8-1-2012

Value of Information in Design of Groundwater Quality Monitoring Network under Uncertainty

Abdelhaleem I. Khader
Utah State University

VALUE OF INFORMATION IN DESIGN OF GROUNDWATER QUALITY

MONITORING NETWORK UNDER UNCERTAINTY

by

Abdelhaleem Khader

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Civil and Environmental Engineering

Approved:

Dr. Mac McKee Major Professor	Dr. Jagath Kaluaratchchi Committee Member
Dr. David Stevens Committee Member	Dr. David Rosenberg Committee Member
Dr. Arthur Caplan Committee Member	Dr. Mark McLellan Vice President for Research and Dean of School of the Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2012

ABSTRACT

Value of Information in Design of Groundwater Quality Monitoring Network under Uncertainty

by

Abdelhaleem I. Khader, Doctor of Philosophy

Utah State University, 2012

Major Professor: Dr. Mac McKee
Department: Civil and Environmental Engineering

The increasing need for groundwater as a source for fresh water and the continuous deterioration in many places around the world of that precious source as a result of anthropogenic sources of pollution highlights the need for efficient groundwater resources management. To be efficient, groundwater resources management requires efficient access to reliable information that can be acquired through monitoring. Due to the limited resources to implement a monitoring program, a groundwater quality monitoring network design should identify what is an optimal network from the point of view of cost, the value of information collected, and the amount of uncertainty that will exist about the quality of groundwater. When considering the potential social impact of monitoring, the design of a network should involve all stakeholders including people who are consuming the groundwater.

This research introduces a methodology for groundwater quality monitoring network design that utilizes state-of-the-art learning machines that have been developed from the general area of statistical learning theory. The methodology takes into account uncertainties in aquifer properties, pollution transport processes, and climate. To check the feasibility of the network design, the research introduces a methodology to estimate the value of information (VOI) provided by the network using a decision tree model. Finally, the research presents the results of a survey administered in the study area to determine whether the implementation of the monitoring network design could be supported.

Applying these methodologies on the Eocene Aquifer, Palestine indicates that statistical learning machines can be most effectively used to design a groundwater quality monitoring network in real-life aquifers. On the other hand, VOI analysis indicates that for the value of monitoring to exceed the cost of monitoring, more work is needed to improve the accuracy of the network and to increase people's awareness of the pollution problem and the available alternatives.

(109 pages)

PUBLIC ABSTRACT

Value of Information in Design of Groundwater Quality Monitoring Network under Uncertainty

by

Abdelhaleem I. Khader, Doctor of Philosophy

Utah State University, 2012

Major Professor: Dr. Mac McKee
Department: Civil and Environmental Engineering

Due to variations in rainfall and limited surface water resources, groundwater is considered the main source for freshwater in many places throughout the word. But this precious resource is being jeopardized by pollution from human activities such as: industry, agriculture, and untreated wastewater, which highlight the need for efficient groundwater resources management. To be efficient, groundwater resources management requires efficient access to reliable information that can be acquired through monitoring. On the other hand, the complicated nature of groundwater aquifers and the uncertainties in the data and the models used to understand the aquifer and its behavior require more powerful and sophisticated tools to handle monitoring problems. Another problem is the limited resources for monitoring, which requires cost-efficient designs.

This research introduces a methodology for groundwater quality monitoring network design that utilizes state-of-the-art learning machines that have been developed from the general area of statistical learning theory. The methodology takes into account uncertainties in aquifer properties, pollution transport processes, and climate. To check the feasibility of the network design, the research introduces a methodology to estimate the value of information (VOI) provided by the network using a decision tree model. Finally, the research presents the results of a survey administered in the study area to determine whether the implementation of the monitoring network design could be supported.

Applying these methodologies on the Eocene Aquifer, Palestine indicates that statistical learning machines can be most effectively used to design a groundwater quality monitoring network in real-life aquifers. On the other hand, VOI analysis indicates that for the value of monitoring to exceed the cost of monitoring, more work is needed to improve the accuracy of the network and to increase people's awareness of the pollution problem and the available alternatives.

To my parents Ibraheem and Amal,

my wife Roba,

and my brothers and sister

Ala Eddin, Abeer, Mohammed, and Basheer

ACKNOWLEDGMENTS

First, I thank God for his countless bounties and endless blessings on me and on my family.

I would like to express my deep gratitude and sincere appreciation to my advisor, Dr. Mac McKee. Since I met him in Nablus on that summer day of 2006, Dr. McKee has been nothing but a great mentor, a helpful supporter, and a close friend. He stood with me in bad times as well as in good times. And I believe I could not have done it without his encouragement, guidance, and moral support.

I really appreciate the help and support of Dr. David Rosenberg especially in the second and third papers. I would also like to thank the other members of my committee, Dr. David Stevens, Dr. Jagath Kaluarachchi, and Dr. Arthur Caplan, for their valuable help.

Finally I would like to express special thanks to my family and friends in Logan and back home in Palestine for their help and love.

Abdelhaleem Khader

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 General Introduction

Groundwater is considered the only reliable source for fresh water in many places throughout the world due to limited rainfall, oftentimes with large variations, and limited surface water resources. In many places, this precious resource is jeopardized by anthropogenic sources of pollution including wastes from agriculture, industry, and municipal discharges which contribute to the degradation of groundwater quality, limit the use of these resources, and lead to health-risk consequences.

In many cases, nitrate is the main pollutant of groundwater. Nitrate pollution may cause health related problems. To address these problems, the need for intensive and efficient management of groundwater has become a necessity. To be effective, groundwater management requires a reliable source of information about the quantity and quality of water available in an aquifer. This information can be acquired through monitoring.

The complicated nature of groundwater aquifers and the uncertainties in the data and models used to understand the aquifer and its behavior require more powerful and sophisticated tools to handle monitoring problems. For this reason, statistical learning machines, which are characterized by their ability to provide predictions of system behavior, have been utilized in this research.

Since information is not free and people have limited resources to pay for it, value of information (VOI) analysis has been conducted on the groundwater quality monitoring design to ensure its economical feasibility. A decision tree model is utilized for this purpose.

Finally, the decision to implement a monitoring system requires the involvement of all stakeholders including the people who are consuming the water. For this purpose, a survey was administered in the study area to infer people's perception about the current situation of groundwater quality and quantity, their expected reactions to a situation where monitoring is implemented, and the implications of that response towards the feasibility of the monitoring network design.

Nitrate pollution and Methemoglobinemia: Agriculture is one of the main culprits in nitrate pollution. Nitrogen is considered a vital nutrient to enhance plant growth, but when nitrogen-rich fertilizer and manure applications exceed the plant demand and the denitrification capacity of soil, nitrogen can leach to groundwater, usually in the form of nitrate (Almasri and Kaluarachchi 2005). Other sources of nitrogen such as septic tanks and dairy lagoons have been shown to contribute to nitrate pollution of groundwater (Almasri and Kaluarachchi 2005; MacQuarrie et al. 2001).

Nitrate has been implicated in Methemoglobinemia (blue baby syndrome) and also inconclusive to a number of other health outcomes. These include proposed effects such as cancer (via the bacterial production of *N*-nitroso compounds), hypertension, increased infant mortality, central nervous system birth defects, diabetes, spontaneous abortions, respiratory tract infections, and changes to the immune system. Although the role of *N*-nitroso compounds and nitrite in the promotion of cancer would appear to be incontrovertible, the evidence relating to the role of nitrates is less clear (Lorna 2004). Thus, Methemoglobinemia is the only health impact that will be considered for discussion in this research.

Methemoglobinemia is caused by decreased ability of blood to carry oxygen, resulting in oxygen deficiency in different body parts. Infants are more susceptible than adults. The disease can be caused by intake of water and vegetables high in nitrate, exposure to chemicals containing nitrate, or can even be hereditary (Majumdar 2003). The toxicity of nitrate in humans is an end result of the reduction of nitrate to nitrite in the intestine by intestinal bacteria. Nitrite reacts with hemoglobin to form methemoglobin (MHb), a substance that does not bind and transport oxygen to tissues, thereby causing asphyxia (lack of oxygen), resulting in cyanosis of body tissues. Infected infants show blueness around the mouth, hands, and feet and is the reason for the common name 'blue baby syndrome' (Majumdar 2003). The most common treatment for methemoglobinemia is methylene blue. This treatment converts MHb to hemoglobin and gives immediate relief. Other treatments will depend on the severity of the case and could include ascorbic acid, vitamins C and E, emergency exchange blood transfusion, and administration of high flow oxygen (Majumdar 2003).

Monitoring Network Design: In many aquifers the need for intensive groundwater resources management has become urgent to address nitrate pollution. More intensive management will require greater investment in monitoring.

The design of groundwater monitoring networks entails the selection of sampling points (spatial) and sampling frequency (temporal) to determine physical, chemical, and biological characteristics of groundwater (Loaiciga et al. 1992). In designing a monitoring network, special consideration must be given to: spatial and temporal coverage of the monitoring sites; possibly competing objectives of the monitoring program; complex nature of geologic, hydrologic, and other environmental factors; stochastic character of transport parameters (geologic, hydrologic, environmental) used in the design process; and risk posed to society (failure to detect, poor characterization, etc.) (Asefa et al. 2005).

Groundwater monitoring networks can be categorized based on the design objective as (Asefa et al. 2005): (1) leak detection; (2) characterization, and (3) long-term monitoring. Furthermore, network design is typically an iterative process, whereby the sampling program must be revised and updated in response to changes in information needs and information gathered through time from the data (Loaiciga et al. 1992).

Learning Machines: The complicated physical, chemical, and biological characteristics of groundwater aquifers, together with our limited and imprecise knowledge of them, present serious challenges for groundwater quality monitoring network design. Another challenge is the variability and uncertainty in climate conditions, pollutant interactions, and future human activities. All of these uncertainties present a challenge to our ability to monitor and manage groundwater quality. State-of-the-art learning machines have been utilized as modeling tools in recent years (Asefa et al. 2004, 2005, 2006; Gill et al. 2006; Khalil et al. 2005a, 2005b; Ticlavilca 2010; Zaman 2010) to address these problems and the uncertainties they present. Learning machines are data driven methods which are characterized by their ability to quickly capture the underlying physics and provide predictions of system behavior (Khalil et al. 2005a) when presented with sufficient data describing system inputs and outputs. Some machines are also able to capture information about the uncertainty in both data and output.

Khalil et al. (2005a) utilized four learning machines as surrogates for a relatively complex and time-consuming mathematical model to simulate nitrate concentration in groundwater at specified receptors. The algorithms are: artificial neural networks (ANNs), support vector machines (SVMs), locally weighted projection regression (LWPR), and relevance vector machines (RVMs). Their prediction results showed the ability of learning machines to build accurate models with strong predictive capabilities and hence constitute a valuable means for saving effort in groundwater contamination modeling and improving model performance. Moreover, the results proved that the RVM is efficient in producing an excellent generalization level while maintaining the sparsest structure.

Sometimes the objective of monitoring network design is to reduce the redundancy in an existing large monitoring network. For this objective, redundancy reduction, Ammar et al. (2008) introduced a methodology based on the application of relevance vector machines. The methodology was employed to reduce redundancy in the network for monitoring nitrate in the West Bank, Palestine. The results indicate that only 32% of the existing monitoring sites in the aquifer are sufficient to characterize the nitrate state without increasing the uncertainty in the characterization, and the other wells are redundant.

This research addresses the monitoring problems by extending a methodology that is based on Bayesian modeling approaches from statistical learning theory. This methodology uses a RVM model that captures the uncertainties in data and predictions about possible present or future aquifer conditions, and does so with a sparcity in model formulation that yields efficiency in the network design.

Value of information (VOI) analysis: Monitoring can be expensive, so at some level the monitoring system must be efficient, as well as dependable, in providing information about the condition of the aquifer.

Information is not free. Money and time are needed to search for and acquire (Sakalaki and Kazi 2006). VOI analysis evaluates the benefit of collecting additional information to reduce or eliminate uncertainty associated with the outcome of a decision. VOI makes explicit any expected potential losses from errors in decision-making due to uncertainty and identifies the "best" information collection strategy as one that leads to the greatest expected net benefit to the decision-maker (Yokota and Thompson 2004a).

To estimate how rational[1] individuals should value the information, expected utility theory provides a normative of information valuation (Delquié 2008). In economics, utility is a real-valued function that reflects consumer satisfaction from receiving a good or service. Expected utility (EU) is the probability-weighted average of the utility from each possible outcome (Perloff 2008).

Expected utility theory can be supported by a decision tree model (Fig. 1.1) that describes the logical structure of the decision. Each tree branch represents a different choice or outcome (Lund 2008). Boxes denote choice nodes, where a decision must be made. Circles denote chance nodes, where outcomes are uncertain. Each branch emanating from a choice node is an alternative, and each branch emanating from a chance node is a possible outcome, with a probability attached. The consequence of each outcome is shown at the far right of the tree. In Fig. 1.1, the decision maker (DM) is deciding whether to make uninformed decisions (Branches 1 or 2) or acquire more information about a system in order to make a more informed decision (Branch 3).

The VOI is measured *ex-ante* as the difference between the EUs of the informed and uninformed branches (Delquié 2008; LaValle 1968). For public policy decisions where consequences are small compared to the scale of the overall enterprise, we can substitute expected value (EV) for EU (Arrow and Lind 1970).

[1] Rational individuals: those who are balancing cost against benefits to arrive at action that maximizes personal advantage (Friedman 1966).

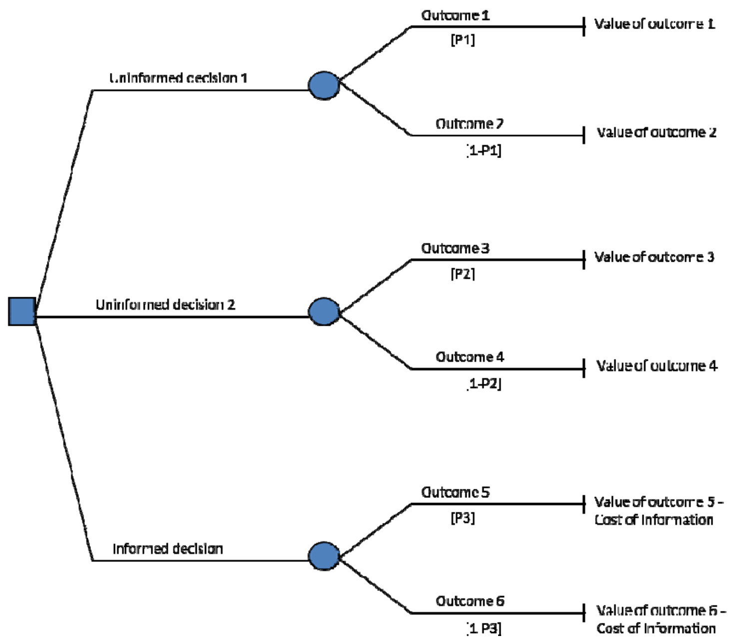


Fig. 1.1 Decision tree example that shows the structure of the tree and the different options

Expected value of each branch is the weighted average of the values of each outcome from that branch. The weights here correspond to the probabilities of each outcome. In this case the VOI is the difference between the EVs of the informed and uninformed branches. To acquire more information, the VOI for the informed decision must exceed the cost of acquiring information.

Willingness to pay (WTP) is another widely used method for VOI estimation (Alberini et al. 2006; DeShazo and Cameron 2005; Dickie and Gerking 2002; Engle-Warnick et al. 2009; Latvala and Jukka 2004; Molin and Timmermans 2006; Roe and Antonovitz 1985; Sakalaki and Kazi 2006). WTP can be defined as the maximum amount a person or a DM is willing to pay in order to receive a good or to avoid something undesirable (Perloff 2008). In this method contingent valuation surveys should be conducted to ask individuals how much are they actually willing to pay (for information in this case) (Alberini et al. 2006; Atkins et al. 2007; Pattanayak et al. 2003). Although this WTP analysis can estimate how much people are actually willing to pay to acquire more information, it can only be done by actually asking people who should be well informed about the problem. On the other hand, the cheaper and easier EU method can estimate how rational people should value information which is sufficient for VOI analysis.

1.2 Objectives

The purpose of the research is to develop a methodology for groundwater quality monitoring network design that is reliable and efficient. The objectives of the research are to:

1. Introduce a methodology for groundwater quality monitoring network design that takes into account the uncertainties in aquifer properties, climate, and pollutant reaction process. This methodology uses groundwater flow modeling, pollutant fate and transport modeling, Monte Carlo simulations, and RVMs to design an optimal[1] monitoring network in terms of the number of monitoring sites and their locations.
2. Estimate the value of information in design of groundwater quality monitoring networks using decision tree analysis.
3. Study the implications of social aspects[2] of a groundwater quality monitoring network design on the feasibility of the design.

---

[1] The optimality here comes from the nature of the RVM model which provides a sparse solution that avoids over-fitting.

[2] Social aspects here refer to the responses of individuals to different design options

1.3 Study Area (the Eocene Aquifer)

The Eocene Aquifer, Palestine, is an unconfined aquifer located in the northern part of the West Bank (Fig. 1.2). The total area of the aquifer is 526 km^2. The geological formation consists mainly of carbonate rocks of limestone and chalky limestone with thickness ranging from 300 to 500 m. The annual rainfall in the area ranges from 400 to 642 mm and the estimated recharge from rainfall ranges from 45 to 65 mcm/yr.

The Eocene Aquifer is used to meet domestic and agricultural demands for 207,000 Palestinians living in 27 communities (Fig. 1.2). The water is obtained from wells and springs. There are 67 wells located within the Eocene aquifer boundary (Fig. 1.2). The annual long-term average abstraction from the Eocene aquifer is about 18.2 mcm. The wells are owned by municipalities or private farmers. There are 25 springs in the aquifer that have a total annual average discharge of about 10.4 mcm (Kharmah 2007; Najem 2008; Tubaileh 2003).

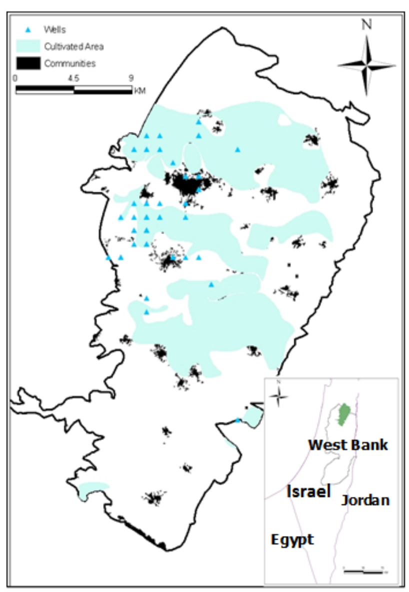


Fig. 1.2 Palestinian communities, abstraction wells, and cultivated areas in the Eocene Aquifer boundaries

Nitrate is the main pollutant in the Eocene Aquifer. The main sources of nitrate pollution in the aquifer are the excessive use of nitrogen-rich fertilizers and the lack of sewer networks[1] (Najem 2008).

1.4 Contribution

This research is presenting a method that uses groundwater flow modeling, pollutant fate and transport modeling, Monte Carlo simulations, RVMs, VOI analysis, and survey statistics all together for the purpose of monitoring network design.

This is performed by introducing a new methodology for groundwater quality monitoring network design that takes into account the uncertainties in aquifer properties, pollutant reaction processes, and climate and estimates the economic and social value of information. This methodology could serve as a tool for decision-makers to design new optimal monitoring networks or to assess existing ones in terms of redundancy. What is new about this methodology is the capability of designing brand-new monitoring network and testing the feasibility of the design in a Probabilistic framework.

The technical contributions of this research include:

1. A methodology that uses a relevance vector machine for groundwater quality monitoring network design under uncertainty, and the application of that methodology to the Eocene Aquifer, Palestine.
2. A Bayesian framework for estimating the value of information from a groundwater quality monitoring network using a decision tree model.

[1] Wastewater treatment and better agricultural practices could mitigate nitrate pollution problem in the long run, but these solutions are expensive and beyond the scope of this study which focusses on groundwater quality monitoring network design.

3. Studying the implications of public involvement on implementing the groundwater quality monitoring network design.

1.5 Dissertation Organization

Chapter 2 presents a methodology for groundwater quality monitoring network design that takes into account uncertainties in aquifer properties, pollution transport process, and climate using a relevance vector machine. Chapter 3 presents a methodology to estimate the value of information provided by a groundwater quality monitoring network using a decision tree model. Chapter 4 presents the results of a survey that was administered to people living in the study area to support a decision to implement a groundwater quality monitoring network. Finally, Chapter 5 summarizes the results of the research and presents the conclusions and recommendations.

The structure of this document follows the multiple-paper dissertation format. As a result, the reader might find some redundancies and repetition of materials, especially in the background and the description of the study area.

CHAPTER 2

USE OF A RELEVANCE VECTOR MACHINE FOR GROUNDWATER QUALITY MONITORING NETWORK DESIGN UNDER UNCERTAINTY[1]

Abstract

This paper presents a methodology for groundwater quality monitoring network design that takes into account uncertainties in aquifer properties, pollution transport processes, and climate. The methodology utilizes a statistical learning algorithm called a relevance vector machine (RVM), which is a sparse Bayesian framework that can be used for obtaining solutions to regression and classification tasks. Application of the methodology is illustrated using the Eocene Aquifer in the northern part of the West Bank, Palestine. The procedure presented in this paper captures the uncertainties in recharge, hydraulic conductivity, and nitrate reaction processes through the application of a groundwater flow model and a nitrate fate and transport model following a Monte Carlo (MC) simulation process. This MC modeling approach provides several thousand realizations of nitrate distribution in the aquifer. Subsets of these realizations are then used to design the monitoring network. This is done by building a best-fit RVM model of nitrate concentration distribution everywhere in the aquifer for each Monte Carlo subset. The outputs from the RVM model are the distribution of nitrate concentration everywhere in the aquifer, the uncertainty in the characterization of those concentrations, and the number and locations of "relevance vectors" (RVs). The RVs form the basis of the optimal characterization of nitrate throughout the aquifer and represent the optimal locations of monitoring wells. In this paper, the number of monitoring wells and their locations where chosen based on the

---

[1] Coauthored by Abdelhaleem Khader and Mac McKee

performance of the RVM model runs. The results from 100 model runs show the consistency of the model in selecting the number and locations of RVs. After implementing the design, the data collected from the monitoring sites can be used to estimate nitrate concentration distribution throughout the entire aquifer and to quantify the uncertainty in those estimates.

2.1 Introduction

Due to large variations in rainfall and limited surface water resources, groundwater is considered the sole reliable source of fresh water in many places in the world. Anthropogenic sources of pollution such as agriculture, industry, and production of municipal waste, contribute to the degradation of groundwater quality, which may limit the use of these resources and lead to health-risk consequences. For these reasons, the need for intensive groundwater resources management has become more urgent. To become more effective, groundwater resources management requires a reliable information system to provide data about the system being managed. However, monitoring can be expensive, so at some level the monitoring system must be economically efficient, as well as dependaple, in providing information about the condition of the aquifer.

The design of groundwater pollution monitoring networks entails the selection of sampling points (spatial) and sampling frequency (temporal) to determine physical, chemical, and biological characteristics of groundwater (Loaiciga et al. 1992). In designing a monitoring network, special consideration must be given to: spatial and temporal coverage of the monitoring sites; potentially competing objectives of the monitoring program; the complex nature of geologic, hydrologic, and other environmental factors; the stochastic character of transport parameters (geologic, hydrologic, environmental) used in the design process; and the risk posed to society (failure to detect, poor characterization, etc.) (Asefa et al. 2005). Monitoring networks can be categorized on the basis of the design objective they are to address (Asefa et al. 2005) (1) leak detection; (2) characterization; or (3) long-term monitoring.

The complicated physical, chemical, and biological characteristics of groundwater aquifers present serious challenges for groundwater quality monitoring network design. Another challenge is the variability and uncertainty in climate conditions, pollutant interactions, and future human activities. All of these uncertainties present a challenge to our ability to monitor and manage groundwater quality. State-of-the-art statistical learning machines have been utilized in recent years to address these problems (Asefa et al. 2004, 2005, 2006; Gill et al. 2006; Khalil et al. 2005a, 2005b; Ticlavilca 2010; Zaman 2010). Statistical learning machines are characterized by their ability to capture the underlying physics of the system to be modeled and provide predictions for system behavior (Khalil et al. 2005a).

Khalil et al. (2005a) utilized four statistical learning algorithms as surrogates for a relatively complex and time-consuming mathematical model to simulate nitrate concentration in groundwater at specified receptors. The algorithms are: artificial neural networks (ANNs), support vector machines (SVMs), locally weighted projection regression (LWPR), and relevance vector machines (RVMs). Their prediction results showed the ability of learning machines to build accurate models with strong predictive capabilities and hence constitute a valuable means for saving effort in groundwater contamination modeling and improving model performance. Moreover, the results proved that the RVM is efficient in producing an excellent generalization level while maintaining the sparsest structure.

Sometimes the objective of monitoring network design is to reduce the redundancy in an existing large monitoring network. For this objective, redundancy reduction, Ammar et al. (2008) introduced a methodology based on the application of relevance vector machines. The methodology was employed to reduce redundancy in the network for monitoring nitrate in the West Bank, Palestine. The results indicate that only 32% of the existing monitoring sites in the aquifer are sufficient to characterize the nitrate state without increasing the uncertainty in the characterization and the other wells are redundant.

This paper addresses groundwater monitoring problems by extending a methodology that is based on Bayesian modeling approaches from statistical learning theory. This methodology uses a RVM model that captures the uncertainties in data and predictions about possible present or future aquifer conditions, and does so with a sparcity in model formulation that yields efficiency in the network design. The conceptual framework of the paper proceeds by first quantifying the uncertainties in recharge, hydraulic conductivity, and nitrate reaction processes by applying conventional groundwater flow and nitrate fate and transport models in a Monte Carlo (MC) simulation process. After that, an optimal monitoring network that takes into account the uncertainties revealed in the MC simulations is designed by developing the RVM model.

The conceptual framework is discussed in Section 2.2 followed by a brief description of the study area in Section 2.3. After that, the model development is presented in Section 2.4. Model results are discussed in Section 2.5 followed by concluding remarks in Section 2.6.

2.2 Conceptual Framework

The conceptual framework of the approach, illustrated in Fig. 2.1, is divided into three modules (1) uncertainty analysis, (2) groundwater flow and fate and transport modeling, and (3) monitoring network design.

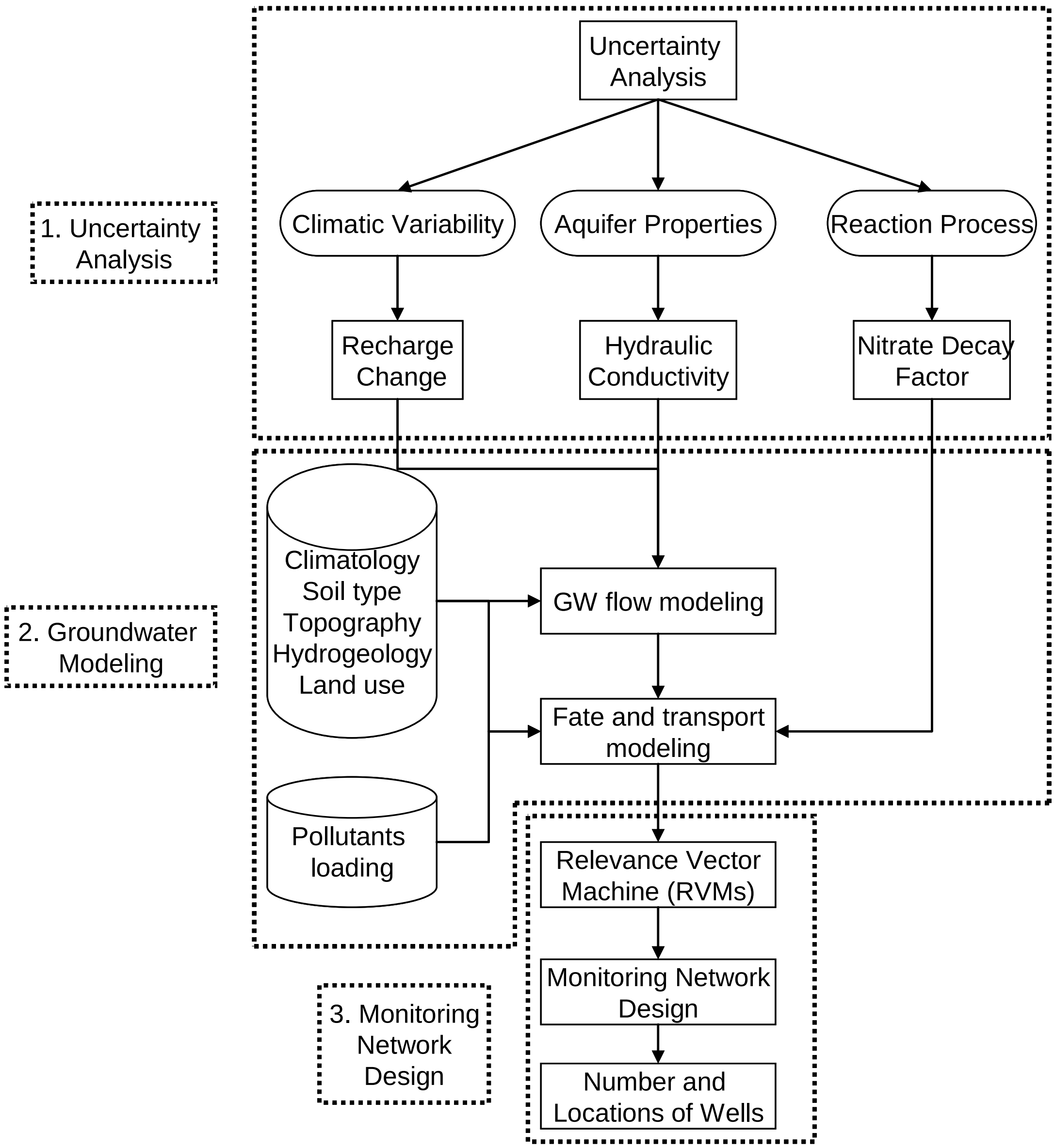


Fig. 2.1 Schematic of the proposed conceptual framework of this study

2.2.1 Module 1: Uncertainty Analysis

Groundwater flow and fate and transport modeling require sufficient data about the input parameters of hydrology and physical properties of the aquifer that is being modeled and the characteristics of pollutants under concern. In this module, the input parameters that contribute to uncertainty of the condition of the physical groundwater system are analyzed. Variability in recharge comes from climatic variability. Climate change is another factor that could add to the uncertainty of the future behavior of aquifer systems, and as a result of climate change, Palestine is among the regions in which drier climates have been observed and are expected to increase (Meehl et al. 2007). Variations in geologic materials and processes result in highly spatially variable hydraulic properties. This variability adds to the uncertainty of the state of groundwater flow and piezometric head. The last input parameter to analyze is the nitrate decay factor which represents uncertainty in nitrate chemical reaction process. The probability distributions of these input parameters are acquired from the histograms in Fig. 2.2. These histograms were obtained from available data in the literature (Kharmeh 2007; Najem 2008; Tubaileh 2003).

Due to the lack of sufficient data about uncertainty in recharge, spatial and temporal variability in precipitation are used to estimate the probability distributions of recharge variability.

There are other factors that may add to uncertainty such as the uncertainty in future human activities and their impact on nitrate loading. However, due to the lack of data, these factors are not analyzed and will be kept for future work. Their deletion from this analysis does not detract from the development of the approach for monitoring network design.

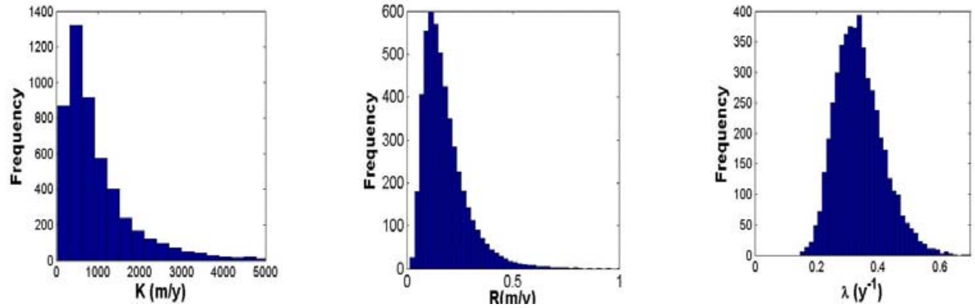


Fig. 2.2 Histograms of hydraulic conductivity (K), recharge (R), and nitrate decay factor (λ)

2.2.2 Module 2: Groundwater Flow and Transport Modeling

In the second module, groundwater flow is modeled using MODFLOW (Harbaugh et al. 2000), and a fate and transport model, MT3DMS (Zheng and Wang 1998), is utilized to simulate the nitrate concentration in the Eocene Aquifer.

Groundwater Flow Modeling: MODFLOW-2000 (Harbaugh et al. 2000) is a computer program that simulates three-dimensional ground-water flow through a porous medium by using a finite-difference method. Kharmah (2007) developed a steady-state groundwater flow model for the Eocene Aquifer using MODFLOW-2000. This study uses that model to simulate the impact of uncertainty in aquifer properties (hydraulic conductivity) and climatic variability (recharge) on groundwater flow in the Eocene Aquifer.

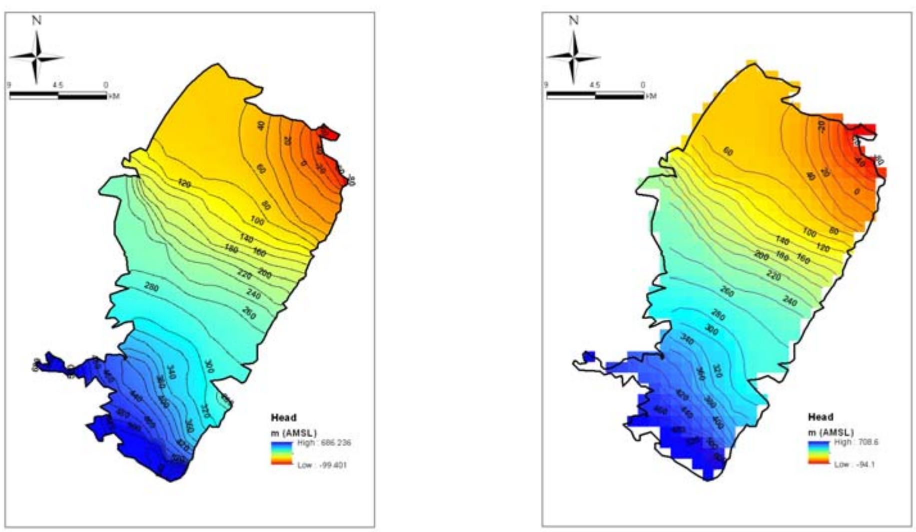
The data needed for the MODFLOW model were obtained from the data bases of the Palestinian Water Authority (PWA) and the British Geological Survey (BGS). The model domain was divided into a 100 m by 100 m finite-deference grid. The total number of cells (active and inactive) is 111,168. The number of active cells is 52,495. Based on the aquifer stratigraphy, the vertical discretization of the model consists mainly of one layer, which represents the Eocene formation. The simulated system is therefore represented as a single-layer two-dimensional groundwater flow situation.

The top layer elevation ranges from 50 m to 950 m above mean sea level, while the bottom layer elevation ranges from -200 m to -600 m.

Two types of boundary conditions were used: a general-head boundary to represent the faulting system in the northeast, and a no-flow boundary to represent the other boundaries which are structurally separated from the other formations in the area. All springs in the area were modeled using the DRAIN package in MODFLOW.

The main sources of recharge are rainfall (93%) and return flow from irrigation and water supply network losses (7%) (Kharmeh 2007). On the other hand, the main sources of discharge are abstraction wells (22%), springs (12%), and the general-head boundary on the northeast side of the aquifer (66%) (Kharmeh 2007).

The model was calibrated using available data. The calibration process was done by tuning the hydraulic conductivity and the transmissivity until the simulated groundwater elevations and spring flows approximated the observations. Fig. 2.3 shows the groundwater elevations that result from the calibrated model.



(a) (b)

Fig. 2.3 Groundwater elevations distribution with contour lines from the calibrated MODFLOW-2000 Model in 100 m grid cell size (a) and 1000 m grid cell size (b).

In this study, the MODFLOW model was recalibrated based on more recent data available from the Palestinian Water Authority (PWA). After that, the model grid was resized into a 1000 m by 1000 m grid. This was done for the purpose of reducing the time needed to run Monte Carlo simulations and the RVM model (discussed in the next section). Fig. 2.3 shows that resizing the grid did not change the resulting groundwater elevation values.

Nitrate Fate and Transport Modeling: This study uses a quasi-steady-state nitrate fate and transport model for the Eocene Aquifer developed by Najem (2008) using MT3DMS (Zheng and Wang 1998). MT3DMS is a modular three-dimensional multispecies transport model for simulation of advection, dispersion, and chemical reactions of contaminants in groundwater systems.

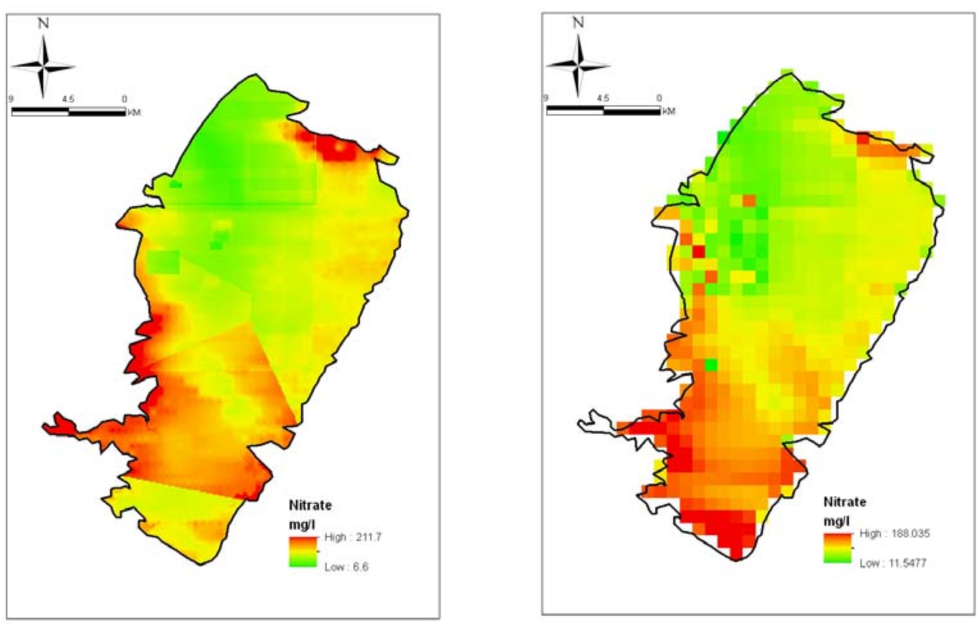
The first step in the development of this fate and transport model was to analyze the on-ground nitrogen loading to the aquifer. The principal sources of nitrogen are: fertilizers, cesspits, atmospheric deposition, and mineralization of soil organic matter. The next step was to estimate the net nitrogen mass that reaches the groundwater after allowing for transformations in the soil, and then modeling the nitrogen fate and transport in the groundwater.

MT3DMS does not have a groundwater flow component, but it has a package that can link the transport model with MODFLOW. Najem (2008) linked his model with the MODFLOW model developed by Kharmah (2007) (see the previous section). In this case, the model discretization was the same as in the MODFLOW model, i.e., a 100 m by 100 m finite deference grid.

Finally, the model was calibrated under quasi-state conditions. The calibration process was performed by refining the model parameters (nitrate decay factor in this case) so that the simulated nitrate concentrations approximate the observed ones. Fig. 2.4 shows the resulting nitrate concentrations from the calibrated model.

As with the case of the groundwater flow model, the fate and transport model grid was also resized in this study 1000 m by 1000 m grid to reduce the run time needed for Monte Carlo simulations and RVM modeling.

Monte Carlo simulations: Ten thousand Monte Carlo simulations are used to describe the effects of the uncertainty in the abovementioned input parameters as indicated in Fig. 2.5. The outputs from this module are 10,000 instances of the spatial distributions of groundwater heads and nitrate concentrations that take into account the variability in the input parameters. These distributions will be used later in the RVM model (Module 3). Fig. 2.6 shows the resulting mean and variance of nitrate concentrations from the Monte Carlo simulations.



(a) (b)

Fig. 2.4 Expected nitrate concentrations spatial distribution in 100 m grid cell size (a) and 1000 m grid cell size (b).

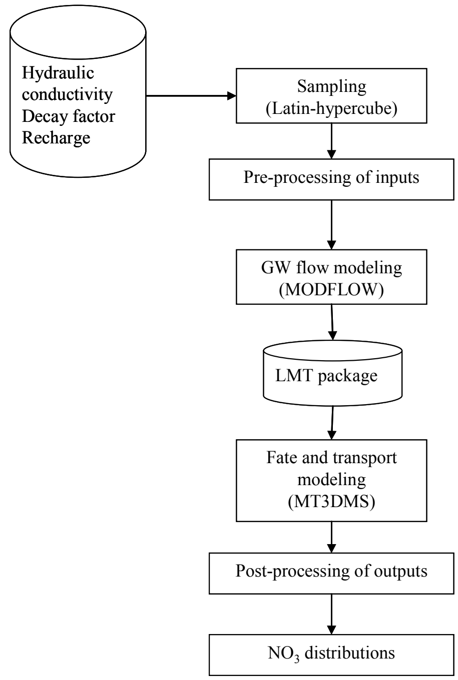
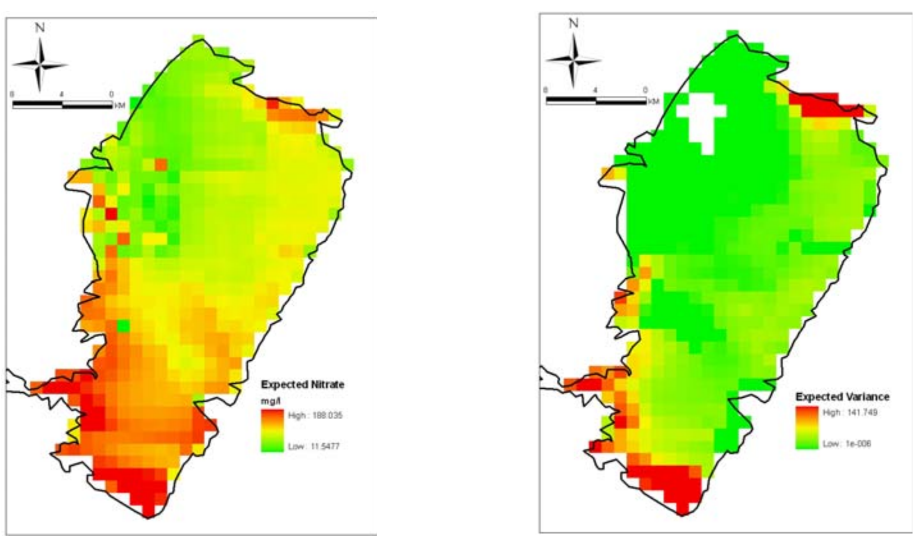


Fig. 2.5 Schematic of Monte Carlo simulations



(a) (b)

Fig. 2.6 Expected nitrate concentration (a) and variance (b) from the 10,000 Monte Carlo simulations

2.2.3 Module 3: Monitoring Network Design

In the third module, an optimal monitoring network that takes into account the uncertainties in the input parameters is designed by utilizing RVMs. The use of RVM modeling in this research is motivated by the fact that many studies have shown that RVMs very often perform better than other statistical learning machines (Ammar et al. 2008; Khalil et al. 2005a; Tipping 2001). The main strength of RVMs is their ability to generate sparse models and to infer information about relationships between inputs and outputs contained in the data because of their Bayesian framework. In particular, RVM models capture both model and data uncertainty and, as a result, lend themselves to characterization of the behavior of nonlinear systems for which uncertainty is of key interest. The theoretical background of RVM modeling is shown in Appendix 1.

The input to the RVM (Fig. 2.7) consists of all the possible locations of monitoring wells. The model output represents the corresponding nitrate concentrations acquired from the Monte Carlo simulations. The RVM model discovers the non-linear relationships between the inputs and the outputs and finds the locations where monitoring can be done that are most relevant for prediction of nitrate concentrations everywhere in the aquifer (hence the name “relevance vector machine”).

Unlike Ammar et al. (2008), which only considered reduction of unnecessary wells from an existing network, the methodology proposed here allows the design of a new monitoring network. This was made possible by the use of distributed groundwater flow and nitrate fate and transport models (Module 2) and the Monte Carlo simulations based on the uncertainties from Module 1.

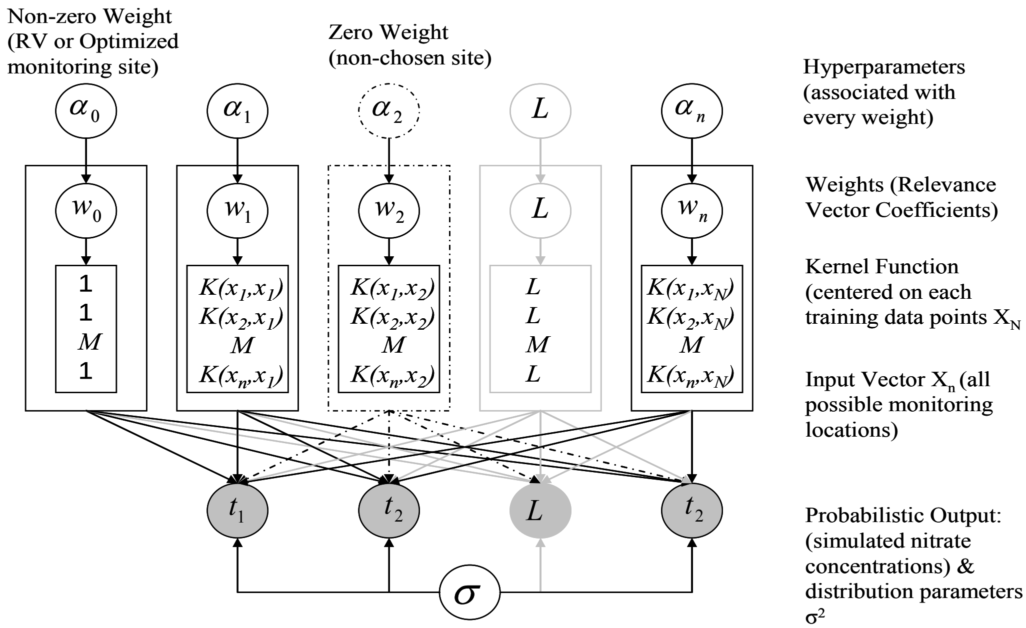


Fig. 2.7 RVM model inputs, outputs, parameters, and hyperparameters. For more details see Appendix 1 (Ammar et al. 2008)

2.3 Study Area -The Eocene Aquifer

Due to the large variations in rainfall and limited surface water resources, groundwater is considered the sole reliable source of water in Palestine. There are three groundwater basins in the West Bank (Abu Zahra, 2001): The Western Basin, The Northeastern Basin, and the Eastern Basin.

This research focuses on the Eocene Aquifer which is located within the Northeastern Basin. It is referred to as the Jenin sub-series (see Fig. 2.8). The geological formation consists mainly of carbonate rocks of limestone and chalky limestone with thickness ranging from 300-500 m (Tubaileh 2003).

There are three major soil associations:

1. Terra Rossa, Brown Rendzinas, and Pale Rendzinas (63%)
2. Brown Rendzinas, and Pale Rendzinas (9%)
3. Grumsols (28%)

In terms of climate, the area falls in the Mediterranean climate zone in which two climatic seasons are defined, a wet winter and a dry summer. The winter extends from October to May. The annual average rainfall in the study area varies sharply from 600 mm to 150 mm, and the average number of rainy days per year ranges from 25 to 60. The estimated recharge from rainfall ranges from 45 to 65 mcm/yr.

In winter, the minimum temperature is around 7 ºC and the maximum is 15 ºC. Temperatures below the freezing point are rare. In summer, the average maximum temperature is 33 ºC and the average minimum is 20 ºC. Evaporation ranges from 1850 mm to 2100 mm (Kharmah 2007).

The Eocene Aquifer is used to meet domestic and agricultural demands for 128,000 Palestinians living in 66 communities, about 51,000 of them are living in the City of Jenin (Fig. 2.8).

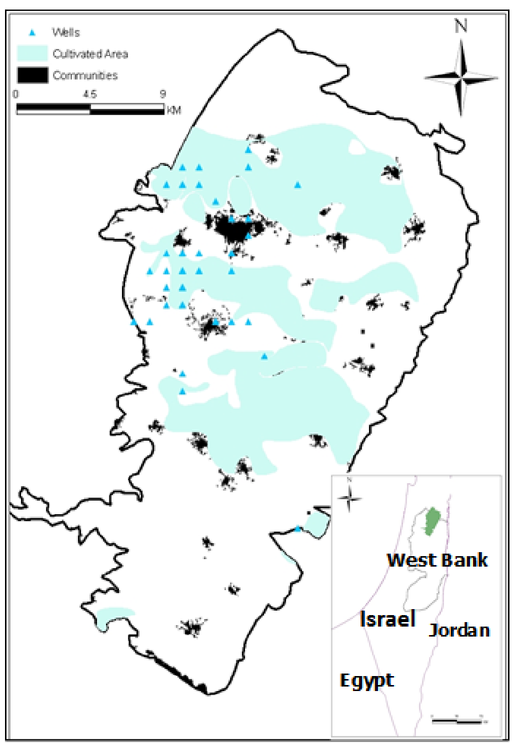


Fig. 2.8 Palestinian communities, abstraction wells, and cultivated areas in the Eocene Aquifer boundaries

Water is obtained from wells and springs. There are 67 wells located within the Eocene aquifer boundary (Fig. 2.8). The annual long-term average abstraction from the Eocene aquifer is about 18.2 mcm. Wells are owned by municipalities or privately by farmers. There are 25 springs in the aquifer that have a total annual average discharge of about 10.4 mcm (Kharmah 2007; Najem 2008; Tubaileh 2003).

2.4 Model Development

2.4.1 Model Inputs and Outputs

As stated previously, the inputs to the RVM model consist of all possible monitoring locations. In other words, the inputs are the x- and y-coordinates of the centers of each active cell in the model domain. The targets are the nitrate concentrations in each cell acquired from the Monte Carlo simulations. This means that we have a distribution from the Monte Carlo simulations of 10,000 nitrate concentration values for each cell. The available RVM modeling tools cannot handle a problem of this size, so to deal with this large number of targets 100 RVM model runs were performed. In each of these runs, the targets were 100 nitrate concentration values for each cell randomly sampled from the total population by keeping the spatial correlation between cells. This process is illustrated in Fig. 2.9. The output of each of these 100 runs was the optimal location of monitoring wells, a model that could use data from those locations to predict nitrate concentration everywhere in the aquifer, and information about the uncertainty that would result from those predictions.

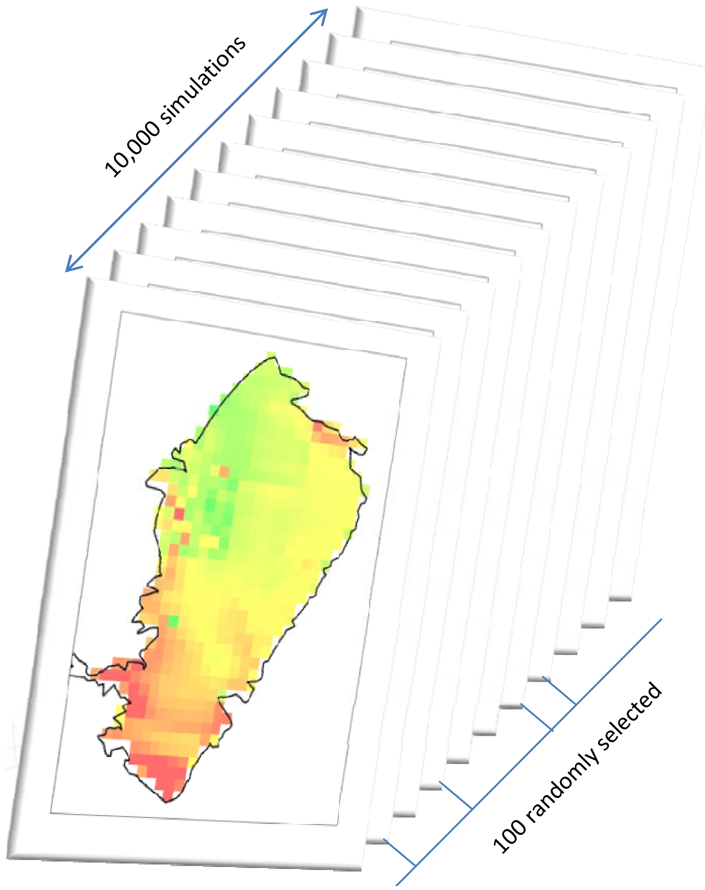


Fig. 2.9 Randomly selecting 100 maps out of 10,000 in each run

2.4.2 Model Calibration

Two parameters are needed to calibrate the RVM model: the kernel type and the kernel width. Fig. 2.10 shows the model performance using different kernel types. Two criteria were used to evaluate this performance: root mean square error (RMSE) (Armstrong and Collopy 1992) and the Nash-Sutcliffe coefficient of efficiency (E) (Nash and Sutcliffe 1970). The model performance is considered to be better if RMSE is low and E is high. Fig. 2.10 indicates that the Laplace kernel has the best performance for both criteria. Therefore, the Laplace kernel type is adopted for use in this case study. After selecting the kernel type, the model performance was tested again under different kernel width (w) conditions. Fig. 2.11 shows the results of these tests. Based on both criteria, the best performance is when w equals 0.3.

2.5 Results and Discussion

After calibrating the model, it was run 100 times as described in the previous section. Fig. 2.12 displays the frequency of how many times each cell was chosen to be the location of an RV in all of the 100 runs. It is clear from Fig. 2.12 that some cells were chosen over and over again, which indicates the consistency of the model. To select the best set of monitoring locations the 100 runs where investigated to find the run in which the RVM has the best performance. Fig. 2.13 shows the performance of the model in these runs based on the RMSE performance criteria. Since the objective here is to have lower RMSE, Run 57 is chosen for design because it satisfies this objective. Fig. 2.14 shows the locations of the cells that were chosen as RVs in Run 57. This indicates that a RVM model based on RVs located at these cells is optimal in terms of representing nitrate distribution in the aquifer. This means that the RV locations are the most suitable for groundwater quality monitoring.

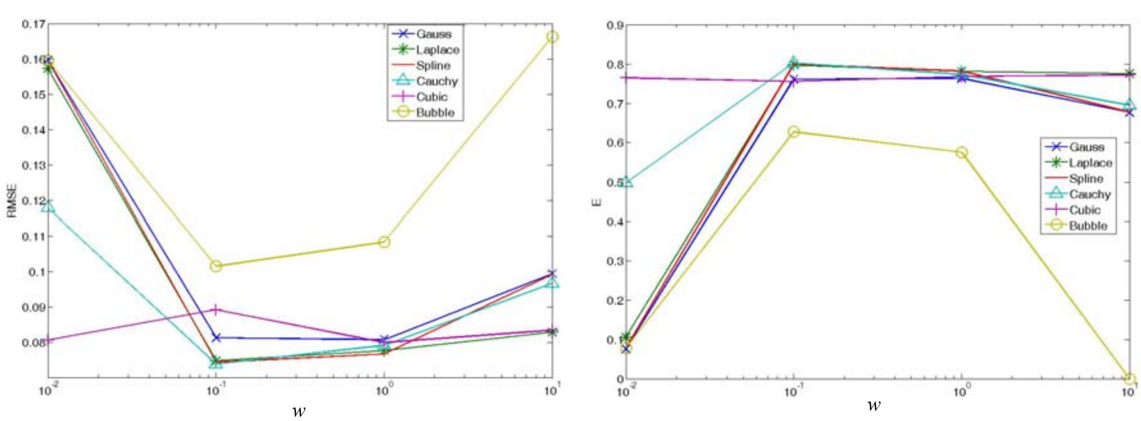


Fig. 2.10 Kernel type selection based on RMSE (left) and E (right)

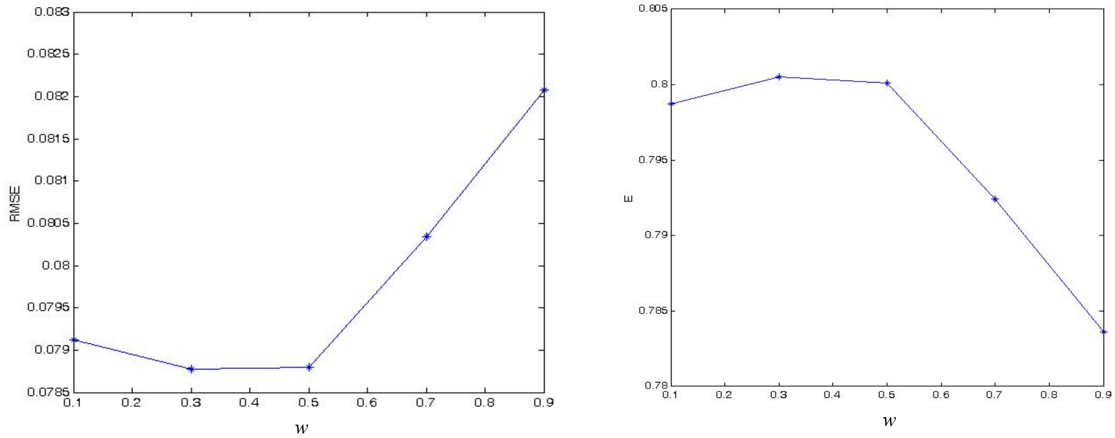


Fig. 2.11 Kernel width selection based on RMSE (left) and E (right)

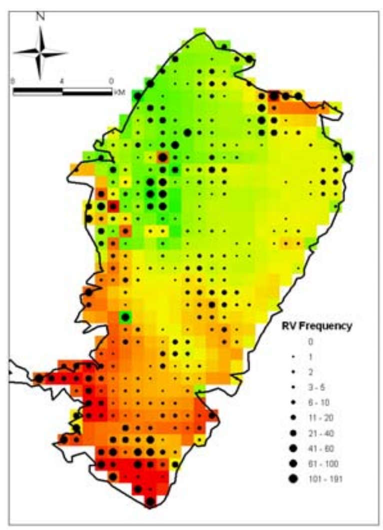


Fig. 2.12 RVM results

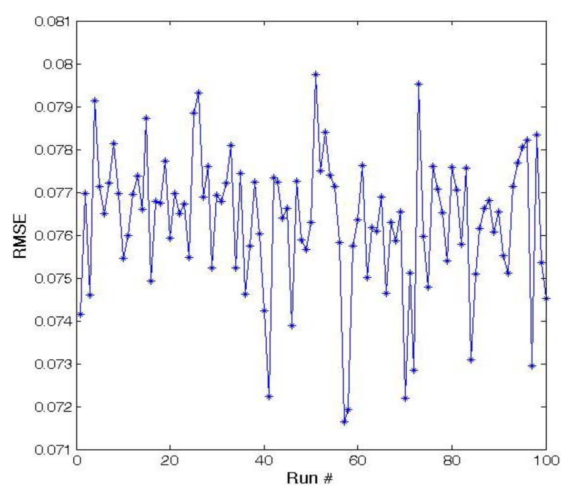


Fig. 2.13 RMSE values for the 100 runs

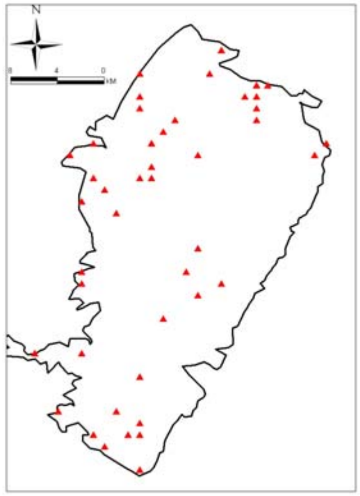


Fig. 2.14 Locations of the RV's in Run# 57

To test the functionality of the designed network, a hypothetical scenario is introduced in which a 5% increase in nitrate concentration is placed in each of the monitoring sites. Using only this information (the updated concentration in the monitoring locations) the RVM model is able to estimate the distribution of nitrate concentration all over the aquifer and to characterize the uncertainty in that distribution estimate. Fig. 2.15 shows the difference in expected nitrate concentration and variance in the aquifer as estimated by the RVM model.

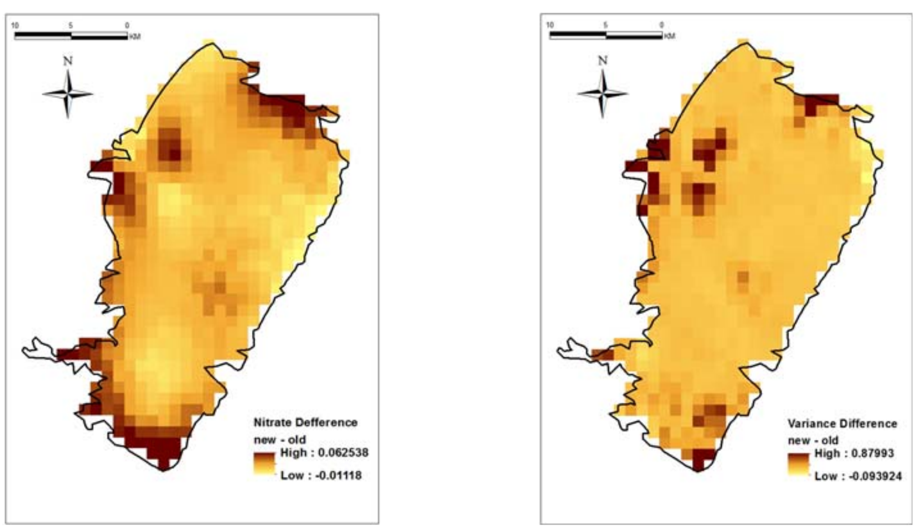


Fig. 2.15 Difference in expected nitrate concentration (left) and variance (right) after increasing nitrate by 5% in the monitoring locations

2.6 Conclusions

The purpose of this paper is to introduce a new methodology for groundwater quality monitoring network design that takes into account uncertainties in climate and aquifer properties. The paper has shown that groundwater flow modeling and pollutant fate and transport modeling can be used to quantify the uncertainties in the inputs through MC simulation method. It has also shown that RVM modeling is a powerful tool that can be used in monitoring network design. The main advantage of RVMs here is their ability to capture the uncertainty in the data and the model due to their Bayesian nature. Because of their sparse nature, we are able to design a monitoring network with the fewest number of monitoring locations.

Limitations to this methodology include the computational effort needed to run the RVM model. A MicrowayWhisperstation (http://www.microway.com/whisperstation) with 24 cores and 64 GB of RAM was utilized. Even that powerful machine was not able to run the RVM model with all realizations acquired from the MC simulations and so a random sampling approach was needed to characterize the distribution of possible RVM solutions. Another limitation is the extensive need for data about the inputs to the flow and fate and transport models. Future work could include examination of other sources of uncertainty such as human activities and on-ground nitrate loading. These sources of uncertainty can be incorporated in the model by sampling from their distributions in the MC simulation process (Fig. 2.5). A significant improvement could be the addition of a temporal dimension to the monitoring network, i.e. the sampling frequency. But adding this option is conditioned on the availability of temporal data about nitrate pollution.

CHAPTER 3

DECISION TREE MODEL FOR ESTIMATING THE VALUE OF INFORMATION PROVIDED BY A GROUNDWATER QUALITY MONITORING NETWORK[1]

Abstract

This paper presents a methodology to estimate the value of information provided by a groundwater quality monitoring network located in an aquifer whose water poses an uncertain health risk. A decision tree model describes the structure of the decision alternatives facing the decision maker (DM) and the expected outcomes from these alternatives. This model is used to estimate the value of information (VOI) of designing and implementing the monitoring network. There are three alternatives to choose from: (i) "do nothing" alternative which ignores the pollution problem, (ii) "not using the aquifer" alternative, and (iii) "monitoring network" alternative. VOI is estimated by evaluating the expected value (EV) of each alternative in the decision tree. The method is illustrated for the Eocene Aquifer in the northern part of the West Bank, Palestine. Nitrate is the main pollutant in the Eocene Aquifer and Methemoglobinemia is the main health problem associated with nitrate pollution in groundwater. The design options in this case study are: (i) ignoring the health risk of nitrate contaminated water, (ii) using alternative water sources such as bottled water or installing home treatment units, or (iii) establishing a groundwater quality monitoring network recommended previously (Chapter 2). The EV of each option was estimated as the weighted average cost of potential outcomes where costs include healthcare for methemoglobinemia, purchase of bottled water, purchase and operation

[1] Coauthored by Abdelhaleem Khader, David Rosenberg, and Mac McKee

of home treatment units, and installation and maintenance of the groundwater monitoring system. These costs are weighted by the probability (likelihood) of each outcome with probabilities reflecting the expected responses of people who live in the Eocene aquifer's area to follow the DM's recommendations to use or not use aquifer water as measured through a survey. The decision tree results show that the value of establishing the proposed groundwater quality monitoring network does not exceed the expected cost of establishing the network. More work is needed to improve the accuracy of the network and to increase people's awareness of the pollution problem and of the available alternatives.

3.1 Introduction

In many places throughout the world, groundwater is considered the only reliable source of fresh water. This important source is being jeopardized by nitrate (NO_3^-) and other pollution due to human activities such as agriculture, industry, municipal waste, septic tanks, cesspits, and dairy lagoons (Almasri and Kaluarachchi 2005). When ingested, nitrate decreases the ability of human blood to carry oxygen, which can result in oxygen deficiency and can cause Methemoglobinemia (blue baby syndrome) and other health problems like dizziness, headache, loss of muscular strength, hemolysis, seizures, or, in the most extreme cases, death (Majumdar 2003). Infants are more susceptible than adults (Lorna 2004) with susceptibility depending on the nitrate concentration in polluted water (Walton 1951). For example, infants who drink water with NO_3^-concentrations less than 45 mg/l are unlikely to get the disease, while 57% of infants who drink water with NO_3^- concentrations between 45-225 mg/l will experience methemoglobinemia, and almost all infants who drink water with NO_3^-

concentrations more than 225 mg/l will be affected. Due to these health risks, there is urgent need to intensively monitor and manage groundwater resources.

Effective groundwater monitoring and management must provide efficient and reliable information about groundwater quality, likelihood of different groundwater quality outcomes, and the costs and consequences of potential outcomes and actions. This information coupled with a value of information (VOI) analysis (Chia-Yu Lin et al. 1999; Dakins 1999; Dakins et al. 1994, 1996; Delquié 2008; Rajagopal 1986; Repo 1989; Sakalaki and Kazi 2006; Yokota and Thompson 2004a, 2004b) can help inform decisions regarding whether to ignore the pollution problem, use alternative sources of water, or design and implement a groundwater quality monitoring network.

Information is not free; it requires money and time to acquire (Sakalaki and Kazi 2006). Thus, VOI analysis evaluates the benefit of collecting additional information to reduce or eliminate uncertainty associated with the outcome of a decision. VOI makes explicit any expected losses from errors in decision-making due to uncertainty and identifies the "best" information collection strategy as one that leads to the greatest expected net benefit to the decision-maker (Yokota and Thompson 2004a).

To estimate how rational individuals should value the information, expected utility (EU) theory provides a normative of information valuation (Delquié 2008). In economics, utility is a set of numerical values that reflect consumer satisfaction from receiving a good or service. Expected utility is the probability-weighted average of the utility from each possible outcome (Perloff 2008).

Expected Utility Theory can be supported by a decision tree model (Fig. 3.1) that describes the logical structure of the decision. Each tree branch represents a different choice or outcome (Lund 2008). Boxes denote choice nodes, where a decision must be made. Circles denote chance nodes, where outcomes are uncertain. Each branch emanating from a choice node is an alternative, and each branch emanating from a chance node is a possible outcome, with a probability attached. The consequence of each outcome is shown at the far right of the tree. In Fig. 3.1, the DM is deciding whether to make uninformed decisions (Branches 1 or 2) or acquire more information about a system in order to make a better informed decision (Branch 3).

The VOI is measured *ex-ante* as the difference between the EUs of the informed and uninformed branches (Delquié 2008; LaValle 1968). For public policy decisions where consequences are small compared to the scale of the overall enterprise, we can substitute expected value (EV) for EU (Arrow and Lind 1970). The EV of each branch is the weighted average of the values of each outcome from that branch. The weights here correspond to the probabilities of each outcome. In this case the VOI is the difference between the EVs of the informed and uninformed branches. To acquire more information, the VOI for the informed decision must exceed the cost of acquiring the information.

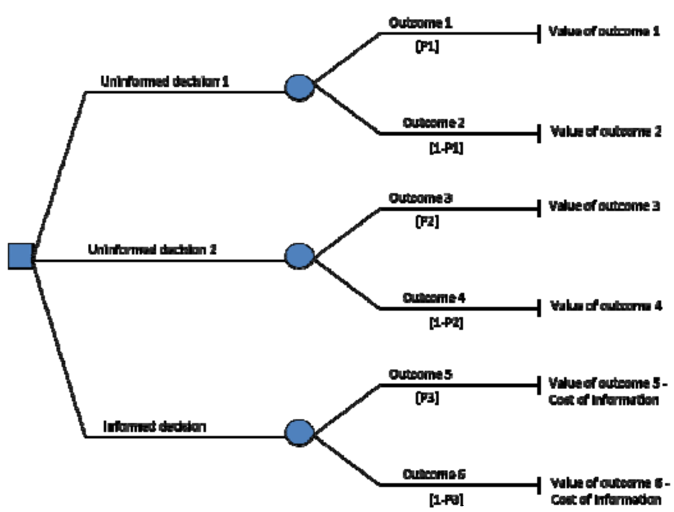


Fig. 3.1 Decision tree example that shows the structure of the tree and the different options

Willingness to pay (WTP) is another widely used method for VOI (Alberini et al. 2006; DeShazo and Cameron 2005; Dickie and Gerking 2002; Engle-Warnick et al. 2009; Latvala and Jukka 2004; Molin and Timmermans 2006; Roe and Antonovitz1985; Sakalaki and Kazi 2006). WTP is defined as the maximum amount a person or a DM is willing to pay in order to receive a good or to avoid something undesirable (Perloff 2008). In this method contingent valuation surveys should be conducted to ask individuals how much are they actually willing to pay (for information in this case) (Alberini et al. 2006; Atkins et al. 2007; Pattanayak et al. 2003). Although this WTP analysis can estimate how much people are actually willing to pay to acquire more information, it can only be done by asking people who should be well informed about the problem. On the other hand, the cheaper and easier EU method can estimate how rational people should value information. This is sufficient for VOI analysis.

This paper uses a decision tree model to estimate the value of information provided by a nitrate groundwater quality monitoring network presented in Chapter 2 which is an application to an actual management decision problem. Past VOI research in fields like general environmental health, water contamination, and toxicology applications tends to focus on demonstrating the usefulness of the VOI approach rather than on applications to actual management decisions (Yokota and Thompson 2004b).

This paper presents a probabilistic framework that depicts the logic and the structure of the choices faced by an aquifer manager concerned about nitrate contamination. These choices are to: (i) ignore the problem and not test for nitrate pollution and face the possibility of methemoglobinemia, (ii) recommend using alternative water sources such as bottled water and home treatment units without monitoring, and (iii) implement the groundwater quality monitoring design. The consequences of these alternatives include the probability of getting sick with methemoglobinemia.

The most common treatment for methemoglobinemia is methylene blue. This treatment converts MHb to hemoglobin and gives immediate relief. The cost of the treatment is about $150 per case (http://www.revolutionhealth.com/drugs-treatments/methylene-blue), which is considered a high cost for people living in the Eocene Aquifers area. Other treatments include (depending on the severity of the case) ascorbic acid, vitamins C and E, emergency exchange blood transfusion, and administration of high flow oxygen (Majumdar 2003). Other consequences are the cost of bottled water, home treatment units, and monitoring network.

The main contribution of this paper is the use of the decision tree framework to estimate the value of implementing a groundwater quality monitoring network by comparing the expected cost of the monitoring alternative with the expected costs of the uninformed options.

The next section describes briefly the study area which is the Eocene Aquifer, Palestine. The expected cost of monitoring is estimated in Section 3. After that, Section 4 discusses the decision tree components. Results from the VOI calculations are discussed in Section 5. And finally, concluding remarks are included in Section 6.

3.2 Study Area

The methodology of this research is demonstrated using the Eocene Aquifer, which is an unconfined aquifer located in the northern part of the West Bank, Palestine (Fig. 3.2). Nitrate is the main pollutant in the Eocene Aquifer. The main reasons for nitrate pollution in the aquifer are the excessive use of nitrogen-rich fertilizers and the lack of sewer networks (Najem 2008). Nitrate pollution may cause methemoglobinemia for people living in the area of the Eocene Aquifer, and this paper presents a decision tree model that describes the alternatives for a DM and clarifies the consequences of these alternatives in terms of methemoglobinemia treatment costs or costs of using alternative sources of water.

The Eocene Aquifer is used to meet domestic and agricultural demands for more than 207,000 Palestinians living in 66 communities, including 53,000 in the City of Jenin (PCBS 2009a). Annual population growth in the area is 3.0% and the average household size is 5.5 (PCBS 2008). More information about the Eocene Aquifer can be found in Chapter 2.

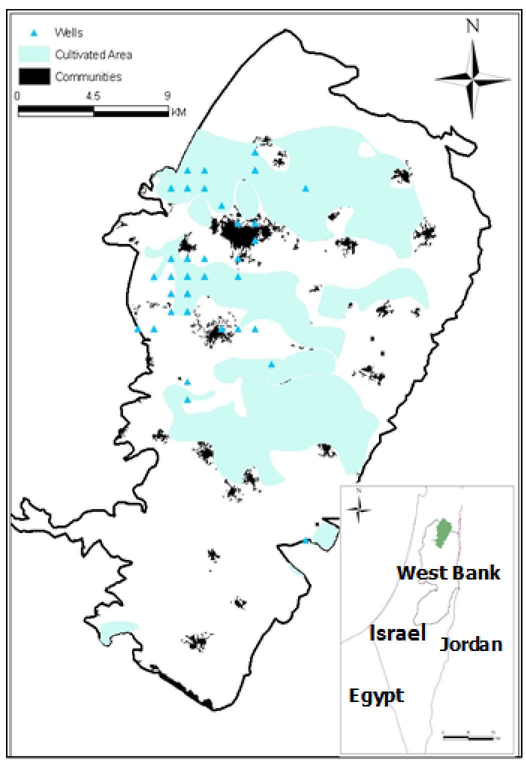


Fig. 3.2 Palestinian communities, abstraction wells, and cultivated areas in the Eocene Aquifer boundaries

3.3 Expected Cost of Monitoring

In Chapter 2 we designed a groundwater nitrate monitoring network for the Eocene Aquifer. The design shows the proposed locations of monitoring wells and takes into account uncertainties in climate and aquifer properties (Fig. 3.3). The network design captures the uncertainties in recharge, hydraulic conductivity, and nitrate reaction process through the application of a groundwater flow model and a nitrate fate and transport model following a Monte Carlo simulation process. A best-fit model of nitrate concentration distribution everywhere in the aquifer for each Monte Carlo subset is built using a relevance vector machine (RVM). The outputs from the RVM model are the distribution of nitrate concentration everywhere in the aquifer, the uncertainty in the characterization of those concentrations, and the number and locations of "relevance vectors" (RVs). The RVs form the basis of the optimal characterization of nitrate throughout the aquifer and represent the optimal locations of monitoring wells.

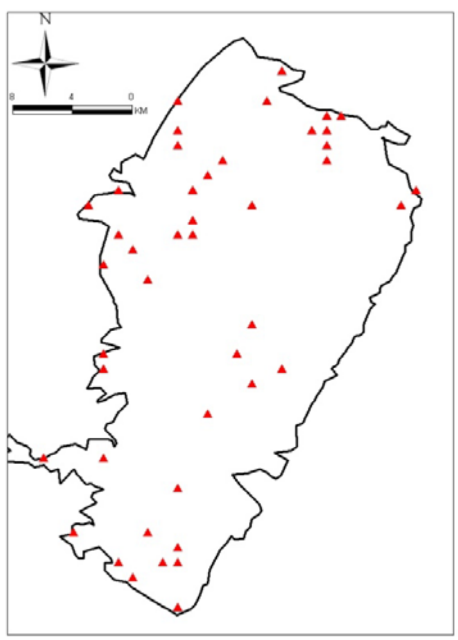


Fig. 3.3 Monitoring wells locations

The expected cost of monitoring from these wells consists of three components (CDLE 2001):

1. Drilling cost: \$53.89/m (for wells <15 m deep) and \$60.45/m (for wells >15 m deep)
2. Finishing cost: \$49.72/m, and
3. Nitrate sampling cost: \$12/year

The depth to ground water at each location is estimated using the groundwater flow model developed in Chapter 2. Total present value monitoring system costs are \$US 2.7 million and include drilling, finishing, and sampling costs for each well, a 30 year project life, and interest rate of 5% (Appendix 2).

3.4 Decision Tree components

3.4.1 Overview

The decision tree depicts the structure of the decision-making problem at hand, which here is whether to ignore the nitrate pollution problem, use alternative sources of water, or implement a groundwater quality monitoring network (Fig. 3.4). There are three branches emanating from the choice node (the box). These branches denote the options or alternatives from which the DM is choosing.

The first option is to ignore the problem and not test for nitrate pollution. In this case the DM will encourage people to use the aquifer and face the health risk if the aquifer water is contaminated ($NO_3^- > 45$ mg/l). In this case, there is a cost associated with methemoglobinemia treatment (section 3.4.2)

In the second option, the DM can recommend not using water from the aquifer without monitoring and use alternative sources of water such as bottled water or installing home treatment units.

The third option is to monitor groundwater quality. Since the monitoring network is imperfect, there is a probability that a reported concentration is different than the actual concentration. If the reported concentration is less than 45 mg/l, People will use the aquifer. But in this case there is a probability that the concentration is higher than 45 mg/l, which means that they might face a health risk. On the other hand, if the reported concentration is higher than 45 mg/l, people will use alternative sources of water.

The decision tree structure can vary depending on how the DM values the response of individuals to decisions regarding drinking water. Two scenarios are considered here: in the first scenario (Fig. 3.4) the DM does not take people's response into account. This means that the DM assumes that people will abide with all the recommendations. In the second scenario (Fig. 3.5) people's response is important in all the options. In this scenario people have the choice to abide with, or to ignore the DM's recommendations.

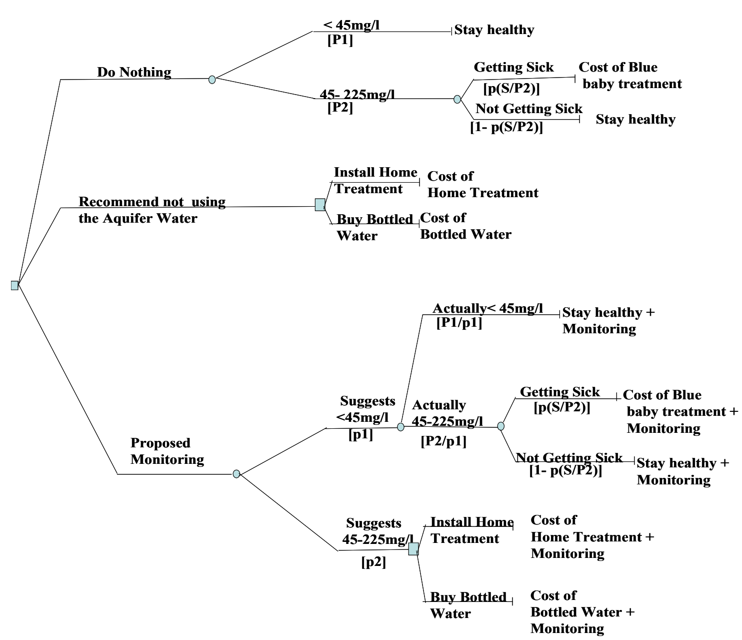


Fig. 3.4 Decision tree model (First scenario: without people's response)

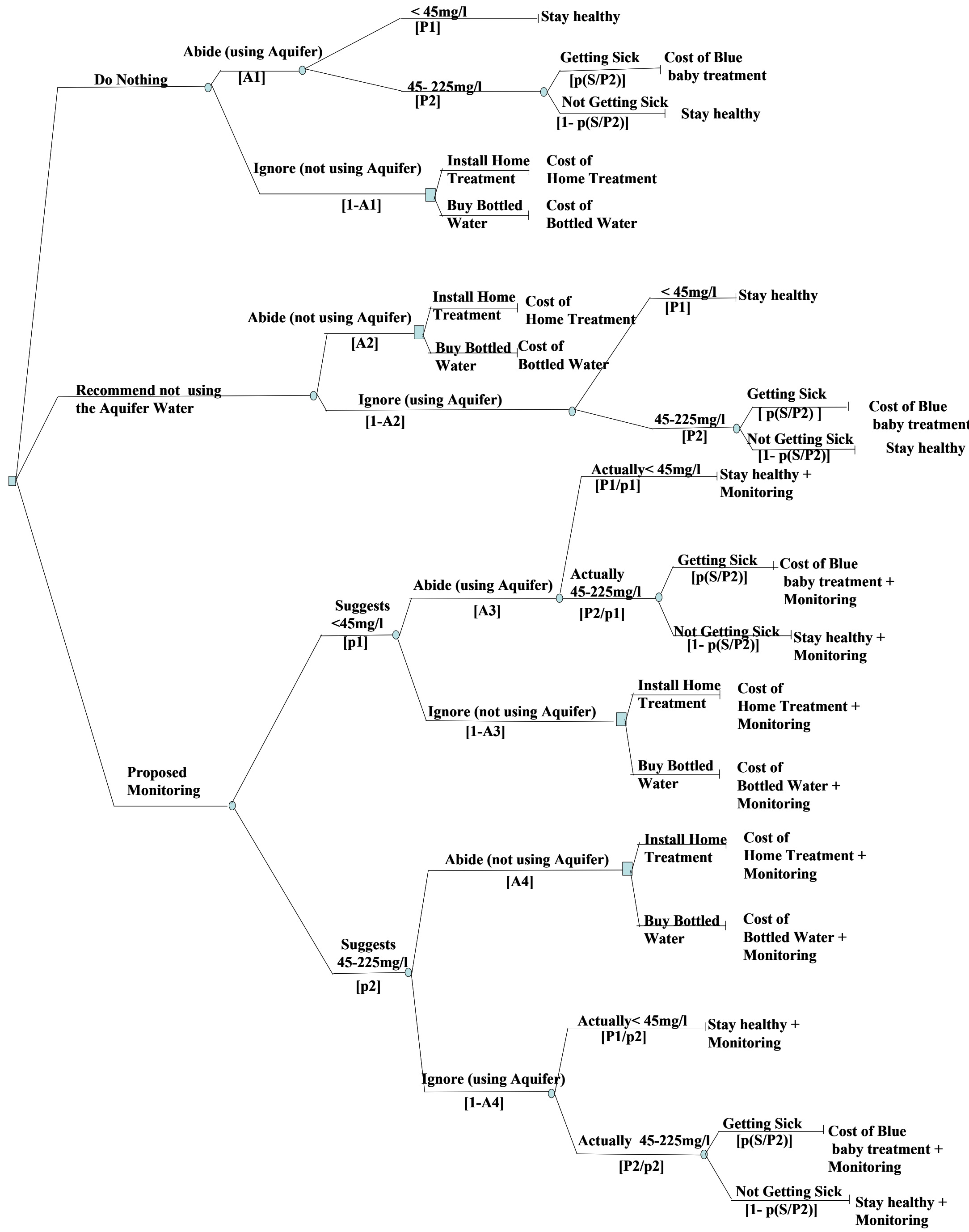


Fig. 3.5 Decision tree model (second scenario: with people's response)

3.4.2 Cost of Alternatives

As shown in the decision tree (Fig. 3.4), there are costs associated with each branch. These costs include:

1. Methemoglobinemia treatment: The most common treatment for methemoglobinemia is methylene blue (Majumdar 2003). The estimated cost of methylene blue treatment is $150 (http://www.revolutionhealth.com/drugs-treatments/methylene-blue). Since the affected population here is infants, we assume that both parents are working and at least one parent will stay home to take care of the infant, as would be commonly the case in the West Bank. The associated cost with the outcome of getting sick in the decision tree is six work days ($50 /day salary).
2. Home treatment units: one option to deal with polluted groundwater is to install home treatment units. Nitrate is easily dissolved in water, which means that it is difficult to remove. Three water treatment systems that remove nitrate are distillation, reverse osmosis (RO), and ion exchange (Jennings and Sneed 1996). RO is more common for home treatment in the West Bank. It costs about $750[1] for the unit and about $150/year for maintenance and running costs (Omour 2011).
3. Bottled water: in this option people make infant formula from bottled water rather than polluted groundwater. About 30% of infants in the West Bank drink formula rather than breast milk (Ammar et al. 2008). On average it costs about $0.6/day/infant to substitute bottled water to prepare formula.

[1] This amount will be considered as initial cost that will be distributed over the life of the project

3.4.3 Public Response

As shown in Fig. 3.5, responses to the DM's recommendations (whether to abide by or ignore them) are important factors that determine the likelihood of outcomes in the second scenario (with people's responses). To understand these responses and estimate their likelihoods (probabilities A1-A4 in Fig. 3.5), a survey was administered in the region. One hundred ninety-six participants were asked how they would respond to a water manager's recommendations to use or not to use the Eocene Aquifer's water based on four hypothetical scenarios. In the first two scenarios, respectively, the government has not tested the aquifer but declares it safe or not safe to drink. In the third and fourth scenarios, the aquifer has been tested properly and then the government declares it safe or not safe to drink (Chapter 4).

Statistical analysis of the responses to four scenarios provides estimates of the probabilities A1-A4, as shown in Table 3.1.

3.4.4 Probability Estimation

Calculating the expected cost of all the alternatives in the decision tree is based on the existing network of pumping wells in the Eocene Aquifer. As shown in Fig. 3.1, there are 44 pumping wells located in the area of the Eocene Aquifer. Due to lack of information about the distribution network, it is assumed here that water from these wells is distributed to the people according to the pumping rate from each well.

Table 3.1 Probabilities of participant's abidance to DM's recommendations

DM recommendation		Probability of abidance			
		Index	**Mean Value**	**Standard Deviation**	**95% C.I**
without monitoring	Do nothing	A1	0.294	0.457	0.230-0.358
	Use other sources	A2	0.959	0.199	0.931-0.987
with monitoring	Use the aquifer	A3	0.624	0.486	0.556-0.692
	Use other sources	A4	0.969	0.174	0.945-0.993

In designing the monitoring network (Chapter 2), Monte Carlo (MC) simulations were used to capture the uncertainty in recharge, hydraulic conductivity, and nitrate reaction process through the application of a groundwater flow model and a nitrate fate and transport model. A RVM model for the nitrate concentration distributions from the MC simulations was used to design the network. To estimate the probabilities needed in the decision tree (Fig.s 3.4 and 3.5), MC simulations and the RVM model results are used as follows:

- [P1]: probability that nitrate concentration is less than 45mg/l. This probability is estimated by considering the number of MC simulations where concentration was less than 45mg/l divided by the total number of simulations.
- [P2]: probability that nitrate concentration is in the range 45-225 mg/l. This probability is also estimated from MC simulations.
- [P3]: probability that nitrate concentration is greater than 225 mg/l. MC results show that the concentration will not likely exceed this limit. Thus, P3 is not considered an outcome in the decision tree.
- [S/P1]: probability of getting sick with methemoglobinemia given the concentration is less than 45 mg/l. This probability is zero (Walton 1951).
- [S/P2]: probability of getting sick with methemoglobinemia given the concentration is in the range 45-225 mg/l. This probability is 57% (Walton 1951).
- [p1]: probability that the monitoring network suggests nitrate concentration less than 45 mg/l. This probability is estimated from the RVM model by considering the number of RVM runs where concentration was less than 45mg/l divided by the total number of runs.

- [p2]: probability that the monitoring network suggests nitrate concentration in the range 45-225 mg/l. This probability is also estimated from the RVM model.

[p1/P1] and [p2/P2] are prior probabilities that represent the probability that the monitoring network suggest a concentration given that the actual concentration is the same. They can be estimated from Monte Carlo simulations and RVM results. Bayes Theorem let us use these prior probabilities to calculate the posterior probabilities needed in the decision tree model as follows:

- $$[P1/p1] = \frac{[P1][p1/P1]}{[p1]} \quad (1)$$
- $$[P2/p1] = 1 - [P1/p1] \quad (2)$$
- $$[P2/p2] = \frac{[P2][p2/P2]}{[p2]} \quad (3)$$
- $$[P1/p2] = 1 - [P2/p2] \quad (4)$$

where:

- [P1/p1]: probability that the actual concentration is less than 45mg/l when the monitoring network suggests it is less than 45 mg/l, and can be estimated from Equation (1)
- [P2/p1]: probability that the concentration is in the range 45-225 mg/l given that the monitoring network suggests it is less than 45 mg/l, and can be estimated from Equation (2)
- [P2/p2]: probability that the concentration is in the range 45-225 mg/l given that the monitoring network suggests it is in the range 45-225 mg/l, and can be estimated from Equation (3)

- [P1/p2]: probability that the concentration is less than 45mg/l given that the monitoring network suggests it is in the range 45-225 mg/l, and can be estimated from Equation (4).

3.4.5 Expected Cost Estimation

As stated earlier, calculations of the expected cost in the decision tree are based on the existing pumping wells in the Eocene Aquifer. Appendix 4 shows the pumping rate from each well and the percentage of total pumping from the aquifer. Based on the assumption that water is distributed proportional to the pumping rate, the number of households and the affected population from each well can be estimated as follow:

$$\text{Number of households / well} = \frac{\text{Total population}}{\text{Average family size}} \times \text{\% of total pumping}$$

where,

Total population: 207,000

Average family size: 5.5

% of total pumping: Appendix 5

$$\begin{aligned}\text{Affected population} &= \text{Total population} \times \text{Natural increase rate} \times \text{\% Using formula} \\ &\quad \times \text{\% of total pumping}\end{aligned}$$

where,

Natural increase rate: 3.0%

% using formula: 30%

The expected cost of each branch in the decision tree is the weighted average of the costs of all possible outcomes from that branch. The weights used in computing this average correspond to the probabilities in the decision tree (Fig.s 3.4 and 3.5) which were estimated in the previous section (Section 4.4)

3.5 Results and Discussion

Based on the expected cost calculations (Appendix 3), Fig. 3.6 shows the different expected costs associated with each of the options in the decision tree model in both scenarios: with and without people's response. Perfect monitoring as a fourth option is considered here, which is a hypothetical scenario used here for comparison. The meaning of "perfect" here is that when the monitoring network suggests nitrate concentrations equal to the actual ones. Again people will abide recommendations based on perfect monitoring results in the first scenario and they have the chance to abide or ignore in the second one. Do nothing is the option with the highest expected cost due to the high cost of methemoglobinemia treatment. This cost takes into consideration the health risk consequences of only one pollutant, nitrate, so actual costs are likely higher. The expected cost of not using the aquifer is still high due to the high cost of other alternatives such as bottled water. The expected cost of the perfect monitoring branch is lower than that of the proposed monitoring system.

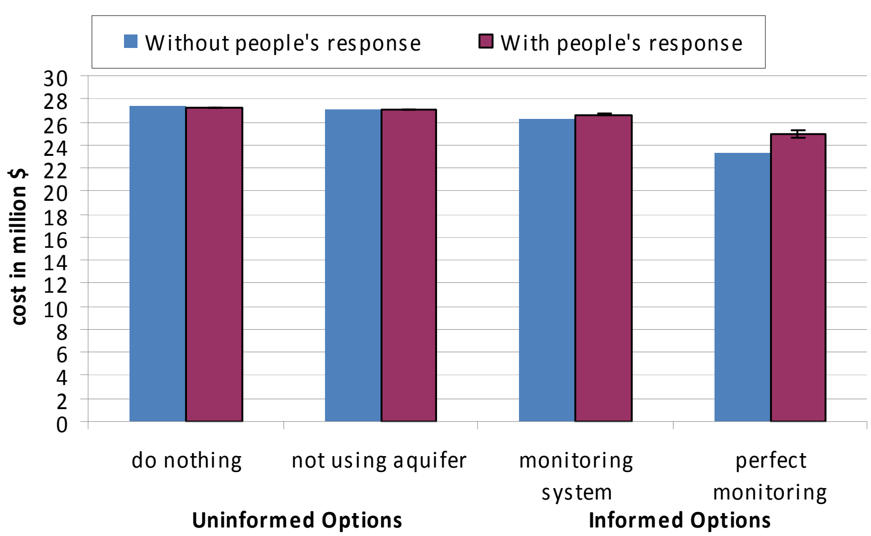


Fig. 3.6 Expected costs of different options in the decision tree model (error bars represent 95% confidence intervals)

The value of information provided by monitoring can be calculated by subtracting the expected cost of the monitoring branch from the expected cost of the best uninformed branch, which is the “not using the aquifer branch” in this case. The results show that in the first scenario, where people’s responses are not important, the value of perfect monitoring exceeds the cost of monitoring while the value of proposed monitoring is less than that cost (Fig. 3.7). It also shows that in the second scenario, where people’s responses are important, the cost of monitoring exceeds the values of both the proposed and perfect monitoring. Analysts often suggest that if the value of perfect monitoring is less than the cost of monitoring (as in the second scenario), then the DM should not invest in monitoring (Yokota and Thompson 2004b). By comparing the first scenario (without people’s responses) with the second scenario (with people’s responses) and the proposed monitoring branch with the perfect monitoring branch, it is seen that the proposed monitoring network is not economically viable because (i) the accuracy is not sufficiently high and (ii) people do not reliably follow recommendations that stem from the monitoring system results. The first problem can be addressed by improving the accuracy of the monitoring system. Chapter 2 recommended including other sources of uncertainty such as human activities and on-ground nitrate loading. The second problem can be addressed by adopting some education and awareness programs that explain monitoring system results and encourage people to act according to them. These awareness programs may include town hall meetings in local communities, advertisement in the media, and education campaigns in schools and universities in the region.

Fig. 3.7 Expected value and cost of monitoring (error bars represent 95% confidence intervals)

3.6 Conclusions

This paper presented a methodology to estimate the value of monitoring groundwater quality, and used the nitrate-polluted Eocene Aquifer in the West Bank, Palestine, as a demonstration case. A decision tree model was used to estimate the value of information from a previously designed groundwater quality monitoring network for the Eocene Aquifer. The options available to the DM are: (i) ignore the problem and use water from the aquifer, (ii) use water from alternative sources, and (iii) establish the monitoring network. Two scenarios were considered: in the first one, the responses of people to the DM's decisions were not taken into account, while in the second one these responses were important.

By comparing the expected cost of monitoring with the value of monitoring in the first scenario (without people's responses), it is found that the value of perfect monitoring exceeds the cost of monitoring, but the value of proposed monitoring is less than the cost of monitoring. In the second scenario, the cost of monitoring exceeds the values of both the proposed and perfect monitoring. The value of monitoring in this paper takes into account only the health risk associated with nitrate pollution in groundwater. Considering that the same monitoring network could be used for other pollutants and health problems, this value underestimates the true societal value of monitoring. More work is needed toward improving the accuracy of the monitoring network and toward increasing people's awareness of the monitoring system.

Another conclusion that could be drawn from the results in Fig. 3.6 is that even with the option which has the least expected cost "the perfect monitoring option", there is still high cost associated with that considering the size of the study area and its small economy. This high coping cost is an indication of how poor the water situation is in the area of the Eocene Aquifer.

CHAPTER 4

SOCIAL ACCEPTANCE OF GROUNDWATER QUALITY MONITORING NETWORK IN THE EOCENE AQUIFER, PALESTINE[1]

Abstract

This paper presents the results of a survey that was administered to people living in the area of the Eocene Aquifer, Palestine to support a decision regarding implementation of a potential groundwater quality monitoring network. One hundred ninety-five participants were asked questions to infer their perception about the current situation of water quality and quantity and the sources of water delivered to them. They were also asked about their expected responses to use or not use the aquifer following decision maker's (DM's) recommendations in four hypothetical scenarios. In the first two scenarios, the government has not tested the aquifer but declares it, respectively, either safe or not safe to drink. In the third and fourth scenarios, respectively, the aquifer has been tested properly and then the government declares it safe or not safe to drink. The results show that most participants use groundwater for their indoor and outdoor uses and that they are generally unsatisfied about water quality and quantity. The results also show that people in general do not trust a government statement that is not based on fact, they are skeptical, and they are willing to spend more on alternative sources of water to reduce health risks in the face of poor information regarding the actual health risk of the aquifer water. Finally, the results show that these responses are consistent regardless of the type of community (urban vs. rural) or the presence of infants in the household.

---

[1] Coauthored by Abdelhaleem Khader, David Rosenberg, and Mac McKee

4.1 Introduction

Groundwater is the main source of freshwater in the Palestinian Territory, where water is considered to be an important and sensitive issue (PCBS 2009b). Palestinians suffer from water deficiency and have limited control over their water resources. Anthropogenic sources of pollution, such as agriculture, industry, and municipal waste, contribute to the degradation of groundwater quality, which may limit the use of these resources and lead to health-risk consequences. For these reasons, the need for intensive groundwater resources management has become more urgent. To be effective, groundwater resources management requires reliable information about the system being managed (Chapter 3). However, the decision to implement a monitoring system requires the involvement of all stakeholders including the people who are consuming the water.

Chapter 2 proposed a groundwater quality monitoring network design for the Eocene Aquifer, Palestine. This aquifer provides agricultural and domestic supplies for approximately 207,000 Palestinians living in 66 communities. The aquifer is polluted by nitrate from the excessive use of nitrogen-rich fertilizers and the lack of sewer networks (Najem 2008). Chapter 3 studied the value of information provided by the proposed monitoring network design by utilizing a decision tree model that can help guide a decision maker's (DM's) decision to implement the design. This paper continues in the same context by studying the social aspect of that decision by inferring people's perceptions of the current situation of groundwater quantity and quality and the expected response of people to a DM's recommendations.

Chapter 3 recommended that more work should be done toward increasing people's awareness of the nitrate pollution problem and the available groundwater management alternatives, including implementing the monitoring network. To understand the target population, a survey was administered in the area in which 250 participants were invited to answer questions regarding the current water situation and their expected response to DM's recommendations to use or not to use the aquifer's water based on four hypothetical scenarios. In the first two scenarios, the government has not tested the aquifer but declares it, respectively, safe or not safe to drink. In the third and fourth scenarios, respectively, the aquifer has been tested properly and then the government declares it safe or not safe to drink. The survey questions are shown in Appendix 1 along with the letter of information for participants.

There are many factors that might affect people's responses to these questions. Among these factors is having an infant in the household. Infants are usually more susceptible to diseases like methemoglobinemia (Majumdar 2003). Another factor is living in urban areas, where access to services such as water supply is generally good, versus living in rural areas where people have less access to clean water. In this paper the survey results are statistically analyzed to detect whether different groups of people respond differently to the DM's recommendations. If differences were detected, the value of monitoring could be improved by targeting specific groups in awareness campaigns.

The methodology of the research is presented in the next section. The different survey results are presented in Section 4.3. Finally, conclusions are presented in Section 4.4.

4.2 Methodology

The target group for the survey consists of people living in the area of the Eocene Aquifer, which is an unconfined aquifer located in the northern part of the West Bank

(Fig. 4.1). There are 66 communities in the region, ranging from small villages to the city of Jenin (population 53,000). Two hundred fifty subjects were invited to participate in the study, and 195 subjects living in 26 communities responded (the response rate is 78%). Fig. 4.1 shows the locations of the Eocene communities. The highlighted communities in Fig. 4.2 depict the ones where respondents were sought. The communities were chosen so that the sample would be spatially representative, as shown in Fig. 4.2. In each community, participants were randomly selected in the centers where people usually pay their utility bills, which include electricity and water. Participants were giving the choice to fill out the survey immediately or to take home and contact the researcher to collect it when it is ready, which explains the 22% non-response rate. The demographic characteristics of the study sample are shown in Table 4.1.

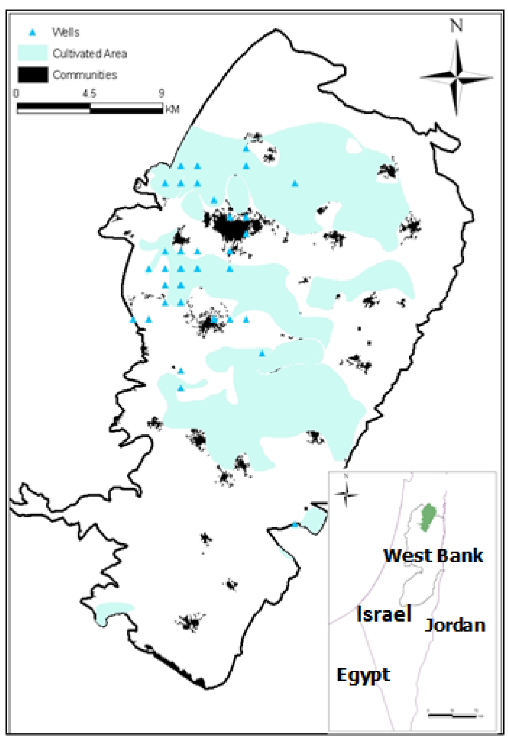


Fig. 4.1 Palestinian communities, abstraction wells, and cultivated areas in the Eocene Aquifer boundaries

The main tool of this study is a two-page questionnaire that contains 12 questions (see Appendix 5). The first group of the questions asks the participants about demographics, e.g., the number of residents in their households and the village/town in which they live (Questions 2 and 12). The second group asks about the sources of water the participants use for indoor and outdoor purposes (Question 1). The third group asks about the participants' awareness about the sources of water delivered to them through the network and their rate of satisfaction about water quantity and quality (Questions 3-5). The final group asks about the participants' likely responses to a DM's recommendations to use or not to use the water based on hypothetical scenarios (questions 6-11). The survey was conducted during May 2011 by the first author.

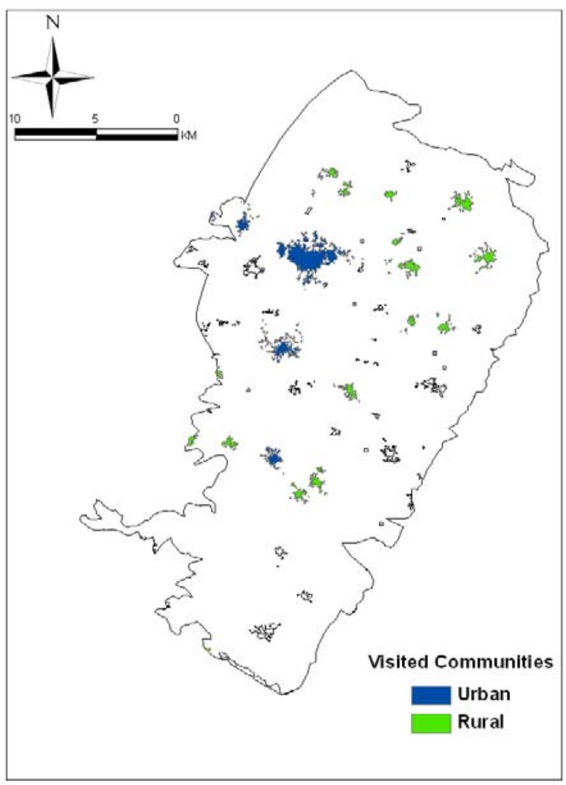


Fig. 4.2 Eocene communities where respondents were sought

Table 4.1 Demographic characteristics of the sample[1]

	Urban	Rural	Total
# of Participants	72	123	195
Average Household Size	6.7	6.4	6.5
# of Households with infants	24	41	65

4.3 Survey Results

Sources of water the participants use: the participants were asked to specify the sources of water they use for indoor and outdoor uses. Indoor uses include: drinking, cooking, bathing, and house cleaning, while outdoor uses include: landscaping, car cleaning, and livestock. The choices were: Pipe Network/Tap, Tanker Truck, Rain Water, Bottled Water, Home Treatment, and Other sources.

As shown in Fig. 4.3, for indoor purposes, the majority of respondents (77%) use the pipe network for indoor purposes, but only 30% of them indicated that they use only this source. Participants complained about insufficiency in water quantities provided to them through the network, especially in summer. As a result, many participants indicated that they rely on other sources, such as: rainwater (50%) and tanker trucks (27%). What makes that possible is the common practice in rural communities in the West Bank to have cisterns in most of the houses. They usually use these cisterns to collect rainwater during the rainy season (October to April) and if needed, they use them during summer as tanks to store water they buy from tanker trucks.

[1] Other demographic information such as age, gender, education, etc., were not collected because they are not important for the objective of the study and to keep the survey as short and simple as possible

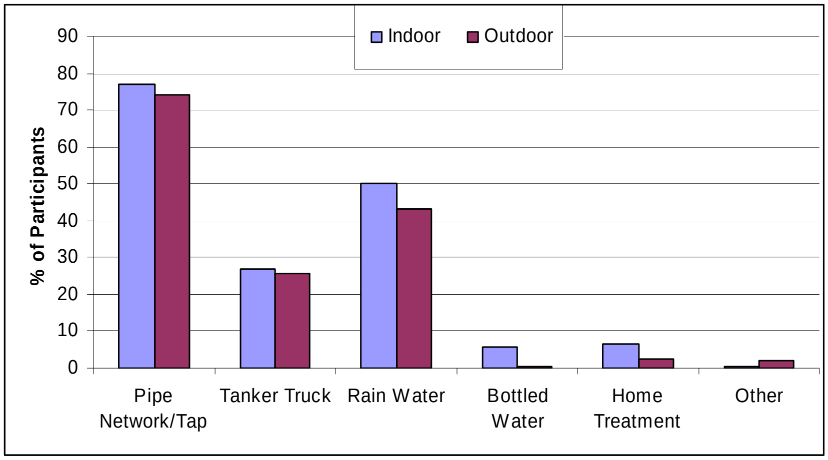


Fig. 4.3 Sources of water Participants use for indoor and outdoor purposes

Due to water scarcity in the region, outdoor uses of water are limited. The common practice for landscaping (if any) is to have a few native fruit trees, which are generally rain fed, and some seasonal vegetables. The other two practices (car washing and livestock) are not too common. People usually use the same sources for these purposes as they use for indoor purposes as seen in Fig. 4.3.

Although the Eocene is the main source of freshwater in the area, there are other sources as well. Participants were asked to specify the source of water they think the government is providing to them through the pipe network. The options were: well-Eocene Aquifer, well-other aquifers, spring, other, and don't know. Fig. 4.4 shows that most participants know that they are getting water from groundwater aquifers, but they are not sure whether it is coming from the Eocene or from other Aquifers.

Rate of satisfaction about water quality and quantity: Participants were asked to rate their satisfaction about water quantity and quality between 1 (satisfied) and 5 (unsatisfied). Fig. 4.5 shows that about 45% of participants are unsatisfied about water quantity (4 and 5) versus 30% who are satisfied. The average of the satisfaction rate is 3.31 and the standard deviation is 1.46. The variability is due to the fact that the water situation is different from town to town.

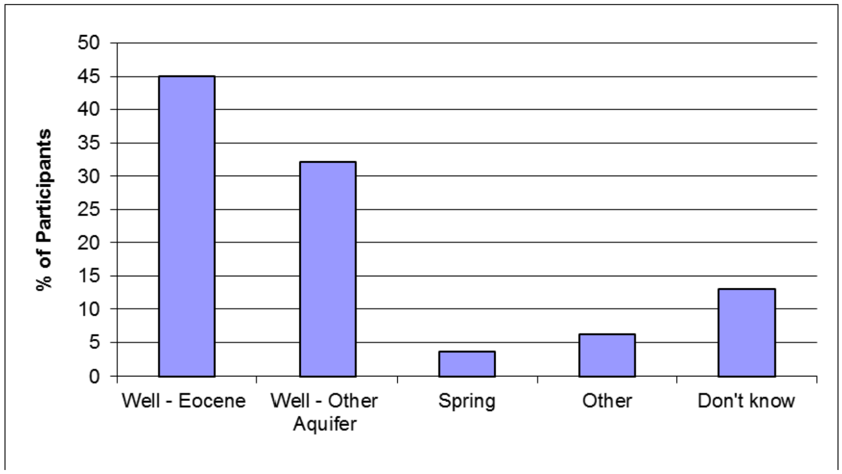


Fig. 4.4 Sources of water participants think the government provide to them through the network

Participants were then asked to explain why they answered about their rate of satisfaction this way. Those who were not satisfied explained that quantities provided to them are not sufficient and spotty. They also complained about the high prices of water.

The average rate of satisfaction about water quality is 3.18 and the standard deviation is 1.40. Fig. 4.5 shows that about 39% were unsatisfied (4 and 5) versus 30% satisfied (1 and 2). Comparing satisfaction about quantity versus quality shows that participants are more worried about water quantity than water quality. The reason for this is insufficiency of water provided to them and the fact that there is no history of major health problems related to water quality.

Participants' responses to DM's recommendations: participants were asked how they would respond to a water manager's recommendations to use or not to use water from the Eocene Aquifer based on four hypothetical scenarios. In the first two scenarios, the government has not tested the aquifer but declares it safe or not safe to drink, respectively. In the third and fourth scenarios, the aquifer has been tested properly and then the government declares it safe or not safe to drink.

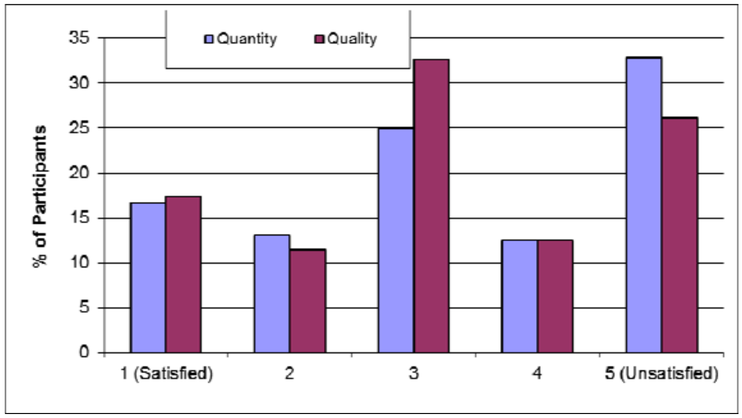


Fig. 4.5 Rate of Satisfaction about water quantity and quality (1 satisfied and 5 unsatisfied)

When the participants were asked about what they will do if government officials have not tested the aquifer but declare it safe to drink (the first scenario), less than 30% of them said they would abide with the government's recommendation. Fig. 4.6 shows that most of the participants would use home treatment units or buy bottled water. On the other hand, when asked about the second scenario (the government declared the aquifer is not safe and recommend people to use other sources), most participants said they would follow the government's recommendation. These results suggest participants only trust a government statement that are based on data, are skeptical, and are willing to spend more to install home treatment or to buy bottled water to reduce health risks in the face of poor information regarding the actual health risk of the aquifer water. When they were asked about what the government could do to make them change their mind and abide with their recommendations, most respondents suggested periodic testing of the water from the aquifer. A few respondents asked for more honest and transparent decisions. Some respondents emphasized on public awareness, and the rest asked for more treatment of drinking water or helping them in buying home treatment units.

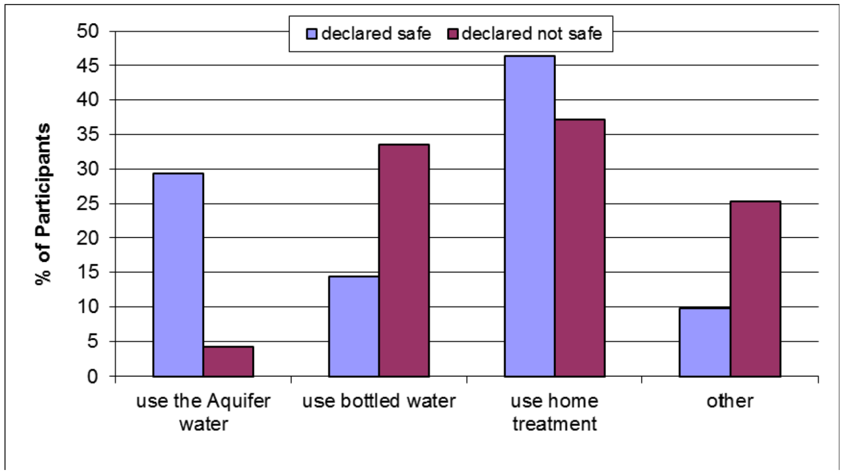


Fig. 4.6 Participants' response to DM's recommendations without monitoring

In the third and fourth scenarios, participants were told that the government has tested the aquifer, and then declares it safe to drink and use. In this case the majority of participants (about 62%) said they would abide by the recommendation and use water from the aquifer (Fig. 4.7). Furthermore, under the same scenario, if government declares the aquifer not safe and recommended that people use other sources, a very small minority of participants (about 3%) said they would use water from the aquifer.

By comparing responses to the four scenarios, we conclude that participants are more likely to follow recommendations that are supported by data.

Infants versus no infants: Sixty-five participants said they have at least one infant in their household (Table 1). Fig. 4.8 shows the responses of participants with and without infants.

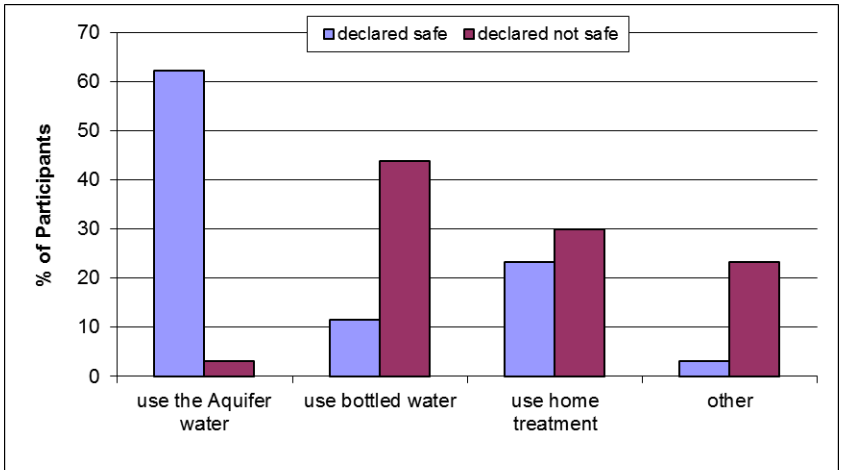


Fig. 4.7 Participants' response to DM's recommendations with monitoring

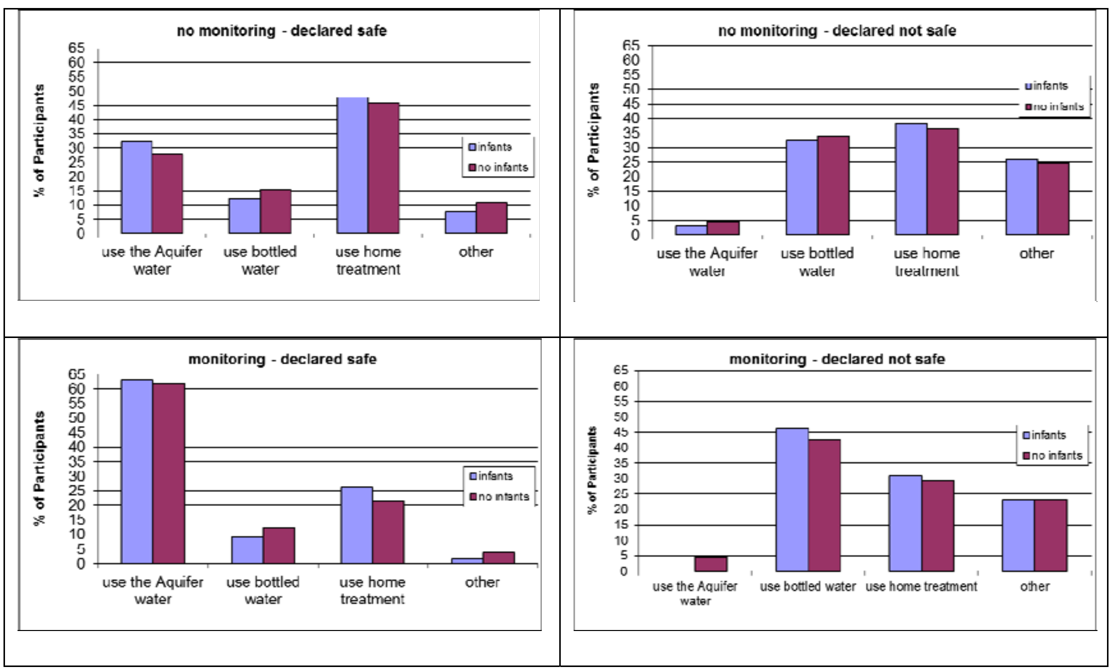


Fig. 4.8 Responses of participants with infants vs. participants without infants to DM's recommendations

Table 4.2 shows the results of statistical tests which were performed to estimate the significance of the differences between the two groups (participants with infants vs. participants without infants). Although infants are more vulnerable to water quality related health issues such as Methemoglobinemia, the data in Table 4.2 do not indicate significant differences in responses of people with or without infants.

Table 4.2 Statistical significance tests results

Proposed Scenarios		Infants vs. no Infants		Urban vs. Rural	
		Significance Level (P-value)	Likelihood of a difference (%)	Significance Level (P-value)	Likelihood of a difference (%)
No monitoring	declared **safe**	0.714	28.6	0.218	78.2
	declared **not safe**	0.289	71.1	0.991	0.9
Monitoring	declared **safe**	0.480	52.0	0.918	8.2
	declared **not safe**	0.995	0.5	0.841	15.9

Participants in both groups follow the same trend: they are skeptical and only trust DM's statements when based on data, and having infants in the households did not have an impact.

Urban versus rural: Seventy-two participants in eight communities (Table 4.1 and Fig. 4.2) are living in urban areas that have bigger municipalities. These municipalities can afford more comprehensive water distribution systems. As in the case of infants versus no infants, living in an urban area in the Eocene Aquifer did not affect the response of participants to DM's recommendations (Table 4.2).

4.4 Conclusions

This paper continues the work of the first and second papers by studying the social aspect of groundwater quality monitoring network design. A survey was administered in the Eocene Aquifer area to ask people about their perception of the current situation of groundwater quality and quantity and their expected response to DM recommendations in four hypothetical scenarios. In the first two scenarios, the government has not tested the aquifer but declares it either safe or not safe to drink. In the third and fourth scenarios, the aquifer has been tested properly and then the government declares it either safe or not safe to drink.

The results show that groundwater is the main source of freshwater for indoor and outdoor uses in the Eocene Aquifer area. The results also show that the participants are generally unsatisfied with water quality and quantity. When asked about their responses to a DM's recommendations in the above scenarios, the results show that the participants in general do not trust a government statement that is not based on data, they are skeptical, and they are willing to spend more to install home treatment or to buy bottled water to reduce health risks in the face of poor information regarding the actual health risk of the aquifer water. Finally, the results show that these responses are consistent regardless of the type of community (urban versus rural) or whether there are infants in the household.

These results suggests that the participants perceive groundwater problems and are likely to support the decision to implement a groundwater quality monitoring network design given the feasibility of that design.

CHAPTER 5

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

5.1 Summary and Conclusions

This research has introduced a methodology for groundwater quality monitoring network design that takes into account uncertainties in climate and aquifer properties, the economic value of information, and the social aspects of network design.

While groundwater is considered a precious resource for many people, the access to this resource is limited by the increasing trend of pollution, especially in terms of nitrate in many locations. For this reason, groundwater resources management has become more and more important. On the other hand, management requires reliable information that can be acquired through monitoring. Due to the complicated nature and uncertainties in the behavior of water in aquifers, more powerful and efficient tools are needed to address monitoring problems and design of monitoring systems. Statistical learning machines which are characterized by their ability to quickly capture the underlying physics and provide predictions of system behavior are utilized in this research.

Due to limited resources, a groundwater quality monitoring network design should be economically efficient. To insure the feasibility of the design, value of information (VOI) analysis is performed by utilizing a decision tree model. Finally, all stakeholders should be involved in the design including people who are consuming water from the aquifer. For that purpose, a survey was administered in the study area to infer people's perceptions of the current situation and their expected reaction to monitoring scenarios.

In Chapter 2 a methodology for groundwater quality monitoring network design was presented that utilizes a statistical learning algorithm called relevance vector machine (RVM). The methodology takes into account uncertainties in aquifer properties, pollution transport processes, and climate.

The procedure starts by quantifying the uncertainties in recharge, hydraulic conductivity, and nitrate reaction processes by applying conventional groundwater flow and nitrate fate and transport models in a Monte Carlo (MC) simulation process. After that, an optimal monitoring network that takes into account the uncertainties revealed in the MC simulations is designed by utilizing a RVM model.

The input to the RVM consists of all the possible locations of monitoring wells. The model output represents the corresponding nitrate concentrations acquired from the Monte Carlo simulations. The RVM model discovers the non-linear relationships between the inputs and the outputs and finds the locations where monitoring can be done that are most relevant for prediction of nitrate concentrations everywhere in the aquifer (hence the name “relevance vector machine”). The RVM results show that a network of 49 monitoring wells (Fig. 5.1) is optimal in terms of representing nitrate distribution in the aquifer and capturing information about the uncertainty of nitrate concentration estimates.

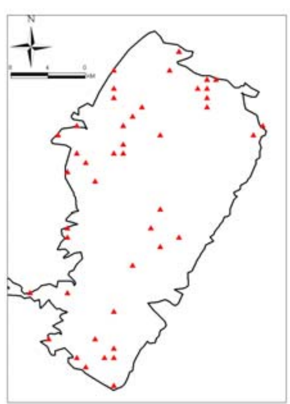


Fig. 5.1 Monitoring wells distribution

In Chapter 3 the VOI provided by the groundwater quality monitoring network design presented in the Chapter 2 was estimated through a decision tree model that describes the structure of the alternatives facing the decision maker (DM). There are three alternatives to choose from: (i) a "do nothing" alternative which ignores the pollution problem, (ii) a "do not use the aquifer" alternative, and (iii) a "monitoring network" alternative. VOI is estimated by evaluating the expected value (EV) of each alternative in the decision tree. The EV of each option was estimated as the weighted average cost of potential outcomes where costs include healthcare for methemoglobinemia, the cost to purchase bottled water, the purchase cost of home treatment units, and the cost to install and maintain the groundwater monitoring system. These costs are weighted by the probability (likelihood) of each outcome, with probabilities reflecting the expected responses of people who live in the area of the Eocene to follow DM's recommendations to use or not use aquifer water as measured through a survey (Chapter 4).

Two scenarios are considered here: in the first scenario the DM does not take people's responses into account, which means that the DM assumes that people will abide with all the recommendations regarding water use. In the second scenario people's responses are important in all the options. In this scenario people have the choice to abide by, or to ignore the DM's recommendations.

The results show that in the first scenario, where people response's are not important, the value of perfect monitoring exceeds the cost of monitoring, while the value of proposed monitoring is less than that cost (Fig. 5.2). It also shows that in the second scenario, where people's responses are important, the cost of monitoring exceeds the values of both the proposed and perfect monitoring.

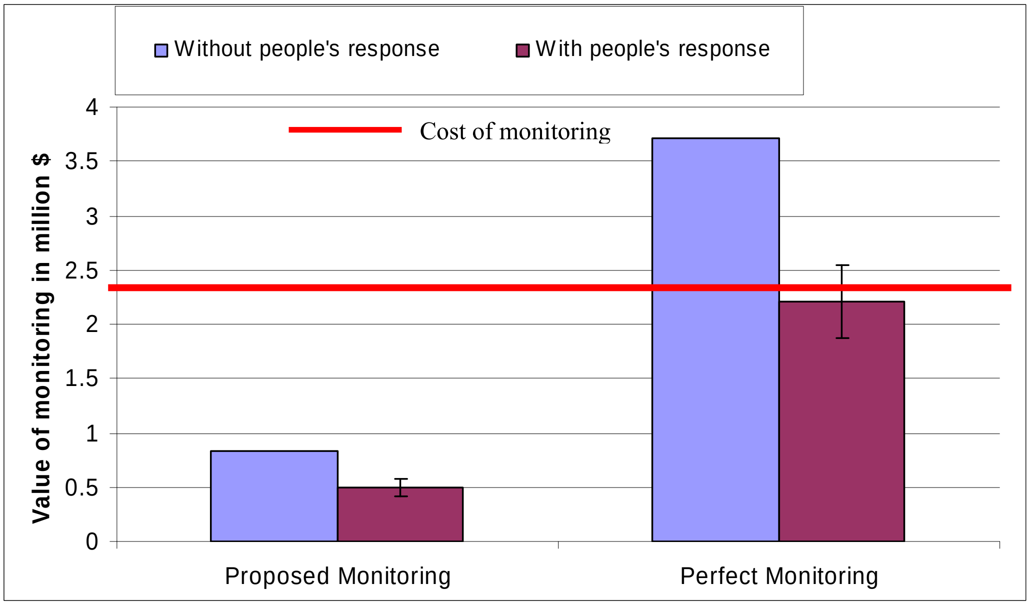


Fig. 5.2 Expected value and cost of monitoring (error bars represent 95% confidence intervals)

As reported in Chapter 4 a survey was administered on people living in the the area of the Eocene Aquifer, Palestine to determine whether a decision regarding implementing a groundwater quality monitoring network could be supported. Participants were asked questions to infer their perception about the current situation of water quality and quantity. They were also asked about their expected responses to a decision maker's recommendations in four hypothetical scenarios. In the first two scenarios, the government has not tested the aquifer but declares it either safe or not safe to drink. In the third and fourth scenarios, the aquifer has been tested properly and then the government declares it either safe or not safe to drink.

The results show that groundwater is the main source of freshwater for indoor and outdoor uses in the Eocene Aquifer area. The results also show that the participants are generally unsatisfied about water quality and quantity. When they were asked about their responses to a DM's recommendations under the above mentioned scenarios, the results show that the participants in general do not trust a government statement that is not based on data, they are skeptical, and they are willing to spend more to install home treatment or to buy bottled water to reduce health risks in the face of poor information regarding the actual health risk of the aquifer water.

In conclusion, this research provided a methodology for designing an optimal groundwater quality monitoring network under uncertainties in aquifer properties, climate, and pollutant reaction processes. The methodology was able to check the economical optimality of the design in terms of the VOI and the implications of people reaction to the network towards its feasibility.

5.2 Recommendations for Future Work

The methodologies and concepts presented in Chapters 2-4 could be improved by the following ideas for future research:

1. The methodology presented in Chapter 2 investigated uncertainties in recharge, hydraulic conductivity, and nitrate reaction process. This methodology could be improved by examination of other sources of uncertainty such as future human activities and on-ground nitrate loading.
2. A significant improvement could be the addition of a temporal dimension to the monitoring network, i.e., the sampling frequency. Adding this option is conditioned on the availability of temporal data about nitrate pollution.
3. The decision tree model developed in Chapter 3 could be improved by including information about incidents of blue baby syndrome, which are not available at the moment.
4. The VOI estimated in Chapter 3 could be improved by adopting some education and awareness programs that explain monitoring system results and encourage people to act according to them. These awareness programs may

include town hall meetings in local communities, advertisement in the media, and education campaigns in schools and universities in the region.

REFERENCES

Abu Zahra, B.A.A. (2001). "Water crisis in Palestine." *Desal.,* 136(1-3) 93-99.

Alberini, A., Hunt, A., Markandya, A. (2006). "Willingness to pay to reduce mortality risks: Evidence from a three-country contingent caluation study." *Enviro. and Res. Econ.,* 33(2) 251-264.

Almasri, M.N., and Kaluarachchi, J.J. (2005). "Multi-criteria decision analysis for the optimal management of nitrate contamination of aquifers." *J. Environ. Man.,* 74(4) 365-381.

Ammar, K., Khalil, A., McKee, M., Kaluarachchi, J. (2008). "Bayesian deduction for redundancy detection in groundwater quality monitoring networks." *Water Resour. Res.*, 44(8) W08412.

Armstrong, J.S., and Collopy, F. (1992). "Error measures for generalizing about forecasting methods: Empirical comparisons." *Int. J. Forecasting,* 8(1) 69-80.

Arrow, K.J., and Lind, R.C. (1970)."Uncertainty and the evaluation of public investment decisions." *The American Econ. Rev.,* 60(3) 364-378.

Asefa, T., Kemblowski, M., McKee, M., Khalil, A. (2006). "Multi-time scale stream flow predictions: The support vector machines approach." *J. Hydrology,* 318(1-4) 7-16.

Asefa, T., Kemblowski, M., Urroz, G., McKee, M. (2005)."Support vector machines (SVMs) for monitoring network design." *Ground Water,* 43(3) 413-422.

Asefa, T., Kemblowski, M.W., Urroz, G., McKee, M., Khalil, A. (2004)."Support vectors–based groundwater head observation networks design." *Water Resour. Res.,* 40(11) W11509.

Atkins, J.P., Burdon, D., Allen, J.H. (2007)." An application of contingent valuation and decision tree analysis to water quality improvements." *Marine Poll. Bulletin,* 55(10-12) 591-602.

Chia-Yu Lin, A., Andrew, G., Phillip, N.P., David, H.K. (1999). "Analysis of local decisions using hierarchical modeling, applied to home radon measurement and remediation." *Stat. Science,* 14(3) 305-337.

Colorado Department of Labor and Employment (CDLE) (2001)." Maximum reasonable costs - drilling."

Dakins, M.E. (1999)."The value of the value of information." *Human and Eco. Risk Assessment,* 5 281-289.

Dakins, M.E., Toll, J.E., Small, M.J. (1994). "Risk-based environmental remediation: Decision framework and role of uncertainty." *Environ. Tox. and Chem.,* 13(12) 1907-1915.

Dakins, M.E., Toll, J.E., Small, M.J., Brand, K.P. (1996). "Risk-based environmental remediation: Bayesian Monte Carlo analysis and the expected value of sample information." *Risk Anal.,* 16(1) 67-79.

Delquié, P. (2008). "Valuing information and options: an experimental study." *J. Behavioral Decision Making,* 21(1) 91-109.

DeShazo, J.R., and Cameron, T.A. (2005). "The effect of health status on willingness to pay for morbidity and mortality risk reductions." *On-Line Working Paper Series, California Center for Population Research, UC Los Angeles*, CCPR-050-05

Dickie, M., and Gerking, S. (2002). "Willingness to pay for reduced morbidity." Paper presented at the workshop, “Economic Valuation of Health for

Environmental Policy: Assessing Alternative Approaches," at the University of Central Florida, Orlando, Fla., Mar. 18–19.

Engle-Warnick, J., Héroux, J., Montmarquette, C. (2009). "Willingness to pay to reduce future risk." *CIRANO Scientific Series* 2009s-27, Montreal, Canada, 2009.

Friedman, M. (1966). *Essays in Positive Econ.* University Of Chicago Press.

Gill, M.K., Asefa, T., Kemblowski, M.W., McKee, M. (2006). "Soil moisture prediction using support vector machines". *JAWRA,* 42(4) 1033-1046.

Harbaugh, A.W., Banta, E.R., Hill, M.C., McDonald, M.G. (2000). "MODFLOW-2000, The US Geological Survey modular ground-water model—user guide to modularization concepts and the ground-water flow process." *Open-File Report* 00–92. US Geological Survey: Reston, VA.

Jennings, G.D., Sneed, R.E. (1996). "Nitrate in Drinking Water." *North Carolina Co. Ext.,* AG 473-474.

Khalil, A., Almasri, M.N., McKee, M., Kaluarachchi, J.J. (2005a). "Applicability of statistical learning algorithms in groundwater quality modeling." *Water Resour. Res.,* 41(5) W05010.

Khalil, A., McKee, M., Kemblowski, M., Asefa, T. (2005b)."Sparse Bayesian learning machine for real-time management of reservoir releases." *Water Resour. Res.,* 41(11) W11401.

Kharmah, R. (2007). "Optimal management of groundwater pumping the case of the Eocene Aquifer, Palestine" M.S. thesis, An-Najah National University, Nablus, Palestine.

Latvala, T., and Jukka, K. (2004). "Consumers' Willingness to pay for additional information on food quality and safety." *Conference Paper, 84th EAAE Seminar*.

LaValle, I.H. (1968). "On cash equivalents and information evaluation in decisions under uncertainty: Part I: Basic theory." *J. American Stat. Assn.,* 63(321) 252-276.

Loaiciga, H.A., Charbeneau, R.J., Everett, L.G., Fogg, G.E., Hobbs, B.F., Rouhani, S., (1992). "Review of ground-water quality monitoring network design." *J. Hydraulic Eng.,* 118(1) 11-37.

Lorna, F. (2004). "Drinking-water nitrate, methemoglobinemia, and global burden of disease: A discussion." *US Department of Health and Human Services*, Research Triangle Park, NC, ETATS-UNIS.

Lund, J.R. (2008). "Probabilistic design and optimization class notes." *Class Notes,* ECI 249 Winter, 2009.

MacQuarrie, K.T.B., Sudicky, E.A., Robertson, W.D. (2001). "Numerical simulation of a fine-grained denitrification layer for removing septic system nitrate from shallow groundwater." *J. Contaminant Hydrology,* 52(1-4) 29-55.

Majumdar, D. (2003). "The blue baby syndrome." *Resonance,* 8(10) 20-30.

Meehl, G.A., Stocker, T.F., Collins, W.D., Friedlingstein, P., A.T. Gaye, Gregory, J.M., Kitoh, A., Knutti, R., Murphy, J.M., Noda, A., Raper, S.C.B., Watterson, I.G., Weaver, A.J., Zhao, Z.-C. (2007). "Global climate projections.", In: Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (Ed.), Climate Change 2007: The physical science basis. Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Molin, E.J.E., and Timmermans, H.J.P. (2006). "Traveler expectations and willingness-to-pay for Web-enabled public transport information services." *Transportation Research Part C: Emerging Technologies,* 14(2) 57-67.

Najem, A. (2008). "Modeling nitrate contamination of the Eocene Aquifer, Palestine" M.S. thesis, An-Najah National University, Nablus, Palestine.

Nash, J.E., and Sutcliffe, J.V. (1970). "River flow forecasting through conceptual models part I - A discussion of principles." *J. Hydrology,* 10(3) 282-290.

Omour, A. (2011). "Reverse osmoses units' costs." Personal communication

Palestinian Central Bureau of Statistics (PCBS) (2008). "Palestine in Fig.s 2007."

Palestinian Central Bureau of Statistics (PCBS) (2009a). "Population, housing and establishment census 2007. Main indicators by locality Type."

Palestinian Central Bureau of Statistics (PCBS) (2009b). "Water statistics in the Palestinian territory. Annual Report, 2008."

Pattanayak, S.K., Smith, V.K., Van Houtven, G. (2003). "Valuing environmental health risks: from preference calibration to estimation." Working Paper 03_04, RTI International.

Perloff, J.M. (2008). "Microeconomics: Theory & applications with calculus." Pearson Addison Wesley.

Rajagopal, R. (1986). "The Effect of sampling frequency on ground water quality characterization." *Ground Water Mon. & Remediation,* 6(4) 65-73.

Repo, A.J. (1989). "The value of information: Approaches in economics, accounting, and management science." *J. American Society for Infor. Science,* 40(2) 68-85.

Roe, T., and Antonovitz, F. (1985). "A producer's willingness to pay for information under price uncertainty: Theory and application." *Southern Econ. J.,* 52(2) 382-391

Sakalaki, M., and Kazi, S. (2006). "How much is information worth? Willingness to pay for expert and non-expert informational goods compared to material goods in lay economic thinking." *J. Information Science,* 33(315).

Ticlavilca, A.M. (2010). "Multivariate Bayesian machine learning regression for operation and management of multiple reservoir, irrigation canal, and river systems" PhD dissertation. Utah State University, Logan, UT.

Tipping, M.E. (2001). "Sparse Bayesian learning and the relevance vector machine." *J. of Machine Learning Res.*,1 211-244.

Tubaileh, H. (2003). "Graundawater flow modeling - case study of the Eocene Aquifer in the West Bank, Palestine" M.S. thesis, An-Najah National University, Nablus, Palestine.

Walton, G. (1951). "Survey of literature relating to infant methemoglobinemia due to nitrate-contaminated water." *American J. Public Health,* 41 986-996.

Yokota, F., and Thompson, K.M. (2004a). "Value of information analysis in environmental health risk management decisions: Past, present, and future." *Risk Anal.,* 24(3) 635-650.

Yokota, F., and Thompson, K.M. (2004b). "Value of information literature analysis: A review of applications in health risk management." *Medical Decision Making,* 24(3) 287-298.

Zaman, B. (2010). "Remotely sensed data assimilation technique to develop machine learning models for use in water management" PhD dissertation. Utah State University, Logan, UT.

Zheng , C., and Wang , P. (1998). "MT3DMS: A modular three-dimensional multispecies transport model for simulation of advection, dispersion, and chemical reactions of contaminants in groundwater systems." Technical Report. Waterways Experiment Station, US Army Corps of Engineers, Vicksburg, MS.

APPENDICES

Appendix 1: RVM model background

Tipping (2001) introduced a general Bayesian framework for obtaining sparse solutions to regression and classification tasks utilizing models linear in the parameters. RVM is a particular specialization of the general framework.

In this model, the training set consists of input vectors $\{\mathbf{x}_n\}_{n=1}^{N}$ and corresponding targets $\{\mathbf{t}_n\}_{n=1}^{N}$. The targets might be real values (in regression) or class labels (in classification). The idea is to make accurate predictions of t for unseen values of x by learning a model of the dependency of the targets on the inputs in the training set. The principal modeling challenge here is to avoid over-fitting of the training set due to the presence of noise in real-word data.

The prediction is based on some function y(x) defined over the input space, and learning is the process of finding the parameters of this function.

$$\mathrm{y}(\mathbf{x};\boldsymbol{w}) = \sum_{i=1}^{M} w_i \boldsymbol{\varphi}_i(\mathbf{x}) = \boldsymbol{w}^{\mathrm{T}} \boldsymbol{\varphi}(\mathbf{x}) \tag{A1}$$

where the output $\mathrm{y}(\mathbf{x};\boldsymbol{w})$ is a linearly weighted sum of M, generally nonlinear and fixed basis functions $\boldsymbol{\varphi}(\mathbf{x}) = (\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x}), \ldots \varphi_{\mathrm{M}}(\mathbf{x}))^{\mathrm{T}}$ and $\boldsymbol{w} = (w_1, w_2, \ldots, w_M)^{\mathrm{T}}$, called weights, are adjustable parameters. The weights appear linearly, and the objective is to estimate `good' values for them.

For regression models, the standard probabilistic formulation for a given input-target pair $\{\mathbf{x}_n, y_n\}_{n=1}^{N}$ assumes that the targets are samples of the model with additive noise:

$$t_n = y(\mathbf{x}_n; \boldsymbol{w}) + \varepsilon_n \tag{A2}$$

where ε_n are independent samples from some noise process, which is typically assumed to be mean-zero Gaussian with variance σ^2. Thus $p(t_n | \mathbf{x}) = N(t_n | y(\mathbf{x}_n), \sigma^2)$, where the notation specifies a Gaussian distribution over t_n, with mean $y(\mathbf{x}_n)$ and variance σ^2.

Due to the assumption of independence of the t_n, the likelihood of the complete data set can be written as:

$$\boldsymbol{p}(\boldsymbol{t}|\boldsymbol{w},\sigma^2) = (2\pi\sigma^2)^{-N/2}\exp\{-\frac{1}{2\sigma^2}\|\boldsymbol{t} - \boldsymbol{\varphi w}\|^2\} \tag{A3}$$

where $\boldsymbol{t} = (t_1 \dots\ t_N)$, $\boldsymbol{w} = (w_0 \dots w_N)$ and $\boldsymbol{\varphi}$ is the Nx$(N+1)$ 'design' matrix with $\boldsymbol{\varphi}_{nm}=k(\mathbf{x}_n,\mathbf{x}_{m-1})$ and $\boldsymbol{\varphi}_{n1} = 1$.

To avoid over-fitting, a common approach is to impose some additional constraint on the parameters through the addition of a `complexity' penalty term to the likelihood or error function. This is done by making the popular choice of a zero-mean Gaussian prior distribution over w:

$$\mathrm{p}(w|\alpha) = \prod_{i=0}^{N} N(w_i|0,\alpha_i^{-1}) \tag{A4}$$

where $\boldsymbol{\alpha}$ is a vector of $N+1$ hyperparameters. So, the key feature of this model is the introduction of this hyperparameter to every weight, which ultimately leads to a sparse model.

Gamma distributions are suitable hyperpriors over $\boldsymbol{\alpha}$, as well as over the final remaining parameter in the model, the noise variance σ^2:

$$p(\alpha) = \prod_{i=0}^{N} Gamma(\alpha_i|a,b),$$
$$p(\beta) = Gamma(\beta|c,d)$$

with $\beta \equiv \sigma^2$ and where

$$Gamma(\alpha|a,b) = \Gamma(a)^{-1} b^a \alpha^{a-1} e^{-b\alpha}, \tag{A5}$$

with $\Gamma(a) = \int_0^\infty t^{a-1}e^{-t}dt$, the 'gamma function'.

After defining the priors, Bayesian inference proceeds by computing, from Bayes' rule, the posterior over all unknowns given the data:

$$p(w,\alpha,\sigma^2 \,|t) = \frac{p(t\,|w,\alpha,\sigma^2)\,p(w,\alpha,\sigma^2)}{p(t)} \tag{A6}$$

Then, given a new test point, x_*, predictions are made for the corresponding target t_*, in terms of the predictive distribution:

$$p(t_*|t) = \int p(t_*|w,\alpha,\sigma^2)p(w,\alpha,\sigma^2|t)dwd\alpha d\sigma^2 \tag{A7}$$

To facilitate the computation, the posterior distribution term is decomposed as:

$$p(w,\alpha,\sigma^2|t) = p(w|t,\alpha,\sigma^2)\,p(\alpha,\sigma^2|t) \tag{A8}$$

The posterior distribution over the weights is given by:

$$\begin{aligned} p(w|t,\alpha,\sigma^2) &= \frac{p(t|w,\sigma^2).p(w|\alpha)}{p(t|\alpha,\sigma^2)} \\ &= (2\pi)^{-(N+1)/2}|\Sigma\ |^{-1/2}.\exp\{-\frac{1}{2}(w-\mu)^T\Sigma^{-1}(w-\mu)\} \end{aligned} \tag{A9}$$

where, the posterior covariance and mean are respectively:

$$\Sigma = (\sigma^{-2}\Phi^T\Phi + A)^{-1}, \tag{A10}$$

$$\mu = \Sigma\Phi^T t\sigma^{-2} \tag{A11}$$

with $A = diag(\alpha_0,\alpha_1,........,\alpha_N)$

For prediction purposes, the hyperparameters posterior p(α,σ2|t) could be approximated by a delta function at its mode i.e. at its most probable values α_{MP}, $\sigma2_{MP}$ (Tipping 2001), given that p(t|α, σ2) p(α) p(σ-2) is peaked around its mode:

$$\int \mathrm{p}(\mathrm{t}_*|\alpha,\sigma)\delta(\alpha_{MP},\sigma_{MP})\,\mathrm{d}\alpha\,\mathrm{d}\sigma^2 \approx \int \mathrm{p}(\mathrm{t}_*|\alpha,\sigma)\mathrm{p}(\alpha,\sigma|t)\,\mathrm{d}\alpha\,\mathrm{d}\sigma^2 \tag{A12}$$

hence, learning becomes a search for the most probable hyperparameter posterior mode, i.e., with respect to $\boldsymbol{\alpha}$ and σ^2 the maximization of $p(\boldsymbol{\alpha}, \sigma^2|\boldsymbol{t}) \propto p(\boldsymbol{t}|\boldsymbol{\alpha}, \sigma^2)\, p(\boldsymbol{\alpha})\, p(\sigma^2)$. For uniform hyperpriors, one need only to maximize the term $p(\boldsymbol{t}|\boldsymbol{\alpha},\sigma^2)$ which is a convolution of two normal distributions, namely $p(\boldsymbol{t}|\boldsymbol{w},\sigma^2)$ and $p(\boldsymbol{w}|\boldsymbol{\alpha})$. Thus the corresponding variances add up as follows:

$$\mathrm{p}(t|\alpha,\sigma^2) = (2\pi)^{-N/2}\left|\sigma^2\mathbf{I} + \boldsymbol{\varphi}A^{-1}\boldsymbol{\varphi}^T\right|^{-1/2} \exp\{-\frac{1}{2}t^{\mathrm{T}}(\sigma^2\mathbf{I} + \boldsymbol{\varphi}A^{-1}\boldsymbol{\varphi}^{\mathrm{T}})^{-1}t\} \qquad \text{(A13)}$$

Differentiating with respect to α and σ, setting to zero and rearranging, ultimately gives iterative re-estimation formulae:

$$\alpha_i^{new} = \frac{\gamma_i}{\mu_i^2} \qquad \text{(A14)}$$

$$(\sigma^2)^{new} = \frac{\|t - \Phi\mu\|^2}{N - \sum_{i=1}^{M}\gamma_i} \qquad \text{(A15)}$$

where μ_i is the i-th posterior mean weight from (A11), and $\gamma_i \equiv 1 - \alpha_i N_{ii}$

Most of posterior probabilities of the weights are zero and the corresponding inputs are “irrelevant”. The non-zero elements are “Relevance Vectors”.

Predictions are based on the posterior distribution over the weights, conditioned on the maximizing value. Predictive distribution is given by:

$$p(t_*|t,\alpha_{MP},\sigma^2{}_{MP}) = \int p(t_*|w,\sigma^2{}_{MP})\,p(w|t,\alpha_{MP},\sigma^2{}_{MP})\,dw \qquad \text{(A16)}$$

Since both terms in the integral are Gaussian, this is readily computed, giving:

$$p(t_*|t,\alpha_{MP},\sigma^2{}_{MP}) = N(t_*|y_*,\sigma^2{}_*) \qquad \text{(A17)}$$

with:

$$y_* = \mu^T\phi(x_*) \qquad \text{(A18)}$$

$$\sigma^2{}_* = \sigma^2{}_{MP} + \phi(x_*)^T\Sigma\phi(x_*) \qquad \text{(A19)}$$

Appendix 2: Monitoring Cost Calculations

Well #	Depth (m)	Drilling ($US)	finishing ($US)	sampling ($US/y)
1	84.00	5077.80	4176.48	12.00
2	83.00	5017.35	4126.76	12.00
3	82.00	4956.90	4077.04	12.00
4	81.00	4896.45	4027.32	12.00
5	80.00	4836.00	3977.60	12.00
6	55.00	3324.75	2734.60	12.00
7	0.00	0.00	0.00	12.00
8	0.00	0.00	0.00	12.00
9	84.00	5077.80	4176.48	12.00
10	55.00	3324.75	2734.60	12.00
11	85.00	5138.25	4226.20	12.00
12	170.00	10276.50	8452.40	12.00
13	83.00	5017.35	4126.76	12.00
14	82.00	4956.90	4077.04	12.00
15	76.00	4594.20	3778.72	12.00
16	74.00	4473.30	3679.28	12.00
17	56.00	3385.20	2784.32	12.00
18	90.00	5440.50	4474.80	12.00
19	111.00	6709.95	5518.92	12.00
20	24.00	1450.80	1193.28	12.00
21	99.00	5984.55	4922.28	12.00
22	47.00	2841.15	2336.84	12.00
23	8.00	431.12	397.76	12.00
24	12.00	646.68	596.64	12.00
25	376.00	22729.20	18694.72	12.00
26	272.00	16442.40	13523.84	12.00
27	186.00	11243.70	9247.92	12.00
28	118.00	7133.10	5866.96	12.00
29	28.00	1692.60	1392.16	12.00
30	82.00	4956.90	4077.04	12.00
31	156.00	9430.20	7756.32	12.00
32	156.00	9430.20	7756.32	12.00
33	44.00	2659.80	2187.68	12.00
34	129.00	7798.05	6413.88	12.00
35	135.00	8160.75	6712.20	12.00
36	0.00	0.00	0.00	12.00
37	154.00	9309.30	7656.88	12.00
38	150.00	9067.50	7458.00	12.00
39	263.00	15898.35	13076.36	12.00
40	184.00	11122.80	9148.48	12.00
41	309.00	18679.05	15363.48	12.00
42	326.00	19706.70	16208.72	12.00
43	347.00	20976.15	17252.84	12.00
44	49.00	2962.05	2436.28	12.00
45	222.00	13419.90	11037.84	12.00
46	162.00	9792.90	8054.64	12.00
47	42.00	2538.90	2088.24	12.00
48	40.00	2418.00	1988.80	12.00
49	84.00	5077.80	4176.48	12.00
Present value		340,504.55	280,172.20	588.00
Future Value (30y)		1,471,641.04	1,210,888.10	39,066.04

Appendix 3: Expected Costs Calculations

Expected costs associated with "Do nothing" branch

Well index	[P1]	[P2]	Healthcare cost ($/year)	Bottled water cost ($/year)	Home treatment cost	
					RO units ($)	M&R ($/year)
1	0	1	34294.79	29280.93	2025801.11	405160.22
2	0	1	34294.79	29280.93	2025801.11	405160.22
3	0	1	34294.79	29280.93	2025801.11	405160.22
4	0	1	34294.79	29280.93	2025801.11	405160.22
5	1	0	0.00	29280.93	2025801.11	405160.22
6	1	0	0.00	328.22	22707.67	4541.53
7	0	1	34294.79	29280.93	2025801.11	405160.22
8	0	1	68589.57	58561.86	4051602.23	810320.45
9	1	0	0.00	11868.84	821146.01	164229.20
10	0	1	933.29	796.85	55129.84	11025.97
11	0	1	34294.79	29280.93	2025801.11	405160.22
12	0	1	9112.92	7780.62	538302.10	107660.42
13	1	0	0.00	1869.68	129353.63	25870.73
14	0	1	7967.13	6802.34	470620.09	94124.02
15	0	1	11308.49	9655.21	667995.45	133599.09
16	0	1	2013.68	1719.29	118948.81	23789.76
17	1	0	0.00	295.87	20469.94	4093.99
18	0	1	3.85	3.29	227.51	45.50
19	0	1	15196.44	12974.74	897657.41	179531.48
20	1	0	0.00	2578.28	178378.02	35675.60
21	0	1	498.94	425.99	29472.29	5894.46
22	0	1	1288.67	1100.26	76121.81	15224.36
23	0	1	544.92	465.25	32188.42	6437.68
24	0	1	1675.67	1430.69	98982.20	19796.44
25	0	1	1387.33	1184.50	81949.89	16389.98
26	0	1	761.71	650.35	44994.60	8998.92
27	1	0	0.00	2643.89	182917.37	36583.47
28	0	1	1516.49	1294.78	89579.37	17915.87
29	0	1	5098.21	4352.86	301152.48	60230.50
30	0	1	1802.51	1538.98	106474.55	21294.91
31	0	1	2683.67	2291.32	158525.17	31705.03
32	0	1	5921.84	5056.07	349804.43	69960.89
33	0	1	316.96	270.62	18723.08	3744.62
34	0	1	4995.96	4265.56	295112.47	59022.49
35	1	0	0.00	7028.55	486270.18	97254.04
36	0	1	3222.29	2751.19	190341.16	38068.23
37	0	1	1364.40	1164.93	80595.72	16119.14
38	0	1	5227.55	4463.29	308792.86	61758.57
39	0	1	13304.29	11359.22	785887.73	157177.55
40	0	1	151.50	129.35	8949.37	1789.87
41	0	1	7376.84	6298.35	435751.38	87150.28
42	0	1	6812.24	5816.30	402400.46	80480.09
43	0	1	6916.07	5904.95	408533.96	81706.79
44	0	1	18632.15	15908.15	1100605.28	220121.06
Total	Present value		412,394.31	407,997.00	28,227,272.73	5,645,454.55
	Future value		27,399,002.9	27,106,850.46	121,996,646.1	375,077,493.6

Expected costs associated with "monitoring system" branch

Well index	[P2/p1]	[P2/p2]	[p1]	[p2]	Healthcare * [P2/p1] * [p1]	Bottled * [p1]	Bottled * [p2]	Healthcare * [P2/p2] * [p2]
1	1	1	0	1	0.00	0.00	29280.93	34294.79
2	1	1	0	1	0.00	0.00	29280.93	34294.79
3	1	1	0	1	0.00	0.00	29280.93	34294.79
4	1	1	0.39	0.61	13374.97	11419.56	17861.37	20919.82
5	0	0	0.33	0.67	0.00	9662.71	19618.22	0.00
6	0	0	0.3	0.7	0.00	98.47	229.75	0.00
7	1	1	0	1	0.00	0.00	29280.93	34294.79
8	1	1	0	1	0.00	0.00	58561.86	68589.57
9	0	0	0.2	0.8	0.00	2373.77	9495.08	0.00
10	1	1	0.1	0.9	93.33	79.68	717.16	839.96
11	1	1	0	1	0.00	0.00	29280.93	34294.79
12	1	1	0	1	0.00	0.00	7780.62	9112.92
13	1	0	0	1	0.00	0.00	1869.68	0.00
14	1	1	0	1	0.00	0.00	6802.34	7967.13
15	1	1	0	1	0.00	0.00	9655.21	11308.49
16	1	1	0	1	0.00	0.00	1719.29	2013.68
17	1	0	0	1	0.00	0.00	295.87	0.00
18	1	1	0	1	0.00	0.00	3.29	3.85
19	1	1	0	1	0.00	0.00	12974.74	15196.44
20	1	0	0	1	0.00	0.00	2578.28	0.00
21	1	1	0	1	0.00	0.00	425.99	498.94
22	1	1	0	1	0.00	0.00	1100.26	1288.67
23	1	1	0	1	0.00	0.00	465.25	544.92
24	1	1	0	1	0.00	0.00	1430.69	1675.67
25	1	1	0	1	0.00	0.00	1184.50	1387.33
26	1	1	0	1	0.00	0.00	650.35	761.71
27	0	0	0.22	0.78	0.00	581.66	2062.23	0.00
28	1	1	0	1	0.00	0.00	1294.78	1516.49
29	1	1	0	1	0.00	0.00	4352.86	5098.21
30	1	1	0.16	0.84	288.40	246.24	1292.75	1514.11
31	1	1	0	1	0.00	0.00	2291.32	2683.67
32	1	1	0	1	0.00	0.00	5056.07	5921.84
33	1	1	0	1	0.00	0.00	270.62	316.96
34	1	1	0.29	0.71	1448.83	1237.01	3028.54	3547.13
35	0	0	0.29	0.71	0.00	2038.28	4990.27	0.00
36	1	1	0	1	0.00	0.00	2751.19	3222.29
37	1	1	0	1	0.00	0.00	1164.93	1364.40
38	1	1	0	1	0.00	0.00	4463.29	5227.55
39	1	1	0	1	0.00	0.00	11359.22	13304.29
40	1	1	0	1	0.00	0.00	129.35	151.50
41	1	1	0	1	0.00	0.00	6298.35	7376.84
42	1	1	0	1	0.00	0.00	5816.30	6812.24
43	1	1	0	1	0.00	0.00	5904.95	6916.07
44	1	1	0	1	0.00	0.00	15908.15	18632.15

Expected costs associated with "perfect monitoring" branch

Well index	[P1]	[P2]	Bottled water cost ($/year)
1	0	1	29280.93
2	0	1	29280.93
3	0	1	29280.93
4	0	1	29280.93
5	1	0	0.00
6	1	0	0.00
7	0	1	29280.93
8	0	1	58561.86
9	1	0	0.00
10	0	1	796.85
11	0	1	29280.93
12	0	1	7780.62
13	1	0	0.00
14	0	1	6802.34
15	0	1	9655.21
16	0	1	1719.29
17	1	0	0.00
18	0	1	3.29
19	0	1	12974.74
20	1	0	0.00
21	0	1	425.99
22	0	1	1100.26
23	0	1	465.25
24	0	1	1430.69
25	0	1	1184.50
26	0	1	650.35
27	1	0	0.00
28	0	1	1294.78
29	0	1	4352.86
30	0	1	1538.98
31	0	1	2291.32
32	0	1	5056.07
33	0	1	270.62
34	0	1	4265.56
35	1	0	0.00
36	0	1	2751.19
37	0	1	1164.93
38	0	1	4463.29
39	0	1	11359.22
40	0	1	129.35
41	0	1	6298.35
42	0	1	5816.30
43	0	1	5904.95
44	0	1	15908.15
Total	Present value		352,102.75
	Future value		23,393,300.69

Appendix 4: Pumping wells and associated population

Well index	Pumping rate mcm/year	% of total pumping	Number of households	Affected population
1	1.30	7.18	2701	134
2	1.30	7.18	2701	134
3	1.30	7.18	2701	134
4	1.30	7.18	2701	134
5	1.30	7.18	2701	134
6	0.01	0.08	30	1
7	1.30	7.18	2701	134
8	2.60	14.35	5402	267
9	0.53	2.91	1095	54
10	0.04	0.20	74	4
11	1.30	7.18	2701	134
12	0.35	1.91	718	36
13	0.08	0.46	172	9
14	0.30	1.67	627	31
15	0.43	2.37	891	44
16	0.08	0.42	159	8
17	0.01	0.07	27	1
18	0.00	0.00	0	0
19	0.58	3.18	1197	59
20	0.11	0.63	238	12
21	0.02	0.10	39	2
22	0.05	0.27	101	5
23	0.02	0.11	43	2
24	0.06	0.35	132	7
25	0.05	0.29	109	5
26	0.03	0.16	60	3
27	0.12	0.65	244	12
28	0.06	0.32	119	6
29	0.19	1.07	402	20
30	0.07	0.38	142	7
31	0.10	0.56	211	10
32	0.22	1.24	466	23
33	0.01	0.07	25	1
34	0.19	1.05	393	19
35	0.31	1.72	648	32
36	0.12	0.67	254	13
37	0.05	0.29	107	5
38	0.20	1.09	412	20
39	0.50	2.78	1048	52
40	0.01	0.03	12	1
41	0.28	1.54	581	29
42	0.26	1.43	537	27
43	0.26	1.45	545	27
44	0.71	3.90	1467	73

Appendix 5: Survey Questions and Letter of Information

1. What water sources do you currently use indoors and outdoors for drinking, cooking, bathing, house cleaning, irrigating, and livestock? Circle all that apply.

Indoor use **(drinking, cooking, bathing, house cleaning)**	**Outdoor use** **(landscaping, car cleaning, livestock)**
Pipe network / Tap	Pipe network / Tap
Tanker truck	Tanker truck
Rainwater	Rainwater
Bottled water	Bottled water
Home treatment unit	Home treatment unit
Other ___________________________	Other ___________________________

2. How many people live in your house and use the sources you identified in question #1?

   Total: _________________

   Aged 0 - 2 years: _________ Aged 2 - 16 years: ___________ 17 and above:__________

3. What is the source(s) of the water that the government delivers to you through the pipe network?

   A) Well - Eocene aquifer B) Well - another aquifer C) Spring D) Other (please specify) _________ E) Don't know

4. Rate your satisfaction with the water quantify and quality the government delivers to you through the pipe network

   Quantity: 1 (Satisfied) 2 3 4 5 (unsatisfied).

   Quality: 1 (Satisfied) 2 3 4 5 (unsatisfied).

5. Please explain your answers to question #4. Why did you respond that way?

   ____
   ____
   ____
   ____

6. Government officials have not tested the Eocene Aquifer but declare it safe to drink and use. What will you do?

   A) Use the aquifer water B) Use bottled water C) Use home treatment

   D) Other (please specify) ________

7. If you answered B, C, or D to question 6, what testing would government officials need to do for you to follow their recommendations?

   ____
   ____
   ____

8. If government officials have not tested the Eocene Aquifer water, declare that the aquifer is **not** safe, _and recommend consumers use other water sources (such as bottled water or home treatment units), what will you do?
   A) Use the aquifer water B) Use bottled water C) Use home treatment

   D) Other (please specify) ________

9. If you answered A to question 8, what testing would government officials need to do for you to follow their recommendations?

   ____
   ____
   ____

10. If government officials collect samples from selected monitoring wells in the aquifer and analyze these samples for groundwater pollutants such as nitrate. After comparing the results with WHO standards they declare that they think that the aquifer is safe and they recommend consumers to use the aquifer water, what will be your response?
    A) Use the aquifer water B) Use bottled water C) Use home treatment

    D) Other (please specify) ________

11. If government officials test the water as described above in Question #10, declare based on the test results that they think the aquifer is **not** safe, and recommend consumers to use other sources of water (such as bottled water or home treatment units), what will be your response?

A) Use the aquifer water B) Use bottled water C) Use home treatment

D) Other (please specify) _________

12. City/Village in which you live? _______________

LETTER OF INFORMATION

Value of information from monitoring network design in the Eocene Aquifer, Palestine

Professor Mac McKee and doctoral degree candidate Abdelhaleem I. Khader in the Department of Civil and Environmental Engineering at Utah State University are conducting research to study the design of a monitoring network in the Eocene Aquifer, Palestine. **Mr. Khader is asking you to complete and return the attached survey** because the Eocene Aquifer is one of the sources for freshwater in your area. The questions in the survey ask you about the water sources you currently use and seek your responses to potential recommendations by the water supply network managers on whether to use water delivered to you from the aquifer indoors or substitute water from alternative sources. The survey should take approximately 10 minutes to complete.

Your participation in this research is entirely voluntary. You may refuse to participate or withdraw at any time without consequence or loss of benefits. And we will consider **your returning a completed survey as providing your informed consent to participate in the study**.

By participating, you face a small risk of loss of confidentiality. However, we will reduce this risk by not including nor ask you to provide any information that can identify you individually. Further, all research records will be kept confidential, consistent with federal and state regulations. Only the investigator will have access to the data which will be kept in a locked file cabinet or on a password protected computer in a locked room.

Your responses to this survey will help decision makers to decide whether to implement a groundwater monitoring network. This decision will affect the communities using the Eocene Aquifer including your household. The effects of this decision may include reducing health risk due to water quality contamination and lowering costs to supply water.

After completing the survey, please return it to the researcher (Abdelhaleem Khader). If you need more time or you prefer to return it later, please call the Mr. Khader at 059-9758363 and he will coordinate with you to pick it up.

The Institutional Review Board for the protection of human participants at Utah State University has approved this research study. If you have any questions or concerns about your rights or a research-related injury and would like to contact someone other than the research team, you may contact the IRB Administrator at (435) 797-0567 or email irb@usu.edu to obtain information or to offer input.

"I certify that the research study has been explained to the individual, by me or my research staff, and that the individual understands the nature and purpose, the possible risks and benefits associated with taking part in this research study. Any questions that have been raised have been answered."

Abdelhaleem Khader
001 (435) 881-0737 (United States)
00970 (59) 9758363 (Palestine)
Abdelhaleem.khader@aggiemail.usu.edu

____

Mac McKee
001(435) 797-3188
mac.mckee@usu.edu

____

David Rosenberg
001(435) 797-8689
david.rosenberg@usu.edu

Appendix 6: Permission Letter

Date: 6/14/12
Abdelhaleem Khader
35 Aggie Village Apt L
Logan, UT 84341
(435) 881-0737

Dear Dr. David Rosenberg,
I am in the process of preparing my dissertation in Civil and Environmental Engineering at Utah State University. I hope to complete in the summer of 2012.
I am requesting your permission to include Chapter 3 and Chapter 4 which were coauthored by you. I will include acknowledgments and appropriate citations to your work.
Please indicate your approval of this request by signing in the space provided. If you have any questions please call me at the number above.
Thank you for your cooperation,
Abdelhaleem Khader

I hereby give my permission to Abdelhaleem Khader to reprint Chapter 3 and Chapter 4 in his dissertation.

David Rosenberg
Signed____

CURRUCULUM VITAE

ABDELHALEEM I. KHADER
Department of Civil and Environmental Engineering
Utah State University, Logan, UT 84322
(435) 881-0737
abdelhaleem.khader@aggiemail.usu.edu

EDUCATION

PhD, Civil and Environmental Engineering May 2012
Utah State University, Logan, Utah, USA **GPA 3.90**
Dissertation: "Value of information from groundwater quality monitoring network design under uncertainty in climate and aquifer properties" Advisor: Dr. Mac McKee

M.S, Water and Environmental Engineering April 2007
An-Najah National University, Nablus, Palestine **Average 90.9%**
Thesis: "Impact of Pumping on Saltwater Intrusion in Gaza Coastal Aquifer, Palestine" Advisor: Dr. Mohammad Almasri

B.S, Civil Engineering January 2004
An-Najah National University, Nablus, Palestine **Average 89.1%**
Project: "Seismic and Structural Design of the new Engineering Building, An-Najah National University " Advisor: Dr. Abdel-Razzaq Touqan

RESEARCH INTERESTS

Monitoring Network Design	Statistical Learning Machines
Value of Information Analysis	Groundwater Modeling
Water/Groundwater Quality	Decision Tree Models

HONORS AND AWARDS

First Place, J. Paul Riley AWRA-Utah Section Student Water Resources Conference and Paper Competition April 10, 2012 Logan, UT

TEACHING AND RESEARCH EXPERIENCE

Graduate Student / Research Assistant 2007-present
Utah Water Research Laboratory, Utah State University Logan, UT

- Groundwater flow modeling for the Eocene Aquifer, Palestine using MODFLOW
- Nitrate fate and transport modeling for the Eocene Aquifer using MT3DMS
- Uncertainty analysis using Monte Carlo Simulations
- Monitoring network design using statistical learning machines
- Studying the health risk consequences of nitrate pollution
- Value of information analysis for optimal monitoring network design
- Pre-mining Groundwater analysis for potash mining in Lisbon Valley in southeastern Utah (undergoing project)

Instructor, Iraqi Agriculture Extension Revitalization Program October 2009
Utah State University Logan, UT

- Conducted lectures in water quality and hydrology
- Prepared tests, grading, and evaluations
- Translated for Arabic speakers and I led discussions in field trips to southern Utah

Master Student, Water and Environmental Studies Institute 2004-2007
An-Najah National University Nablus, Palestine

- Worked on saltwater intrusion modeling using MODFLOW, SEAWAT, and GWM

Teaching Assistant, Civil Engineering Department 2004
An-Najah National University Nablus, Palestine

- Instruction, grading, and preparing tests for Construction Materials Lab and Structural Analysis II

WORK EXPERIENCE

Site Engineer, Engineering Works Department 2004-2007
An-Najah National University Nablus, Palestine

- Worked in supervising the new science building. Total cost of the project: $8,000,000. Total area: 18,000 m^2
- Prepared bills of quantities and as-built maps
- Modified structural designs when needed
- Supervised daily activities of 120 construction workers
- Reviewed monthly bills by the contractor
- Reviewed and signed monthly payments for the contractor

TRAINING COURSES

Getting Started as a Successful Proposal Writer and Academician April 2012 Logan, Utah
An intensive one-day workshop for beginning concepts in grant writing
Office of Research and Graduate Studies, Utah State University

Integrated Water Resources Management (IWRM) September 2005
Water Studies Institute, Birzeit University, Palestine. Birzeit, Palestine

Seismic Design of Buildings December 2003
Engineers Association – Jerusalem Center Nablus, Palestine

CONFERENCE PRESENTATIONS

- A.Khader, M. McKee (2010). "Value of information analysis for groundwater quality monitoring network design". American Geophysical Union (AGU) Fall meeting. San Francisco, CA 2010.
- A.Khader, M. McKee (2010). "Groundwater Monitoring Network Design Under Uncertainty in Climate and Aquifer Properties". Utah State University Spring runoff conference. Logan, UT 2010.

- A.Khader, M. McKee (2010). "Analyzing the Impacts of Climate Change on Groundwater Monitoring Network Design Using GIS". American water resources association (AWRA) spring specialty conference. Orlando, FL 2010.
- A.Khader, M. Amasri (2008). "Impact of Pumping on Saltwater Intrusion in the Gaza Coastal Aquifer, Palestine". Universities council on water resources (UCOWR) conference. Durham, NC 2008.

PROFESSIONAL AFFILIATIONS

American Geophysical Union
American Society of Civil Engineers
Jordanian Engineers Association – Jerusalem Center

LANGUAGES

English (fluent)
Arabic (fluent)

COMPUTER SKILLS

MATLAB	R
LINGO	ArcGIS
ERDAS	MODFLOW
MT3DMS	SEAWAT
EPANET	GWM
MS PROJECT	AUTOCAD
MS OFFICE	

RELEVANT COURSES

PhD:

Physical Hydrology	Groundwater Engineering
Environmental Data Analysis	Remote Sensing
Microeconomics	Water Resources Systems Analysis
Statistical Learning	Evaluation of Hydrological Models
Environmental Quality Analysis	Physical/Chemical Environmental Processes

Masters:

Water Chemistry (with Lab.)	Groundwater
Water Treatment	Advanced Hydrology
Environmental Impact Assessment	GIS Applications
Solid Waste Management	Design of Hydraulic Structures
Public Health and Sanitation	Probability and Statistics