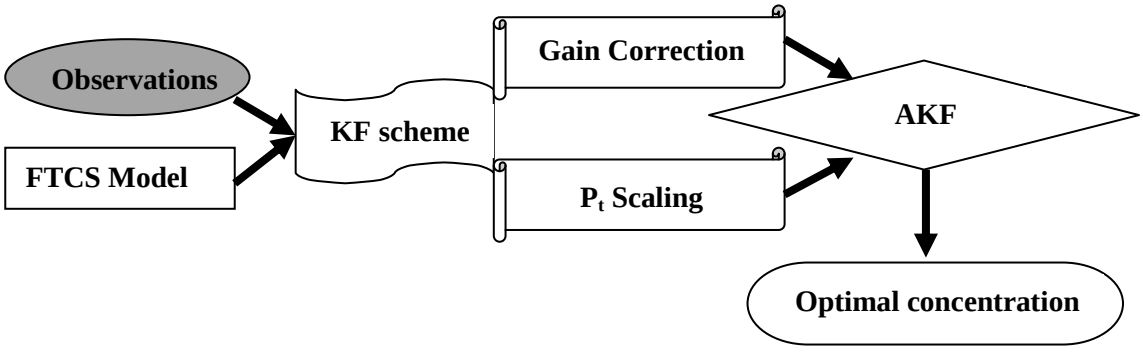

Figure 3.1. Schematic representation of the Innovation Covariance and Gain Correction (ICS-GC) Adaptive Scheme

3.4 "True Field" Simulation

In order to implement the data assimilation schemes, observation information is required to guide the system model to project the optimal benzene concentration. Various studies have simulated synthetic observation data by creating "true field" conditions (McLaughlin 2002; Goegebeur and Pauwel 2007; Chang and Latif 2010). In this study the observation data, which represents the true state of the benzene concentration, was simulated based on a random "true field". A random noise scheme was created by introducing an observation error of 2.5% into the "true field" which was obtained from the analytical solution of the advection-dispersion equation. In generating this true field, the linear velocity was further reduced by 15% since it's considered inaccurate in the deterministic model. The analytical solution governing the advection-dispersion partial differential equation is given below

$$C(x,y,t) = \frac{M_0}{4\pi A \eta t \sqrt{D_x D_y}} e^{\left(-\frac{(x-Vt/R)^2}{4D_x t/R} - \frac{y^2}{4D_y t/R} - kt\right)} \tag{3.30}$$

where M_0 is the instantaneous mass input (g), η is the porosity of the medium and b is the aquifer thickness (m) (Schwartz and Zhang 1994); all other notations have the same meaning as previously defined.

3.5 Experimental System

A two-dimensional plane with 400 grid points was defined as the testing environment for the numerical scheme. The model grid had n_x grid points in the x direction and n_y grid points in the y direction. The spatial step sizes in the x and y

directions were estimated to be $\Delta x = \Delta y = 1.5$m based on the stability and convergence criterion in equation (3.5a). The time step was also estimated to be $\Delta t = 0.2$ day using equation (3.5b). The simulation was carried out for a period of 10 days using a basic time unit of days. The linear velocity was given as V= 0.85 m/d and the retardation factor was estimated as R=1.20 using a benzene partition coefficient of 0.095 L/kg and soil bulk density of 1.99 kg/cm^3 (Gomez et al. 2002). The dispersion coefficient in the x and y directions were estimated as $D_x = 0.75$ m^2/d and $D_y = 0.25$ m^2/d respectively, using a molecular diffusion coefficient of 0.95 cm^2/day, a tortuosity of 0.6 and a dispersivity of 0.9 (Gomez et al. 2002; Chang and Latif 2010). The first-order decay rate for benzene was specified as 0.015/day (Batlle-Aguilar et al. 2009). In the analytical model, the aquifer thickness and porosity were set at 2.5 m and 0.3 (Gomez et al. 2002) respectively. A 5% Gaussian process noise was injected into the process model to create a random noise condition in the modeling process.

In the numerical model grid, twelve (12) observation locations were selected as shown in Figure 3.2. In practice, it is not feasible to construct observation wells at every point in the model domain because direct data acquisition in the subsurface is typically very expensive and tedious; hence field data sets are usually sparse. As a result, approximately 3% of the 400 grid points were assigned observation values in the model domain. The observation data pattern matrix, **H** was then constructed as a 12-by-400 sparse matrix to reflect the incompleteness of the field observations. Although the selection of observation sites can affect model results, the selection process can be carried out through detailed optimization procedure which is beyond the scope of this study.

For demonstration purposes, a benzene tracer of mass 464.5g, which corresponds to an initial benzene concentration of 5000 mg/L was injected as a pulse input at the coordinates (5, 10) resulting in a two-dimensional plume. Figure 3.2 shows the locations of the initial benzene contaminant and the observation sites in the model domain.

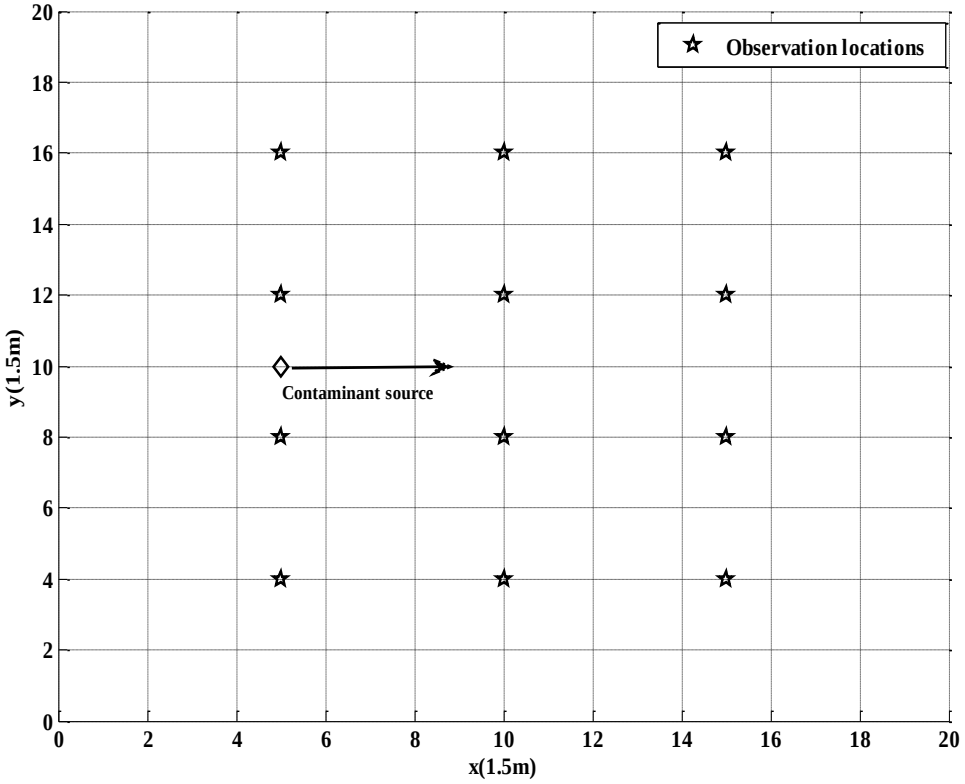

Figure 3.2. Observation locations and initial benzene source

3.6 Prediction Accuracy Examination

Extensive validation of the KF data assimilation has been shown in various applications using different statistical performance indices. In this study, the Root-Mean-Squared-Error (RMSE) index is used to examine the effectiveness of the KF and AKF

schemes. The model predictions are compared with a simulated true field to estimate the error parameter (RMSE).

$$RMSE(t) = \sqrt{\frac{1}{n_{ij}} \sum [C^E(i, j, t) - C^P(i, j, t)]^2} \tag{3.31}$$

where, C^E is the true field concentration at time t (mg/L); C^P is the predicted benzene concentration at time t (mg/L). The performance of the deterministic model and the KF scheme as well as the impact of the adaptive filter are illustrated and discussed below.

CHAPTER 4

Results and Discussion

4.1 Deterministic Model Prediction

The experiments were designed and implemented in a two-dimensional model domain using MATLAB routines. The simulation was carried out for a period of 10 days using a basic time unit of days. The deterministic model (FTCS) prediction was solely based on the specified hydraulic parameters, boundary conditions and initialization data. Figure 4.1 shows the benzene concentration prediction by the FTCS model at day 4. As shown in the figure, the prediction by the deterministic model clearly departs from the true field; this could be attributed to errors in the numerical scheme, model mechanisms and the uncertainties in the hydraulic parameters and initial conditions.

Hydraulic parameters and initialization data obtained from field observations vary markedly in space and as a result, field-scale models cannot be accurately validated on strictly deterministic grounds. This variability in model parameters is usually unavoidable due to the heterogeneous nature of the subsurface. Also, field data acquisition in the subsurface is very expensive and in most cases, obtaining hydraulic parameters with field equipment may be time consuming or even erroneous. The complex nature of the subsurface makes it extremely difficult to predict the fate and transport of contaminants by the use of numerical models. Although numerical models can provide information about contaminant migration in cases where field measurements may be infeasible, the results in most cases do not give an accurate representation of the transport process.

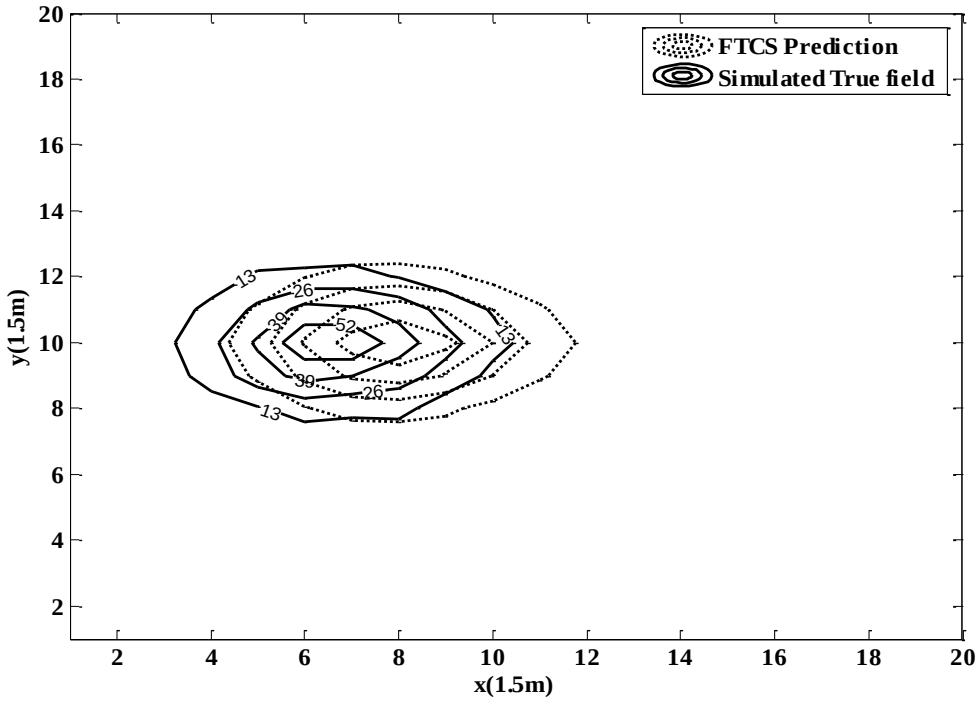

Figure 4.1. Comparison of FTCS prediction and true field concentration at day 4

The deterministic model provides an approximate solution to the transport equation without any information about the true field. Due to its inability to simulate true field conditions, the result of numerical models is usually idealistic for practical applications. Figure 4.2 shows the benzene contaminant plume at day 8 as predicted by the FTCS model. The benzene plume has spread over a wider area relative to the profile at day 4 due to the advection of contaminant and hydrodynamic dispersion. Furthermore, the smooth Gaussian-shaped contour produced by the deterministic model is as a result of the underlying advection-dispersion transport process and the non-stochastic nature of the model. The benzene contaminant plume predictions in Figure 4.1 and Figure 4.2 can be further improved by assimilating observation information into the system dynamics to guide the state estimation in order to give much accurate predictions.

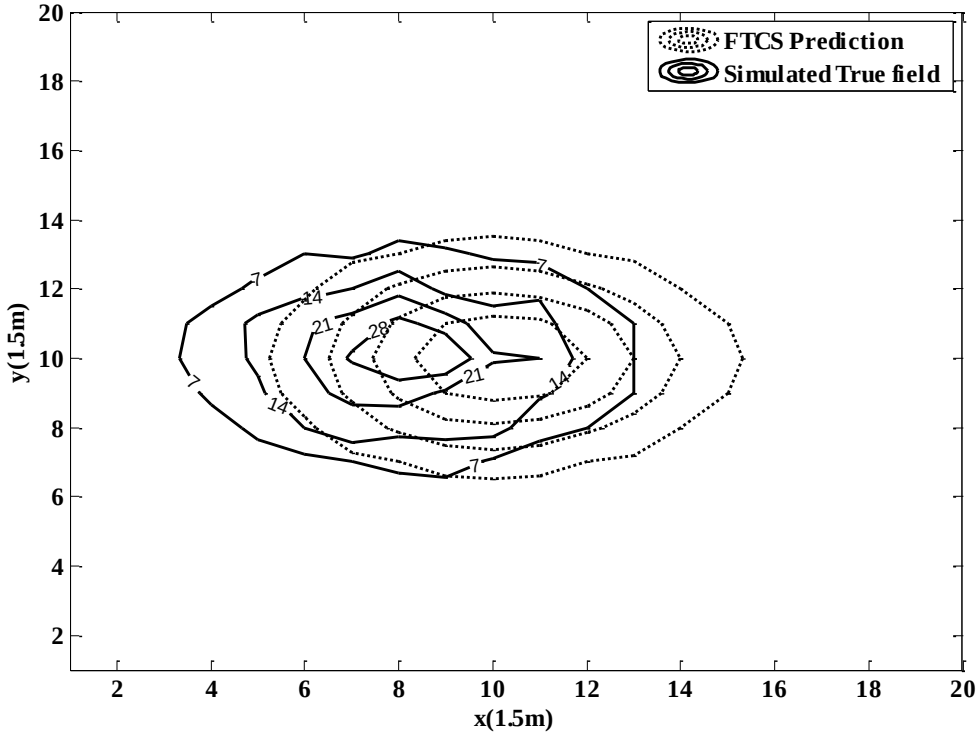

Figure 4.2. Comparison of FTCS prediction and true field concentration at day 8

4.2 Predictions by KF and AKF Data Assimilation Schemes

Stochastic data assimilation schemes can integrate observation information into the system model to simulate real-world transport processes with uncertain sources. The observation data can guide the system model to give accurate predictions that are much closer to the true field. In this study, the results for the stochastic models were generated based on twelve observation data points in the model domain. The twelve observations were sampled at each time step to project the optimal benzene concentration. A comparison in Figure 4.3 shows that the KF has better prediction accuracy than the deterministic model in that the KF gives a much better prediction that is much closer to the true field than the pure deterministic model. Furthermore, unlike the deterministic

model which is unable to predict natural stochastic conditions, the KF has the ability to predict irregular contour shapes which is typical of the true field. The stochastic nature of the KF accounts for the variability in the hydraulic parameters and initial conditions thereby reducing the effect of errors in the parameters and initialization data. This variability has been taken into account by injecting a 5% Gaussian noise into the system model during the KF implementation. Figure 4.4 shows the KF prediction at day 8 which further indicates a better representation of the true field in comparison with the FTCS model. As indicated in the figure, the contours predicted by the KF scheme coincided more accurately with the true field concentration whereas the FTCS had a faster moving plume in the x-direction due to the inaccurate linear velocity used in the model.

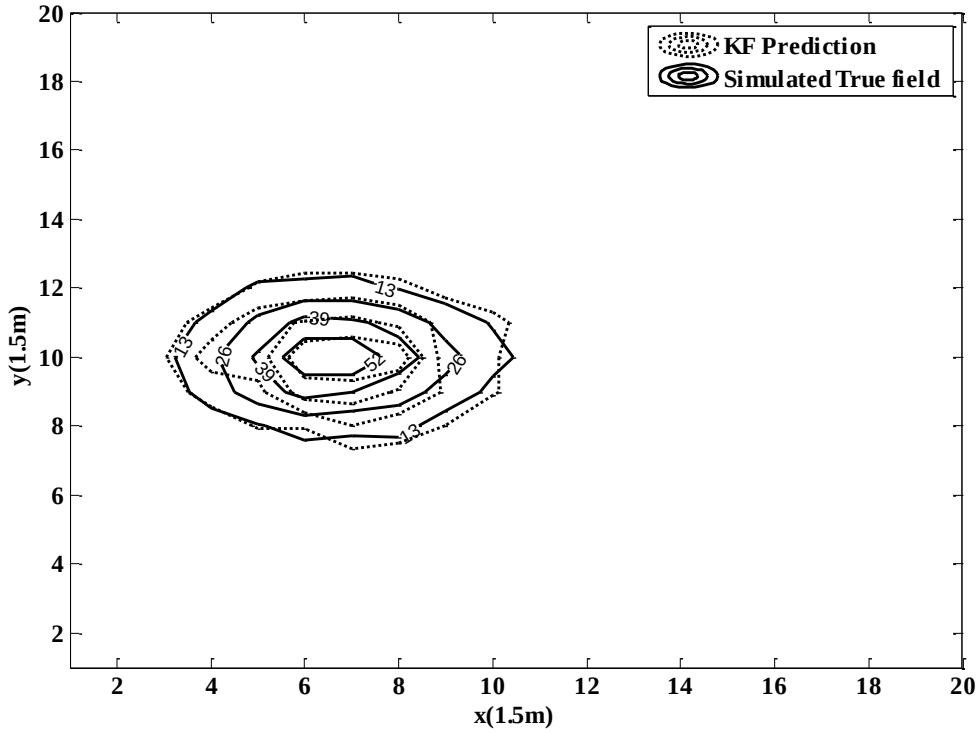

Figure 4.3. Comparison of KF prediction and true field concentration at day 4

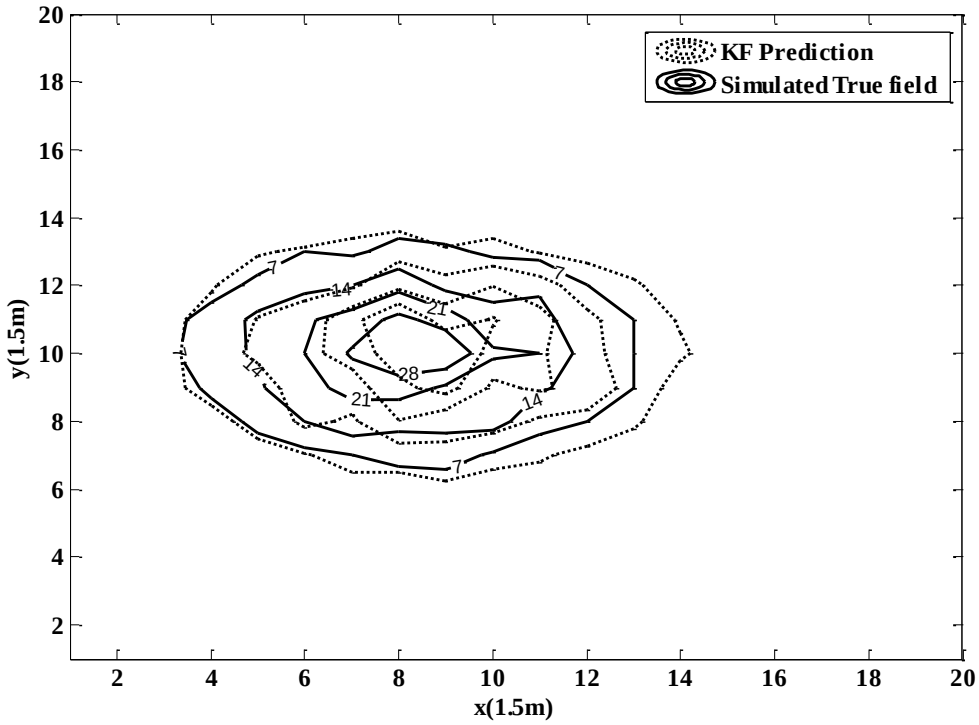

Figure 4.4. Comparison of KF prediction and true field concentration at day 8

As indicated in Figure 4.5, the introduction of the adaptive filtering scheme was successful in further reducing the prediction error of the KF scheme. The AKF had a much better prediction accuracy due to its ability to trace the irregular plume shape of the true field more accurately than both deterministic and KF models. As already discussed in previous sections, the adaptive scheme is introduced to adjust the dynamics of the system at each time step in order to give an accurate representation of the contaminant migration process. The limitation of the conventional KF is mainly attributed to inaccurate estimate of the process and observation noise covariances which must be specified before filter operation. The ICS-GC adaptive scheme improves the filter prediction accuracy by manipulating the system error covariance, $\mathbf{P}_t$ and the optimal Kalman gain, $\mathbf{K}_t$ by the use

of appropriate adaptive and forgetting factors.

The optimality of the KF filtering process depends significantly on the optimal Kalman gain ($\mathbf{K}_t$) which determines the weight assigned to the existing process and observation information. The estimation equation for $\mathbf{K}_t$ depends on a known $\mathbf{P}_t$, the observation error structure, $\mathbf{R}_t$ as well as the observation data pattern, $\mathbf{H}$; therefore accurate estimation of the error covariance can greatly improve filter performance. A further comparison in Figure 4.6 indicates an improved plume prediction relative to the true field. The effectiveness of the KF and AKF data assimilation schemes is mainly due to the fact that, both schemes utilize observation information to predict the state whereas the FTCS prediction is solely based on the deterministic model system of equations.

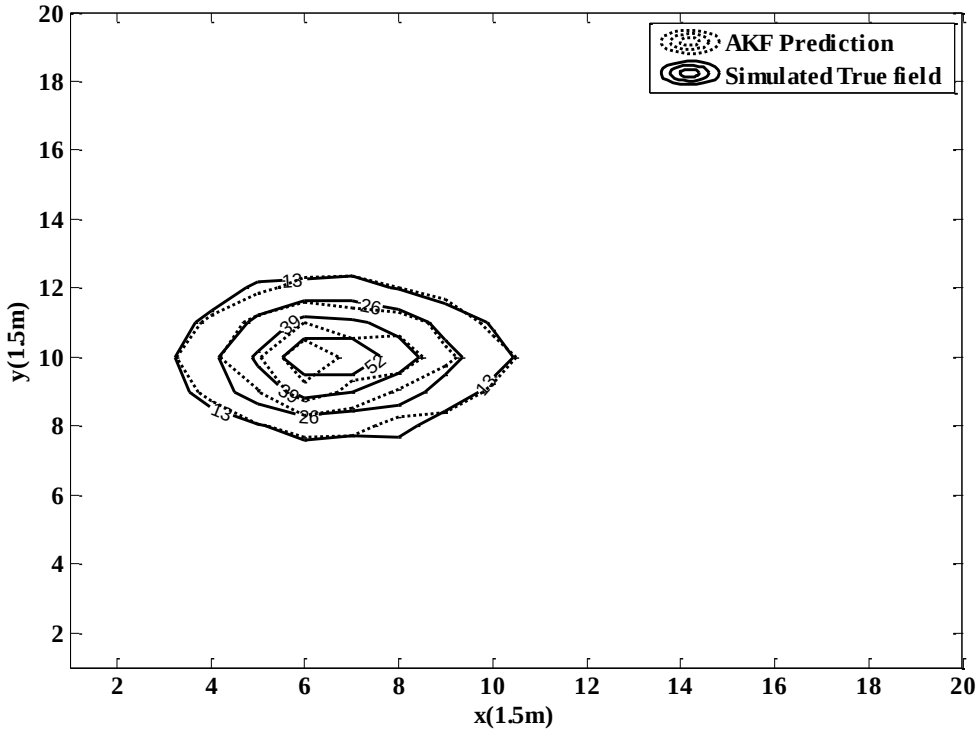

Figure 4.5. Comparison of AKF prediction and true field concentration at day 4

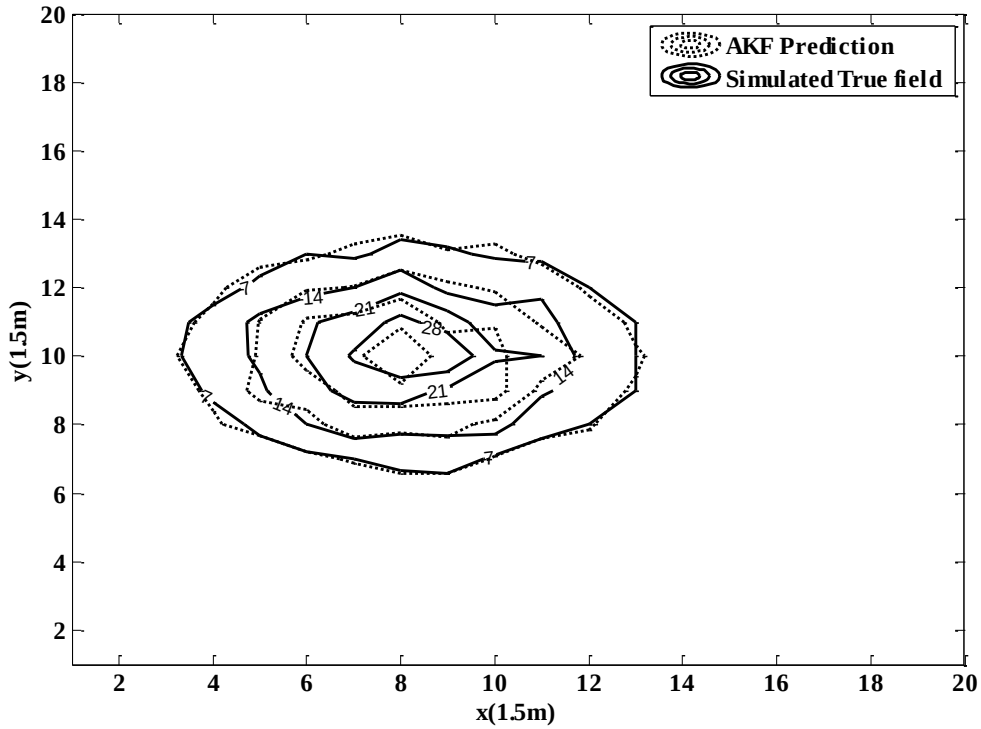

Figure 4.6. Comparison of AKF prediction and true field concentration at day 8

The benzene plume predictions by the various models are also depicted in three dimensional figures to give a much clearer representation of the concentration profiles. Figure 4.7 shows the concentration profiles by the various models. As already discussed, the AKF prediction was much closer to the true field concentration than both the deterministic and the KF predictions. At day 6, the true peak concentration was about 28 mg/L which was closely predicted by the AKF to be 25 mg/L. The peak concentration predicted by the KF was about 21 mg/L while that predicted by the deterministic model was approximately 18 mg/L indicating 16% and 28% prediction errors respectively. Furthermore, the deterministic model had a relatively lower peak concentration as a result of the inaccurate hydrologic parameter used in the model. In order to demonstrate the

effectives of the AKF, the deterministic model was assigned an inaccurate velocity of 0.85 m/d, which was 15% higher than the velocity used in the true field, 0.72 m/d. Although other hydraulic parameters and initial conditions may be inaccurate, the AKF data assimilation scheme can utilize observation information to project the optimal benzene contaminant plume with minimum deviation.

Figure 4.8 shows the AKF concentration profiles at various days of the simulation period; the trend indicates a rapid reduction in peak concentration as the benzene plume migrates along the flow path. As indicated in the figure, the initial benzene concentration of 5000 mg/L reduced to 82 mg/L after two days. This further reduced to about 15 mg/L at the end of the simulation period representing 0.3% of the initial benzene concentration. The reduction in aqueous benzene concentration can be attributed to factors such as advection of contaminant, molecular diffusion, decay, sorption and other mechanisms which affect the attenuation of benzene in the subsurface. The attenuation process may also depend significantly on the organic carbon content and the oxygen levels in the contaminated aquifer which characterize the assimilative capacity of the aquifer. (Kim et al. 2006). Advection of the benzene contaminant plume results from unidirectional fluid flow due to hydraulic gradients while molecular diffusion occurs as a result of chemical concentration gradient. In a saturated media, diffusion mainly occurs as a result of the random Brownian motion of water molecules or eddies on a larger scale.

For ethanol-blended petroleum products, the biodegradation of benzene may be inhibited by the presence of ethanol. Studies have shown that, benzene plume lengths are likely to increase by 16-34% in the presence of ethanol (Deeb et. al. 2002).

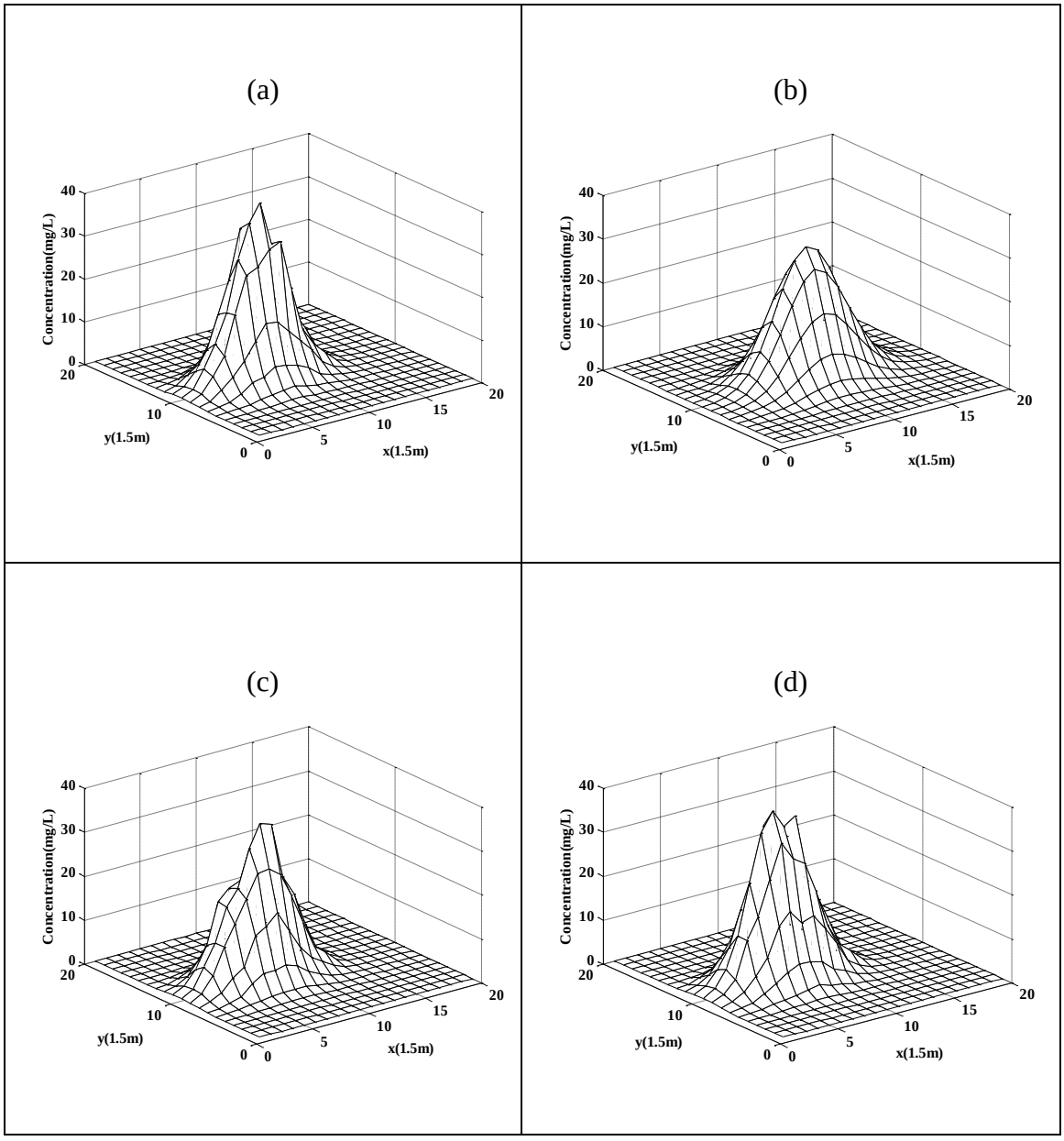

Figure 4.7. Comparison of benzene plume prediction by various models at day 6: (a) True field (b) FTCS model (c) KF model (d) AKF model

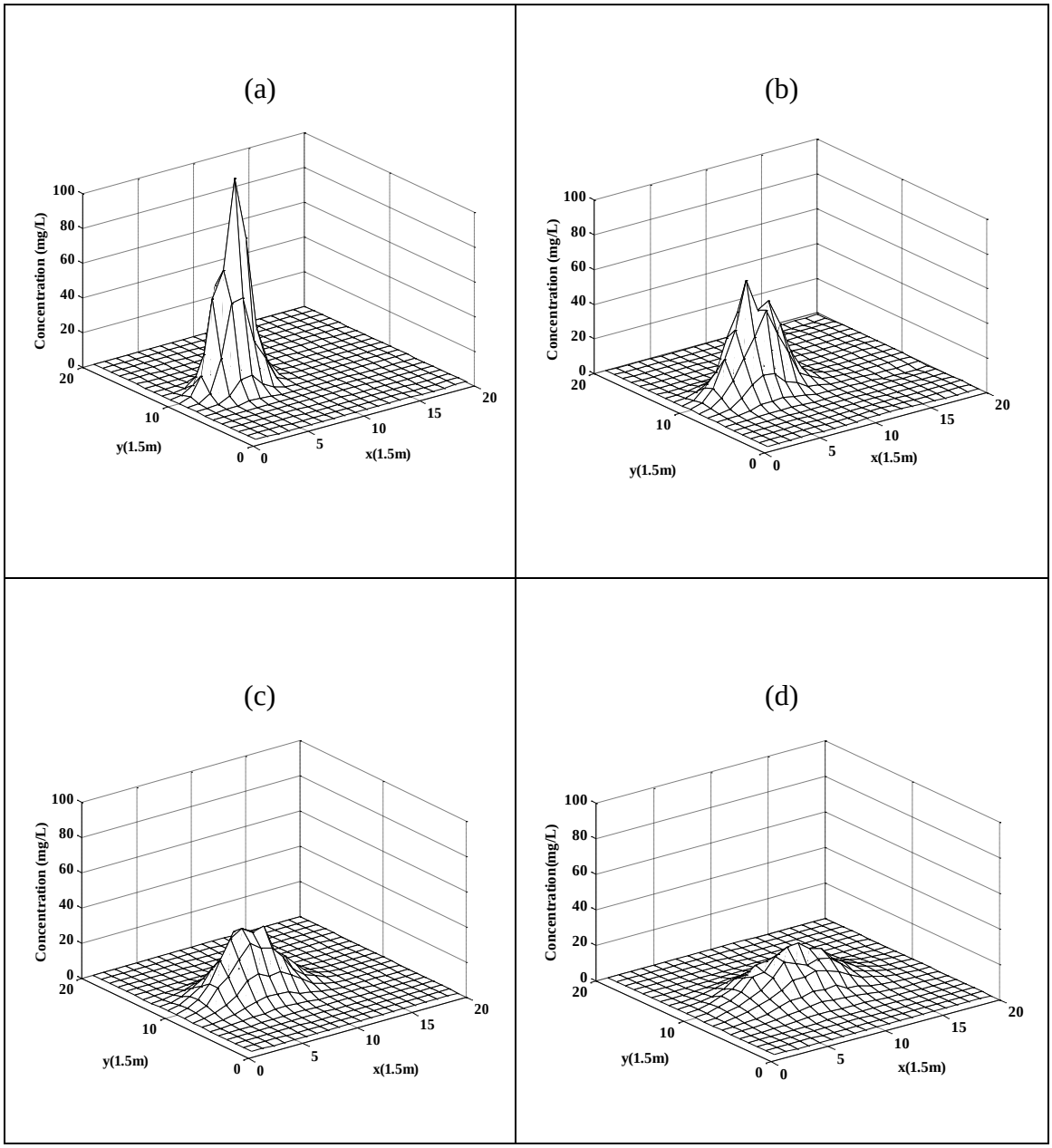

Figure 4.8. AKF prediction at various days of simulation: (a) Day 2 (b) Day 4 (c) Day 8 (d) Day 10

4.3 Effectiveness of Prediction Models

The effectiveness of the prediction models was further validated by the RMSE profile as shown in Figure 4.9. The RMSE has been chosen as the average standard error between the simulated true field and model predictions. For the entire simulation period, the FTCS model had an average prediction error of about 7.0 mg/L when compared with the true field. This, however, reduced to 5 mg/L at the end of the simulation. The average prediction error of the KF model was about 2.5 mg/L while that of the AKF scheme was approximately 1.6 mg/L for the entire simulation period. With the initialization of the adaptive process at time step five, the prediction error of the AKF further reduced from 8.4 mg/L to about 2.2 mg/L. The models had higher prediction errors at the initial time steps due to incomplete observation information and poor knowledge of the flow field. As filtering proceeded, both KF and AKF models successfully assimilated observation information into the system to give a much accurate prediction of the benzene plume. It is evident in Figure 4.9 that the introduction of the KF scheme improved the deterministic model prediction by reducing the model error from 5.0 mg/L to 1.1 mg/L, thus improving the prediction accuracy by about 78%.

Although the KF was successful in reducing the errors in the numerical model, its prediction accuracy was limited by the uncertainties associated with the process and observation noise statistics. A further comparison in Figure 4.9 indicates the AKF had the least prediction error. Using observation information from just 3% of the model grid points, the AKF scheme further reduced the prediction error of the KF from 1.1 mg/L to 0.9 mg/L at the end of the simulation period thus improving prediction accuracy by 18%.

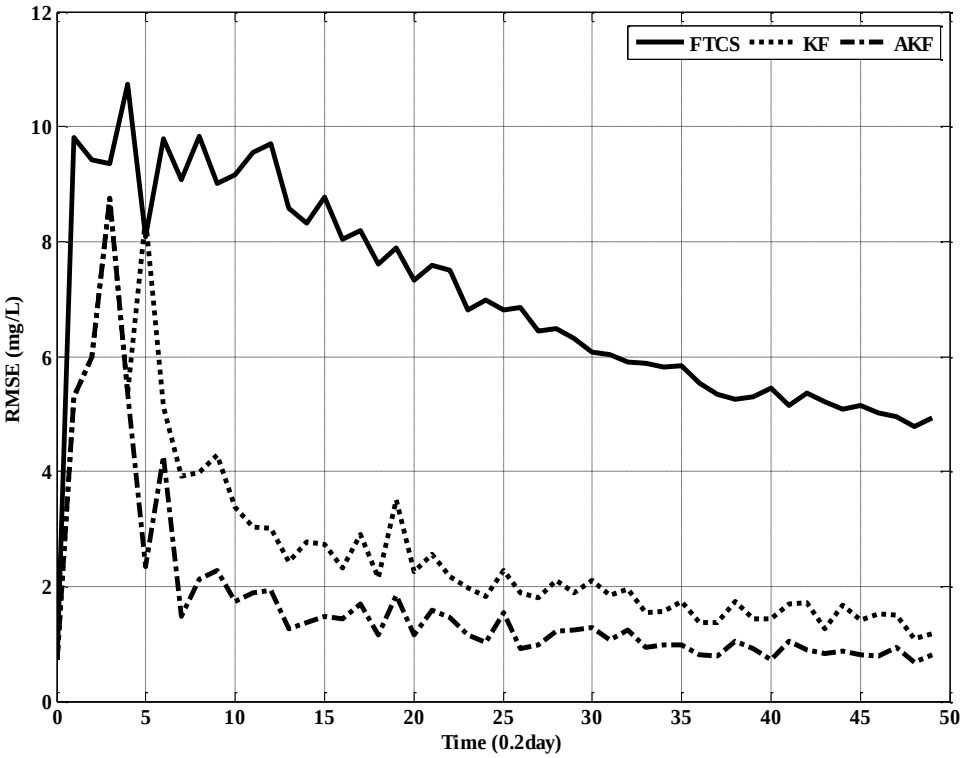

Figure 4.9. Root-Mean-Squared-Error (RMSE) profile for FTCS, KF and AKF

4.4 AKF Stability Examination

The stability of the KF depends significantly on the process and observation noise covariances. Inaccurate estimates of these noise statistics may lead to poor convergence performance and divergence in certain cases. As a result, the AKF has been constructed to further boost the filter stability and convergence performance by reducing the effects of associated errors in the noise statistics in order to avoid filter divergence. The stability of the proposed AKF scheme was examined by conducting multiple runs of the RMSE profile. Figure 4.10 shows the variability of ten repeated runs of the RMSE profile. It can be inferred from the figure that the AKF had a reasonable level of stability and tends to converge after the 20^{th} time step of the simulation period.

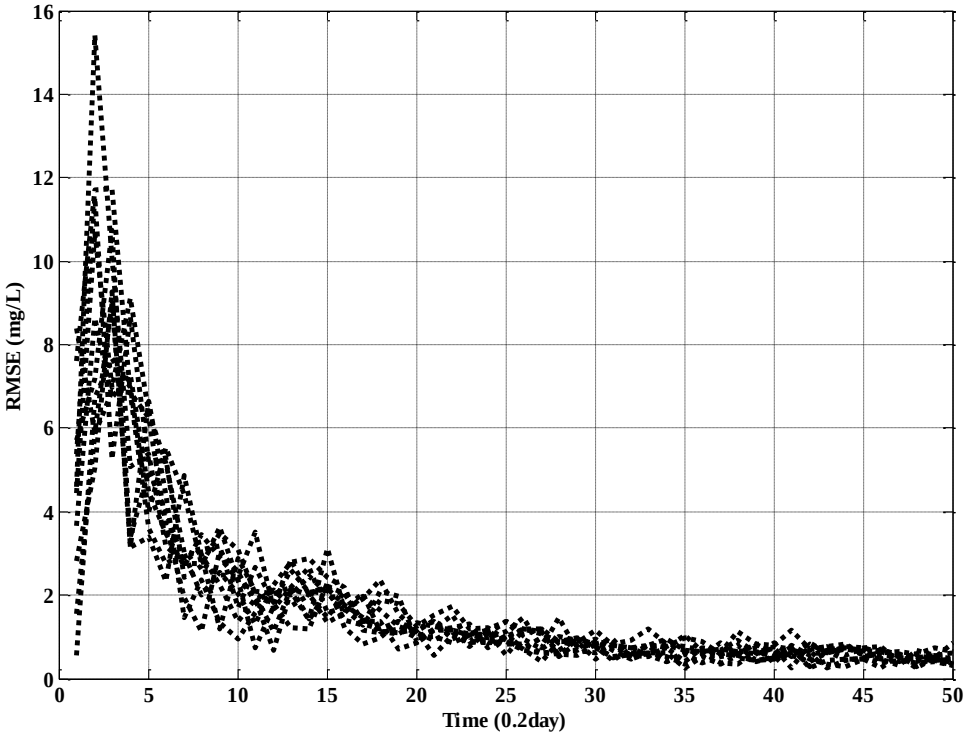

Figure 4.10. Repeated runs of RMSE profile for AKF

Figure 4.11 shows the error bars for five repeated runs of the RMSE profile. Error bars often indicate one standard deviation of uncertainty although they may also indicate the standard error. The figure shows the variability of the various runs; the upper and lower ranges and the mean error value. The upper values represent the summation of the mean and the highest value whereas the lower values indicate the difference between the mean and the lowest value. As shown in the figure, the trend indicated slight instability in the profiles at the initial time steps which was characterized with higher prediction errors. The maximum prediction error was estimated to be about 9.9 mg/L which further reduced to 5.8 mg/L at the beginning of the adaptive filtering process. The profile however converged to an average prediction error of about 1.3 mg/L after the 30^{th} time step.

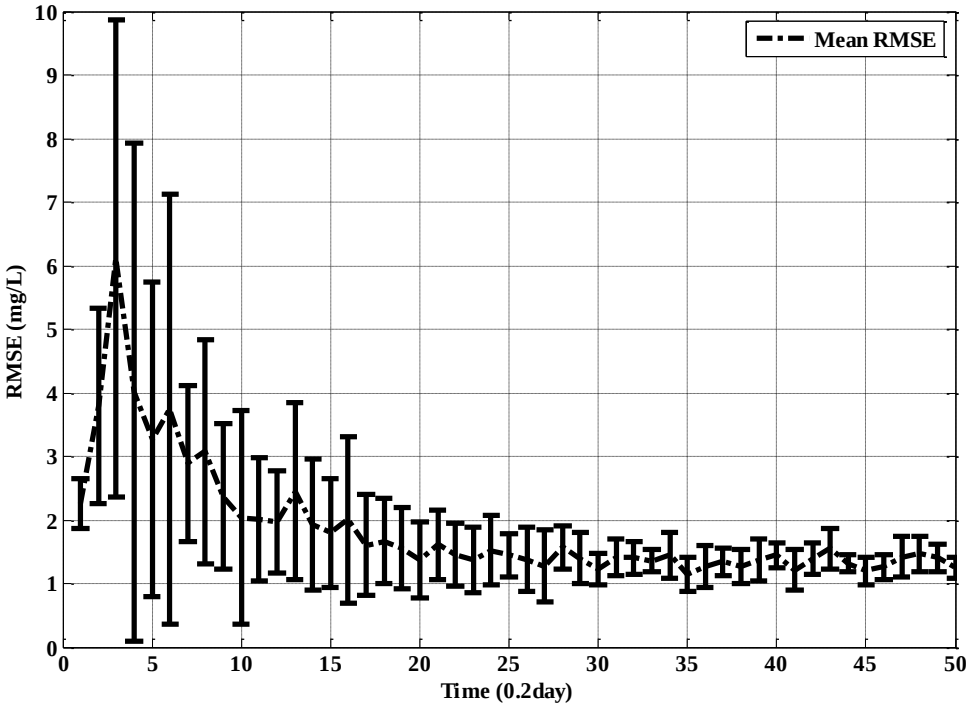

Figure 4.11. Error bar profile for AKF prediction

4.5 Sensitivity Tests on Model Parameters

The ICS-GC scheme manipulates the error covariance and the optimal Kalman gain by the use of the adaptive and forgetting scalar variables. The range of these scalar factors has a significant effect on the performance of the AKF scheme. In this study, a scalar factor greater than one was required to make a significant impact on the error covariance and the optimal Kalman gain. Increasing the size of the adaptive factor can make the system less responsive to the dynamic model and predict the state solely on the basis of the observation data. Similarly, using a very large forgetting factor can make the AKF scheme less responsive to the observations. Therefore, working within the appropriate scalar factor range can ensure an efficient adaptive system which will

assimilate the required observation data into the system model to optimize filter performance. Figure 4.12 shows the adaptive factor sequence for the total simulation period. The window size used for estimating the true innovation covariance was 5; hence, the rapid increase in the scalar factor value at time step 5 indicates the beginning of the adaptive filtering process.

Sensitivity tests were further conducted to examine the impact of the window size on the model results by estimating the innovation covariance within different window sizes. Figure 4.13 shows the effect of different window sizes on the performance of the ICS-GC AKF scheme. The trend in the result suggests that increasing the window size tends to degrade filter performance and stability.

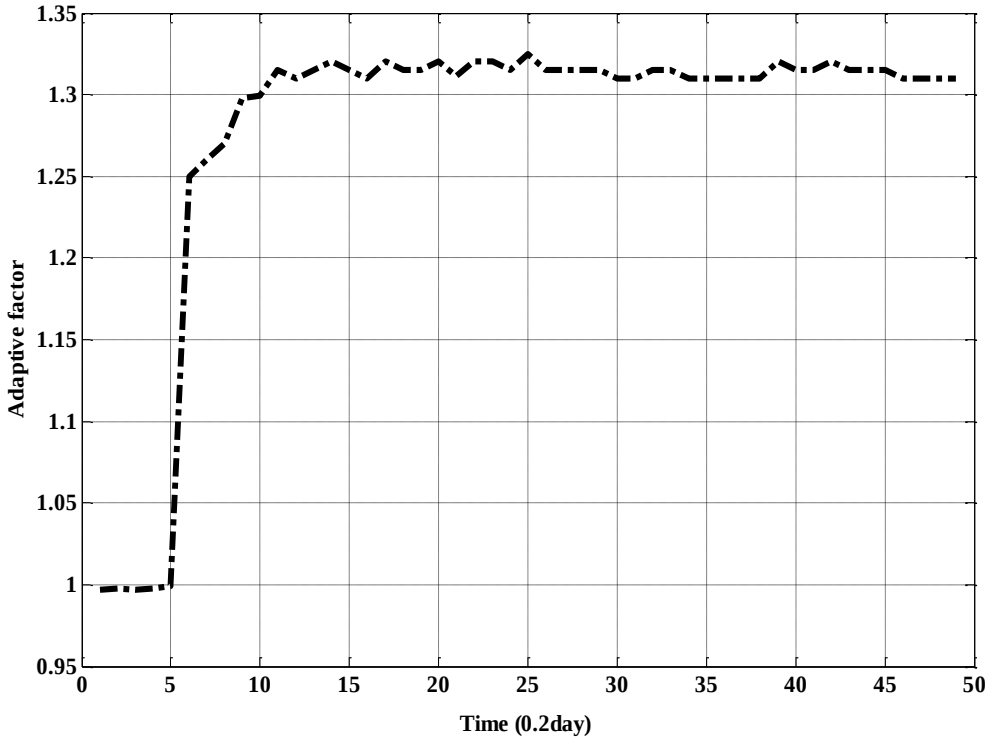

Figure 4.12. Adaptive factor sequence for total simulation period

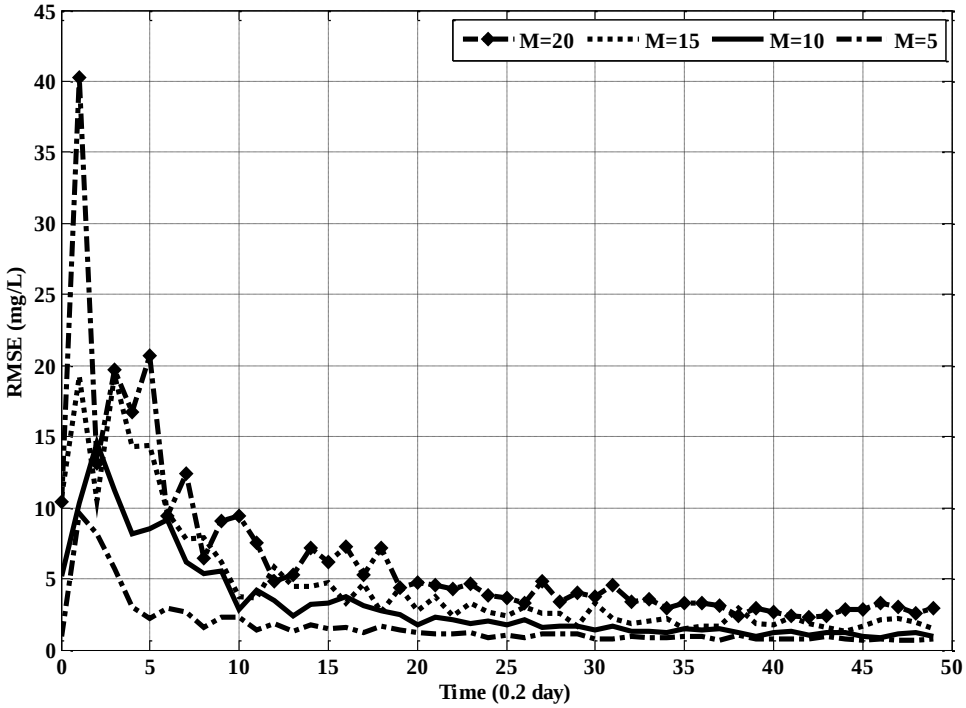

Figure 4.13. Comparison of AKF performance with different window sizes

As the window size increases, the adaptive filtering process is delayed which makes the AKF scheme less effective. As shown, a window size of 5 had the least prediction error of 0.9 mg/L whereas a window size of 20 was characterized with an unstable profile and a much higher prediction error (3.5 mg/L). Although a window size of 10 had a lower prediction error, the profile had slight instability at the initial stages. It must however be noted that when the window size becomes too small, the true innovation covariance may not be accurately estimated and consequently the performance of the AKF becomes less effective. On the other hand, using a large window size may affect the optimality of the filter because within a large window, the covariance does not depend greatly on the past predicted covariance since it changes at every time step.

Therefore, in implementing the AKF scheme, choosing the appropriate window size can significantly improve filter stability and accuracy.

The performance of the ICS-GC adaptive filter scheme greatly depends on the size of the moving window and the range of the scalar factors. In certain adaptive filtering mechanisms, the scalar factor can either be greater or smaller than one which provides a full range of options to tune the error covariance and the optimal Kalman gain. In some instances, constant adaptive factors have been used to avoid the repetitive estimation of the scale factors at each time step. However, such schemes have had limited success due to their inability to adjust the filter to the dynamics of the system.

The KF in general has a lower computational effort due to its ability to predict the current state without storing information about the previous states. The AKF scheme proposed in this study requires a relatively lower computational effort in the MATLAB environment; in its implementation, the average simulation time was approximately 60 seconds on a 2.0 GHz computer with 3.0 GB RAM.

CHAPTER 5

Conclusion

5.1 Summary and Conclusion

The effectiveness of conventional numerical models in contaminant transport modeling is usually limited by the presence of errors which mostly result from uncertainties in the hydraulic parameters and errors in the numerical scheme. The KF scheme is a very effective data assimilation tool in subsurface contaminant transport modeling which assimilates observation data into the system dynamics to improve predictions of traditional numerical models. This study has implemented and tested the effectiveness of an Adaptive Kalman Filter (AKF) scheme based on an innovation covariance scaling and gain correction approach. The AKF has demonstrated its ability to further improve the prediction accuracy of the KF by reducing the effect of errors in the process and observation noise covariances.

The simulation results indicated that the prediction error of the deterministic model was significantly higher than that of the KF and AKF schemes. At the end of the simulation period, the average prediction error of the deterministic model was reduced from 5.0 mg/L to 1.1 mg/L when the KF scheme was implemented. However, the introduction of the AKF scheme further reduced the prediction error of the KF from 1.1 mg/L to 0.9 mg/L thus improving accuracy by 18%. Overall, the implementation of the AKF successfully improved the prediction accuracy of the deterministic model by about 82% which indicates a significant improvement in benzene plume prediction. The results

of sensitivity tests on stability and convergence performance have also demonstrated the robustness of the ICS-GC AKF scheme. Furthermore, a comparison of the AKF performance using different window sizes suggest that using a window size of five can give a much improved filter stability and accuracy. Thus, choosing the appropriate window size can have a significant impact on filter performance.

The results of this study have demonstrated that accurate estimate of the process and observation noise covariances may greatly improve filter stability and accuracy. The optimality of the KF greatly relies on the accuracy of these noise statistics. Due to its statistical filtering nature, the KF scheme requires prior complete knowledge of the process and observation noise statistical structures. In practice, these statistics are difficult to obtain, therefore an adaptive formulation of the KF can help to adjust the filter dynamics to reduce the effect of errors in the noise statistics.

Although the AKF scheme has been applied to simulate the transport of benzene, the model also has the capacity to predict the concentration of other hydrocarbons and radioactive pollutants in the subsurface. Conservative pollutants such as chlorides can also be simulated by the proposed scheme by eliminating the chemical reaction term in the deterministic model.

The implementation of remediation systems for contaminated sites is essential for protecting groundwater resources and reducing risks to local communities. However, effective design and management of NAPL-contaminated aquifers still remains a challenging problem to practitioners. The use of subsurface transport model, coupled with stochastic data assimilation schemes can provide accurate information about

contaminant transport which can assist engineers and policy makers in making long term management plans for the remediation of contaminated sites.

5.2 Recommendations for Future Study

While the performance of the AKF is satisfactory, it has been developed based on the assumption of linearity in the process and observation models. In practice however, optimal contaminant plume prediction can be obtained if these models are treated as non-linear. Therefore, the Adaptive Extended Kalman Filter (AEKF) is expected to be constructed in future research for application in systems with nonlinear process and observation models. Furthermore, the use of a multi-component reactive model, capable of simulating the individual species of BTEX compounds is recommended for future extension of this study. Such a scheme can account for the effect of interaction of the various components of BTEX on benzene concentration in the subsurface.

REFERENCES

Avis, S.P., and Hutton, C.P. (1993). "Acute Benzene Poisoning: A Report of three fatalities." *Journal of Forensic Sci.*, 38(3), 599-602.

Batlle-Aguilar, J., Brouyère, S., Dassargues, A., Morasch, B., Hunkeler, D., Höhener, P., Diels, L.,Vanbroekhoven, K., Seuntjens, P., and Halen, H. (2009). "Benzene dispersion and natural attenuation in an alluvial aquifer with strong interactions with surface water." *Journal of Hydrology*, 369, 305-317.

Chang, S.Y., and Jin, A. (2005). "Kalman filtering with regional noise to improve accuracy of contaminant transport models." *Journal of Environmental Engineering*, 131(6), 971-982.

Chang, S.Y., and Latif, S. M. I. (2010). "Extended Kalman filtering to improve accuracy of subsurface contaminant transport models." *Journal of Environmental Engineering*, 136 (5), 466-474.

Chang, S.Y., and Latif, S. M. I. (2007). "Use of Kalman filtering and Particle filtering in a one dimensional leachate transport model." *Proc. Third National Conference on Environmental Science and Technology*, Sep 12-14, 2007, Greensboro, NC.

Chang, S.Y., and Li, X.P. (2006). "Modeling Chlorobenzene leaching from a Landfill into a Soil Environment Using Particle Filter Approach." *Journal of Environmental Informatics*, JEI 12(2),88-95.

Chen, B.H., Wang, C.Y., and Chiu, C.P. (1996). "Evaluation of Analysis of Polycyclic Aromatic Hydrocarbons in Meat Products by Liquid Chromatography." *Journal of Agriculture and Food Chemistry*, 44, 2244-2251.

Choi, J-W., Ha, H.C., Kim, S-B., and Kim, D-J. (2005). "Analysis of benzene transport in a two dimensional aquifer model." *Hydrol. Process*., 19, 2481-2489.

Corseuil, H. X., Hunt, C.S., Dos Santos, R.C.F and Alvarez, P.J.J (1998).''The influence of the gasoline oxygenate ethanol on aerobic and anaerobic BTEX biodegradation." *Water Resources*, 32 (7), 2065-2072.

Da Silva, M. L. B., and Alvarez, P.J. J. (2002). ''Effects of ethanol versus MTBE on benzene, toluene, ethylbenzene, and xylene natural attenuation in aquifer columns.'' *Journal of Environmental Engineering*, 128 (9), 862-867.

Dean, B. J. (1985). ‘‘Recent findings on the genetic toxicology of benzene, toluene, xylenes and phenols.’’ *Mutation Research*, 145, 153-181.

Deeb, R.A., Sharp, J.O., Stocking, A., McDonald, S., West, K.A., Laugier, M., Alvarez, P.J.J., Kavanaugh, M.C., and Alvarez-Cohen, L. (2002). "Impact of Ethanol on Benzene Plume Lengths: Microbial and Modeling Studies." *Journal of Environmental Engineering*, 128 (9), 868-875.

Delshad, M., Pope, G.A., and Sepehrnoori, K. (1996). "A Compositional Simulator for modeling surfactant enhanced aquifer remediation," 1. Formulation, *Journal of Contaminant Hydrology*, 23,303-327.

Fathabadi, V., Shahbazian, M., Salahshour, K., and Jargani, L. (2009). "Comparison of Adaptive Kalman Filter Methods in State Estimation of a Nonlinear System Using Asynchronous Measurements." *Proceedings of the World Congress on Eng. and Computer Sc.*, 2, October, 20-22,2009, San Francisco, CA.

Fischer, A., Theuerkorn, K., Stelzer, N., Gehre, M., Thullner, M., and Richnow, H.H. (2007). "Applicability of Stable Isotope Fractionation Analysis for the Characterization of Benzene Biodegradation in a BTEX-contaminated Aquifer." *Environ. Sci. Technology*, 41 (10), 3689-3696.

Franssen, H. H., Kuhlman, U., Kaiser, H., Stauffer, F., and Kinzelbach, W. (2008). "The Ensemble Kalman Filter for real-time Groundwater Flow Modeling of the upper Limmat aquifer in Zurich (Switzerland)." *Geophysical Research Abstracts*,10, EGU2008-A-04366.

Freeze, R.A., and Cherry, J.A., (1976). "*Groundwater*." Prentice-Hill Inc., Englewood Cliffs, New Jersey.

Fu, J.K., and Luthy, R.G. (1986). "Effect of Organic solvent on Sorption of Aromatic solutes onto soils." *Journal of Environmental Engineering*, 111 (2), 346-366.

Fu, W., Fu, H., Skøtt, K., and Yang, M. (2008). "Modeling the Spill in the Songhua River after the explosion in the Petrochemical Plant in Jilin." *Env. Sci. Pollut. Res*., 15(3), 178-181.

Gary, S., and Rixey, W.G. (1999). "The dissolution of benzene, toluene, m-xylene and naphthalene from a residually trapped non-aqueous phase liquid under mass transfer limited conditions." *Journal of Contaminant Hydrology*, 36, 313-331.

Goegebeur, M., and Pauwels, V. R. N. (2007). "Improvement of the PEST parameter estimation algorithm through extended Kalman filtering." *Journal of Hydrology*, 337, 436-451.

Gomez, D.E. and Alvarez, P.J.J. (2009). "Modeling the natural attenuation of benzene in groundwater impacted by ethanol-blended fuels: Effect of ethanol content on the lifespan and maximum length of benzene plumes." *Water Resources Research*, 45, W03409, doi: 10.1029/2008WR007159

Grewal, M.S. and Weil, L.R. (2001). "*Global Positioning Systems, Inertial Navigation and Integration*." John Wiley & Sons, USA.

Guo, W.J., and Wang, W.X. (2008). "A numerical Oil Spill Model based on a hybrid Method." *Marine Pollution Bulletin*, 58, 726-734.

Hajiyev, C. (2007). "Adaptive filtration algorithm with the filter gain correction applied to integrated INS/radar altimeter." *Proc. of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, 221(5), 847-855.

He, L., Huang, G.H., and Lu, H.W. (2011). "Characterization of Petroleum-Hydrocarbon Fate and Transport in Homogeneous and Heterogeneous Aquifers Using a Generalized Uncertainty Estimation Method." *Journal of Environmental Engineering*, 137 (1), 1-8.

Heermann, S. E., and Powers, S. E. (1996). ''The dissolution of BTEX compounds from oxygenated gasoline.'' *ACS Symp. Petroleum Contamination in the Environment*: Assessment and Remediation, New Orleans.

Hide, C., Michaud, F., and Smith, M. (2004)."Adaptive Kalman Filtering Algorithms for integrating GPS and low cost INS." *IEEE Position Location and Navigation Symposium*, Monterey, CA, 227-233.

Hide, C., Moore, T. and Smith, M. (2004b). "Multiple model Kalman filtering for GPS and low-cost INS integration." *ION GNSS 17th international technical meeting of the satellite division*, Long Beach California, 1096-1103.

Hu, C., Chen, W., Chen, Y., and Liu, D. (2003). "Adaptive Kaman Filtering for Vehicle Navigation." *Journal of Global Positioning System*, 2 (1), 42-47.

Jin, A. (2009). "Ensemble Based Kalman Filtering to Improve Accuracy of Subsurface Contaminant Models." *PhD Dissertation*, North Carolina A&T State University, Energy and Environmental Studies, July, 2009.

Johnson, J., Gourley, A., Baxter, E., and Vieux (2006). "A model for identifying sources of model uncertainty in rainfall-runoff simulations." *Journal of Hydrology*, 327, 68-80.

Kalman, R. E. (1960). "A New Approach to Linear Filtering and Prediction Problems." Transaction of the ASME, *Journal of Basic Engineering*, 35-45.

Kao, C.M., Chien, H.Y., Surampalli, R.Y., Chien, C.C., and Chen, C.Y. (2010). "Assessing of Natural Attenuation and Intrinsic Bioremediation Rates at a Petroleum-Hydrocarbon Spill Site: Laboratory and Field Studies." *Journal of Environmental Engineering*, 136 (1), 54-67.

Kerr, R. A., Kintisch, E., Schenkman, L. ,and Stokstad, E. (2010). "Gulf oil disaster: five questions on the spill." *Science* 328 (5891), 962-963.

Kim, K.H., and Lee, J.G. (2006). "Adaptive Two-Stage EKF for INS-GPS Loosely Coupled System with Unknown Fault Bias." *Journal of Global Positioning Systems*, 5(1-2), 62-69.

Kim, S.B., Kim, D.J., and Yun. S.T. (2006). "Attenuation of Aqueous Benzene in Soils under Saturated Flow Conditions." *Environmental Technology*, 27 (1), 33-40.

Law, A., and Kelton, W.D., (2000). *Simulation Modeling and Analysis*, Third Edition. New York, McGraw Hill.

Li, L., and Wendy, D. G. (1999). "Stochastic analysis of solute transport in heterogeneous aquifers subject to spatiotemporal random recharge." *Water Resources Research*, 35, 953-971.

Li, X. (2006). "State and parameter estimation using Particle filter: Application on organic pollutant transport in groundwater." *MSc. Thesis*, North Carolina A&T State University, Greensboro, NC.

Mackay, D.M., Roberts, P.V., and Cherry, J.A. (1985). "Transport of organic contaminants in groundwater." *Environ. Sci. and Technology*, 19, 384-392.

Maraqa, M.A., Wallace, R.B., and Voice, T.C. (1999). "Effect of Water Saturation on Retardation of Ground-Water Contaminants" *Journal of Environmental Engineering*, 125 (8), 697-704.

McLaughlin, D. (2002). "An integrated approach to hydrologic data assimilation: interpolation, smoothing and forecasting." *Advances in Water Resources*, 25, 1275-1286.

Mor, S., Ravindra, K., Dahiya, R.P., Chandra, A., 2006. Leachate characterization and assessment of groundwater pollution near municipal solid waste landfill site. Environmental Monitoring and Assessment 118, 435-456.

National Ground Water Association (NGWA) and United States Geological Survey (USGS) (2003). "*Ground Water Statistics*." www.ngwa.org (assessed on: June 15, 2010).

Nerger, L., Danilov, S., Kivman, G., Hiller, W., and Schroter, J. (2007). "Data assimilation with the Ensemble Kalman Filter and the SEIK filter to a finite element model of the North Atlantic." *Journal of Marine Systems*, 65, 288-298.

Prommer, H., Barry, D.A., and Davis, G.B. (1998). "A one-dimensional reactive multi-component transport model for biodegradation of petroleum hydrocarbons in groundwater." *Environ. Modeling and Software*, 14(2-3), 213-223.

Qin, X.S., Huang, G.H., and Hea, L. (2009). "Simulation and optimization technologies for petroleum waste management and remediation process control." *Journal of Environmental Management*, 90, 54-76.

Reisch, M. (2003). "Brownfields and superfund issues in the 108th Congress." *Congressional Research Service Issue Brief IB10114*, The Library of Congress, Washington, DC., www.ncseonline.org/nle/crsreports/03Jun/IB10114.pdf.

Saad, G.A. (2007). "Stochastic Data Assimilation with Application to multi-phase flow and health monitoring problems." *Doctoral dissertation*, University of Southern California, Los Angeles, CA.

Schwartz, F. W., and Zhang, H. (1994). *Fundamentals of Groundwater*. John Wiley & Sons, Inc., 605 Third Avenue, NY.

Singh, R.K., j Datta, M., and Nema, A.K. (2009). "A new system for groundwater contamination hazard rating of landfills." *Journal of Environmental Management*, 91, 344-357.

Smith, J.A., and Jaffe, P.R. (1994). "Benzene Transport through Landfill Liners containing Organophilic Bentonite." *Journal of Environmental Engineering*, 120 (6), 1559-1577.

Tam, E.K.L., and Beyer, P.H. (2002). "Remediation of contaminated lands: A decision methodology for site owners." *Journal of Environmental Management*, 64, 387-400.

USEPA (2009). "*End of Year Activity Report*." U.S. Environmental Protection Agency, Office of Underground Storage Tanks, Washington, DC.

USEPA (2010) U.S. Environmental Protection Agency. "Drinking Water Contaminants." Groundwater and Drinking Water, http://www.epa.gov/safewater/hfacts.html (assessed: May 22, 2010).

Van Geer, F.C. (1982). "An Equation Based Theoretical Approach to Network Design for Groundwater Levels using Kalman Filters*." Int. Assoc. Hydrol. Sci*., 136, 241-250.

Weidong, D., Wang, J., and Rizos, C. (2006). "Improving Covariance Based Adaptive Estimation for GPS/INS Integration." School of Surveying and Spatial Information Systems, University of New South Wales, Sydney, Australia.

Welch, G., and Bishop, G. (1995). "An introduction to the Kalman Filter." University of North Carolina–Chapel Hill, Department of Computer Science, Tr95-041, http://www.Cs.Unc.Edu/~ Tracker/Ref/S2001/Kalman (assessed: June 15, 2010).

Yang, Y. and Gao, W. (2006). "An optimal adaptive Kalman filter." *Journal of Geodesy*, 80(4), 177-183.

Yang, Y. and Xu, T. (2003). "An adaptive Kalman filter based on sage windowing weights and variance components." *Journal of Navigation*, 56(3), 231-240.

Zhang, Y., Person, M., Merino, E., and Szpakiewcz, M. (2005). "Evaluation of Soluble Benzene Migration in the Uinta Basin." *Journal of Geofluids*, 5, 106-123.

Zou, S., and Parr, A. (1995). "Optimal Estimation of two-dimensional Contaminant Transport." *Ground Water*, 33(2), 319-325.