Title

Data-Worth Assessment for a Three-Dimensional Optimal Design in Nonlinear Groundwater Systems

Journal

Ground Water, 57(4)

ISSN

0017-467X

Authors

Safi, Amir
Vilhelmsen, Troels N
Alameddine, Ibrahim
et al.

Publication Date

2019-07-01

DOI

10.1111/gwat.12835

Peer reviewed

Data-worth assessment for a 3D optimal design in non-linear groundwater systems

Amir Safi
Department of Civil and Environmental Engineering, American University of Beirut, Beirut, Lebanon

Troels Norvin Vilhelmsen
Department of Geoscience, Aarhus University, Aarhus, Denmark

Ibrahim Alameddine
Department of Civil and Environmental Engineering, American University of Beirut, Beirut, Lebanon

Majdi Abou Najm
Department of Civil and Environmental Engineering, American University of Beirut, Beirut, Lebanon; Now at the Department of Land, Air, and Water Resources, University of California, Davis

Mutasem El-Fadel
Corresponding author: mfadel@aub.edu.lb
Department of Civil and Environmental Engineering, American University of Beirut, Bliss Street, PO Box 11-0236, Beirut, Lebanon 1107 2020; Telephone +9613228338

Conflict of interest: None

Keywords: Groundwater monitoring networks, Prediction Uncertainty, Optimal Design, Data Worth, Bayesian model averaging.

Impact statement: A new method is developed to identify optimal groundwater sampling locations with minimal observation wells in space and depth.

Abstract

Groundwater model predictions are often uncertain due to inherent uncertainties in model input data. Monitored field-data are commonly used to assess the performance of a model and reduce its prediction uncertainty. Given the high cost of data collection, it is imperative to identify the minimum number of required observation wells and to define the optimal locations of sampling points in space and depth. This study proposes a design methodology to optimize the number and location of additional observation wells that will effectively measure multiple hydrogeological parameters at different depths. For this purpose, we

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/gwat.12835

incorporated Bayesian model averaging (BMA) and genetic algorithms (GA) into a linear data-worth (DW) analysis in order to conduct a three-dimensional location search for new sampling locations. We evaluated the methodology by applying it along a heterogeneous coastal aquifer with limited hydrogeological data that is experiencing saltwater intrusion. The aim of the model was to identify the best locations for sampling head and salinity data, while reducing uncertainty when predicting multiple variables of saltwater intrusion. The resulting optimal locations for new observation wells varied with the defined design constraints. The optimal design depended on the ratio of the start-up cost of the monitoring program and the installation cost of the first observation well. The proposed methodology can contribute towards reducing the uncertainties associated with predicting multiple variables in a groundwater system.

Introduction

Groundwater models are commonly used in conjunction with field monitoring to assess the physical processes representing subsurface flow and solute transport. Such models simulate the groundwater dynamics in an aquifer by translating its physical, chemical, and biological characteristics into mathematical equations by simplifying assumptions (Holzbecher and Sorek 2006). These equations require data about aquifer characteristics- such as hydraulic properties, geological borders, boundary conditions, sources and sinks- that will sufficiently aid in understanding groundwater dynamics (Bakalowicz 2005). However, the complexity of subsurface conditions may lead to a paucity in data describing the control parameters; this in turn will result in increased uncertainties with model simulations (El-Fiky 2010). The lack of data coupled with model prediction uncertainty makes it difficult for water resources managers and decision makers to plan a management strategy to secure the quantity and quality of groundwater (Tribbia and Moser 2008; Comte et al. 2006). Therefore, it is imperative to design a monitoring network that would reduce prediction uncertainties in order

to improve the protection and management of aquifer systems (Storck 1997). In this context, models can be used as test beds to identify new (optimal) monitoring locations that would increase the reliability of model simulations. This technique is generally referred to as an optimal design (OD) with several reported methods to guide the design of monitoring networks towards reducing uncertainties in model predictions (Tiedeman et al. 2003; Herrera et al. 2000; Reed et al. 2000; Cieniawski et al. 1995; Wagner 1995; Andricevic and Foufoula-Georgiou 1991; Loaiciga 1989; Rouhani and Hall 1988).

A recently developed method by Moore and Doherty (2005), and later extended within the Bayesian context by Christensen and Doherty (2008), evaluated the variance of prediction uncertainty using a linear propagation of uncertainties associated with parameters that are formulated for distributed models. Using this method, an existing calibration dataset is augmented by adding new observations. The worth of such an addition (subsequently referred to as data worth (DW)) on reducing model prediction uncertainty is then evaluated. Dausman et al. (2010) applied the DW-based OD on the Henry problem to define the optimal locations of salinity concentration and temperature that would reduce the uncertainty of predicting the displacement of a salt/fresh water interface caused by a change in the inflow rate to the system. Wallis et al. (2014) extended the DW-based OD for selecting multiple observations and Wöhling et al. (2016) extended it further by using a genetic algorithm (GA) to incorporate multiple new observations of head and/or hydraulic conductivity to decrease the predictive uncertainty. Vilhemson and Ferre (2017) carried out yet another extension to simultaneously select multiple new measurements targeting multiple forecasts of interest. Note that the applications of the DW-based OD are largely restricted to a single or multiple observation location(s) in two dimensions. The simultaneous DW-based optimization of monitoring design with measurements in three dimensions has not been reported. Expanding

the DW approach into three-dimensional space is particularly important for monitoring contaminant concentrations and to design a solute transport system.

In practice, groundwater contamination distribution varies spatially and vertically over a geologic domain. Three dimensions (3D) optimal design aids in identifying optimal locations for monitoring contaminant concentration such as saltwater intrusion (SWI). 3D models are especially useful when attempting to understand SWI, where the spread of intrusion typically occurs in three dimensions through lateral and vertical displacement (or upconing) of the interface. Neglecting three-dimensionality in geologic input data (i.e. hydraulic conductivity) can result in large uncertainties with regards to model predictions (Werner et al. 2013), influencing both the magnitude and the trend of the intrusion (Kerrou and Renard 2010). It can also lead to the overestimation of the toe penetration length (Lu et al. 2009), and subsequently to the misevaluating the DW of a proposed design when the OD target is to increase the reliability of a model in predicting the future position of the interface. To avoid such situations, ideally hydraulic conductivity values should be collected in all spatial directions in order to reduce prediction uncertainty. However, financial constraints and/or spatial limitations (e.g. in urbanized aquifers) reduce to the ability to directly measure the hydraulic conductivity from deeper parts of an aquifer (Hartmann, et. al. 2014). Using inverse modeling, hydraulic conductivity values can be estimated through the inverse solution of groundwater flow and/or solute transport equations for the value of an observed dependent variable (or an indirect observation) such as hydraulic head and/or contaminant concentration (Hoeksema and Kitanidis, 1984). Dausman et al. (2010) recognized that salinity concentration defines the interface, and that measuring salinity is crucial to understanding how the interface moves.

Total cost and utility of measuring head and/or salinity depends on the operation, implementation and start-up costs of the monitoring plan, types of monitoring equipment, resolution of monitoring data, and data processing costs (Hericks et al. 2017). In practice, nearby existing supply wells are usually used as observation wells to monitor head and/or contaminant concentration data (Sen 2015). Using existing wells will result in avoiding extra costs for drilling new observation wells.

The DW methodology assumes that an initial baseline can be established using historical data and available information about the main hydrogeological characteristics of the aquifer under study (e.g. boundary conditions, and source/sinks). The initial baseline is expected to provide insight into the overall water flow system and serve as a testbed to estimate hydraulic heads, conduct model simulations, and calculate the sensitivity of head and predictions to model parameters. However, uncertainties in model predictions are common when attempting to predict beyond the range of available input data, for example in highly parameterized models with more unknown parameters than observations. In such nonlinear models, the entire range of possible values of observational data should be considered during the OD in order to calculate a wide range for the sensitivities of observations to the model parameters because the actual values of the observation data are unknown prior to collection (Leube et al. 2012). If the model non-linearity is high, it may result in multiple plausible observation locations.

Several Monte Carlo (MC) based techniques have been developed to account for model uncertainty. MC techniques can be used to generate a range of measurement values obtained by using different parameter sets (multiple realizations) that are conditioned by the calibration dataset (Keating et al. 2010). Compared with other techniques such as the Markov Chain Monte Carlo (MCMC) method (Harvey and Gorelick, 1995), the generalized likelihood uncertainty estimation (GLUE) (Beven and Binley, 1992), and calibration-

constrained Monte Carlo methods (Tavakoli et al., 2013), the subspace technique for calibration-constrained Monte Carlo analysis referred to as the null-space Monte Carlo (NSMC) method (Tonkin and Doherty 2005) requires less computational time for generating a large number of calibration-constrained parameter fields. The NSMC method is best suited for groundwater modeling in highly parameterized systems due to the long model runtime and the large number of required model runs for such systems (Herckenrath et al. 2011). In this method, a set of random parameter realizations is first generated from a probability distribution defined from available prior information, for example about the hydraulic conductivity. The generated random realizations are then projected onto the null-space, and adjusted through model re-calibration. This results in a set of calibration-constrained realizations (for details see Tonkin and Doherty 2009). For the OD, calibration-constrained realizations can be ranked and assigned weights according to their goodness of fit with observed data. Using Bayesian model averaging (BMA), model weights are determined via Bayes' theorem from the likelihood that the calibration dataset are generated from realizations (Hoeting et al. 1999). When applying the OD, the optimal location of a new observation can be determined by averaging all possible locations obtained using calibration-constrained realizations (Freeze et al. 1992). In this manner, parameter and prediction uncertainties are both considered when attempting to find the optimal locations for new observations.

An ideal DW-based OD should provide flexibility concerning model dimensionality, allow for any desired task-oriented formulation, target any measurement type (direct and indirect), account for various sources of uncertainty (e.g. geologic structure, heterogeneity, boundary condition, and source/sink) while also ensuring that it is cost-effective. Existing DW-based OD methodologies fall short of simultaneously providing these criteria for the design of a monitoring network in a groundwater system.

In this study, we expand the DW-based OD method to optimize simultaneous measurements of various data types collected at different depths at a single and multiple spatial locations, while considering model non-linearity through a Bayesian model averaging (BMA) framework and minimizing costs. The cost-effective solution involves obtaining adequate hydrogeological information with a minimum number of observation wells. In what follows, we describe the theory underpinning the design methodology followed by the design method that is applied to a case study that reduces the uncertainty in the predictions of SWI by determining the optimal location(s) that would allow for efficiently obtaining data within the model domain.

Methods and Materials

The proposed methodology involves a Bayesian model averaging (BMA) framework with a three-dimensional data worth (DW)-based optimal design (OD) analysis that is implemented to select an optimal observation dataset that would reduce model uncertainty (Figure 1).

Figure 1: Optimal Design framework
X_{old}: Sensitivity matrix of existing observations to parameters; Y: Sensitivity matrix of predictions to parameters; C(e): Covariance matrix of measurement noise (e); C(P): Covariance matrix of parameters (P) innate variability; X: Jacobian matrix; ∂^2_{base} is the base predictive uncertainty variance; X_{new}: Sensitivity matrix of new observation wells to parameters; L: Sensitivity matrix of an observation well with measurements at multiple depths to parameters; D_k is a proposed design using model M_k; and Po, N is operation cost of a monitoring project for N new observations

Bayesian model averaging (BMA) framework

We denote a set of distributed groundwater models M_k: k=1, …, K for predicting flow and/or solute transport over a geologic domain. Each model is a probability distribution model comprising the likelihood function $P(h_0|P_k, M_k)$ of the observed data h_0 and the model parameters P_k (e.g. hydraulic conductivity). The posterior predictive distribution of the forecast of interest Δ is determined as a weighted averaged individual prediction as expressed in Equation 1, where weights can be determined using Bayes' theorem on the basis of the likelihood that the observed data h_0 are generated using equation 2 (Hoeting et al. 1999):

$$P(\Delta|h_0) = \sum_{k=1}^{m} P(\Delta|h_0, M_k)P(M_k|h_0) \quad (1)$$

$$P(M_k|h_0) \propto P(h_0|M_k)P(M_k) \quad (2)$$

where $P(M_k)$ is a probability mass function over the model M_k. The use of a defuse prior $P(M_k) = 1/m$ ensures that there is no subjective preference for any of the model M_k (Wöhling et al. 2015). One can normalize the weights of the models by applying Bayes theorem as expressed below:

$$P(M_k|h_0) = \frac{P(h_0|M_k)P(M_k)}{\sum_{k=1}^{m} P(h_0|M_k)P(M_k)} \quad (3)$$

where $P(h_0|M_k)$ is the likelihood of observing the calibration dataset h_0 under model M_k. It can be determined based on its prior parameter distribution using equation 4:

$$P(h_0|M_k) = \int P(h_0|M_k, P_k)\, P(P_k|M_k) dP_k \quad (4)$$

where $P(P_k|M_k)$ is created by generating random parameter fields that meet calibration constraints. We use the NSMC method to create random parameter fields (P_k). Prior to incorporating the NSMC method, the hydraulic conductivity field is parameterized by defining a large number of pilot points (P_k) that cover the geologic domain. The model is then calibrated to estimate the values of the pilot points. The extent to which a pilot point parameter can be informed (identified) by the existing observations can be measured by a singular value decomposition of the Jacobian matrix that represents the sensitivity of observations to the pilot point parameters (for details see e.g. Doherty and Hunt 2010). The pilot points corresponding to the singular values that are larger than a given user-defined "truncated" value (5.0×10^{-6} in this work) span the calibration solution space. These pilot points are deemed to be estimable on the basis of existing observations. In contrast, pilot points that contain low or zero singular values (that span the calibration null space) are considered inestimable. Using the NSMC method, a set of random values is first generated

from a prior probability distribution of hydraulic conductivity fields. The random values are placed on the pilot points (which is called a random realization). The generated random realization is then projected onto the null-space through differencing the random values and the calibrated pilot points' values, and then re-adding the projected difference onto the calibrated values. The projected parameter set is then adjusted through a model re-calibration in order to respect calibration constraints made by existing observations. The result is a calibrated-constrained realization that respects both the stochastic variability of the hydraulic conductivity field as well as the calibration constraints (for details see Tonkin and Doherty 2005). The posterior probability distribution of a model prediction is then computed on the basis of the generated calibrate-constrained realizations.

Linear model calibration

We assume that each Bayesian model M_k is a linear model that defines a relationship between its parameters and its predictions using equation (5) (Doherty 2015):

$$M_k: h = XP + \varepsilon \quad (5)$$

where h denotes a $m \times 1$ vector of head observations comprising the calibration dataset that are contaminated with noise ε (i.e. error in field measurement), P represents a $n \times 1$ vector of model (pilot points) parameters in the conceptual model, and X is the action of model or model sensitivity (or Jacobian matrix). The unknown model parameters can be estimated by minimizing an objective function that is defined based on the sum of the squared weighted residuals between the model results and (potential) observed data (or model-to-measurement misfit) as shown in equation 6:

$$\varphi = (h - X\bar{P})^T \, Q \, (h - X\bar{P}) \quad (6)$$

where h is a $m \times 1$ vector of potential head observations, $\bar{P}$ is the vector comprised of unknown parameters (or parameter estimates), T stands for the matrix transpose operation,

and Q is a diagonal matrix with squared observations weights (ω_i) that is defined to be proportional to the inverse of the covariance matrix of the observations noise. The parameter vector minimizing the objective function (in equation 6) can be determined using equation 7:

$$\bar{P} = (X^T Q X)^{-1} X^T Q \bar{h} \quad (7)$$

where Q is the matrix with the squared (calculated) weights of observations. The potential wrongness (or error) of the estimated parameters $\bar{P}$ compared with the true parameter fields P can be evaluated by equation 8:

$$P - \bar{P} = IP - (X^T Q X)^{-1} X^T Q h \quad (8)$$

where I is the identity matrix. Let us further assume that s denotes a true model prediction. Then, the relationship between s and P is estimable using equation 9:

$$s = y^T P \quad (9)$$

where y is a $n \times 1$ vector representing the sensitivity of the predictions to model parameters. If $\bar{s}$ is a model prediction that is computed from $\bar{P}$ using equation 6, then the potential error in the computed prediction can be expressed by equation 10:

$$s - \bar{s} = y^T (P - \bar{P}) \quad (10)$$

However, the true parameter fields (P) and prediction (s) are unknown. Therefore, none of the potential wrongness of (or error in) the estimated parameter (in equation 8) and the computed prediction (in equation 10) can be calculable. If we assume that P and ε are independent and their covariance matrices are known, then the covariance of the parameter error can be expressed by equation 11:

$$C(P - \bar{P}) = (I - R) C(P) (I - R)^T + E C(\varepsilon) E^T \quad (11)$$

where $C(P)$ is a $n \times n$ parameter covariance matrix representing innate parameter variability, which can be created using a Kriging variogram that is defined to represent the spatial distribution of the hydraulic conductivity fields, $C(\varepsilon)$ is a $m \times m$ matrix of measurement

noise ε, $R = (X^T X)^{-1} X^T \bar{Q} X$ is a so-called resolution matrix that describes the relationship between the estimated and true parameters, and $E = (X^T \bar{Q} X)^{-1} X^T \bar{Q}$.

Prediction uncertainty variance

Combing equations 10 and 11 leads to the expression of the variance of model prediction uncertainty as shown in equation 12 (Christensen and Doherty 2008):

$$\sigma^2{}_s = y^T C(P) y - y^T C(P) X^T [X C(P) X^T + C(\varepsilon)]^{-1} X C(P) y \qquad (12)$$

The first term on the right-hand side of the equation is the precalibration uncertainty for the predictions. The second term shows the amount that the prediction uncertainty is reduced by calibrating the model using measurements comprising the calibration dataset.

Data Worth (DW) analysis (prediction of single variable)

The variance of the model prediction uncertainty does not account for the values of parameters, measurements or prediction. Instead, it comprises only of the sensitivity of the model's observations and predictions to the parameters, which are included in the X and y matrices respectively. For the purpose of an optimal design, the change in the prediction uncertainty can be evaluated when a new observation (set) is added to the existing calibration dataset. In general, when adding an observation, it reduces the model prediction uncertainty, while increasing the DW that this observation has on the calibration dataset. The DW is measured using equation 13 (More details can be found in Vilhelmsen and Ferre 2017):

$$DW = \frac{\sigma^2{}_{dec}}{\sigma^2{}_{base}} \qquad (13)$$

where the $\sigma^2{}_{dec}$ is the decrease in the prediction uncertainty when adding a new observation point, and $\sigma^2{}_{base}$ is the predictive uncertainty pertaining to the existing calibration dataset. The DW is represented by a value ranging between 0 and 1; it reflects the impact of additional observations on the base predictive uncertainty (reduction). For example, the DW

of 1 denotes that the prediction uncertainty is completely diminished by adding a new observation whereas a DW of 0 indicates that a new observation does not reduce the prediction uncertainty.

Value Index (VI) analysis (prediction of multiple variables)

In practice, multiple variables are often of interest to simulate using a groundwater model, e.g. predictions of stream flow, flow velocity, or contaminant migration. The DW for an observational well can be further evaluated if the target of an OD is to minimize the uncertainty associated with predicting multiple variables in a given groundwater system. For this purpose, the DW is first calculated for each prediction variable. A weight is then defined for each variable, and subsequently applied to each DW. The weighted DWs are then combined into a single value index (VI) that indicates the value of monitoring each observational well according to the priority of the prediction variables (equation 14) (See Vilhelmsen and Ferre (2017) for more details):

$$VI_j = \sum_{i=1}^{n} w_i \, DW_{i,j} \qquad (14)$$

where *j* corresponds to an observation set, *n* stands for the number of prediction variables, w_i is the weight of the *i*th prediction variable, and $DW_{i,j}$ is the *DW* of *j*th observation set to *i*th variable. Weighting prediction variables is a subjective choice of the modeler/manager, and can be based on various factors such as economic worth of making predictions, or prioritizing a prediction when making management plans.

Data Worth (DW)-based 3D Optimal Design (OD)

As outlined in Figure 1, the methodology consists of five interrelated steps outlined below.

(A) For each model M_k, a series of matrices are formed to represent the following: 1) the sensitivity of the existing calibration dataset to model parameters (X_{old}); 2) the sensitivity of predictions to model parameters (Y); 3) the innate parameter variability C(p); 4) the measurement noise C(e); and 5) the Jacobian matrix X that consists of Y and X_{old}. These matrices are used to calculate the base predictive uncertainty. In this work, the PREDUNC program in the PEST suites of utilities was used to calculate the base predictive uncertainty (Doherty 2015).

(B) A set of arbitrary 3D locations for potential (yet to be collected) observations is specified and used as input to the DW analysis. In order to perform a 3D design, we create an X_{new} matrix containing the spatial locations of observation wells, where each row of the matrix represents a single observation location. Each observation location (i.e., an element of the X_{new} matrix) contains a corresponding sensitivity matrix L with *m* rows and *l* columns, where *m* is the maximum number of sampling (or measurement) depths and *l* is the number of model parameters. The L matrix comprises the sensitivity of a single observation location to all model parameters with respect to the depth at which the measurements were taken. Each row of the L matrix contains the sensitivity of a measurement depth to all model parameters. Each column includes the sensitivity of a certain parameter to all measurement depths. All new measurements pertaining to the *L* matrices are added to the X_{old}, and then the sensitivities to the parameters are estimated by calculating the Jacobian matrix X using PEST (Doherty 2015). The new measurements are then detached from the X_{old}.

Note that for temporal monitoring design, the X_{new} matrix can be simply expanded by additional *L* matrices corresponding to different model stress periods. For this purpose, an L_i matrix is generated for each model stress period.

(C) The third step is to find an optimal set of *n* potential observation wells. Combinations of *n* rows are selected from the X_{new} matrix. The Genetic Algorithm (GA) is then used to select multiple rows corresponding to the combinations in the X_{new} matrix (for details about GA see Wohling et al., 2016). A p × 1 vector of randomly sampled N rows is then generated, where p is a user defined population size (p=50 in this work). This vector forms the initial population of designs (i = 1). Each element of this vector contains a random combination of *n* rows in the X_{new}. The L matrices pertaining to these *n* rows are added to the X_{old}, and subsequently the value of information (VI) of combinations are evaluated using PREDUNC5 (Doherty 2015). Note that each combination comprises *n* number of new observation locations, and each new observation contains measurements at multiple depths. Therefore, the X_{old} matrix is expanded by adding the maximum $n \times m$ measurements. In the next step, a new population of design (size *N*) is generated by applying the standard GA selection schemes, i.e. selection, mutation, and crossover (Wohling et al., 2016). In the present study, we retained 40% of the population for the next generation (i = i + 1), muted 5% of population (which was allowed to increase if the population was too uniform), and allowed a 15% chance of selecting outside of the admissible location (similar to Vilhelmsen and Ferre 2017 and Wohling et al., 2016). The new proposal design is then compared with the previously generated designs. We ended the loop when the highest VI and the proposed designs were similar in the last ten subsequent trials. We allowed a maximum of 1,000 trials to re-populate the designs in order to achieve convergence. The converged design is recorded as the proposed design of size N (i.e. D_k) for model M_k.

(D) The X_{old} matrix is amplified by the proposed design (D_k) of size N obtained using one of the models M_k. The reduction in prediction uncertainty is then evaluated in all models M_k using the amplified X_{old} matrix, and then the VI is calculated for each model. The estimated VIs are multiplied by the models weights (estimated through the BMA), and then they are

averaged. The averaged VI is the impact of the proposed design (D_k) on the prediction uncertainty reduction with respect to model non-linearity.

The proposed design of size N is re-created for all models M_k (k<K). Afterwards, the averaged VI is calculated. Among all proposed designs for models M_k (k=1,2…K), the design that has the highest averaged VI is selected as the optimal (or best) design for placing *N* observations with respect to the uncertainty of the model input.

(E) After completing the previous steps, one can conduct a cost-effective analysis to determine the optimal number of new observation wells (i.e. optimal size of design). A design is considered cost-effective when the reduction of prediction uncertainty (i.e. increase in DW) outweighs its cost. The cost is herein defined by the number of required observation wells. The cost-effective analysis is accomplished by calculating the cost (Po) of each design of size N, and the design that provides the most information at the smallest operation cost is considered the optimal design. This is estimated by the ratio: DW/Po. The design that has the highest DW/Po ratio is deemed the most cost-effective design.

Application

Description of study area

The performance of the proposed method was evaluated by simulating flow and solute transport in an actual aquifer system. Located along the Eastern Mediterranean (Figure 2), the pilot aquifer (covering an area of approximately 42 Km^2) underlies Beirut city (Lebanon) and its suburbs. The study area has a 16.5 km of shorelines encompassing rocky beaches, sandy shores, and cliffs. It is bounded by several faults to the east and south, and partly by an intermittent river to the south.

Figure 2. Location of the pilot aquifer, and its faults and geologic cross-sections

Hydrogeology

The hydrogeology of the pilot aquifer consists of Cretaceous karst limestone overlaid by Upper Tertiary and Quaternary unconsolidated deposits (Peltekian 1980). The ~700m thick, fractured Cenomanian-Quaternary system is dominated by hard and compact limestone and dolomite interbedded with chert, and intercalations of marl (Kahair 1992). According to the available geologic cross-sections (Figure 3), the rock sequence of Cenomanian-Turonian age (Walley 1997) can be divided into three subunits, as follows: 1) the Afqa Dolomite member that consists of crystalline, dolomitic, marly dolomitic and reefal limestone; 2) the Aaqoura member comprises a sequence of thinly bedded limestone, marly limestone, dolomite and marly dolomite strata; and 3) the Mnaitra member that is composed of thick and compact limestone and fossiliferous strata with several chert bands and nodules across different horizons. These are also known as the C4a, C4b, and C4c formations, respectively (Saint Marc 1974). The aquifer can be divided into seven zones to describe its geologic surface and subsurface (Figure 4 and Table 1). The upper geologic layer consists of a mix of C4c and Quaternary formations in the north (zone 1), and a mix of the C4a and Quaternary formations in the middle and to the south (zone 2). The middle layer contains the C4a formation (aquifer) to the east (zone 3), the C4c formation (aquifer) in the north (zone 4), and the C4b formation (aquitard) along the western coastline (zone 5). The lowest geologic layer with a thickness of ~250 meters comprises of the C4c formation (aquifer) in the north (zone 4), the deep C4c formation (aquifer) along the western coastline (zone 6), and the deep C4b formation (aquitard) to the east (zone 7) (Table 1).

The Cenomanian formation and Cenomanian-Quaternary systems are permeable with a specific yield of 0.03 and 0.15, respectively (UNDP 1970). The infiltration rates are high in the quaternary deposits (Khair 1992). The freshwater influx to the aquifer in the year 1969 was primarily through a reportedly high recharge, equivalent to 20-30% of the precipitation

(21% reported by UNDP 1970; 27% reported by Khair et al. 1994; and 30% reported by Ukayli 1971). The increase in urbanization since 1969 has nevertheless decreased the recharge potential to near nil by 2018 (Safi et al. 2018).

Figure 3. Geologic cross-sections CC′ and EE′

Figure 4: Location of historic head observations in the pilot aquifer with geologic zones

Table 1: Geologic formations in the pilot aquifer with corresponding hydraulic conductivity ranges (Ukayli 1969; and Petekian 1980).

Model domain	Thickness (m)	Geologic formation	Type	Log hydraulic conductivity (m/day)	Zone
Layer 1	~100 to 150	C4c-Quaternary	Aquifer	-1.69 to 2.69	1
		C4a-Quaternary	Aquifer	-1.26 to 2.69	2
Layer 2	~150	C4a	Aquifer	-1.3 to 2.69	3
		C4c	Aquifer	-1.69 to 2.69	4
		C4b	Aquitard	-5 to -3	5
Layer 3	~250	C4c	Aquifer	-1.69 to 2.69	4*
		C4c	Aquifer	-1.69 to 2.69	6
		C4b	Aquitard	-5 to -3	7

* The same zone used to characterize the C4c in the second and third layers.

Statement of problem and needs for a monitoring plan

The upper part of the pilot aquifer is highly vulnerable to saltwater intrusion (SWI) with many locations already experiencing high salinity because of groundwater overexploitation (Safi et al. 2018; Rachid et al. 2017). This limits the freshwater resources available in the aquifer and is pushing authorities to consider alternatives such as tapping the deeper parts of the aquifer. In this context, groundwater modeling can guide decision makers towards sustainable abstraction without accelerating SWI. Moreover, it can help protect the deeper parts, where the lack of subsurface characterization will inevitably increase uncertainties associated model predictions. Therefore, it is imperative to design a monitoring network with optimal locations to constrain/ reduce model uncertainties.

To design a monitoring network for the pilot aquifer requires simulating and understanding the dynamics of saltwater intrusion in response to future groundwater abstractions from the deep parts of aquifer. Hence, emulating realistic future conditions of the pilot aquifer is an important step in the OD analysis. In the current application, a scenario was defined whereby it was assumes the authorities will start extracting groundwater from the freshwater resources in the deep aquifer (zone 6) starting in March 2018. Extraction was assumed to occur through

pumping from 50 wells at 200m^3/d (Figure 5). In this scenario, the bottom elevation of the pumping wells reaches to depths of 360 meters below mean sea level (BSL) (in the middle of layer 3 in zone 6). A groundwater model was then used to simulate the future SWI in response to the groundwater exploitation from the upper and deeper parts of aquifer. The objective of the model was to find the best locations that will provide the needed information for model prediction of: (a) the displacement of the salt/fresh water interface in zone 6 caused by groundwater abstraction in the entire aquifer (prediction variable 1), and (b) the increase in the salinity concentration in the newly installed pumping wells in zone 6, which will be caused by the landward displacement of the interface. With regards to the latter, salinity concentrations were predicted at two points (A and B) specified in front of the pumping wells. Points A and B were located at depths of 360 meters (BSL) (similar to the bottom elevation of the pumping wells in the scenario) (Figure 4). Salinity prediction at point A was denoted as prediction variable 2, and salinity prediction at point B was referred to as prediction variable 3.

Figure 5: Potential new observations locations for monitoring head and salinity in zone 6, along with the locations of new pumping wells and model prediction points A and B at the depth of 360 meters below sea level – colored contours are the average of log hydraulic conductivity between all stochastic models – X and Y axes are in units of meters

Model set-up

SEAWAT code (Guo and Langevin 2002) was used to simulate the salinity migration in the pilot aquifer and to perform the OD in zone 6. SEAWAT is a variable density groundwater flow modeling code, representing flow and solute transport processes that are solved jointly by MODFLOW and MT3D. The criteria considered in the code selection process centered on its ability to: 1) simulate the 3D nature of the vertical and lateral encroachment of salinity in confined and unconfined aquifers, 2) characterize various types of time-dependent boundary conditions, 3) simulate steady-state and long-term transient flow and solute transport with the

least numerical instability, 4) link to an inversion code (such as PEST) to quantify uncertainties, and 5) contain reasonable computational resources.

Our model comprised a transient stress period of 50 years subdivided into 50 sub-periods of one-year duration, extending from March 1969 up to March 2019 (the future state). The first stress period (March 1969) was used as the calibration-time period because that period had the most information with regards to head observations within the aquifer, with a total of 35 head observations tapping into the upper geologic layer (Figure 3).

In the set-up of the SEAWAT model, the sea boundary to the north and west was specified as a constant head and concentration boundary condition with an average salinity level of 35 gram per liter (gr/l). The eastern boundary was assumed to be a no-flow boundary due to aquitards and Faults 1 & 3 in the vicinity. The horizontal discretization contained 4,251 active cells designed in 115 rows and 75 columns, where each grid cell represents a square of 100 by 100 meters (Figure 6). Groundwater abstraction for the upper aquifer was estimated for the years 1969 to 2019 assuming a 1.75% population growth and 180 liters per capita per day (l/c/d) domestic consumption rate (MoEW 2010). The population were considered to increase from approximately 0.35 million (M) in 1969 to 1M in 2019, respectively. The freshwater influx to the aquifer in the year 1969 was assumed to be primarily through the recharge equivalent of 30% of precipitation (Ukayli 1971). This rate linearly decreased to zero for 2019 due to increase in impervious pavements over time (Safi et al. 2018). The lateral flow was assumed to be zero due to faults and aquitards in the vicinity of the aquifer.

Figure 6: Model grids, boundary conditions, and locations of abstraction wells in the upper part of the Beirut coastal aquifer

Pilot points parameterization

The geology of the aquifer is poorly characterized due to limited information about its characteristics, but its hydraulic conductivity can vary considerably over small distances (Safi et al., 2018). Therefore, the use of only a few zones to represent the hydraulic conductivity field may not accurately represent the flow conditions. To remedy this problem, a set of pilot points was assigned to each geologic zone to represent the spatial variability in the hydraulic conductivity. Altogether, a total of 564 pilot points was defined for the entire model domain (i.e. the vector P_k in the BMA). We used an exponential Kriging variogram to spatially interpolate the pilot points values over the geologic domain. The value of 500 meters (i.e., the separation distance between two pilot points) was defined for the range of the variogram (Doherty and Hunt 2010). Using PEST, the pilot points parameterization approach was used in conjunction with SVD and Tikhonov regularization to estimate the values of the pilot points using the existing head observations (Doherty 2015).

The calibrated pilot point model was then used to calculate the sensitivities of the observations and model prediction variables (1, 2 and 3) to the pilot point values. According to the sensitivity results, the pilot points parameters were categorized into three subsets: subset (i) contained the pilot points that spanned the solution space or were correlated with the pilot points that lied in the solution space (zones 1 to 4); subset (ii) comprised the null-space located pilot points that inform the predictions (zone 6); and subset (iii) included the pilot points that lied in the null-space and did not inform the predictions (zones 5 and 7).

Bayesian model averaging (BMA)

A random realization of the hydraulic conductivity field was generated for each subset on the basis of the prior probability distribution of the subset's hydraulic conductivity. The generated random realizations of all subsets were then combined into one realization that

contained the full set of the pilot points parameters. The NSMC method was then used to calculate a calibrated-constrained realization on the basis of the combined generated random realization. This process was repeated several times to create multiple calibrated-constrained realizations.

The prior probability distributions were defined according to the available geologic knowledge for each subset, as follows: subset (i) used a probability distribution that was estimated based on the results of the calibrated pilot points model. Subset (ii) used a fuzzy theory set (Zadeh 1969) to create a fuzzy probability distribution for the mean log hydraulic conductivity value. For this subset, a trapezoidal distribution was used to define the membership functions of the log mean hydraulic conductivity values for the pilot points (for details about fuzzy set theory see e.g. Bardossy et al. 1990). Accordingly, the membership function was set to 1 for the range of log hydraulic conductivities between 1.5 and 2m/day, (lower and upper support limits). The membership value was set to zero for the lower and upper limits of the log mean hydraulic conductivity values, which were defined as 0.31 and 2.69 m/day, respectively. Subset (iii) did not use any probability distribution and the hydraulic conductivity value was defined as constant for all pilot points.

Our BMA application encompassed the generation of only 10 calibration-constrained realizations due to the large run-time of the model. The Bayesian models (M_k) constructed on the basis of the realizations were then used to simulate SWI for the near future. For the purpose of model averaging, similar weights were assigned to the models (M_k) because all had almost the same sensitivity to the existing observed data after the model calibration.

Optimal Design of new observation wells

The OD analysis involved defining the best locations for a set of new observation wells in zone 6. The design assumed that the observation wells were available every 150 meters in

zone 6 (i.e. spatial location), and a total of 192 potential locations were specified in that zone (Figure 4). Each observation well was assumed to contain two measurements of field-data, namely: (1) the head and salinity concentration at a depth of 300 meters BSL, and (2) the head and salinity concentration at a depth of 360 meters BSL. Hence, the measurements vary spatially and vertically over zone 6 (i.e. three dimensions). The values of these measurements were extracted from the last stress period (i.e. the year 2019) of the Bayesian models and assigned to the observation wells.

The OD framework was then applied to define: (1) the optimal design to make all predictions under different prediction weights when having one observation well where head and salinity are measured; and (2) the optimal design for multiple locations (i.e. N=1, 2…,5 observation wells) whereby measurements of head and salinity are taken at both single and multiple depths to reduce the uncertainty with the prediction of the salt/fresh water interface. Several scenarios were considered to define a variation in the predictions' weights, listed as follows: scenario a: w1=w2=w3=0.33; scenario b: w1=0.6, w2=0.3, w3=0.1; scenario c: w1=0.6, w2=0.1, w3=0.3; scenario d: w1=0.8, w2=w3=0.1; and scenario e: w1= 1, w2=w3=0; where wi is the weight assigned to the previously defined prediction variables (i=1, 2, and 3).

Finally, a cost-effective analysis was made to find the optimal size of the design, involving new observations, N=1, 2…,5. It was assumed that the scaled start-up cost (Ps) of a monitoring project varies between 0 and 1, regardless of the number of planned observation wells. The cost of implementing the first observation well (P1) was assumed to vary between 0 and 100% of the start-up cost. Variations between start-up costs and implementation costs is captured by the P1/Ps ratio. The implementation cost decreases by 0%, 10%, 20%, and 30% with any additional observation well ($P1^{+}$). This variation is captured by $P1^{+}$/P1. A set of random values were then generated based on the above-mentioned cost criteria for

implementing a given monitoring project. A uniform distribution was used to randomize the costs within their defined ranges. The reduction in predictive uncertainty in relation to the cost of sampling the head and salinity was evaluated by dividing the DW by the cost of the project implementation for simultaneous observation wells (i.e., N=1, 2…,5).

Results and Discussion

Bayesian models

Calibration-constrained realizations were generated for the hydraulic conductivity field. The model-to-measurement misfit ranged from 251 to 258m^2 with an average residual absolute error of ~2m across all of the stochastic models M_k. Figure 7 shows the calibration-constrained log hydraulic conductivity fields along with the position of the salt/fresh water interface in zone 6 after a 50-year simulation based on the 10 generated stochastic models. The stochastic results are only shown for zone 6, where the OD was performed. The uncertainty with the estimated log hydraulic conductivity had a noticeable impact on the computation of the position of the interface. In four of the models (models M1, M5, M6 and M10), the log hydraulic conductivity values were low along the coastline, and therefore the landward displacement of the interface caused by groundwater abstraction was small as compared with that in other models that contained large hydraulic conductivity values. In the other models, it was apparent that the high hydraulic conductivity values tended to exacerbate SWI, subsequently decreasing the depth to the interface. Since the large uncertainties in the estimated hydraulic conductivities of zones 6 increased the uncertainties in the model prediction, it was necessary to quantify the prediction uncertainty.

Figure 7: Log hydraulic conductivity distribution in zone 6 using 10 stochastic models Mk: k=1, 2 … 10; along with the position of the interface with 75% (black), 50% (gray), and 25% (white) of sea water concentration (35gr/l) after 50 years simulation

We used the range of the simulated values obtained using the generated calibration-constrained realizations of the hydraulic conductivity to quantify the uncertainty in predicting the displacement of the salt/fresh water interface after the 50 years simulation. The histogram of the log hydraulic conductivity values approached an almost or a near normal distribution that ranged from 0.85 to 2.33 m/day (Figure 8-a). The distribution was negatively skewed due to the limitations imposed by the low hydraulic conductivity values through a trapezoidal membership function during the randomization process. Figure 8-b shows the histogram of the predicted values for the displacement of the interface where the transitional mixing zone was limited to 1 gr/l (threshold for drinking water). The predicted values are shown as percent displacement of the interface from its initial position with respect to the coastline, which involved calculating the percent increase in the volume of salinity concentration of greater than 1gr/l in freshwater due to groundwater abstraction. The histogram for the model prediction exhibited a large level of uncertainty in the predicted percent change of the position of the interface, ranging from 40 to 70% (Figure 8-b). The prediction histogram had a shape similar to that of the log hydraulic conductivity value (Figure 8-a vs. 8-b). This underlines the importance of estimating or knowing the hydraulic conductivity in the deep parts of a coastal aquifer to compute the displacement of the interface.

Figure 8: Histograms for log hydraulic conductivity values for zone 6 and for the corresponding model prediction of the percent change in the position of the 3D salt/fresh water interface with salinity >1gr/l from the coastline

Optimal design for measurements with single depth at a single observational well

The OD analysis determined if the uncertainty in the model predictions would be affected by adding measurements of head and salinity obtained from a single depth at a single observational well in zone 6. The DW of these measurements was found to be sensitive to both the spatial location of the observational well and to the depth at which the measurements were taken (Figure 9). Head and salinity measurements made near points A and B were found

to be more effective at informing model predictions of salinity levels at these points as compared to measurements made at other locations (Figures 9-b and 9-c). The DW increased slightly (~20% more) when the head and salinity measurement depth increased by 20% (Figure 9-b vs. 9-e and Figure 9-c vs. 9-f). Measurements acquired at depths shallower than points A and B (e.g., 300m BSL) informed the predictions more effectively (i.e. variables 2 and 3) when they were sampled from an observational well located on the seaward side of these points (Figures 9-b and 9-c). Conversely, measurements taken at the same depths as points A and B (e.g. 360m BSL) had more impact on reducing the predictions' uncertainties when they were sampled from an observational well located on the landward side of these points (Figures 9-e and 9-f).

The observation wells that were located parallel to the coastline and within the transition zone predicted better the displacement of the salt/fresh water interface (i.e. prediction variable 1) (Figures 9-a and 9-d). A single observational well that had the highest impact on the uncertainty reduction of prediction variable 1 was located ~500m from the coastline if the measurements were taken at a depth of 300m BSL (Figure 9-a). When the measurement depth increased by 60 meters, an observational well that was located 200 meters further landward provided the most information for predicting variable 1 (Figure 9-a vs. 9-d). However, adding a single observational well had less impact on reducing the uncertainty associated with predicting variable 1 as compared to variables 2 and 3. Our findings showed the importance of considering the three dimensionalities in an OD when predicting SWI.

Figure 9: Data worth (DW) of new (yet to be collected) observation locations with measurements of head and salinity at different depths coresponding to prediction variables:
1 (interface displacement); 2 (salinity increase at point A); and 3 (salinity increase at point B)

Optimal design for measurements with multiple depths at a single observational well

Figure 10 shows the contoured VI averaged over the 10 models for the three prediction variables (the 3D displacement of interface and the salinity levels at points A and B), given a

set of specific weight distributions assigned to these three variables. Different locations for an observational well were specified according to the different weights of the prediction variables. The optimal location of a single observational well lied close to the points A and B when all prediction variables were assigned a similar weight (Figure 10-a). The DW of a single observational well was twice as high when predicting salinity levels at points A and B than when predicting the interface displacement (Figure 10-a vs. 10-b and 10-c). Although increasing the weight assigned to predicting variable 1 (w1) two-fold increased the VI in the observational wells located within points A and B, the best location for a single observational well was found to be still close to these points (Figures 10-b and 10-c). An eight-fold increase in w1 as compared to w2 and w3 shifted the best location ~500 meters upward (Figure 10-d). It was found that it was better to collect head and salinity data in the middle of zone 6 within points A and B; that would provide a more accurate prediction of the interface displacement (variable 1) (Figure 10-e) because most models showed the highest intrusion occurring along the middle of the coastline (Figure 7). Compared with observational wells located elsewhere, the observational wells located in the middle of zone 6 sensed more concentration changes as the interface approached landward. Our results suggest that designing the location of only one observational well did not noticeably reduce the uncertainty in the prediction of the interface displacement (i.e. variable 1). Moreover, the uncertainty with predicting salinity levels at points A and B appears to approach nil as the number of new observational wells reaches the number of prediction variables (i.e. two observation wells for two point-source prediction variables). In contrast, more observational wells provided substantial information on estimating the interface displacement.

Figure 10: Black solid triangle represents the optimal design location for a single observation with measurement of head and salinity at multiple depths for multiple prediction variables - a) w1=w2=w3=0.33; (b) w1=0.6, w2=0.3, w3=0.1; (c) w1=0.6, w2=0.1, w3=0.3; (d) w1=0.8, w2=w3=0.1; (e) w1= 1, w2=w3=0. wi is prediction weight - colored contours are VI that was averaged over the models Mk- locations of points A and B are shown in black rectangles

Optimal design for measurements with multiple depths at multiple observational wells

The performance of the methodology was examined for a larger number of observations in the case where the weight was given solely to the model prediction of the 3D displacement of the interface (i.e. prediction variable 1). The optimal locations for N=2, … and 5 simultaneous new observations (i.e. proposed design) were specified separately for each of the 10 models M_k. Altogether, 10 designs were proposed for every N. Figure 11 shows the effectiveness of each of the proposed designs in reducing the prediction uncertainty when considering the impact of model non-linearity. Boxplots were used as a means of comparing the proposed designs across the 10 models. In these plots, the averaged DW using all models is shown for each design along with the estimated DW for that model. Each boxplot contains 10 estimated DWs and the highest DW in each boxplot corresponds to the original model M_k for which the design D_k was proposed.

The results suggest that the best proposed design having N=2 observations corresponded to model M_5 (Figure 11-a). The corresponding design (D_5) had the highest mean DW and the smallest variance compared with the designs obtained from other models, suggesting that the uncertainty with regards to the hydraulic conductivity values did not significantly affect identifying the optimal locations when N was equal to 2 observations. For the designs with 3 observations, the highest mean estimate of DW (of 0.72) corresponded to D_1, D_6, D_7, and D_{10} (Figure 11-b). Among these designs, D_7 had the smallest variance, and its proposed locations for 3 observations were selected as the optimal observational locations. The design corresponding to model M_1 (i.e. design D_1) had the highest impact (mean DW of ~0.75) on the reduction of prediction uncertainty in all the models when it was used to propose the optimal locations for 4 new observational wells (Figure 11-c). The mean DW for the designs with 5 observations ranged from 0.68 to 0.78 and the variance of the DW ranged from 0.01 to 0.05 (Figure 11-d). We selected D1 as the best design for proposing the locations for the 5

new observational wells because it had the highest DW (of 0.78) compared with all other designs.

An inspection of all boxplots together (Figure 11) reveals that the median is higher than the mean for most designs, indicating that the DW has a heavy left tail. The DW (i.e. additional uncertainty reduction) per additional observation decreases with increasing the number of observations. Interestingly, the spread of the DW also decreased as the number of observations was increased. The designs with a larger number of observations appeared to have similar centers that exceeded those found for the designs with fewer observations. Overall, an increased number of observations resulted in the reduction of the prediction uncertainty, which is not surprising, however it also increased the number of outliers, indicating that the model non-linearity can strongly affect the estimation of DW for a large-size design. From the estimated DWs, we noticed that the models (M7 and M9) that caused such influence contained a very high mean log hydraulic conductivity (of > 2m/day) as compared with that in other models (Figure 7). In these models, the DW was very low when applying the proposed designs that were specified using other models (M_k).

Figure 11: Estimated DW of proposed designs Dk from models Mk (k=1,2…10) for N=2, 3, 4, and 5 observation wells in plots a to d respectively: x-axis corresponds to a proposed design specified using a model Mk and y-axis is the estimated DW when applying a proposed design on all models – Black dot is the mean estimate of the DW and red line is the median of the estimated DWs for each design using all models.

The observation sets with the highest averaged DW for design sizes N=1,2, …, 5 (Figure 12) showed that the proposed locations changed with the number of planned additional observations. Measurements were found to be spread from north to south parallel to the coastline with an increasing number of observations. In the small size design, the proposed locations were found to be independent of previously proposed locations. For example, the locations for the two observations did not correspond to that associated with the design size of 1 (Figure 12-b vs. 12-a). As the design size increased (e.g. design size of 3 to 5), the

locations of some observation wells were similar to the preceding design size. For example, in the design size of 5, three of the five proposed observation locations were approximately the same as three of the four observation locations in the design size of 4 (Figure 12-e vs. 12-d).

Figure 12: Optimal design for N=1, …5 new observations with measurement of salinity at multiple depths for predicting the displacement of the salt/water interface. ▲ is a proposed location for a new observation. The colored contours represent the DW for a single observation with measurement at multiple depths

Cost-effective analysis

The cost criteria were applied on the proposed design sizes of N=1 … 5 (Figure 13). Each point represents how the reduction of prediction uncertainty (i.e. increase in DW) exceeded its cost. The least and most cost-effective designs were the designs for 5 and 2 observations, respectively (red and green dots in Figure 13). This indicates that although the prediction uncertainty decreases by increasing the number of observations (Figure 11), the optimal design size varies according to the cost criteria. The most effective criterion was found to be P1/Ps (the cost of implementing the first observation according to the cost at start-up). An inspection of the cost-effective results shows that the optimal size of a design should include a maximum of two observations when the cost of implementing the first observation is more than 50% of the start-up cost of the monitoring project (Figures 13-c and 13-d). Under this condition, the optimal size can be increased by decreasing the cost of implementing an additional observation to < 80% of the cost of operating the first observation (i.e. $P1^{+}$/P1) (Figures 13-a and 13-b). With P1/Ps < 30%, the optimal size can be increased up to 4 observations if $P1^{+}$/P1 is <0.8 (Figure 13-a). The implementation of 5 observations seems to be the most cost-effective only if the cost of implementing the first observation is much lower than the start-up cost, which may not be plausible.

Figure 13: Cost-effective analysis for size of proposed design: x-axis shows the ratio of the cost of implementing the first observation (P1) to the start-up cost (Ps) (which is scaled between 0 and 1), y-axis is the reduction of uncertainty (data worth (DW)) with predicting the 3D displacement of interface to the full cost of a

project (Po), dots correspond to an observation set of n=1,2…5 observations – each plot shows the different analysis according to the operation cost of an additional observation ($P1^{+}$) to the cost of implementing the first observation (P1) – DW is constant at each design size and only the cost of design varies

Conclusion

In this study, we extended an existing linear data worth (DW) method that optimizes the process of locating multiple new observational locations (yet to be collected) in order to reduce the uncertainty in predicting multiple variables in a groundwater system. Compared to previous studies that used two dimensional locations for the observations, our method also optimizes simultaneous for measurements occurring at different depths at a single or multiple locations (i.e. three dimensions) at a minimum cost. We also suggested the use of BMA, which was used to define weights for each Bayesian model that contains a set of stochastic parameters. The capability to produce the calibration dataset (also considered as a model prediction) by the stochastic parameters was used to calculate the weight of each Bayesian model. The final outcome of the optimal design (OD) was a set of proposed locations for an observational set that accounts for the non-linearity of the model.

We applied the proposed methodology on a pilot heterogeneous coastal aquifer that lacks hydrogeological information for its deep geologic layers. The target of the design was to find the best locations for placing 1, 2…,5 new observations that could contribute to the reduction of the prediction uncertainties. Two types of prediction were used as the optimization targets: capturing the increase in salinity at two points located in the deep part of the aquifer and the displacement of the interface caused by groundwater abstraction. The types of observations that were accounted for included head and salinity at different locations in three dimensions. The following findings were deduced from the OD results of our case study:

- It is important to sample data at different depths and locations if the target of a monitoring design is to predict a solute transport over a three-dimensional geologic domain.
- Model non-linearity has a slight impact on proposing a single location for an observational well with head and salinity measurements. This impact increases by increasing the design size.
- The DW of an observational well with head and salinity measurements depends primarily on its spatial proximity to the coastline, while the depth of measurement is secondary but also important. The proposed spatial location for an observational well moves toward the coastline when the depth of measurement decreases.
- When the depth of measurement is less than the depth of a point at which the predicted salinity concentration is required, a proposed observation should be located adjacent to that point and seaward (i.e. in the direction of the sea). Conversely, when the depth of the measurement is equal to or greater than the depth of the point, the observation should be located adjacent to the point and landward (i.e. further inland).
- To reduce uncertainty with the future interface, observational wells should be located close to the coastline if the measurements to be taken are located at shallow depths. Conversely, observational wells can be located farther from the coast as measurement depth increases.
- The proposed spatial locations for (head and salinity) observations change with the design size. The locations become more similar when the number of planned additional observations increases.
- The optimal size for the monitoring plan depends mostly on the ratio between the start-up cost of the monitoring project and the cost of drilling the first observation

well, while the implementation cost of additional observational wells is secondary but also important.

– For the current application, a maximum of two observation wells is needed to obtain a cost-effective monitoring plan if the cost of implementing the first observation and the monitoring data is more than 50% of the start-up cost of the monitoring project. If the cost of implementing the first observation is equal or greater than the start-up cost, only one observation is recommended for a cost-effective design. Monitoring data from multiple observation wells are the most cost-effective if the start-up cost is much more than the cost of implementing the first observation (e.g. using nearby pumping wells as observation wells).

This study stresses that the effectiveness of the proposed methodology to secure the optimal results hinges on properly weighing the stochastic models when non-linearity is high. This underlines the importance of *a priori* knowledge of the system, before designing a monitoring network to produce an effective and successful model calibration. While an increase in the number of planned observations can reduce the prediction uncertainties during the model calibration stage, the cost-effectiveness of a monitoring design was found to be mostly contingent on the cost of operating the first observational location. The results of this study can be used for future field-studies to guide adaptations and implementing sampling strategies in aquifers. Due to the long model run-times to simulate SWI, we limited the OD to steady-state measurements of head and salinity. With transient measurements, it is expected that the proposed locations for observation wells will shift landward with the movement of the interface.

Acknowledgements

This study is part of a program on climate change and seawater intrusion along the Eastern Mediterranean funded by the International Development Research Center (IDRC) of Canada at the American University of Beirut Grant No. 106706-001. Special thanks are extended to Dr. Charlotte Macalister at IDRC for her support and feedback in implementing this program and to Dr. Steen Christensen for his invaluable help, which greatly improved this manuscript.

References

Andricevic, R. and E. Foufoula-Georgiou. 1991. A Transfer Function Approach to Sampling Network Design for Groundwater Contamination. *Water Resources Research* 27: 2759-2769.

Badaruddin, S., A. D. Werner and L. K. Morgan. 2015. Water table salinization due to seawater intrusion. *Water Resources Research* 51: 8397-8408.

Bakalowicz, M. 2005. Karst groundwater: A challenge for new resources. *Hydrogeology Journal* 13: 148–160.

Bardossy, A., I. Bogardi and L. Duckstein. 1990. Fuzzy regression in hydrology. *Water Resources Research* 26: 1497-1508.

Beven, K. and A. Binley. 1992. The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Process* 6: 279–298.

Christensen, S. and J. Doherty. 2008. Predictive error dependencies when using pilot points and singular value decomposition in groundwater model calibration. Advances in Water Resources 31, no. 4: 674–700.

Cieniawski, S. E. J. W. Eheart and S. Ranjithan. 1995. Using Genetic Algorithms to Solve a Multiobjective Groundwater Monitoring Problem. *Water Resources Research* 31: 399-409.

Comte, J. C., R. Cassidy, J. Obando, N. Robins, K. Ibrahim, S. Melchioly, H. Shauri, A. Bourhane, I. Mohamed, C. Noe and J. Davies. 2016. Challenges in groundwater resource management in coastal aquifers of East Africa: Investigations and lessons learnt in the Comoros Islands, Kenya and Tanzania. *Journal of Hydrology: Regional Studies* 5: 179-199.

Dausman, A.M., J. Doherty, C. D. Langevin, and M. C. Sukop. 2010. Quantifying data worth toward reducing predictive uncertainty. *Groundwater* 48: 729–740.

Doherty, J.E. and R. J. Hunt. 2010. Approaches to highly parameterized inversion—A guide to using PEST for groundwater-model calibration: *U.S. Geological Survey Scientific Investigations Report* 2010–5169, 59 p.

Doherty, J. 2015. Calibration and Uncertainty Analysis for Complex Environmental Models. *Watermark Numerical Computing*, Brisbane, Australia. ISBN: 978-0-9943786-0-6.

El-Fiky, A. A. 2010. Hydrogeochemical Characteristics and Evolution of Groundwater at the Ras Sudr-Abu Zenima Area, Southwest Sinai, Egypt. *JKAU: Earth Science* 21: 79-109.

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

Gómez-Hernández, J. J., A. Sahuquillo and J. E. Capilla. 1997. Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data: 1. Theory. *Journal of Hydrology* 203: 162–174.

Guo, W. and C. D. Langevin. 2002. User's Guide to SEAWAT: A Computer Program for Simulation of Three-Dimensional Variable-Density Ground-Water Flow. *Techniques of Water-Resources Investigations* Book 6, Chapter A7, 77 p.

Hartmann, A., Goldscheider, N., Wagener, T., Lange, J., and M. Weiler. 2014. Karst water resources in a changing world: Review of hydrological modeling approaches. *Reviews of Geophysics, 52*(3), 218-242.

Harvey, C. F. and S. M. Gorelick. 1995. Mapping Hydraulic Conductivity: Sequential Conditioning with Measurements of Solute Arrival Time, Hydraulic Head, and Local Conductivity. *Water Resources Research* 31: 1615-1626.

Herrera, G., J. Guarnaccia and G. F. Pinder. 2000. A methodology for the design of space-time groundwater quality sampling networks. *Proceedings of the Conference on Computational Methods in Water Resources XIII*, vol. 1, edited by L. R. Bentley et al., pp. 579– 585, A. A. Balkema, Brookfield, Vt.

Hoeksema, R. J., and P. L. Kitanidis. 1984. An Application of the Geostatistical Approach to the Inverse Problem in Two-Dimensional Groundwater Modeling. *Water Resources Research, 20*(7), 1003-1020.

Hoeting, J.A., D. Madigan, A. E. Raftery and C. T. Volinsky. 1999. Bayesian model averaging: A tutorial. Statistical Science 14: 382- 417.

Holzbecher, E. and S. Sorek. 2006. Numerical Models of Groundwater Flow and Transport. *Encyclopedia of Hydrological Sciences*. 13:155.

Kerrou, J. and P. Renard. 2009. A numerical analysis of dimensionality and heterogeneity effects on advective dispersive seawater intrusion processes. *Hydrogeology Journal*, 18: 55-72.

Khair, K. 1992. The Effects of Overexploitation on Coastal Aquifers in Lebanon. *International Association of Hydrogeologists*, 3: 349 - 362.

Leube, P. C., A. Geiges, and W. Nowak. 2012. Bayesian assessment of the expected data impact on prediction confidence in optimal sampling design. *Water Resource Research* 48: W02501.

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

Lu, C., R. K. Kitanidis and J. Luo. 2009. Effects of kinetic mass transfer and transient flow conditions on widening mixing zones in coastal aquifers. *Water Resources Research 45*(12).

MoEW. 2010. National Water Sector Strategy: Supply / Demand Forecasts, draft, Beirut. *Ministry of Energy and Water, Beirut, Lebanon.*

Moore, C. and J. Doherty, J. 2005. Role of the calibration process in reducing model predictive error. *Water Resources Research* 41: W05020.

Peltekian, A. A. 1980. Groundwater quality of Greater Beirutin relation to geologic structure and the extent of seawaterintrusion. Master thesis, *American University of Beirut, Lebanon*. [online] Available at: http://hdl.handle.net/10938/3442

Rachid, G., M. El-Fadel, I. Alameddine and M. Abou Najm. 2017. Towards a Framework for the Assessment of Saltwater Intrusion in Coastal Aquifers. *Environmental Impact Assessment Review* 67: 10-22.

Reed, P., B. Minsker and A. J. Valocchi. 2000. Cost-effective long-term groundwater monitoring design using a genetic algorithm and global mass interpolation. *Water Resources Research* 36: 3731-3741.

Rouhani, S. and T. J. Hall. 1988. Geostatistical schemes for groundwater sampling. *Journal of Hydrology* 103: 85-102.

Rubin, Y. and G. Dagan. 1987. Stochastic identification of transmissivity and effective recharge in steady state groundwater flow: 1. Theory. *Water Resource Research* 23: 1192–1200.

Safi, A., G. Rachid, M. El-Fadel, J. Dummar, M. Abou Najm and I. Alameddine. 2018. Synergy of climate change and local pressures on saltwater intrusion in coastal urban areas: Effective adaptation for policy planning. *Water International* 43: 145–164.

Saint Marc, P. 1974. Etude stratigraphique et micropaléontologique de l'Albien, du Cénomanien et du Turonien Note et mémoires sur le Moyen-Orient – Tome *XIII. Paris, Beirut: CNRS.*

Sen, Z. 2015. Chapter 3 - Groundwater Hydraulics and Confined Aquifers. Practical and Applied Hydrogeology. *Elsevier* Pages 99-208.

Storck, P., J. W. Eheart and A. J. Valocchi. 1997. A method for the optimal location of monitoring wells for detection of groundwater contamination in three-dimensional heterogenous aquifers. *Water Resources Research* 33: 2081-2088.

Tavakoli, R., H. Yoon, M. Delshad, A. H. ElSheikh, M. F. Wheeler and B. W. Arnold. 2013. Comparison of ensemble filtering algorithms and null-space Monte Carlo for parameter estimation and uncertainty quantification using CO2 sequestration data. *Water Resources Research* 49: 8108-8127.

Tiedeman, C. R., M. C. Hill, F. A. D'Agnese and C. C. Faunt. 2003. Methods for using groundwater model predictions to guide hydrogeologic data collection, with application to the Death Valley regional groundwater flow system. *Water Resources Research* 39: 1010.

Tribbia, J. and S. C. Moser. 2008. More than information: what coastal managers need to plan for climate change. *Environmental Science & Policy* 11: 315-328.

Tonkin, M. J. and J. Doherty. 2005. A hybrid regularized inversion methodology for highly parameterized environmental models. *Water Resources Research* 41: 10.

Ukayli, M. 1971. Hydrogeology of Beirut and Vicinity, Master of Science Thesis, *American University of Beirut, Lebanon*. [online] Available at: http://hdl.handle.net/10938/1974.

UNDP. 1970. Etude des Eaux Souterraines. Programme des Nations Unies pour le development. *Departement de la cooperation technique pour le developpment*, New York. 186 p.

Vilhelmsen, T. N. and T. P. A. Ferre. 2017. Extending Data Worth Analyses to Select Multiple Observations Targeting Multiple Forecasts. *Groundwater*, doi: 10.1111/gwat.12595.

Wagner, B. J. 1995. Sampling Design Methods For Groundwater Modeling Under Uncertainty. *Water Resources Research* 31: 2581-2591.

Walley, C. 1997. The Lithostratigraphy of Lebanon. *Lebanese Science Bulletin*, *10*(1).

Wallis, I., C. Moore, V. Post, L. Wolf, E. Martens and H. Promrner. 2014. Using predictive uncertainty analysis to optimise tracer test design and data acquisition. *Journal of Hydrology* 515: 191–204.

Werner, A. D., M. Bakker, V. E. A. Post, A. Vandenbohede, C. Lu, B. Ataie-Ashtiani, C. T. Simmons and D. A. Barry. (2013). Seawater intrusion processes, investigation and management: Recent advances and future challenges. *Advances in Water Resources* 51: 3-26.

Wöhling, T., A. Schöniger, S. Gayler and W. Nowak. 2015. Bayesian model averaging to explore the worth of data for soil-plant model selection and prediction. *Water Recourses Research* 51: 2825–2846.

Wöhling, T., A. Geiges and W. Nowak. 2016. Optimal Design of Multitype Groundwater Monitoring Networks Using Easily Accessible Tools. *Groundwater* 54: 861-870.

Zadeh, L., A. 1965. Fuzzy sets. *In Information and Control* 8: 338-353.

Figure Captions

Figure 1: Optimal Design framework
X_{old}: Sensitivity matrix of existing observations to parameters; Y: Sensitivity matrix of predictions to parameters; C(e): Covariance matrix of measurement noise (e); C(P): Covariance matrix of parameters (P) innate variability; X: Jacobian matrix; ∂^2_{base} is the base predictive uncertainty variance; X_{new}: Sensitivity matrix of new observation wells to parameters; L: Sensitivity matrix of an observation well with measurements at multiple depths to parameters; D_k is a proposed design using model M_k; and Po, N is operation cost of a monitoring project for N new observations

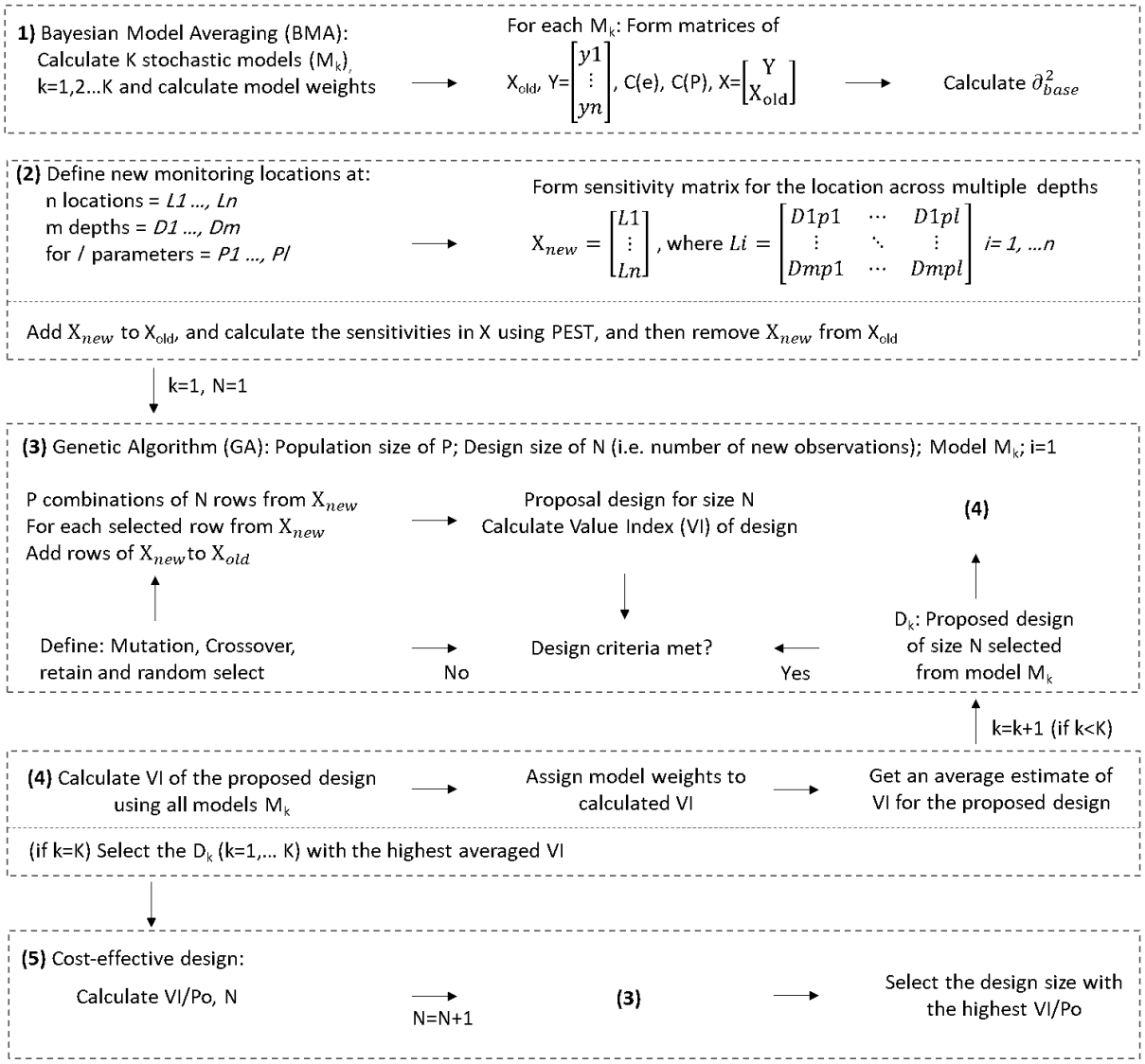

Figure 2. Location of the pilot aquifer, its faults and geologic cross-sections

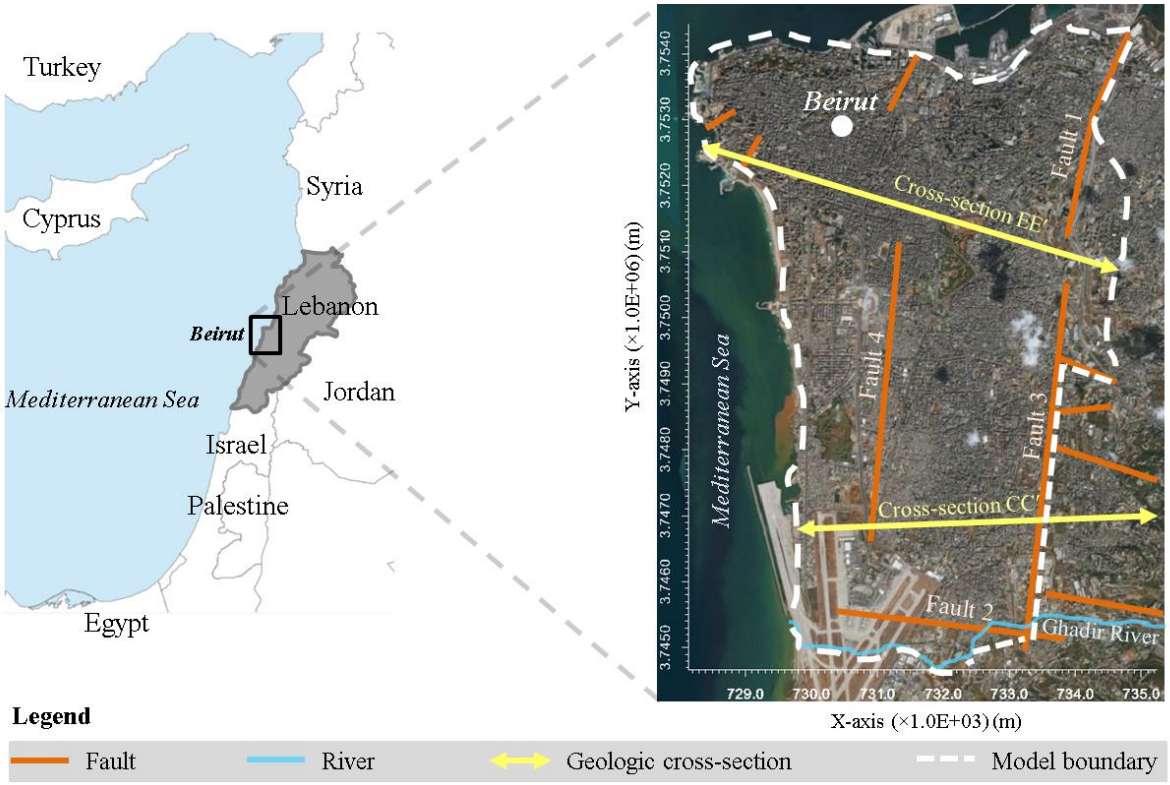

Figure 3. Geologic cross-sections CC′ and EE′

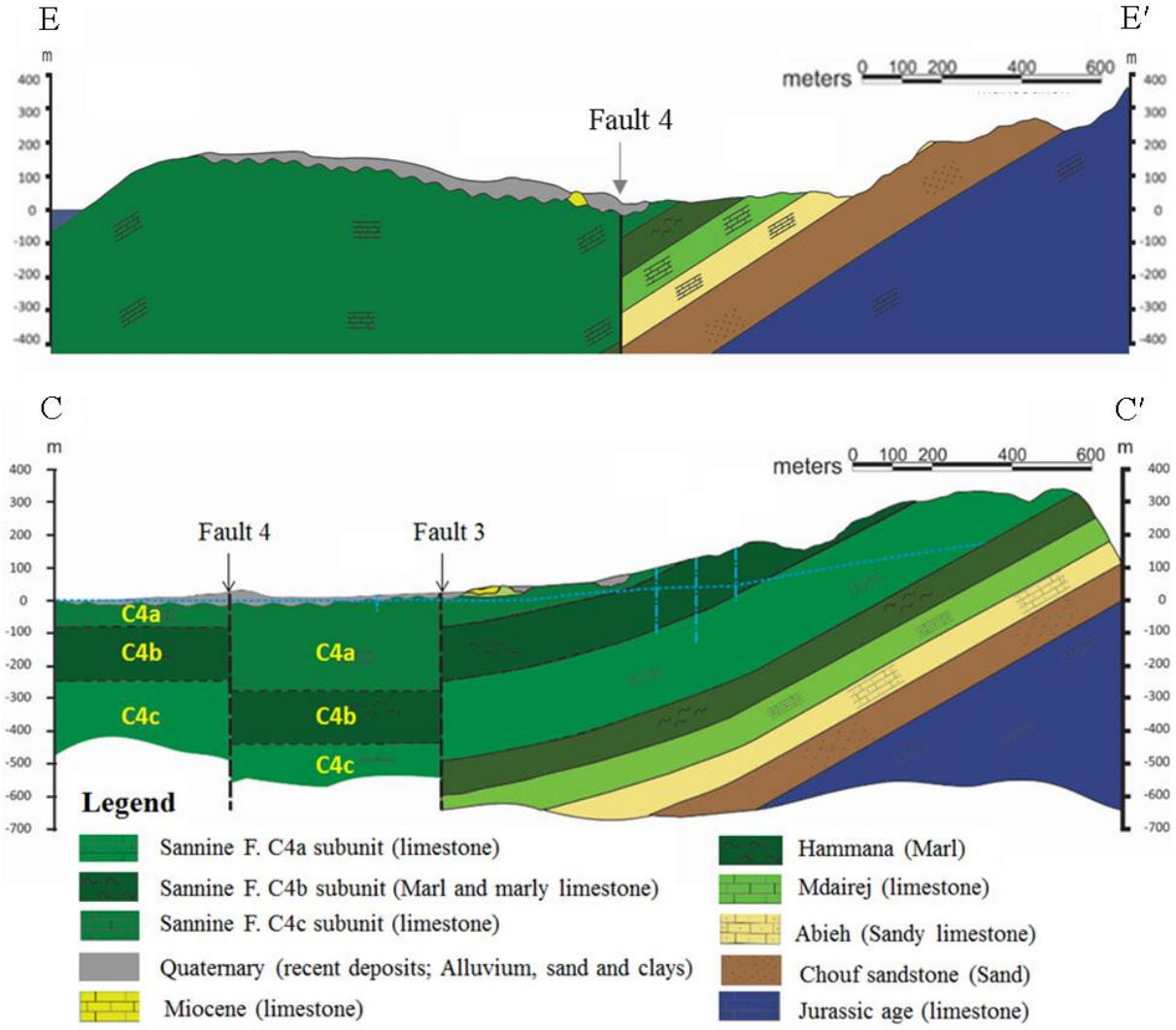

Figure 4: Location of historic head observations in the pilot aquifer with geologic zones

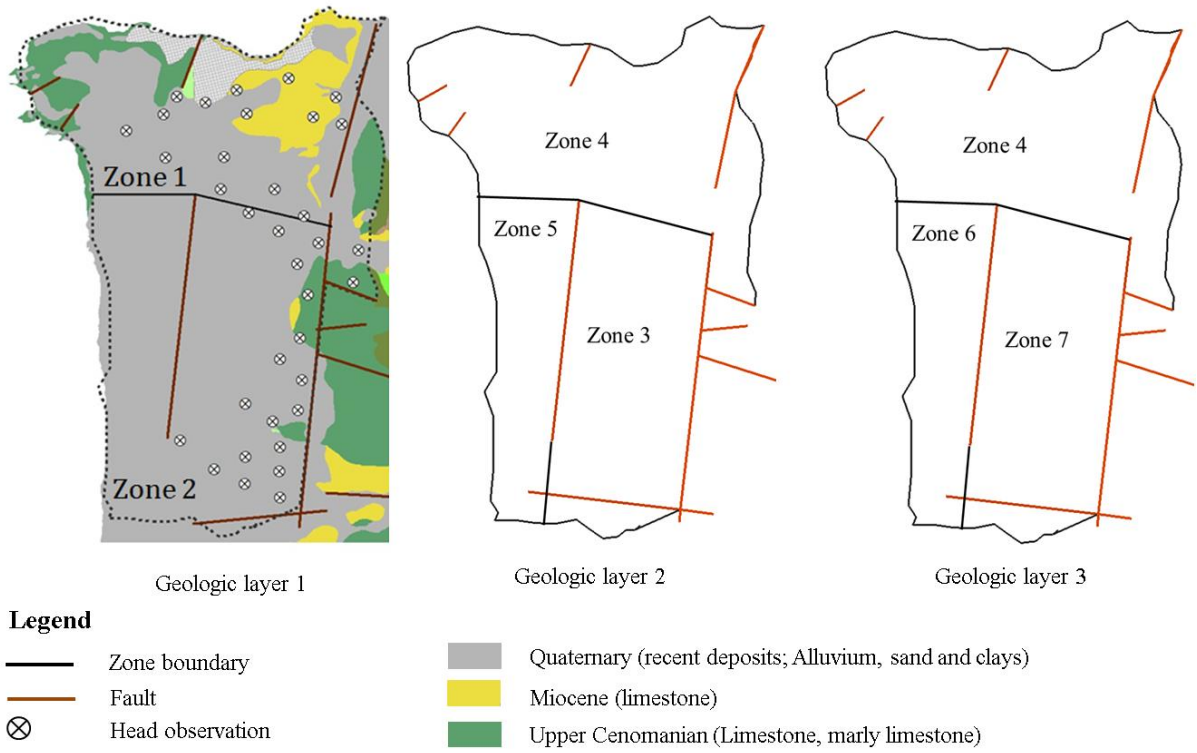

Figure 5: Potential new observations locations for monitoring head and salinity in zone 6, along with the locations of new pumping wells and model prediction points A and B at the depth of 360 meters below sea level – colored contours are the average of log hydraulic conductivity between all stochastic models – X and Y axes are in units of meters

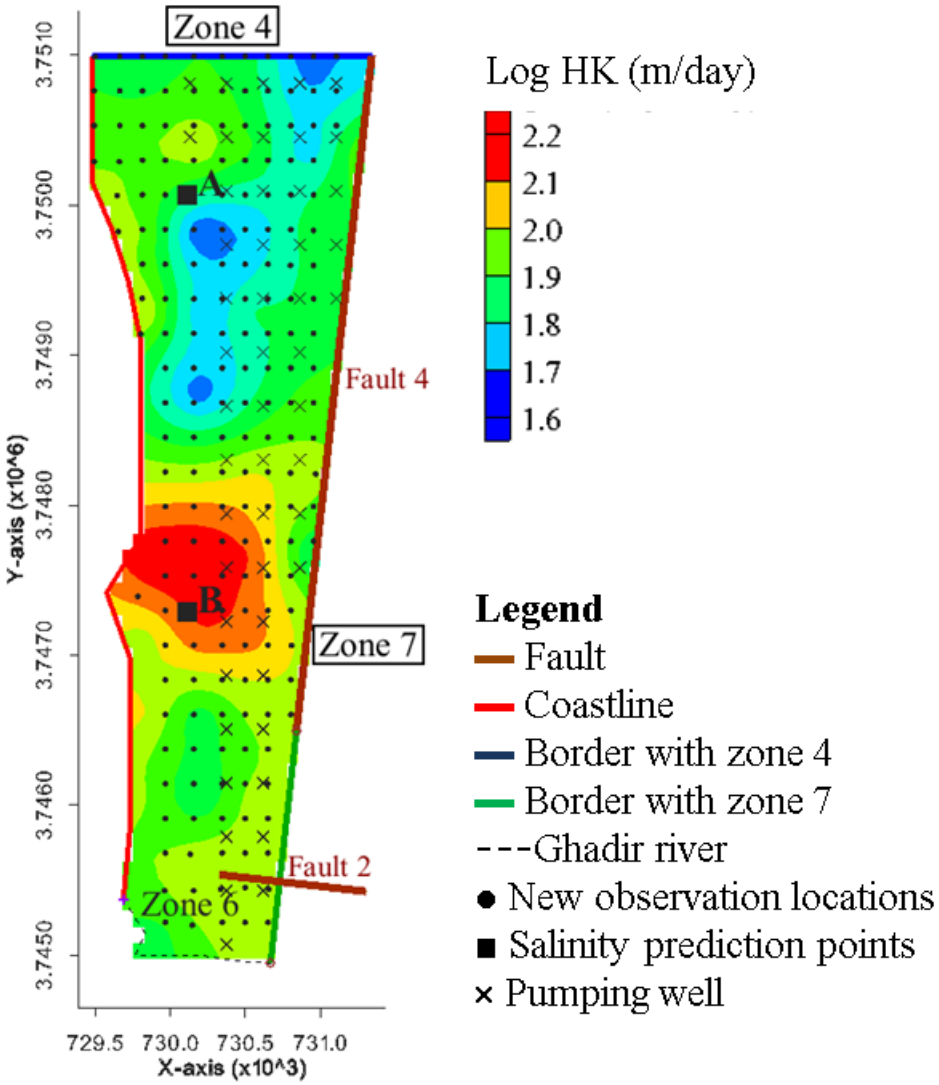

Figure 6: Model grids, boundary conditions, and locations of abstraction wells in the upper part of the Beirut coastal aquifer

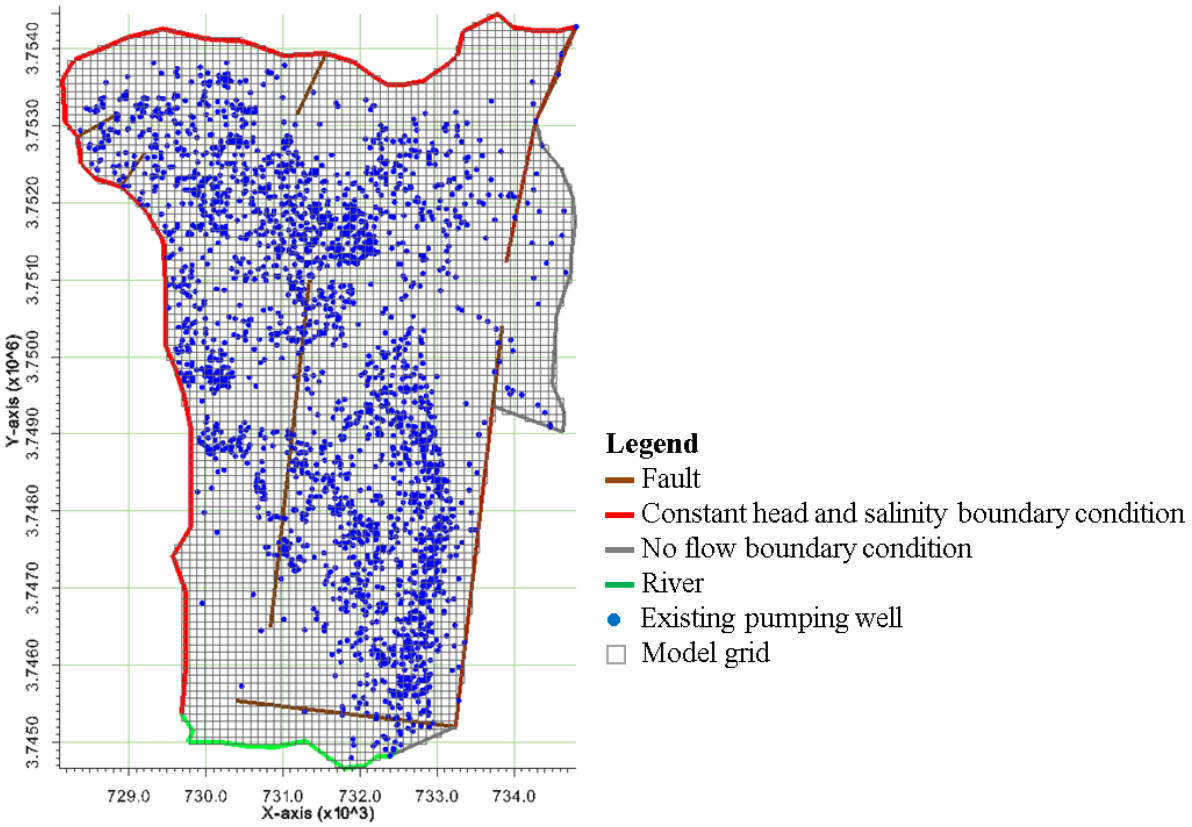

Figure 7: Log hydraulic conductivity distribution in zone 6 using 10 stochastic models Mk: k=1, 2 … 10; along with the position of the interface with 75% (black), 50% (gray), and 25% (white) of sea water concentration (35gr/l) after 50 years simulation

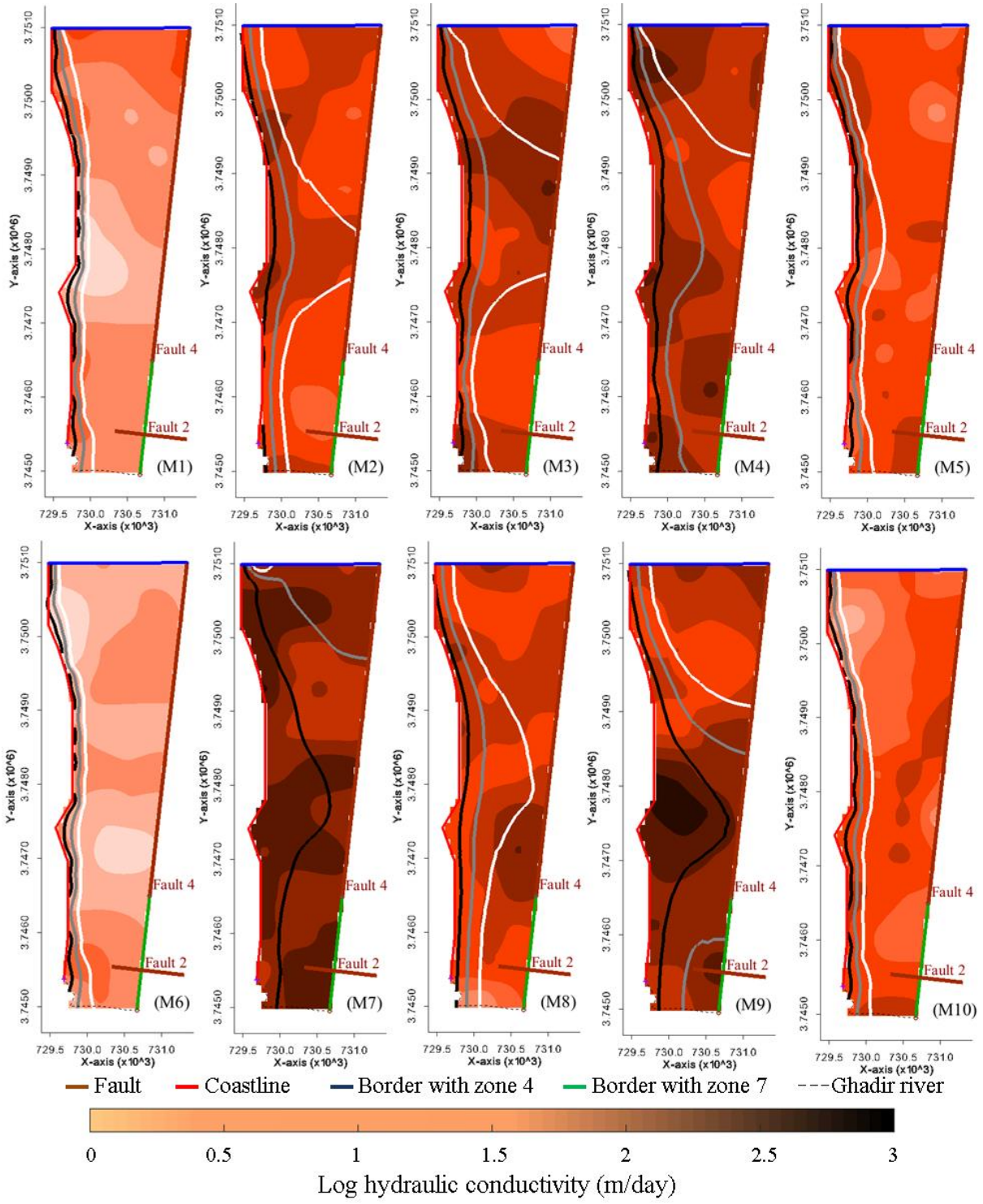

Figure 8: Histograms for log hydraulic conductivity values for zone 6 and for the corresponding model prediction of the percent change in the position of the 3D salt/fresh water interface with salinity >1gr/l from the coastline

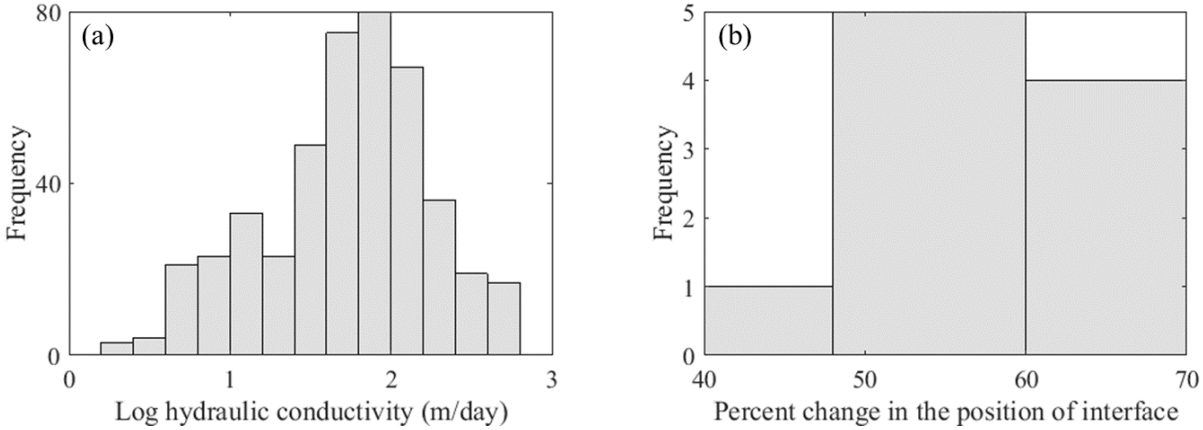

Figure 9: Data worth (DW) of new (yet to be collected) observation locations with measurements of head and salinity at different depths coresponding to prediction variables:
1 (interface displacement); 2 (salinity increase at point A); and 3 (salinity increase at point B)

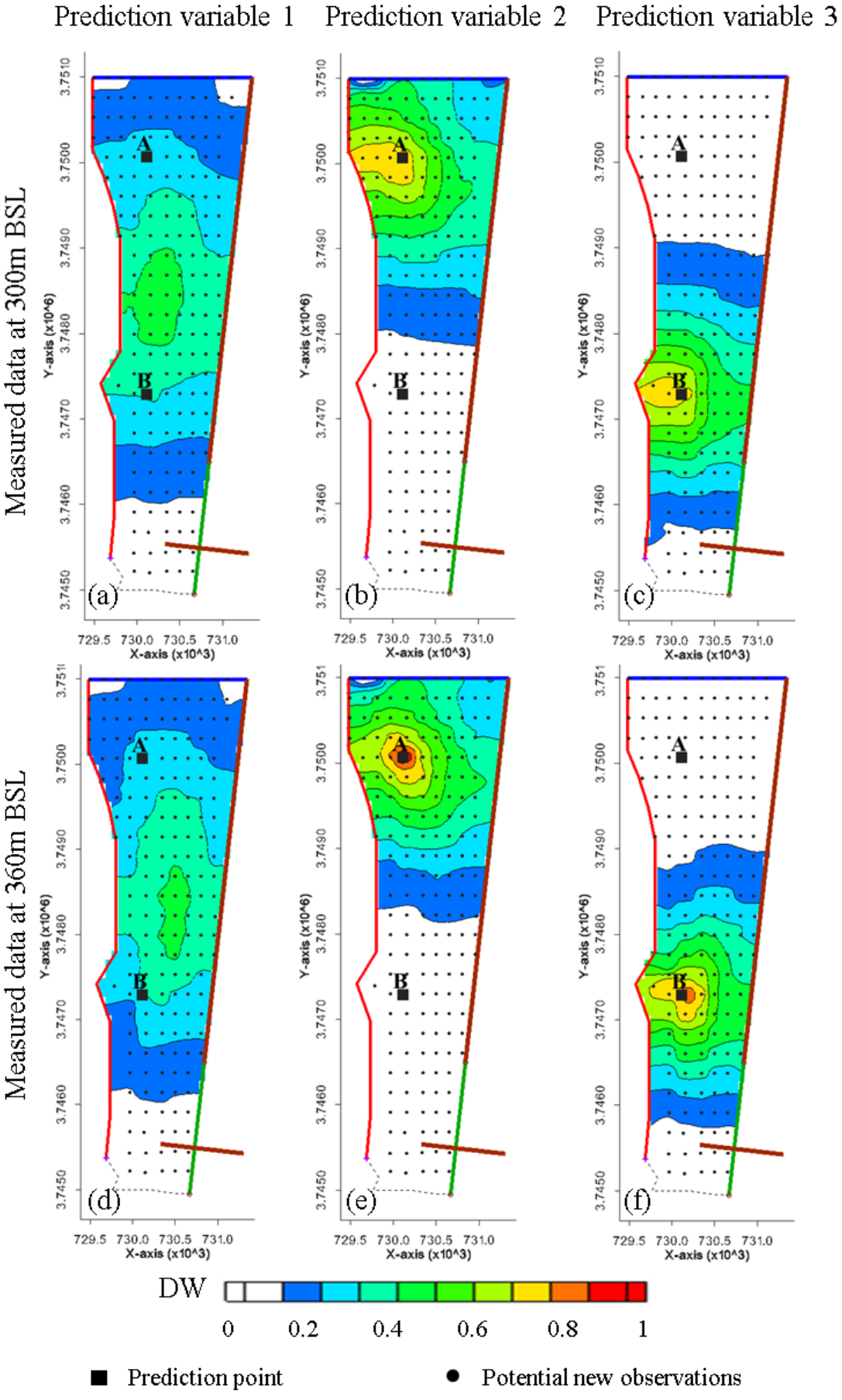

Figure 10: Black solid triangle represents the optimal design location for a single observation with measurement of head and salinity at multiple depths for multiple prediction variables - a) w1=w2=w3=0.33; (b) w1=0.6, w2=0.3, w3=0.1; (c) w1=0.6, w2=0.1, w3=0.3; (d) w1=0.8, w2=w3=0.1; (e) w1= 1, w2=w3=0. wi is prediction weight - colored contours are VI that was averaged over the models Mk- locations of points A and B are shown in black rectangles

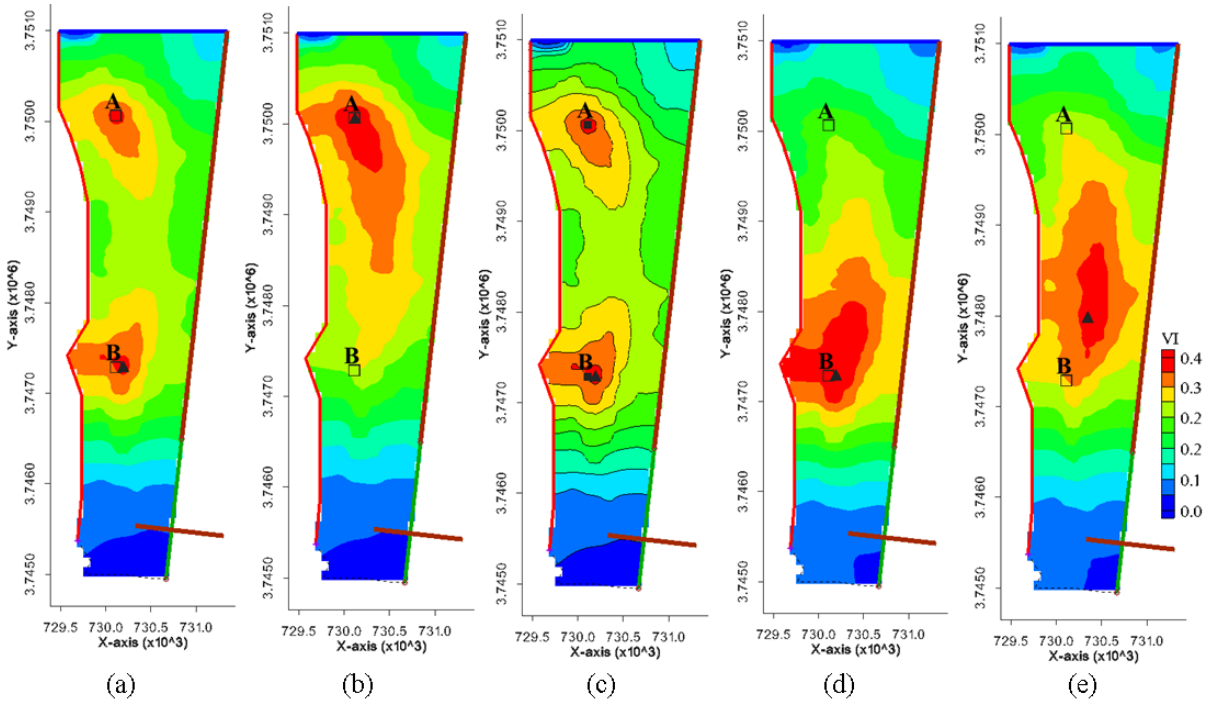

Figure 11: Estimated DW of proposed designs Dk from models Mk (k=1,2…10) for N=2, 3, 4, and 5 observation wells in plots a to d respectively: x-axis corresponds to a proposed design specified using a model Mk and y-axis is the estimated DW when applying a proposed design on all models – Black dot is the mean estimate of the DW and red line is the median of the estimated DWs for each design using all models.

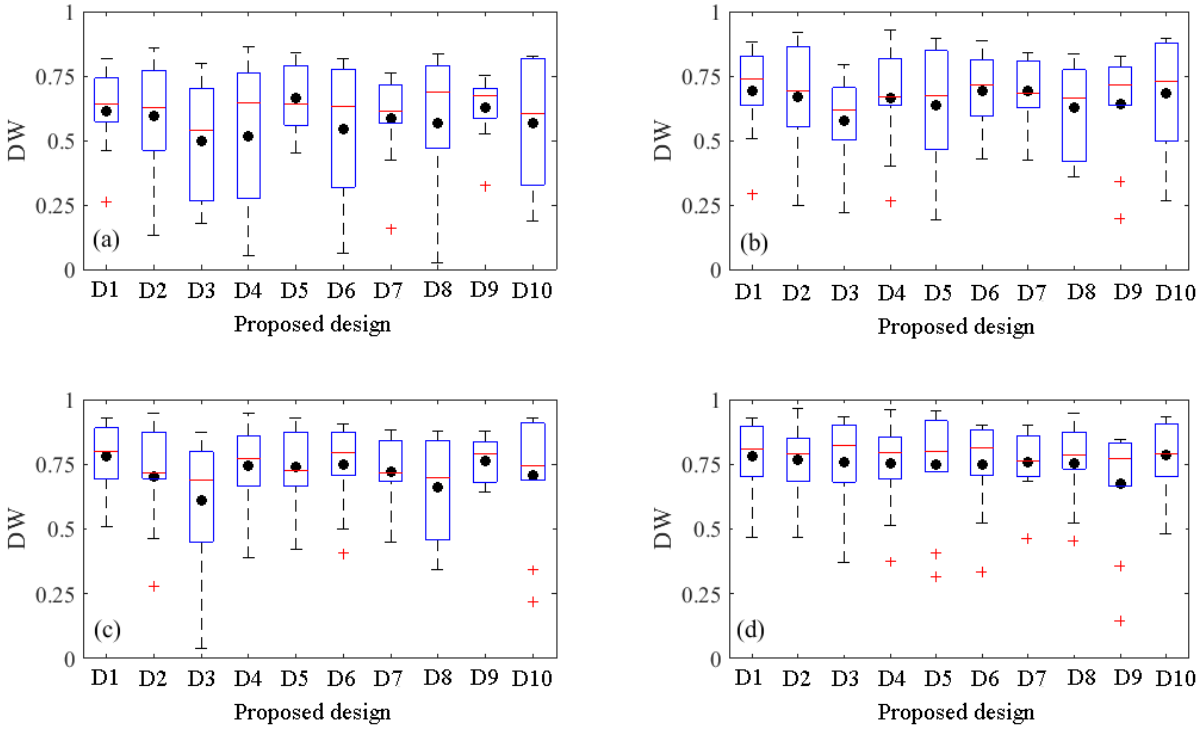

Figure 12: Optimal design for N=1, …5 new observations with measurement of salinity at multiple depths for predicting the displacement of the salt/water interface. ▲ is a proposed location for a new observation. The colored contours represent the DW for a single observation with measurement at multiple depths

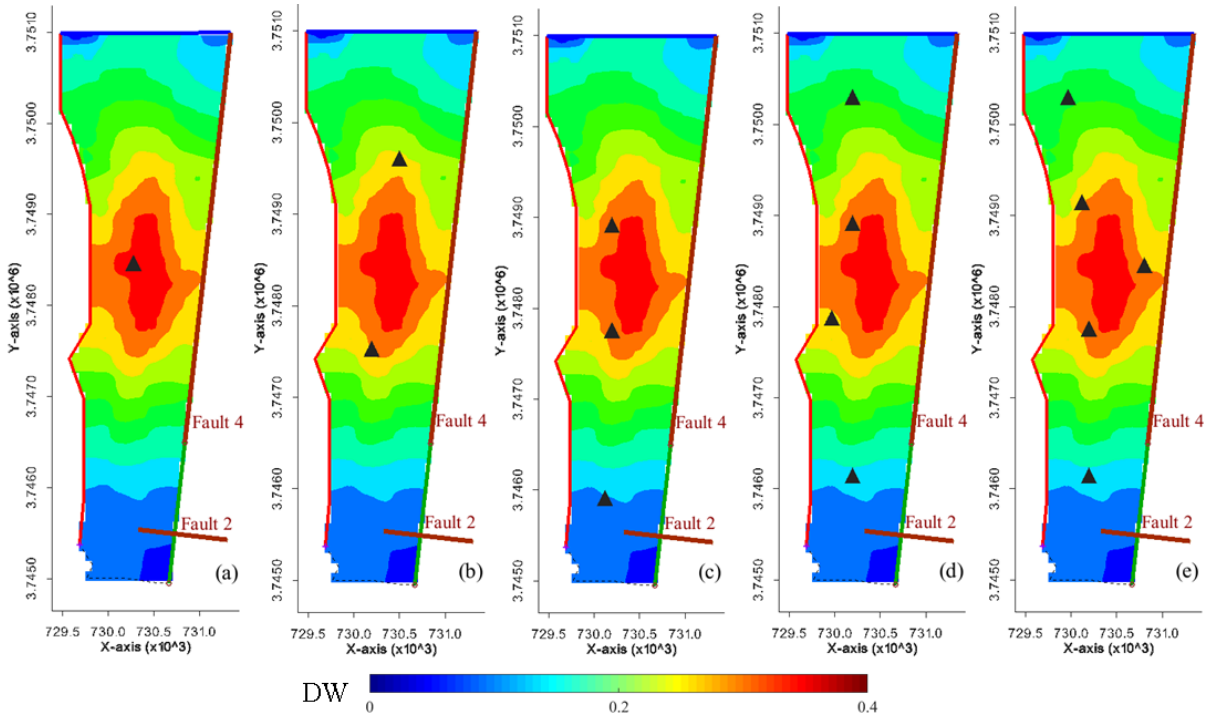

Figure 13: Cost-effective analysis for size of proposed design: x-axis shows the ratio of the cost of implementing the first observation (P1) to the start-up cost (Ps) (which is scaled between 0 and 1), y-axis is the reduction of uncertainty (data worth (DW)) with predicting the 3D displacement of interface to the full cost of a project (Po), dots correspond to an observation set of n=1,2…5 observations – each plot shows the different analysis according to the operation cost of an additional observation ($P1^{+}$) to the cost of implementing the first observation (P1) – DW is constant at each design size and only the cost of design varies