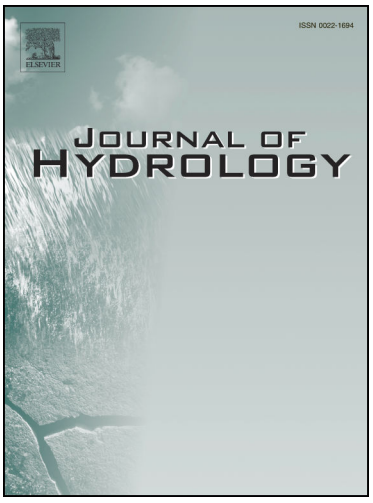
Journal Pre-proofs



Research papers

Efficient Uncertainty Quantification for Seawater Intrusion Prediction using Optimized Sampling and Null Space Monte Carlo Method

Samia Saad, Akbar A. Javadi, Raziyeh Farmani, Mohsen Sherif

PII: S0022-1694(23)00438-9
DOI: https://doi.org/10.1016/j.jhydrol.2023.129496
Reference: HYDROL 129496

To appear in: *Journal of Hydrology*

Received Date: 14 November 2022
Revised Date: 31 March 2023
Accepted Date: 4 April 2023

Please cite this article as: Saad, S., Javadi, A.A., Farmani, R., Sherif, M., Efficient Uncertainty Quantification for Seawater Intrusion Prediction using Optimized Sampling and Null Space Monte Carlo Method, *Journal of Hydrology* (2023), doi: https://doi.org/10.1016/j.jhydrol.2023.129496

Efficient Uncertainty Quantification for Seawater Intrusion Prediction using Optimized Sampling and Null Space Monte Carlo Method

Samia Saad [a,b], Akbar A. Javadi [c], Raziyeh Farmani [d], Mohsen Sherif [e]

[a] Ph.D. student, Department of Engineering, University of Exeter, Exeter, United Kingdom, ss936@exeter.ac.uk

[b] Research associate, Department of Irrigation & Hydraulics, Ain Shams University, Cairo, Egypt, samia.saad@eng.asu.edu.eg

[c] Professor of Geotechnical Engineering, Department of Engineering, University of Exeter, Exeter, United Kingdom, A.A.Javadi@exeter.ac.uk

[d] Professor of Water Engineering, Department of Engineering, University of Exeter, Exeter, United Kingdom, R.Farmani@exeter.ac.uk

[e] Professor of Water Resources, Department of Civil and Environmental Engineering, United Arab Emirates University, Al Ain, UAE, msherif@uaeu.ac.ae

Correspondence to: Samia Saad (ss936@exeter.ac.uk)

Abstract

Uncertainty in environmental modeling predictions, stemming from parameter estimation, is a crucial challenge that must be addressed to ensure effective decision-making. Limited field measurements, high computational costs, and a lack of guidance in estimating measurement uncertainty further compound this challenge, particularly for highly parameterized complex models. In this study, we propose a novel and computationally efficient framework for quantifying predictive uncertainty that can be applied to a range of environmental modeling contexts. The novel components of the framework include efficient parameter space sampling using an Optimized Latin hypercube sampling strategy, and applying the Null Space Monte Carlo method (NSMC) along with a developed filtering technique. The NSMC generates sample sets to calibrate the model while exploring the null space. This space contains parameter combinations that are not sufficiently supported by observations. The filtering technique omits low-potential parameter sets from undergoing model calibration. The framework was tested on the seawater intrusion (SWI) model of Wadi Ham aquifer in the United Arab Emirates (UAE) to investigate aquifer sustainability in 2050. Our results demonstrate the importance of incorporating direct and indirect measurements of heads, salinity, and geophysical survey data into the calibration dataset to reduce uncertainty in salinity predictions. The extent of SWI for multiple calibrated parameter sets varied by 4.5% to 11% relative to their means at two main pumping fields. We conclude, with a moderate to a high degree of certainty, that SWI is a serious threat to these fields, and actions are needed to protect the aquifer from salinization. Additionally, variations in SWI length under different geological conditions illustrate regions of high uncertainty that require further data collection. Our framework effectively reduced and quantified prediction uncertainty and provides decision-makers with critical information to inform risk management strategies.

Keywords

Seawater intrusion, Parameter estimation, Data worth, Parameter identifiability, Optimized Latin hypercube sampling, uncertainty analysis.

1. Introduction

Groundwater is a vital source of freshwater for over two-thirds of the world's population living in coastal regions (Singh, 2014). However, this resource is threatened by seawater intrusion (SWI) (Werner et al., 2013). To mitigate, the negative impacts of SWI, numerical models are used to develop effective management strategies. However, these models are only as reliable as the input. Some data such as aquifer geometry, hydraulic properties, and boundary conditions, are often uncertain due to a lack of field measurements, aquifer heterogeneity, regional

flow systems, and forecasted salinization drivers. This uncertainty leads to uncertain predictions (Doherty, 2016), making it imperative for stakeholders and decision-makers to quantify the prediction uncertainty (Doherty and Moore, 2020) in order to make informed decisions.

To address this issue, predictive uncertainty analysis is required to characterize the potential variability in a prediction (model output) through the identification of multiple parameter sets (input variables) that can be used by the model (Dausman et al., 2010). Table 1 summarizes the main methods for quantifying the prediction uncertainty in SWI. The three methods for predictive uncertainty analysis include: Monte Carlo simulations, Conditional Monte Carlo simulations, and Linear uncertainty analysis. Monte Carlo simulations method, which has been used in both conceptual (e.g., Ketabchi and Jahangir, 2021; Rajabi and Ataie-Ashtiani, 2014; Zhao et al., 2016) and real-world studies (e.g., Miao et al., 2019; Mostafaei-Avandari and Ketabchi, 2020). It repeatedly runs the model with different sets of parameters to generate multiple realizations of the model output. The conditional Monte Carlo simulations method, on the other hand, conditions the input variables to observed data. The approach has two perspectives. The first perspective considers the non-uniqueness of the parameter sets through the model calibration in a stochastic inverse modeling framework, such as Markov Chain Monte Carlo (MCMC) (Rajabi and Ataie-Ashtiani, 2016; Zeng et al., 2016), self-calibrating (Pool et al., 2015), gradual conditioning (Llopis-Albert et al., 2016), and iterative ensemble smoother (iES) (Hugman and Doherty, 2022) methods. The second perspective seeks uniqueness by implementing regularization to incorporate prior information on parameters within model calibration. Depending on the inferred single-calibrated parameter set, the Null space Monte Carlo (NSMC) (Tonkin and Doherty, 2009) method generates multiple realizations to estimate the prediction uncertainty (e.g., Herckenrath et al., 2011). Linear uncertainty analysis assumes linear model behavior and represents the relationship between parameters and predictions with a matrix that is independent of parameter values (e.g., Coulon et al., 2021).

Table 1 also highlights two other key features; the considered number of uncertain parameters and the adopted ways to alleviate the computational burden associated with the prediction uncertainty analysis. Although most of the previous investigations have limited the model parameters to a small number (Ketabchi and Jahangir, 2021; Miao et al., 2019; Mostafaei-Avandari and Ketabchi, 2020; Pool et al., 2015; Rajabi and Ataie-Ashtiani, 2014; Rajabi and Ataie-Ashtiani, 2016; Yang et al., 2021; Zhao et al., 2016) to reduce the computational cost (Rajabi et al., 2018), highly parameterized models are necessary to reduce predictive uncertainty by incorporating parameters that are salient to the model predictions (Doherty and Moore, 2021) and to recognize the small-scale variability of the hydraulic property that greatly influences the hydrodynamic process. For instance, the degree of the aquifer heterogeneity has a considerable effect on the SWI wedge length and seawater volume (Ketabchi and Jahangir, 2021). Among the methods used, NSMC and iES can reduce the computational burden and handle nonlinear highly parameterized models. Despite the computational efficiency of iES, its success is highly dependent on the quality and size of the ensemble of realizations that are sampled from the prior parameter distributions. Thus its prior parameter distribution needs special considerations (Delottier et al., 2022). Additionally for highly non-linear problems, the ensemble-based method may produce instability and run failures (Omagbon et al., 2021).

Previous studies have shown the effectiveness of the NSMC method in a highly parameterized SWI model, and its efficiency in reducing the computational burden. Keating et al. (2010) compared the performance of the NSMC method to the MCMC (a statistically rigorous and computationally demanding method) in terms of model calibration, estimating parameters, and prediction uncertainties on a high dimensional parameter estimation groundwater problem. The outcomes had a high degree of similarity, however, NSMC obtained the results with much less computational time. Herckenrath et al. (2011) proved the efficiency of the NSMC method in quantifying the uncertainty in the prediction of the location of the SWI interface due to the inflow reduction using a synthetic SWI model based on Henry problem. Safi et al. (2019) designed a monitoring network in Beirut aquifer, Lebanon to capture the SWI behavior through the optimization of the best potential locations of the observation wells to reduce the uncertainty in the prediction of the interface location using the realizations generated by NSMC method.

However, further improvement in computational efficiency is needed. Three promising ways are found in the literature to address this issue (Table1) including: 1) parallelization and grid computing (e.g., Mostafaei-Avandari and Ketabchi, 2020), 2) efficient sampling strategies that decrease the required number of simulation runs to obtain a stable statistical mean and variance for the prediction probability distribution (Janssen, 2013) with an acceptable accuracy (e.g., Rajabi and Ataie-Ashtiani, 2016; Zhao et al., 2016), and 3) development of a surrogate model to act as a simplification to the complex model. Lower-fidelity models were used in highly parameterized models (Burrows and Doherty, 2015; Hugman and Doherty, 2022). Although data-driven surrogate models have been used to handle models with a limited number of parameters (e.g., Miao et al., 2019; Rajabi and Ataie-

Ashtiani, 2014), they have limitations for highly parameterized models where the number of samples required to capture the behavior grows prohibitively fast with the number of parameters (Asher et al., 2015) and can increase prediction uncertainty especially for large-scale, and complex models (Mostafaei-Avandari and Ketabchi, 2020).

Decreasing the unnecessary simulation runs in Monte Carlo-based methods, via implementing more efficient sampling strategies that combine the innovative space-filling criteria and optimization schemes, is not widely used in the literature (Rajabi et al., 2015), despite its popularity in limiting the required number of evaluation points to represent the design space (Wang et al., 2020). While simple random sampling (SRS) is the standard method (e.g., Herckenrath et al., 2011; Pool et al., 2015), it does not ensure space-filling , which means that one element may be chosen more than once. Conversely, Latin hypercube sampling (LHS) approach (McKay et al., 1979) divides the range of each input variable into equally probable intervals, the number of which is equal to the sample size. Within each interval, a random sample is selected for each variable, then the drawn samples are paired with each other randomly. Therefore, LHS requires a smaller sample size than SRS for space-filling. Applying space-filling optimality criterion to LHS will: (i) enhance its performance, especially in large dimensions, (ii) ensure the filling of the input space properly avoiding close points (Damblin et al., 2013), and (iii) minimize the number of the Monte Carlo simulation runs to reach the target accuracy (Janssen, 2013). Rajabi and Ataie-Ashtiani (2014) showed that using optimized Latin hypercube sampling (OLHS) based on the enhanced stochastic evolutionary (ESE) optimization and employing the uniformity criterion was the most efficient approach in terms of the consumed time for Monte Carlo simulations. This was concluded based on a comparison of this strategy to SRS, LHS, and eight other OLHS strategies (differ in the optimization method and/or the space-filling criterion) on two test cases involving SWI models. Mostafaei-Avandari and Ketabchi (2020) employed OLHS based on ESE optimization and the uniformity criterion in Monte Carlo simulations to study the uncertainty in permeability and assess its impact on SWI of Ajabshir coastal aquifer, Iran.

Uncertainty quantification is critical for accurately predicting seawater intrusion (SWI) in coastal aquifers but often hindered (Herrera et al., 2022) by the limitations of traditional methods, particularly in terms of computational efficiency. A critical research gap in previous studies (Table 1) that needs to be explored is how to efficiently quantify uncertainty in highly parameterized, strongly nonlinear SWI models caused by uncertain aquifer parameters. This study presents a novel and efficient framework for quantifying uncertainty in SWI prediction by combining the benefits of the Optimized Latin Hypercube Sampling (OLHS) and Null Space Monte Carlo (NSMC) methods. The framework is applied to the coastal aquifer of Wadi Ham in UAE, which is facing severe SWI due to over-exploitation.

To the best of the authors' knowledge, this study is the first to consider the combined advantages of OLHS and NSMC for efficient uncertainty quantification in highly parameterized nonlinear SWI models, particularly in heterogeneous coastal aquifers. The proposed framework includes: (i) Calibrating a highly parameterized model using regularized inversion, (ii) Defining the calibration solution space using matrix decomposition techniques, (iii) Sampling the parameter space efficiently using OLHS, (iv) Applying the NSMC method to generate parameter sets with the same calibration space, (v) Filtering less likely parameter sets from the recalibration using a proposed rejection sampling algorithm, (vi) Generating posterior parameter and predictive probability distributions using the ensemble of the calibration-constrained parameter sets.

This study provides a further improvement in computational efficiency for the NSMC method and makes a valuable contribution to the field by providing a practical, and computationally efficient framework for quantifying uncertainty in SWI prediction.

2. Methodology

A six-step general framework has been developed to quantify the prediction uncertainty. A flowchart showing the overall methodology is presented in Figure 1. Each step is described below.

Step 1: Uncertainties of the calibration dataset

The uncertainty values reflect the noise in the measurements. Their values are assigned based on the used equipment to collect the data, expert knowledge, and/or from other studies with similar site conditions. The sources of the uncertainty in the calibration dataset of groundwater head and salinity concentration at each observation well, and the earth resistivity imaging data were estimated as follows:

- Groundwater head observations

The piezometric heads were collected as the measured depths to the water table at the observation wells (Data were extracted from available sources as discussed in Section 3). The depths were taken manually using a calibrated measuring tape, while satellite mapping was used to identify the elevation of the top of each well casing. This was done to obtain the head of the fluid within each well, with the obtained heads being related to the actual fluid density. To compare with simulated heads, a conversion to equivalent freshwater heads was carried out using Eq.1.

$$h_f = \frac{\rho}{\rho_0} h - \alpha z \quad (1)$$

$$\alpha = \frac{\rho - \rho_0}{\rho_0} \quad (2)$$

where, h_f is piezometric head of the equivalent freshwater head (m), which is given by $h_f = \left({}^{P}/_{\rho_0 g}\right) + z$, P is the fluid pressure (N/m^2), h is the measured head (m), α is the density ratio, ρ, ρ_0 are the actual fluid density, and freshwater density respectively (kg/m^3), z is the elevation at the midpoint of the screen length inside the observation well (m). The datum of the hydraulic heads is the mean sea level.

The total uncertainty of the equivalent freshwater heads was calculated (Eq. 3) based on Eq. 1 assuming that ρ_0 is constant and the errors in the measurements are random and uncorrelated (Post et al., 2018):

$$\sigma_h = \sqrt{\left(\frac{\rho}{\rho_f} h \,.\, \sqrt{\left(\frac{\sigma_\rho}{\rho}\right)^2 + \left(\frac{\sigma_{h_{measured}}}{h}\right)^2}\right)^2 + \left(\alpha z \,.\, \sqrt{\left(\frac{\sigma_\rho}{\rho - \rho_0}\right)^2 + \left(\frac{\sigma_z}{z}\right)^2}\right)^2 + (\sigma_{pump})^2} \quad (3)$$

where, the first and second terms represent the uncertainties propagated from the first and second terms of Eq.1, and the third term considers the modeled to measured misfit at the observation well, which is near a pumping field. Symbols used in Eq.3 are defined in Table 2.

- Salinity concentration observations

The measured electrical conductivity for the water samples collected from the observation wells was converted to salinity concentration data. The uncertainty associated with the measurements is shown in Table 2.

- Earth resistivity tomography (ERT) data

The average length of SWI was estimated from ERT transects data, measured from the coastline to the start of the freshwater zone (total dissolved solids of 1000 mg/l), and was used to constrain the coupled flow and transport model as indirect observations for the salinity distribution. The uncertainty associated with the ERT intrusion length (σ_{ERT}) observations is presented in Table 2. A custom Python script was used to include the indirect observation of the intrusion length in FePEST during the parameter estimation. FePEST links PEST, and FEFLOW (Diersch, 2013). PEST is a model-independent non-linear parameter estimation software (Doherty, 2016). FEFLOW solves the system of partial differential equations for groundwater flow and for variable-density flow coupled with the mass transport equation.

Table 2 shows the standard deviation values for the above-mentioned sources of uncertainty associated with the field measurements. The measurement noise is assumed to show no spatial or temporal correlation and is characterized by a multiGaussian distribution. This distribution is defined by mean function (observed values of the calibration dataset) and covariance function (diagonal elements of the covariance matrix are the weights assigned to the various field measurements, and are the inverse of the standard deviation values).

Since, the calibration dataset, which included observations of heads, salinity, and ERT survey data was used in the parameter estimation problem, a multi-component objective function resulted (Eq.4). Due to the wide range

of values in the heads and concentration measurements, they were partitioned internally, with each hydrograph or salinity concentration forming a separate group that represented a component. To filter out structural noise in groundwater models, a strategy proposed by Doherty and Welter (2010) was adopted. The model-to-measurement misfit is dominated by structural noise (Doherty and Welter, 2010). The strategy gives equal recognition to each component in the objective function, by assigning different weights to each group. The weights were estimated as the squared inverse of the standard deviation $(\frac{1}{\mathbf{stdev}})^2$ (Table 2) multiplied by a coefficient that differed between groups to ensure equality.

Step 2: Data worth assessment

A prior step to the parameter estimation is to evaluate the worth of the information in the calibration dataset to reduce the prediction uncertainty. To examine the worthiness of the data, each observation group from the calibration dataset was removed successively, and the incurred increase in the uncertainty of the predictions was calculated (Doherty et al., 2011). The data worth was assessed using linear uncertainty analysis (first-order, second-moment-based analysis) as an approximate and low computational cost analysis tool. Detailed descriptions of the used equations and their derivations can be found in (Dausman et al., 2010). The assessment of the data worth of each observation group in all observation types was done by processing PEST files, and implementing them in Python using the PyEMU (Python framework for Environmental Modeling Uncertainty analyses) library (White et al., 2016).

Step 3: Model calibration using regularized pilot points method

The mass transport process in the saltwater wedge is influenced by heterogeneities in the aquifer, leading to preferential flow pathways and spatial variations in transport velocity. To account for small-scale variability, the pilot point method was used by distributing predefined points in the model domain on a pseudo-regular grid, with a minimum of five points per correlation range (Alcolea et al., 2006). The parameters were estimated at these points, and then spatially interpolated using a selected geostatistical model (variogram function) to obtain aquifer properties at all model elements. The pilot point method was then incorporated with Tikhonov regularization to penalize the model parameters' deviation from their prior estimates only to an extent that provided an acceptable fit to field measurements (Doherty, 2016). These prior estimates were derived from pumping test results, prior geological knowledge, and previous studies.

The regularized inversion (i.e., model calibration) attempted to minimize the global objective function Φ (Doherty, 2016), which was defined as:

$$\Phi = \Phi_m + \mu \Phi_r \quad \mathbf{4}$$

$$\Phi_m = (c - X\,p)^t\, Q_m\, (c - X\,p) \quad \mathbf{5}$$

$$\Phi_r = (d - R\,p)^t\, Q_r\, (d - R\,p) \quad \mathbf{6}$$

where; Φ_m is the measurement objective function. Φ_r is the regularization objective function. c is the vector of the measurement observations. p is the vector of the parameters requiring estimation. X is the calibration sensitivity matrix (Jacobian matrix) of size (m×n), where m and n are the numbers of field observations used in the calibration dataset and the model parameters respectively. Each element of the matrix expresses the sensitivity of each model output (corresponding to the field observations) to each parameter. Q_m is the matrix of the weights assigned to the different observations. It represents the noise associated with the measurements and induced by model imperfection (structural noise). Q_r is the inverse of the pilot point covariance matrix, whose elements were calculated based on the selected geostatistical model. t denotes the transpose operation. d is the vector of the regularization observations, which expresses a preferred condition. R is the matrix that encapsulates the regularization equations. It represents constraints on parameter values. In the present study, R is the identity matrix for preferred values constraints. μ is the Tikhonov regularization parameter.

Model calibration was conducted using FePEST. BeoPEST was used to achieve further computational efficiency by running a model simultaneously on different cores of the same computer. The sensitivity matrix was filled based on finite differences with a 1.5% parameter perturbation. The Tikhonov regularization parameter (μ) was calculated in every optimization iteration during the inversion process by PEST to achieve a target level of model-to-measurement fit. PEST viewed the problem as a constrained optimization problem, where the objective was to minimize Φ_r subject to the constraint that Φ_m should be less than the maximum acceptable misfit between model outputs and the field data.

Truncated Singular Value Decomposition (SVD) and Tikhonov regularization were combined using hybrid regularization to address their weaknesses and benefit from their strengths (Tonkin and Doherty, 2005). SVD separated inestimable parameter combinations (zero and low singular values) residing them in the calibration null space and created a super parameter set of estimable parameters residing them in the calibration solution space. This reduced dimensionality, stabilized the inversion process, and reduced computation time.

The model calibration performance was evaluated using statistical indices, including Root Mean Square Error (RMSE), Coefficient of determination (RSQ), and The Nash-Sutcliffe Model Efficiency (NSE). These indices provided measures of the difference, similarity, and fit between the observed and simulated datasets, respectively.

Step 4: Parameter identifiability

To reduce the computational time, the parameter estimation was performed in two phases. In phase 1, head data were used and in phase 2 the salinity concentrations and geophysics survey data were used. In phase 1 the parameters which had a high identifiability value (parId >0.75) were identified as adequately estimated. By considering these parameter values as fixed parameters in the next phase, a reduction in the computational burden was achieved.

The parameter identifiability (parId) was estimated to provide insight into how much information was extracted from the observations to inform the model parameters during the calibration using Eq.7 (Doherty and Hunt, 2009).

ParId has a moderate computational burden, considers the parameter correlation, and approximates the relationship between the model outputs and the parameters as linear (Doherty et al., 2010). ParId value varies between zero and 1. Zero indicates completely non–identified parameters and one indicates that they lie in the calibration solution space and are fully identified. The weighted sensitivity matrix ($Q_m^{1/2}X$) was computed based on calibrated parameter values and the inverse of the measurements noise. It was subjected to SVD employing GENLINPRED utility (Doherty, 2016). SVD divides the parameter space into solution subspace (V_1) and null subspace (V_2). The threshold of truncation between the solution and null subspaces was calculated using the SUPCALC utility (Doherty, 2016).

The identifiability of a parameter is calculated as the sum of the squared components of all the Eigenvectors spanning the solution subspace (V_1) corresponding to that parameter as follows:

$$\text{ParId} = \sum (V_1{}^2) \quad 7$$

Step 5: Model validation

The calibrated parameter set was assigned. A different observation dataset than that used in model calibration was considered for the validation. The performance of the model validation was evaluated using the above-mentioned statistical indices of RMSE, RSQ, and NSE.

Step 6: Prediction uncertainty analysis

The NSMC method developed by Tonkin and Doherty (2009) was used for prediction uncertainty analysis. A brief description of the method's steps is presented in the supplementary material S(A). The method is based on generating random fields (p) from the prior parameter probability distributions, using a stochastic parameter generator. SRS is commonly used, however in this study, Optimized Latin hypercube sampling (OLHS) was used to improve the sampling efficiency. The difference between the stochastic parameters (p) and the calibrated

parameters ($\overline{p}$) was projected onto the calibration null-space (V_2). The projected difference field was then added to the calibrated parameters. The new set of parameters (p_{new}) did not produce exactly the calibrated parameter set due to model nonlinearity and the considered truncation threshold between the solution and null subspaces. The model was then executed using p_{new} and re-calibrated if the mismatch exceeded an acceptable level. The existing sensitivities (estimated using the calibrated parameter set) were used in the first optimization iteration to save computational time. If additional iterations were required for the recalibration, the sensitivities were recalculated based on the newly obtained parameter set. The objective function threshold was considered 10% higher than the achieved objective function during the calibration. In the present work, a maximum of two iterations were conducted for each realization. The prediction uncertainty was assessed using the resulting parameter sets after the recalibration. A flowchart of the NSMC method is shown in Figure 1.

Drawing samples from a prior parameter probability distribution which is too wide, may lead to few potential parameter sets that can be accepted while reducing the span of the prior parameter space can sacrifice the integrity of the process (Doherty et al., 2011). In this study, a straightforward filtering technique was proposed to screen out from recalibration the parameter sets that were less likely to reach the acceptable objective function in the second iteration without reducing the parameter space. A flow chart for the filtering process is presented in Figure 2. The details are described below:

1. The achieved objective functions of the parameter sets, that need to undertake additional iterations after the first iteration, are sorted in descending order to form a ranked list.
2. The second iteration is carried out first using the parameter set that corresponded to the objective function in the middle of the list (in this way the whole list is split into upper and lower lists).
3. After the second iteration, if the achieved objective function is acceptable, step 2 is repeated using the parameter set, which was associated with the objective function in the middle of the upper list. Otherwise, this parameter set and the others associated with higher objective functions are excluded from the list. Step 2 is repeated using the parameter set in the middle of the lower list, and so on. This procedure is stopped when reaching the top of the list which had undergone the second iteration and succeeded in satisfying the threshold objective function.
4. The second iteration is performed on all the remaining parameter sets.

Step 6 is divided into six substeps (Figure 1). Further descriptions of substeps 6-iii and 6-vii are shown as follows:

Optimized Latin hypercube sampling (Step 6-iii)

The efficiency of Monte Carlo methods is highly dependent on the space-filling characteristics of the sampling strategy (Janssen, 2013; Rajabi and Ataie-Ashtiani, 2014). Latin hypercube sampling aims to ensure that each of the input variables, has its range well scanned based on a probability distribution. The quality of filling can be further improved by optimizing the distribution of the sample points. The selected space-filling criterion was a point-distance criterion (φ_p criterion) which controlled the distance between the design points themselves or the input domain and the points (Damblin et al., 2013). In order to optimize the selected criterion, the enhanced stochastic evolutionary (ESE) technique was employed.

φ_p criterion was proposed by Morris and Mitchell (1995) and is calculated as follows:

$$\varphi_p = \left[\sum_{i=1}^{s} J_i d_i^{-p}\right]^{1/p} \quad 8$$

By sorting all the inter-sited distances $d_{ij}(1 \le i,j \le n, i \ne j)$ for a given LHS design, a distance list (d_1,d_2,…d_s) and an index list (J_1,J_2,…J_s) are obtained. d_{ij} is the Euclidean distance between two sample points i,j, n is the number of sample points, each sample point represents the parameters vector, d_i represents the distinct distance value, the elements in the distance list are sorted in ascending order ($d_1 < d_s$), J_i represents the number of pairs of sites separated by d_i in the design, and s is the number of the distinct distance values. p was taken as 10.

The enhanced stochastic evolutionary (ESE) algorithm was used to generate an LHS design satisfying the φ_p criterion. The method began with an initial design and tried to find better designs by iteratively changing the current design. The method consisted of an inner loop and an outer loop. In the inner loop, the new designs were

explored, while the outer loop controlled the acceptance of the new design using a threshold-based acceptance criterion. The acceptance criterion was as follows:

$$\Delta f \leq T_h \times random(0,1) \quad 9$$

where Δf is the amount of improvement made by the new design compared to the current design, $random(0,1)$ is a function that generates uniform random numbers between 0 and 1, and T_h is the threshold.

The threshold value was controlled by the outer loop, in such a way that if the improvement made in the inner loop was less than a certain level, the threshold value was first increased sharply to initiate the exploration process, avoid local optimum, and then decreased slowly to discover better designs. On the other hand, if the improvement made was above a certain level, the improvement process was initiated attempting to find a local optimum quickly. Detailed descriptions of the ESE algorithm can be found in Jin et al. (2005).

The parameters of the algorithm were assigned with the values suggested in Jin et al. (2005). The OLHS was implemented in SMT- surrogate modeling toolbox, an open-source Python package (Bouhlel et al., 2019). The run batch file in the PEST software suite was edited to read the sample sets generated by OLHS and to implement the NSMC method (Doherty, 2016).

Generation of posterior parameter uncertainty (Step 6 –vii)

After recalibration, an ensemble of the calibration-constrained parameter sets was obtained. The values of each parameter were used to determine the best-fit probability distribution (pd) that this data followed. After selecting the pd, its distribution parameters (mean and standard deviation) were estimated. Consequently, the posterior pd of each parameter was defined. The magnitude of the distribution parameters was used to 1) assess the degree to which the posterior pd became narrower than its prior pd, and 2) provide some evidence on characterizing the parameter and predictive uncertainty, despite the sample size (number of calibrated parameter sets) being generally small to derive a true posterior parameter pd. Sampling the true posterior parameter pd in the context of highly parameterized, nonlinear, and underdetermined inverse problems is not guaranteed, even by using full-Bayesian approaches and generating thousands of parameter sets (Keating et al., 2010). Bayesian approaches formulate the parameter estimation problem in a probabilistic framework, considering model inputs and outputs as random variables and including prior information and regularization terms in the inference process (Rajabi et al., 2018).

The Anderson-Darling hypothesis test was used to select the best-fit pd. The test compared the fit of the cumulative distribution function of the observed data to the expected cumulative distribution function. Ten probability distributions (Normal, Lognormal, Exponential, 2-parameter exponential, Weibull, Gumbel maximum, Gumbel minimum, Gamma, Logistic, and LogLogistic) were tested as the expected distributions. The distribution test that had p-values less than 0.005 indicated that it was statistically significant to reject the null hypothesis. The null hypothesis stated that the observed data followed this hypothesized distribution. Among the distributions that had strong evidence to be retained, the distribution of the highest p-value was chosen to be the best-fit distribution. Minitab statistical package was used to perform the distribution tests (Arend, 1993).

3. Application of the proposed framework to a real-world case study

3.1. Study site and numerical modeling

Wadi Ham aquifer, a coastal aquifer in an arid region, is located in the northeastern part of the UAE. Limited surface water resources, the prevalence of drought conditions for successive years, and excessive abstractions have placed tremendous pressure on the aquifer and have resulted in severe seawater intrusion problems (Sherif et al., 2014; Sowe et al., 2020). The total area of the aquifer system is 61 km^2. To simulate the groundwater flow and solute transport in the Wadi Ham aquifer, FEFLOW (Diersch, 2013) was employed. The set of governing equations and boundary conditions of density-dependent fluid flow and solute transport in soils is briefly presented in supplementary material S(B2). Table 3 shows the considered hydraulic and transport parameters, hydrogeological and geological information, and the aquifer model discretization details. Table 4 presents the values of the imposed flow and transport boundary conditions in the simulation model. The 3D view of the coastal aquifer model showing the assigned initial and boundary conditions is presented in Figures 3b&c.

3.2. Available observation dataset and estimation of parameters using regularized pilot points method

The available data include transient heads, salinity concentrations, and geophysical survey data. Table 5 presents the available observation dataset for model calibration and validation.

Following the data worth assessment, a decision was taken to conduct the model calibration in two phases (will be discussed in Section 4.1): December 1988 to September 1994 using hydraulic head values with a single fluid constant density model, and October 1994 to March 2005 using salinity concentration and intrusion length estimated from ERT survey data. A coupled flow and solute transport model was used in phase 2. The final parameter set was from phase 2 and adequately estimated parameters with ParId ≥ 0.75 from phase 1 (their values in the second phase of calibration were fixed). The model was validated from October 1994 to April 2005 against the hydraulic heads. Computational time for one run was 10 and 27 minutes for phases 1 and 2, respectively, on a desktop PC with 3.4 GHz Intel(R) Xeon(R) CPU and 192 GB RAM using 10 cores via BeoPEST.

The target objective functions for phase 1 and phase 2 were estimated to be 18236 and 197, respectively, using the assigned weights described in the supplementary materials S(C3), which consider measurement uncertainty and ensure equal visibility for each observation group, and Eq 10.

$$Target\ objective\ function = \sum_{i=1}^{n}\sum_{i=1}^{m}(stdev \times assigned\ weight)^2 \quad 10$$

where $stdev$ is the uncertainty in measurements listed in Table 2, n is the number of the observation groups, and m is the number of readings per group. The standard deviation of the head measurements noise varies slightly between different readings of the same observation group. Average values were considered, while a global standard deviation for the salinity concentration measurements at the observation well was used ($\sigma_{sal} = \sqrt{(\sigma_{sal\_samples})^2 + (\sigma_{sal\_raster})^2}$), since the salinity observation group consisted of readings from two different sources.

The pilot points method was used to describe the spatial characterization of the hydrogeological properties, horizontal hydraulic conductivity (Khz) and the specific yield (Sy) (Table 5). In order to generate spatially correlated fields for them, geostatistical models were initially fitted to the measurement data for both hydraulic conductivity and specific yield (prior information). The geostatistical parameters (sill and range) were optimized so that the interpolated field minimized the misfit between the model outputs and the observation measurements (residuals). The ordinary Kriging interpolation method was used to obtain the parameters values. Table 5 shows the selected geostatistical models and their initial parameters.

The regularization constraints, which were imposed on the parameters were the preferred values deduced from the prior information. The singular value truncation threshold was at a singular value that exhibited the ratio between the lowest to the highest eigenvalue of greater than 1.0×10^{-6}. The mean and the standard deviation describing the prior parameter distributions are listed in Table 5. Figure 4 displays the mean of the prior parameter distributions at the pilot point locations as interpolated fields.

3.3. Future prediction of seawater intrusion

The model predictive uncertainty was quantified for the average seawater intrusion length ($L_{5\%}$). The length is measured from the coastline to the 5% isochlor (The contour line of 2000 mg/l) at the Quaternary aquifer bottom, along the aquifer width. The 5% ischlor was selected as the upper acceptable salinity limit for irrigation, according to the Food and Agriculture Organization of the United Nations (Ayers et al., 1985).

The NSMC method was applied to the SWI model from December 1988 to September 1994 using the transient hydraulic heads (from 1988 to 1994) and the salinity concentrations measurements (in 1994) to construct the calibration dataset. The objective function was estimated at 2455 considering the achieved residuals using the obtained calibrated parameter set, and the assigned weights for each observation group ($\frac{1}{\mathbf{stdev}}$, Table 2). However, for equal visibility for each observation group, salinity weights were multiplied by 20, since the magnitudes of salinity and hydraulic heads were vastly different. A threshold of 2700 was set for the target objective function.

Prior parameters' probability distributions were centered on the calibrated set with standard deviation, bounds, and type as listed in Table 5. 250 stochastic parameter sets were generated. The samples in the OLHS design were drawn from a uniform distribution on [0,1] and then mapped onto the prior distributions respecting the assigned parameters bounds. The chosen number of generated parameter sets was based on computation time, and is comparable to other studies (Pollicino et al., 2021; Rajabi et al., 2015).

The ensemble was used to estimate the uncertainty range in $L_{5\%}$ prediction for 2050. Current abstraction was assumed to persist due to groundwater deterioration and terminated pumping wells (Sowe et al., 2020). Recharge and storage depth at the dam site followed the same trend, while inflow across the boundaries was maintained constant over the prediction period.

4. Results

4.1. Data worth analysis

The data worth analysis was performed on the period of 1988 to 1994 considering prior information on parameter values, the observations data of transient heads from 1988 to 1994 (546 monthly head measurements), salinity measurements in 1994 (11 salinity observations), and the length of intrusion retrieved from ERT survey data. Figure 5 shows the results of this assessment for salinity concentration prediction at 121 uniformly distributed points. The maximum percentage of increase in the prediction uncertainty variance which was incurred by successive removal of each observation group was contoured (Figure 5a). The analysis showed that despite the small number of salinity observations, they had high information content (the predictive uncertainty of 56 points was increased by the maximum compared to the removal of the other observation types), affecting the prediction uncertainty the most in the eastern part of the aquifer (near the coast). The hydraulic head observations had a greater impact on reducing the uncertainty in salinity predictions upstream of the seawater wedge. Generally, across the aquifer domain, the uncertainty in the prediction increased the most near the observation points that were omitted, except near the salinity observations at S15, S16, and S20, and head observations at BHF9A, BHF4, and BHF17. Thus, they did not have information that was relevant to the prediction and/or not repeated in other groups.

It is evident that even though the concentration observations are scarce, it is worth including them in the calibration dataset. This requires combining groundwater equations with salt transport equations by using the variable-density flow and solute transport models. The low values of scaled RMSE (RMSE divided by the range in the simulated heads) between the simulated hydraulic heads using constant-density and variable-density models, ranged from 0.86% to 6.4% for all the observation wells. This suggests that the groundwater flow model could be used in phase 1 for parameter estimation benefitting from its fast run time, with the consideration of the variable density effect in phase 2 using the salinity concentration and ERT survey data.

The standard deviation of the post-calibration predictive uncertainty was reduced compared to its pre-calibration uncertainty (Figure 5b&c), However, the salinity prediction uncertainty was still high, indicating the need for acquiring more complementary data.

4.2. Parameter identifiability of the hydrogeological parameters at different phases

The ability of the calibration dataset to constrain the model parameters in phases 1 and 2, was expressed by the parameter identifiability (ParId). Out of 242 parameters, 86 were adequately identified with a ParId cut-off value of 0.75 (38 were adequately identified in phase 1 and assumed fixed, and an additional 48 were adequately identified in phase 2). Figure 6 shows the ParId of 121 pilot points used to estimate hydraulic conductivity and specific yield in phases 1 and 2.

The transient heads in phase 1, constrained the hydraulic conductivity at the pilot points close to the observation measurements with higher identifiability. 9 out of 121 pilot points were considered adequately identified. The specific yield parameters are more influenced by soil texture. In fine sediments of the western part, 29 out of 121 pilot points were considered adequately identified in phase 1. However, in coarser sediments (east), the hydraulic heads in the calibration dataset could not identify specific yield in phase 1.

The salinity observations and ERT survey data significantly increased the number of identified hydraulic conductivities to 47 (38 in phase 2 and 9 in phase 1). The specific yield was less constrained by salinity observations with a slight increase of 10 parameters, resulting in 39 adequately identified parameters in total (29 in phase 1 and 10 in phase 2).

4.3. Model calibration and validation

The model performance was evaluated using the calibrated parameter dataset after the model calibration. The goodness of fit of the transient head and salinity measurements evaluated using RMSE, NSH, and RSQ showed overall satisfactory results for the model calibration and validation. The error statistics of the model calibration and validation are shown in Figure 7 and Table 6.

The mismatch of all the simulated heads and salinity concentrations, represented by RMSE was located within the estimated 95% confidence interval (CI) of the measurements, except for two observations wells BHF1 under the validation, and BHF15 under the calibration and validation. A high correlation was identified between the observed and simulated measurements at different observation groups (≥ 0.7), indicating a significant and positive relationship. However, for BHF10, the correlation value was 0.12, indicating no similarity in the pattern of variation between the two datasets. Considering NSE coefficients (a value of 1 corresponds to a perfect match, values between 0 and 1 are viewed as acceptable performance, while negative values represent unacceptable performance), good model performance can be seen. The lower values (approaching zero) were reported in the salinity measurements. This is due to the small number of measurements (2 measurements) in each observation group. Approaching zero means that the average value of the observed dataset is better than the predictions simulated by the model.

The performance criteria proved that the extracted information from the calibration datasets and the additional information in the form of the prior information imposed on the parameters were able to efficiently constrain the parameters to achieve an acceptable reproduction for the observations. The estimated parameter values for the horizontal hydraulic conductivity and the specific yield are presented in Figure 4 together with their initial values estimated from the site data. The heterogeneity in the hydraulic conductivity was captured, and the values were reliable with small departures from their initial values. The values of the specific yield varied from their initial values, which was expected due to the high uncertainty in their estimates.

4.4. Analysis of posterior parameter uncertainty

The NSMC method was combined with the OLHS and filtering (rejection sampling) techniques to perform a two-step re-calibration of the SWI model. The first iteration was performed after generating 250 stochastic parameter sets, with 31 sets having an objective function value below the threshold of 2700. These 31 sets were added to the ensemble of calibration-constrained parameter sets, and 120 sets passed the filtering process to proceed to the second iteration. The second iteration resulted in 78 parameter sets that satisfied the objective function threshold, bringing the total number of well-calibrated parameter sets to 110 (1 calibrated set and 31 and 78 sets from the first and second iterations, respectively).

The average value of the multi-component objective function for these 110 sets was reduced from 27000 to 2560, demonstrating the effectiveness of the proposed method. A calculated objective function higher than 50000 was found to be unsatisfactory and was discarded.

The probabilistic coverage of input space of each parameter provided by OLHS strategy was calculated using Eq 11 and compared to that generated by SRS strategy. The probabilistic coverage can be used as guidance (Hardyanto and Merkel, 2007) to choose which strategy is preferable. Eq.11 was used to calculate the probabilistic coverage for each strategy and considered each parameter as a lognormal distributed independent variable. The mean and standard deviation were calculated from drawn samples of size 250.

$$P(L(X_i) \leq X_i \leq U(X_i)) \quad 11$$

In Eq.11, $L(X_i)$ and $U(X_i)$ are the minimum and maximum values for each parameter (X_i). The probabilistic coverage for each parameter generated by the OLHS strategy was approximately 92% and ranged from 71% to 91% for the SRS strategy. OLHS strategy provided better probabilistic coverage compared to the SRS strategy.

The application of the NSMC method resulted in an ensemble of calibration-constrained parameter sets, obtained by exploring the null space where the parameters were not well-constrained by observations, and adjusting the solution-space parameters. The results indicate that the method was effective in constraining the parameters of Khz and Sy in regions near observation measurements, with a reduced ability to constrain in coarse sediments for

Sy. The posterior parameter values were found to be significantly different from their prior values generated by OLHS, particularly in specific regions near the southeast boundary and northwestern part of the aquifer. The highest hydraulic conductivity values were observed near the southeast boundary (ranging from 167 to 209 m/day), while lower values were found in the northwestern part (ranging from 10 to 48 m/day) near the Shaarah pumping field, which represented fine-grained soil. The inversion process also constrained a small number of Sy parameters near observation measurements in fine-grained soil, but with reduced ability to constraint in coarse sediments. Away from the regions with constrained parameters, the parameters were allowed to take different values in order to achieve an acceptable level of misfit. These results emphasize the significance of defining the null space for the integrity of the prediction uncertainty analysis and considering both prior knowledge and observation data in the parameter calibration process. The findings are presented in Figure 8, with five randomly selected posterior realizations assumed to represent samples drawn from the posterior parameter distribution.

The values taken by each parameter in the 110 calibration-constrained parameter sets were used to identify the probability distribution (pd) that this data followed. The posterior parameter pd of the logarithm of the hydraulic conductivity was found to fit a normal distribution with an average p-value of 0.05 (considering 121 pilot points), while the specific yield was fitted best by the Gumbel maximum probability distribution with an average value of 0.082. The maximum likelihood estimates of the parameters describing each distribution (location and scale are equivalent to the mean (μ) and the standard deviation (σ) of the parameter variability) were calculated and the coefficient of variation (Cv) was utilized to represent the posterior variance of the parameters. The Cv was calculated for Lognormal and Gumbel maximum distributions using Eq. 12 and 13, respectively.

$$C_{v_{logmormal}} = \sqrt{e^{\sigma^2} - 1} \qquad 12$$

$$C_{v_{Gumbel\ maximum}} = \frac{\sigma}{\mu + \left(0.5772 \times \left(\frac{\sigma\sqrt{6}}{\pi}\right)\right)} \qquad 13$$

Figure 9 depicts the posterior mean and C_v of the Khz and specific yield. The mean posterior conductivity values (Figure 9a&c) agreed well with the results of the pumping tests and the aquifer geology (Figure 4). The Cv of the logarithm of the conductivity field ranged from 7 to 76% (Figure 9b), while its prior C_v was 112%. A lower dispersion around the mean was shown near the observation data, but higher uncertainty near the Southern boundary and the uncertainty due to the lack of available observations. For the specific yield parameters, the parameter estimation reduced the prior Cv from an average of 315% to 85%. The area of higher certainty was found near the observation data within the fine-grained soil of the western part of the aquifer. The mean of the specific yield posterior pd at each pilot point was significantly different than its value in the parameter set of the minimum error variance obtained by the model calibration. This was due to the large prior uncertainty considered to account for the lack of data. The prior parameter pd at 242 pilot points describing the logarithm of Khz and Sy was assumed normal distribution, and their distribution parameters are shown in Table 5.

The correlation between the identifiability (ParID) and the reduction in the standard deviation of the probability distribution (σ_{red}) was calculated to investigate the extent to which the parameter identifiability is valid for nonlinear model behavior. It was found to be moderate and negative, with correlation coefficients of -0.56 for Khz and -0.68 for Sy. The identifiability and the reduction in standard deviation are approximately consistent. Figure 10 demonstrates the reduction in uncertainty for 8 parameters from different identifiability categories (0–0.25, 0.25-0.5, 0.5-0.75, and 0.75-1). The reduction in uncertainty is visualized by comparing the height of the posterior and prior distributions at the median. The ratio between posterior and prior values indicates the amount of additional knowledge gained from the inversion. The results showed that the identifiability and the reduction in standard deviation are approximately consistent.

4.5. Quantifying the Risk of Seawater Intrusion at Pumping Fields

The 110 calibration-constrained parameter sets were used to delineate the uncertainty range in the prediction of $L_{5\%}$ (2000 mg/l isoline) for the year 2050. The dispersion of the posterior prediction distributions was quantified, with values less than 10% representing high certainty and values between 10-20% representing moderate certainty. High certainty was found near Kalba pumping field ($\frac{\sigma}{\mu} = 4.5\%$) while moderate certainty was found

near Fujyarh pumping field ($\frac{\sigma}{\mu} = 11\%$) (Figure 11a). The wider distribution is the direct outcome of the ill-informed parameters by the measurements of the system state as depicted in Figure 9. The posterior predictive uncertainty distributions of the intrusion length, measured from the coastline, at the two pumping fields are shown in Figures 11 b&c. The 95 percentiles of the two distributions showed that it was possible to advance the $L_{5\%}$ to the centers of the two pumping fields. However, the maximum prediction of $L_{5\%}$ completely encompassed the Kalba pumping field and reached a portion of the Fujayrah field, leading to salinity levels exceeding the acceptable limit for irrigation purposes.

5. Discussion

5.1. Data worth and their impact on prediction uncertainty

The useful feature of analyzing the data worth of the observation data, following the method described by (Dausman et al., 2010), is that the predictive uncertainty is independent of the actual values of both the measurements in the calibration dataset and the model parameters. However, it depends on the stochastic characterization of parameter variability, measurement noise, and model output sensitivity to parameters. Accordingly, the worthiness of ERT survey data was assessed, although the data were not available during the analysis period. The findings highlight the importance of including salinity observations in the calibration dataset to better constrain the predictions, which is consistent with Refsgaard et al. (2012). They argued that the salinity predictions made by a constant-density model are associated with high uncertainty as the predictions are not constrained by the calibration dataset.

The hydraulic head observations had a greater impact on reducing the uncertainty upstream of the seawater wedge, where freshwater inflows (Hald stream and Wadi Ham stream, and the infiltration from the dam reservoir) mainly governed the salinity. The uncertainty in the salinity predictions increased by losing these data by an average of 70%. On the other hand, in the region near the coast, enclosed by the contour line of 5000 mg/l (Figure 5a), the salinity observation data had the higher influence, where the effect of freshwater inflow from the western part of the aquifer decreased and the effect of the inland lateral flow from the sea increased.

The post-calibration uncertainty was reduced indicating the effectiveness of the available information. However, the salinity prediction uncertainty was still high in the eastern part of the aquifer, with an average value of 1000 mg/l and exceeding 8000 mg/l, near Hyal stream, suggesting a need for acquiring more complementary data. Hence, all the available data for the salinity concentrations were used in the model calibration. This analysis suggests areas for further investigation to improve the understanding of the aquifer system, making it an essential prior step to the parameter estimation process.

The groundwater model was used to provide reliable initial parameter estimates to reduce the computational demand of the model calibration in phase 2 using variable density and solute transport model including the salinity observations. The accuracy of its usage was evaluated by obtaining the low values of the scale RMSE between using constant and variable density hydraulic heads. These values can be attributed to the aquifer being shallow and the western part being governed by the freshwater inflow over the calibration period. This study adopts simpler models to reduce computational cost without sacrificing accuracy. Other techniques such as combining fast-running model based on non-physical parameters, and a complex model to estimate equivalent values for the non-physical parameters (Hugman and Doherty, 2022), increasing grid size (Burrows and Doherty, 2015), using 2D models (Sherif et al., 2012b), or assuming a sharp interface (Coulon et al., 2021) have also been used. However, caution should be exercised as these techniques are site-specific and may not be universally applicable.

5.2. Evaluation of the parameter estimability under the linear assumption

Parameter identifiability was used to evaluate the parameter estimability. The analysis showed that 86 out of 242 parameters were adequately identified. The transient heads in the calibration dataset identified the hydraulic conductivities near the observation locations, while soil texture had a larger impact on specific yield. In the western part of the aquifer with fine sediments and steep water level gradients, which was indicated from the field measurements at the observation wells of average fluctuation in the water levels of 18m (Sherif et al., 2014), the specific yield was better identified. In coarser sediments (east), the mild hydraulic head gradient (the average fluctuation in the water levels was 1m) made the model outputs insensitive to specific yield. The salinity and ERT survey data improved the identification of hydraulic conductivities. This can be attributed to the key role of hydraulic conductivity in controlling the extent of the SWI (Ketabchi and Jahangir, 2021), where the model outputs of salinity concentrations were sensitive to those parameters, accordingly they displayed high

identifiability (Figure 6). However, the specific yield was less constrained by the salinity observations. This finding agrees with other studies (Delsman et al., 2017; Zhang et al., 2020).

Previous studies on SWI models have employed a phased approach to parameter estimation to mitigate computational complexity (Abd-Elaty et al., 2021; Sanford and Pope, 2010). In this study, we utilized fixed, adequately-identified parameters in the second phase of the model calibration process, resulting in a further reduction in computational effort. The linearity assumption made by the ParID method was assessed in light of potential nonlinearities in the model. Our results indicate that the ParID methodology provides valuable insights into the informativeness of the calibration data and the extent to which prediction uncertainty is reduced, as shown in Figure 10.

5.3. Model calibration, validation, and parameterization

The model calibration was performed using the observations of transient heads, salinity, and ERT survey data. The Tikhonov regularization method was used to incorporate prior information and SVD to simplify the calibration process by reducing the number of parameters to those that were sensitive to the observations. The results showed an overall reduction in the objective function for head and salinity measurements by 84% and 58%, respectively (Table C3.1, supplementary material).

However, it should be noted that the limited number of salinity measurements in each observation group led to an approach of zero in the NSE coefficients, indicating that further measurements are needed to improve the accuracy of the model. Additionally, the observed misfit in head measurements in two observation wells near the Shaarah pumping field (BHF1 under validation and BHF15 under calibration and validation) may be attributed to the significant fluctuation in pumping schedules than the assigned average values. The proximity of BHF10 to Wadi Hyal stream resulted in a model structural error due to the assigned fixed boundary condition, which did not account for the seasonal fluctuations in the stream water level.

Minimization of the objective function alone is not a sufficient criterion to evaluate the efficiency of the calibration. The reliability of the parameter estimates must also be considered to ensure the geological plausibility of the results. The spatial variability of the hydraulic conductivity was found to be geologically realistic and preserved the aquifer heterogeneity, while the specific yield values varied from their initial estimates but were still acceptable in the geological context. Specific yield values in similar alluvial aquifers have been measured to range from 0.05 to 0.15 (Chen et al., 2010), and in ophiolite formations, the values have ranged from near zero to 0.01 (Lachassagne et al., 2021). The results contribute to the understanding of the geological formation of the system and highlight the importance of considering both the minimization of the objective function and the reliability of the parameter estimates in model calibration.

5.4. Calibration-constrained parameter sets ensemble

The implementation of the NSMC method linked with OLHS and the filtering technique has successfully generated multiple hydrologically reasonable parameter sets that are in agreement with the available head and salinity observations. The filtering technique removed 40% of the generated parameter sets during the recalibration, which greatly reduced the computational time of the optimization process. This novel approach offers a promising solution for the calibration of SWI models and could be applied to other hydrological models.

The complex nature of SWI modeling in heterogeneous aquifers can pose challenges in accurately estimating uncertainty. In cases where the calibration dataset does not provide enough information to estimate model complexity, such as seen in Figure 9 where hydraulic conductivities were only constrained in two regions and a limited number of Sy parameters near observation measurements in fine-grained soil, it is still important to include this complexity in the uncertainty analysis to accurately reflect the uncertainty in the predictions. This study highlights the efficiency of NSMC in exploring the null space, which is where the insetimated parameters reside and hence is the primary source of uncertainty in the predictions (Moore and Doherty, 2005). Exploring the null space is based on the prior realization that spans the entirety of parameter space. The generated prior realizations by OLHS were substantially different from each other (Figure 9), resulting in a more comprehensive probabilistic coverage than the traditionally used sampling strategy (SRS).

A 110 calibration-constrained parameter set has resulted from the application of the NSMC method and was used to derive the posterior parameter probability distribution of log(Khz) and Sy at each of 121 pilot points. The posterior probability distributions of log(Khz) and Sy were found to fit normal and Gumbel maximum distributions, respectively. The derived parameter probability distributions can be used to identify areas of high

uncertainty in predictions and guide the expansion of the calibration dataset. The coefficient of variation was used to compare different types of probability distributions and magnitudes, with lower values indicating lower dispersion around the mean and higher certainty (Figure 10). In general, the lower Cv values were obtained close to the observation data, indicating that the information content in the observations is sufficient to inform/constrain the model parameters. The certainty in parameter estimation is high, and consequently, the uncertainty in the predictions that are sensitive to those parameters is low. On the other hand, the parameters that were not informed by the observation data were constrained by prior information.

5.5. Analysis of the future prediction of SWI

The model's performance in predicting SWI was compared to Hussain et al. (2015), who predicted the inland encroachment of 50% ischlor in 2015 under similar abstraction and recharge conditions. 81% of their prediction fell within the minimum and maximum prediction range (Figure 11d), indicating the model was calibrated correctly and provides reasonable uncertainty estimates. The model was then used to predict SWI in 2050, and further deterioration of water quality was revealed due to limited water availability and continued abstraction. Protective measures must be taken to protect the aquifer's resilience and sustainability despite the associated costs and challenges. This finding is crucial for decision-makers, especially considering the potential exacerbation of future threats from sea level rise and reduced recharge due to rainfall scarcity under the conditions of climate change (Bolleter et al., 2021).

This study offers new insights into the parameter estimation and uncertainty analysis of SWI models, building upon previous studies that have used computationally expensive methods with limited parameters (e.g., Llopis-Albert et al., 2016; Miao et al., 2019; Zeng et al., 2016) and assumed linear model behavior (e.g., Coulon et al., 2021; Dausman et al., 2010). The findings of this study are site-specific and model-dependent, as the results indicate that the parameter uncertainty is influenced by various factors such as the uncertainty in the observations, their number and locations relative to the predictions of interest, and the aquifer geological and hydrogeological parameters. Despite this, some of the findings can be generalized, and the proposed method applies to other cases.

5.6. Limitations of the model and future directions

Uncertainty in SWI models is caused by factors such as tidal effects, recharge variability, geology, solute dispersivity, and anthropogenic influences. This study aimed to quantify uncertainty in two main parameters that impact SWI progression, namely, hydraulic conductivity and specific yield. The results indicate high uncertainty in specific yield estimation due to limited head measurement data in fine-grained soils. Further data collection of head measurements in fine-grained soil or laboratory tests for coarse-grained soil could reduce uncertainty. Uncertainty intervals and prediction distributions were estimated highlighting regions of high uncertainty but may be overestimated due to the small sample size (Zhang and Shields, 2018).

In high-dimensional and nonlinear models, optimization algorithms can get stuck in local minima, which reduces the efficacy of the NSMC method. Using multiple random initial parameters in the calibration process can help overcome this challenge. Another cause of efficacy reduction is fixing calibrated parameters in some cases, which adds uncertainty to the prediction but reduces computational effort.

The NSMC method is preferred for its computational efficiency in highly parametrized SWI models, however, it may not accurately represent Bayesian posterior probabilities and may not accurately estimate uncertainty (Keating et al., 2010). To improve accuracy in both model calibration and uncertainty analysis, a high computational effort is needed. In future studies, the use of surrogate models can be considered as approximations to highly parametrized models reducing the computational effort. One such surrogate model is Gaussian process regression (GPR), which has been demonstrated to accurately capture SWI behavior and estimate approximations and their uncertainty (Saad et al., 2022). Although GPR has been recently adapted to handle high dimensional problems, its application to highly parametrized SWI models has not yet been tested.

6. Conclusions

A novel model-independent framework was successfully developed to quantify the uncertainty associated with the estimation of model parameters and their propagation to the predictions of seawater intrusion (SWI) in the Wadi Ham aquifer, UAE. The parameter estimation process considered the aquifer heterogeneity in terms of horizontal hydraulic conductivity and specific yield, using measurements of transient heads, salinity concentrations, and geophysical survey data, along with prior knowledge of the geological formation. To maximize the computational efficiency, parameter estimation was conducted in two phases using constant and

variable density models. Adequately identified parameters were considered fixed in the second phase. The OLHS was incorporated in the NSMC method and with the aid of the filtering technique, an ensemble of constrained calibrated parameter sets was attained to derive the predictive probability distribution of SWI length.

The following conclusions can be drawn from the findings of this research:

1) Data worth analysis provided a basis for the model calibration phasing and emphasized the importance of including salinity observations and geophysical surveys in the parameter estimation process, to reduce predictive uncertainties, despite their limited number compared to head observations. This analysis is an essential prior step to the parameter estimation process.
2) The transient heads only were able to identify some of the hydraulic conductivities in their proximity (9%). When the salinity observations and the geophysical surveys were included, the adequately identified parameters increased to 39% of the total parameters.
3) For the specific yield, the adequately identified parameters were found within the fine-grained soil and near observation locations. Head observations contained more information to constrain them relative to salinity observations.
4) The significant decrease in parameter and predictive uncertainty in any subregion of the model domain was strongly controlled by the density of the observations located within that subregion. However, constraining the specific yield using head data was more affected by the soil texture.
5) The non-redundant information contained in the calibration dataset constrained the highly identifiable parameters, while the remaining parameters were more strongly constrained by prior knowledge. Reasonable a priori bounds were essential for obtaining acceptable parameter values in the geological context. SVD combined with Tikhonov regularization helped to stabilize and constrain the inversion process to yield a plausible geological distribution.
6) Even though the uncertainties in the seawater intrusion length inferred from the geophysical surveys were considerably high, nonredundant information was encapsulated in them. On their removal, the uncertainty in the salinity prediction increased. Geophysical observations are valuable for their wide spatial coverage and low cost.
7) The space-filling scheme of OLHS based on ESE optimization was more efficient in terms of probabilistic coverage compared to the SRS scheme.
8) The proposed filtering technique is a straightforward procedure to save computational time within the recalibration process and is favored over narrowing the possible parameter space intervals which may lead to the NSMC calibrated parameter sets being significantly constrained to the parameter set obtained through model calibration. 40% of the generated parameter sets were omitted from the recalibration.
9) A moderate correlation was observed between the ParID, which assumed model linearity, and the reduction in the standard deviation of the prior parameter distribution to its posterior that was estimated considering model nonlinearity. ParID can give useful insights into the extracted information from the calibration dataset and reduce parameter uncertainty.
10) The posterior predictive probability distribution determined the maximum and minimum predictions of the 5% isochlor position. The resulting uncertainty intervals near Kalba and Fujyarh pumping fields varied by 4.5% and 11% from their means respectively, which indicates, with a high to moderate certainty, that the aquifer would be vulnerable to salinization, and the water quality at these pumping fields would be unsuitable for irrigation.

The framework presented in this study for probabilistic assessment of seawater intrusion (SWI) due to input uncertainties has broad applicability beyond the site and model studied, where it could be applied to other hydrological models improving predictions of water levels, quantities, and quality. It can help identify data gaps and improve prediction certainty allowing for more accurate assessments of SWI and the impact of management strategies on the aquifer system. The resulting probability distribution can be used to estimate prediction intervals, assess the reliability of management strategies, and evaluate the probability of exceeding critical SWI thresholds. Future research should address uncertain inputs beyond hydraulic conductivity and specific yield, including boundary conditions such as sea level rise, recharge variations, and abstractions. This is crucial for assessing future SWI under climate change conditions. Integrating the proposed framework with simulation-optimization-based groundwater management models will help mitigate SWI under uncertainty and propose a sustainable coastal management strategy.

In summary, the highly parameterized regularized inversion allowed the extraction of information from the historical field measurements achieving reliable parameter estimates and a sufficient level of heterogeneity. The NSMC method linked with OLHS and filtering techniques was computationally efficient and succeeded in

quantifying the prediction uncertainty of 5% isochlor intrusion length. The framework proposed in this study can be generally applied in different coastal areas.

7. ACKNOWLEDGMENTS

The first author is funded by the Ministry of Higher Education of the Arab Republic of Egypt, Netwon Mosharafa scholarship [ID: NMM26/17]. The authors would also like to express their gratitude to the DHI group for providing the free license of FEFLOW. Data and information related to the Wadi Ham aquifer were provided by the National Water and Energy Center, United Arab Emirates University. We would like to take this opportunity to thank the reviewers and the editor for their valuable comments that have certainly improved the quality of the paper.

8. REFERENCES

Abd-Elaty, I., Javadi, A.A., Abd-Elhamid, H., 2021. Management of saltwater intrusion in coastal aquifers using different wells systems: A case study of the Nile Delta aquifer in Egypt. Hydrogeology Journal, 29(5): 1767-1783.

Alcolea, A., Carrera, J., Medina, A., 2006. Pilot points method incorporating prior information for solving the groundwater flow inverse problem. Advances in water resources, 29(11): 1678-1689.

Arend, D.N., 1993. Choices (Version 4.0)[Computer software]. Champaign, IL: US Army Corps of Engineers Research Laboratory.(CERL Report No. CH7-22510).

Asher, M.J., Croke, B.F.W., Jakeman, A.J., Peeters, L.J.M., 2015. A review of surrogate models and their application to groundwater modeling surrogates of groundwater models. Water Resour. Res. Water Resources Research, 51(8): 5957-5973.

Ayers, R.S., Westcot, D.W., Food Agriculture Organization of the United Nations, 1985. Water quality for agriculture. Food and Agriculture Organization of the United Nations, Rome.

Bolleter, J., Grace, B., Hooper, P., Foster, S., 2021. Wet-bulb Temperature and Sea-level Rise in the United Arab Emirates–Planning Responses. Planning Practice & Research, 36(4): 408-429.

Bouhlel, M.A. et al., 2019. A Python surrogate modeling framework with derivatives. Advances in Engineering Software, 135: 102662. DOI:https://doi.org/10.1016/j.advengsoft.2019.03.005.

Brunner, P., Doherty, J., Simmons, C.T., 2012. Uncertainty assessment and implications for data acquisition in support of integrated hydrologic models. Water Resources Research, 48(7).

Burrows, W., Doherty, J., 2015. Efficient calibration/uncertainty analysis using paired complex/surrogate models. Groundwater, 53(4): 531-541.

Carrera, J., Hidalgo, J.J., Slooten, L.J., Vázquez-Suñé, E., 2010. Computational and conceptual issues in the calibration of seawater intrusion models. Hydrogeology Journal, 18(1): 131-145.

Chen, X., Song, J., Wang, W., 2010. Spatial variability of specific yield and vertical hydraulic conductivity in a highly permeable alluvial aquifer. Journal of Hydrology, 388(3-4): 379-388.

Costall, A., Harris, B., Pigois, J., 2018. Electrical resistivity imaging and the saline water interface in high-quality coastal aquifers. Surveys in geophysics, 39(4): 753-816.

Coulon, C. et al., 2021. A framework for parameter estimation using sharp-interface seawater intrusion models. Journal of Hydrology, 600: 126509.

Damblin, G., Couplet, M., Iooss, B., 2013. Numerical studies of space-filling designs: optimization of Latin Hypercube Samples and subprojection properties. Journal of Simulation, 7(4): 276-289.

Dausman, A.M., Doherty, J., Langevin, C.D., Sukop, M.C., 2010. Quantifying data worth toward reducing predictive uncertainty. Groundwater, 48(5): 729-740.

Delottier, H., Therrien, R., Young, N.L., Paradis, D., 2022. A hybrid approach for integrated surface and subsurface hydrologic simulation of baseflow with Iterative Ensemble Smoother. Journal of Hydrology, 606: 127406. DOI:https://doi.org/10.1016/j.jhydrol.2021.127406.

Delsman, J.R., de Louw, P.G.B., de Lange, W.J., Oude Essink, G.H.P., 2017. Fast calculation of groundwater exfiltration salinity in a lowland catchment using a lumped celerity/velocity approach. Environmental Modelling & Software, 96: 323-334. DOI:https://doi.org/10.1016/j.envsoft.2017.07.004.

Diersch, H.-J.G., 2013. FEFLOW: finite element modeling of flow, mass and heat transport in porous and fractured media. Springer Science & Business Media.

Doherty, J., 2016. PEST, model-independent parameter estimation—User manual, with slight additions: Watermark Numerical Computing.

Doherty, J., Hunt, R.J., 2009. Two statistics for evaluating parameter identifiability and error reduction. Journal of Hydrology, 366(1): 119-127. DOI:https://doi.org/10.1016/j.jhydrol.2008.12.018.

Doherty, J., Moore, C., 2020. Decision support modeling: data assimilation, uncertainty quantification, and strategic abstraction. Groundwater, 58(3): 327-337.

Doherty, J., Moore, C., 2021. Decision Support Modelling Viewed through the Lens of Model Complexity. National Centre for, 640.

Doherty, J., Welter, D., 2010. A short exploration of structural noise. Water Resources Research, 46(5).

Doherty, J.E., Hunt, R.J., Geological, S., 2010. Approaches to highly parameterized inversion : a guide to using PEST for groundwater-model calibration.

Doherty, J.E., Hunt, R.J., Tonkin, M.J., Groundwater Resources, P., Geological, S., 2011. Approaches to highly parameterized inversion : a guide to using PEST for model-parameter and predictive-uncertainty analysis.

Elkhrachy, I., 2018. Vertical accuracy assessment for SRTM and ASTER Digital Elevation Models: A case study of Najran city, Saudi Arabia. Ain Shams Engineering Journal, 9(4): 1807-1817. DOI:https://doi.org/10.1016/j.asej.2017.01.007.

FAO, 2018. WaPOR Database Methodology: Level 1 Data. Food and Agriculture Organization of the United Nations Rome, Italy.

Hardyanto, W., Merkel, B., 2007. Introducing probability and uncertainty in groundwater modeling with FEMWATER-LHS. Journal of Hydrology, 332(1): 206-213. DOI:https://doi.org/10.1016/j.jhydrol.2006.06.035.

Herckenrath, D., Langevin, C.D., Doherty, J., 2011. Predictive uncertainty analysis of a saltwater intrusion model using null-space Monte Carlo. Water Resources Research, 47(5).

Herrera, P.A., Marazuela, M.A., Hofmann, T., 2022. Parameter estimation and uncertainty analysis in hydrological modeling. WIREs Water, 9(1): e1569. DOI:https://doi.org/10.1002/wat2.1569

Hugman, R., Doherty, J., 2022. Complex or Simple—Does a Model Have to be One or the Other? Front. Earth Sci, 10: 867379.

Hussain, M.S., Javadi, A.A., Sherif, M.M., 2015. Three Dimensional Simulation of Seawater Intrusion in a Regional Coastal Aquifer in UAE. Procedia Engineering, 119: 1153-1160. DOI:https://doi.org/10.1016/j.proeng.2015.08.965.

Janssen, H., 2013. Monte-Carlo based uncertainty analysis: Sampling efficiency and sampling convergence. Reliability Engineering & System Safety, 109: 123-132.

Jiang, X. et al., 2021. Two-stage surrogate model-assisted Bayesian framework for groundwater contaminant source identification. Journal of Hydrology, 594: 125955. DOI:https://doi.org/10.1016/j.jhydrol.2021.125955.

Jin, R., Chen, W., Sudjianto, A., 2005. An efficient algorithm for constructing optimal design of computer experiments. Journal of Statistical Planning and Inference, 134(1): 268-287. DOI:https://doi.org/10.1016/j.jspi.2004.02.014.

Keating, E.H., Doherty, J., Vrugt, J.A., Kang, Q., 2010. Optimization and uncertainty assessment of strongly nonlinear groundwater models with high parameter dimensionality. Water Resources Research, 46(10).

Ketabchi, H., Jahangir, M.S., 2021. Influence of aquifer heterogeneity on sea level rise-induced seawater intrusion: a probabilistic approach. Journal of Contaminant Hydrology, 236: 103753.

Lachassagne, P., Dewandel, B., Wyns, R., 2021. Hydrogeology of weathered crystalline/hard-rock aquifers—guidelines for the operational survey and management of their groundwater resources. Hydrogeology Journal: 1-34.

Llopis-Albert, C., Merigó, J.M., Xu, Y., 2016. A coupled stochastic inverse/sharp interface seawater intrusion approach for coastal aquifers under groundwater parameter uncertainty. Journal of Hydrology, 540: 774-783.

McKay, M.D., Beckman, R.J., Conover, W.J., 1979. A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. technometrics Technometrics, 21(2): 239-245.

Miao, T., Lu, W., Lin, J., Guo, J., Liu, T., 2019. Modeling and uncertainty analysis of seawater intrusion in coastal aquifers using a surrogate model: a case study in Longkou, China. Arabian Journal of Geosciences, 12(1): 1-11.

Moore, C., Doherty, J., 2005. Role of the calibration process in reducing model predictive error. Water Resources Research, 41(5).

Morris, M.D., Mitchell, T.J., 1995. Exploratory designs for computational experiments. Journal of statistical planning and inference, 43(3): 381-402.

Mostafaei-Avandari, M., Ketabchi, H., 2020. Coastal groundwater management by an uncertainty-based parallel decision model. Journal of Water Resources Planning and Management, 146(6): 04020036.

NASA Shuttle Radar Topography Mission, 2013. Shuttle Radar Topography Mission (SRTM) Global. Distributed by OpenTopography.

Nwankwor, G., Gillham, R., van der Kamp, G., Akindunni, F., 1992. Unsaturated and saturated flow in response to pumping of an unconfined aquifer: Field evidence of delayed drainage. Groundwater, 30(5): 690-700.

Omagbon, J. et al., 2021. Case studies of predictive uncertainty quantification for geothermal models. Geothermics, 97: 102263. DOI:https://doi.org/10.1016/j.geothermics.2021.102263.

Pollicino, L.C., Colombo, L., Formentin, G., Alberti, L., 2021. Stochastic modelling of solute mass discharge to identify potential source zones of groundwater diffuse pollution. Water Research, 200: 117240.

Pool, M., Carrera, J., Alcolea, A., Bocanegra, E., 2015. A comparison of deterministic and stochastic approaches for regional scale inverse modeling on the Mar del Plata aquifer. Journal of Hydrology, 531: 214-229.

Post, V.E., Banks, E., Brunke, M., 2018. Groundwater flow in the transition zone between freshwater and saltwater: a field-based study and analysis of measurement errors. Hydrogeology Journal, 26(6): 1821-1838.

Rajabi, M.M., Ataie-Ashtiani, B., 2014. Sampling efficiency in Monte Carlo based uncertainty propagation strategies: Application in seawater intrusion simulations. Advances in Water Resources, 67: 46-64. DOI:https://doi.org/10.1016/j.advwatres.2014.02.004.

Rajabi, M.M., Ataie-Ashtiani, B., 2016. Efficient fuzzy Bayesian inference algorithms for incorporating expert knowledge in parameter estimation. Journal of Hydrology, 536: 255-272.

Rajabi, M.M., Ataie-Ashtiani, B., Janssen, H., 2015. Efficiency enhancement of optimized Latin hypercube sampling strategies: Application to Monte Carlo uncertainty analysis and meta-modeling. Advances in Water Resources, 76: 127-139. DOI:https://doi.org/10.1016/j.advwatres.2014.12.008.

Rajabi, M.M., Ataie-Ashtiani, B., Simmons, C.T., 2018. Model-data interaction in groundwater studies: Review of methods, applications and future directions. Journal of hydrology, 567: 457-477.

Rau, G.C. et al., 2019. Error in hydraulic head and gradient time-series measurements: a quantitative appraisal. Hydrology and Earth System Sciences, 23(9): 3603-3629.

Refsgaard, J.C. et al., 2012. Review of strategies for handling geological uncertainty in groundwater flow and transport modeling. Advances in Water Resources, 36: 36-50. DOI:https://doi.org/10.1016/j.advwatres.2011.04.006.

Saad, S., Javadi, A., Chugh, T., Farmani, R., 2022. Optimal Management of Mixed Hydraulic Barriers in Coastal Aquifers using Multi-objective Bayesian Optimization. Journal of Hydrology: 128021. DOI:https://doi.org/10.1016/j.jhydrol.2022.128021.

Saaltink, M.W., Carrera, J., Olivella, S., 2004. Mass balance errors when solving the convective form of the transport equation in transient flow problems. WRCR Water Resources Research, 40(5).

Safi, A., Vilhelmsen, T.N., Alameddine, I., Abou Najm, M., El-Fadel, M., 2019. Data-Worth Assessment for a Three-Dimensional Optimal Design in Nonlinear Groundwater Systems. Groundwater, 57(4): 612-631. DOI:https://doi.org/10.1111/gwat.12835.

Sanford, W.E., Pope, J.P., 2010. Current challenges using models to forecast seawater intrusion: lessons from the Eastern Shore of Virginia, USA. Hydrogeology Journal, 18(1): 73.

Sherif, M., Ebraheem, A., Shetty, A., 2017. Groundwater Recharge from Dams in United Arab Emirates, World Environmental and Water Resources Congress 2017, pp. 139-146. DOI:doi:10.1061/9780784480618.014.

Sherif, M., Kacimov, A., Javadi, A., Ebraheem, A.A., 2012a. Modeling Groundwater Flow and Seawater Intrusion in the Coastal Aquifer of Wadi Ham, UAE. Water Resour Manage Water Resources Management : An International Journal - Published for the European Water Resources Association (EWRA), 26(3): 751-774.

Sherif, M. et al., 2006. Geoelectrical and hydrogeochemical studies for delineating seawater intrusion in the outlet of Wadi Ham, UAE. Environmental Geology, 49(4): 536-551. DOI:10.1007/s00254-005-0081-4.

Sherif, M., Mohamed, M., Kacimov, A., Shetty, A., 2011a. Assessment of groundwater quality in the northeastern coastal area of UAE as precursor for desalination. Desalination, 273(2-3): 436-446.

Sherif, M., Sefelnasr, A., Ebraheem Abdel, A., Javadi, A., 2014. Quantitative and Qualitative Assessment of Seawater Intrusion in Wadi Ham under Different Pumping Scenarios. Journal of Hydrologic Engineering, 19(5): 855-866. DOI:10.1061/(ASCE)HE.1943-5584.0000907.

Sherif, M., Sefelnasr, A., Javadi, A., 2012b. Incorporating the concept of equivalent freshwater head in successive horizontal simulations of seawater intrusion in the Nile Delta aquifer, Egypt. Journal of Hydrology, 464-465: 186-198. DOI:https://doi.org/10.1016/j.jhydrol.2012.07.007.

Sherif, M.M., Mohamed, M.M., Shetty, A., Almulla, M., 2011b. Rainfall-runoff modeling of three wadis in the northern area of UAE. Journal of Hydrologic Engineering, 16(1): 10-20.

Singh, A., 2014. Optimization modelling for seawater intrusion management. HYDROL Journal of Hydrology, 508: 43-52.

Sowe, M.A., Sathish, S., Greggio, N., Mohamed, M.M., 2020. Optimized Pumping Strategy for Reducing the Spatial Extent of Saltwater Intrusion along the Coast of Wadi Ham, UAE. Water, 12(5): 1503.

Tonkin, M., Doherty, J., 2009. Calibration-constrained Monte Carlo analysis of highly parameterized models using subspace techniques. Water Resources Research, 45(12).

Tonkin, M., Doherty, J., Moore, C., 2007. Efficient nonlinear predictive error variance for highly parameterized models. Water Resources Research, 43(7).

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

Wang, Q. et al., 2020. Modified Algorithms for Fast Construction of Optimal Latin-Hypercube Design. IEEE Access, 8: 191644-191658. DOI:10.1109/ACCESS.2020.3032122.

Werner, A.D. et al., 2013. Seawater intrusion processes, investigation and management: Recent advances and future challenges. Advances in Water Resources, 51: 3-26. DOI:https://doi.org/10.1016/j.advwatres.2012.03.004.

White, J.T., 2018. A model-independent iterative ensemble smoother for efficient history-matching and uncertainty quantification in very high dimensions. Environmental Modelling & Software, 109: 191-201.

White, J.T., Fienen, M.N., Doherty, J.E., 2016. A python framework for environmental model uncertainty analysis. Environmental Modelling & Software, 85: 217-228. DOI:https://doi.org/10.1016/j.envsoft.2016.08.017.

Yang, Y. et al., 2021. A conjunctive management framework for the optimal design of pumping and injection strategies to mitigate seawater intrusion. Journal of Environmental Management, 282: 111964.

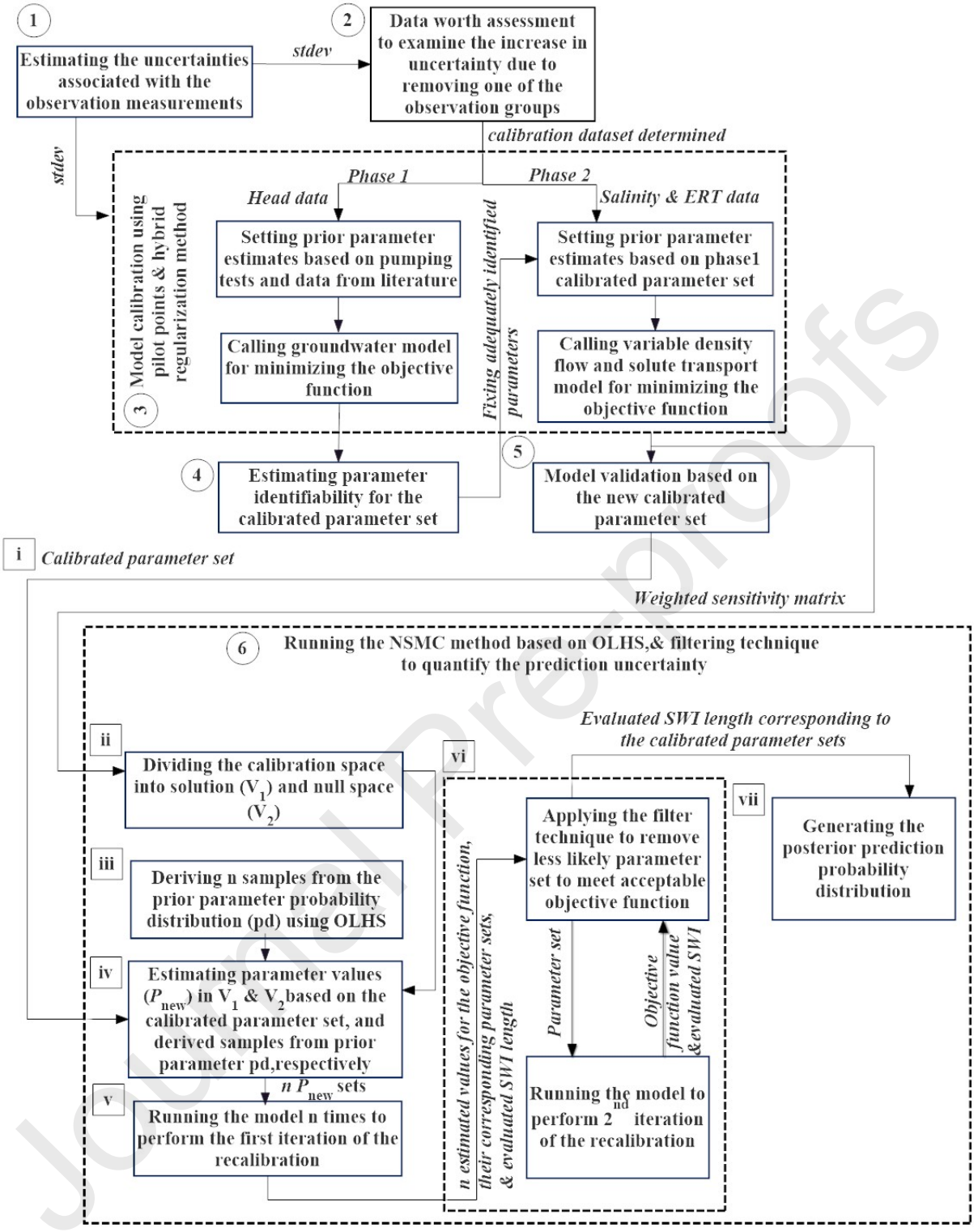
Zeng, X., Wu, J., Wang, D., Zhu, X., 2016. Assessing the pollution risk of a groundwater source field at western Laizhou Bay under seawater intrusion. Environmental research, 148: 586-594.

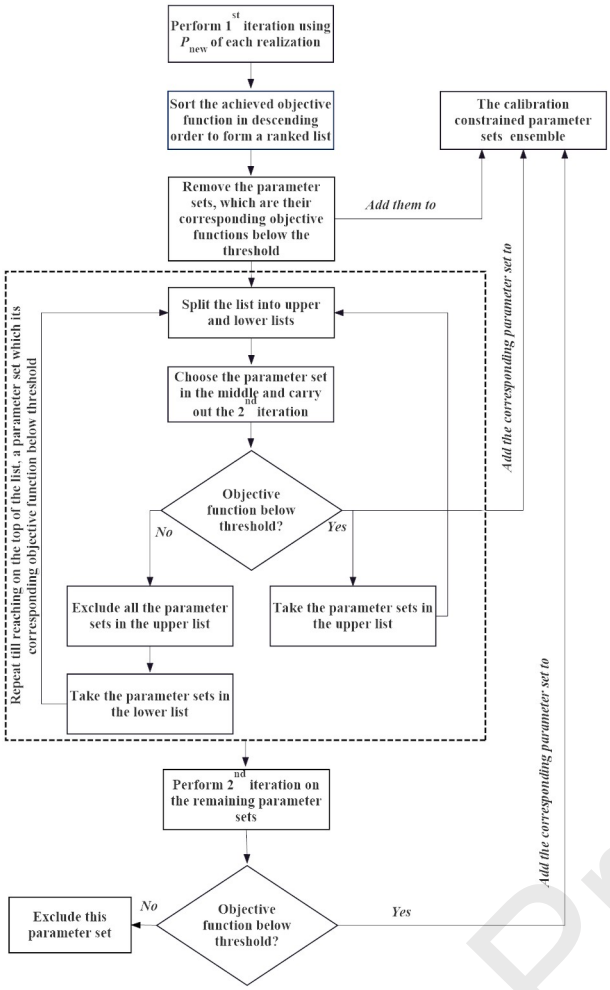
Zeynolabedin, A., Ghiassi, R., Norooz, R., Najib, S., Fadili, A., 2021. Evaluation of geoelectrical models efficiency for coastal seawater intrusion by applying uncertainty analysis. Journal of Hydrology, 603: 127086.

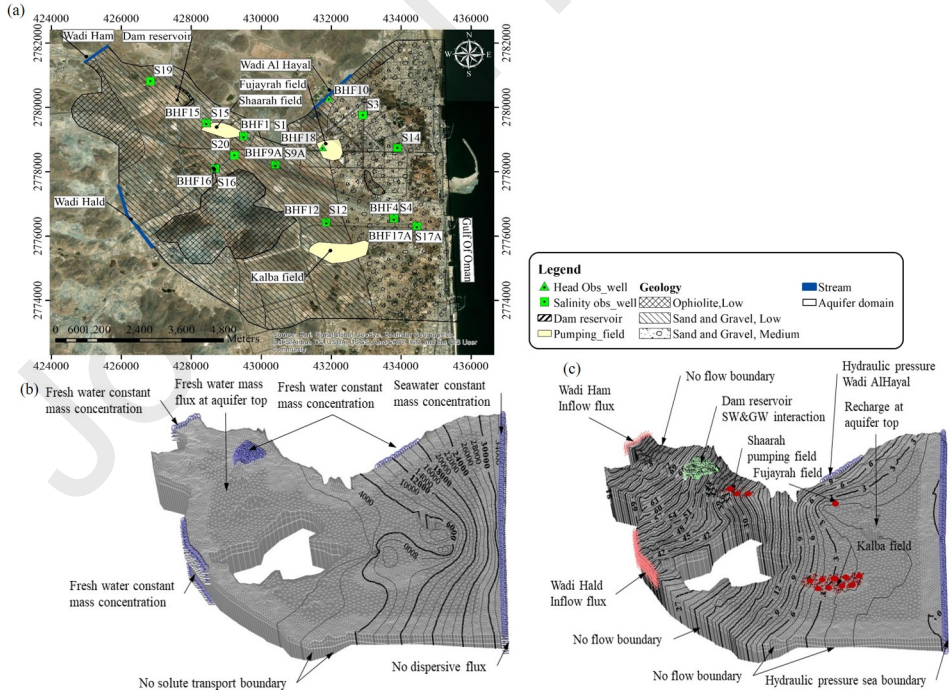
Zhang, D. et al., 2020. Global sensitivity analysis on a numerical model of seawater intrusion and its implications for coastal aquifer management: a case study in Dagu River Basin, Jiaozhou Bay, China. Hydrogeology Journal, 28(7): 2543-2557.

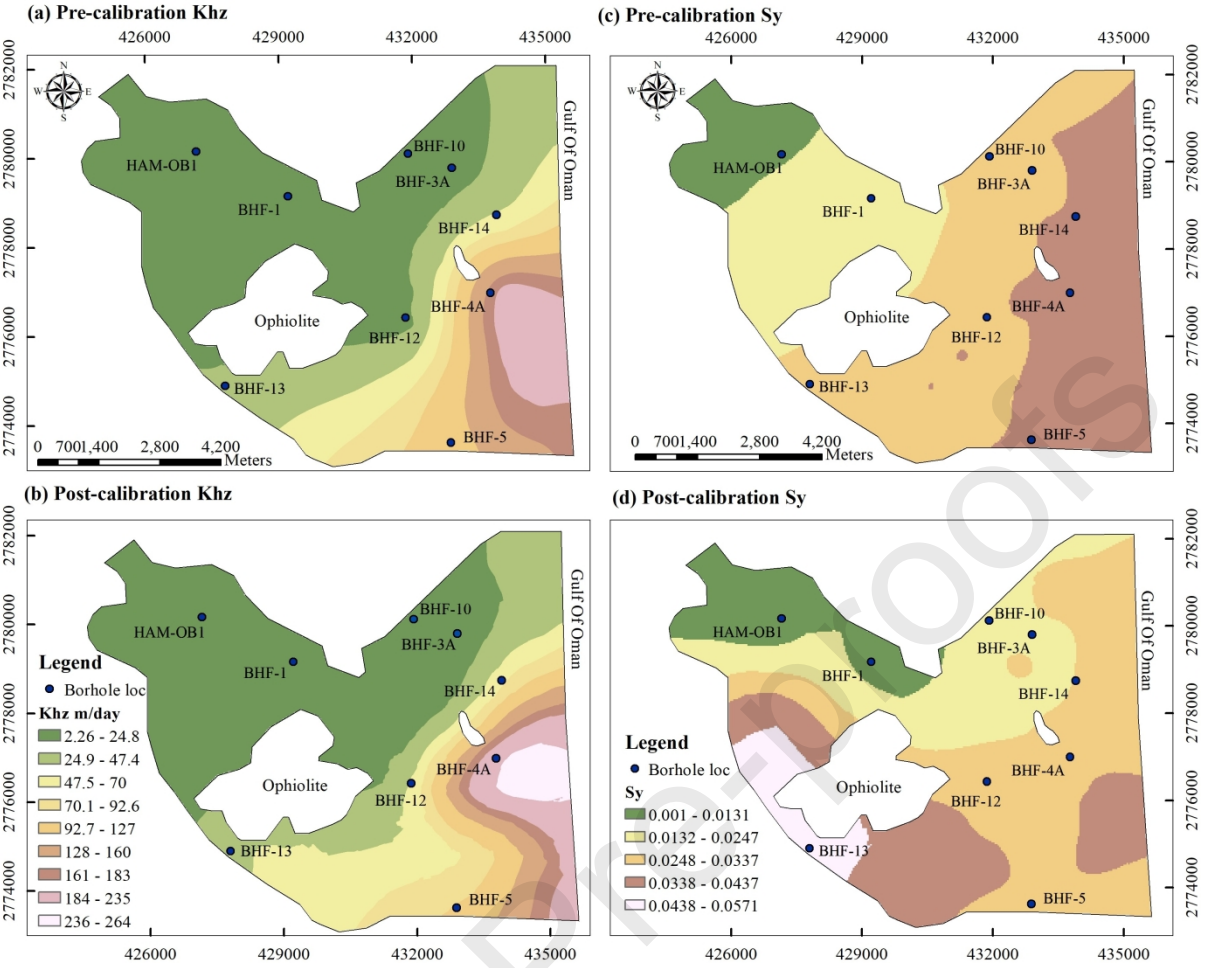
Zhang, J., Shields, M.D., 2018. On the quantification and efficient propagation of imprecise probabilities resulting from small datasets. Mechanical Systems and Signal Processing, 98: 465-483. DOI:https://doi.org/10.1016/j.ymssp.2017.04.042.

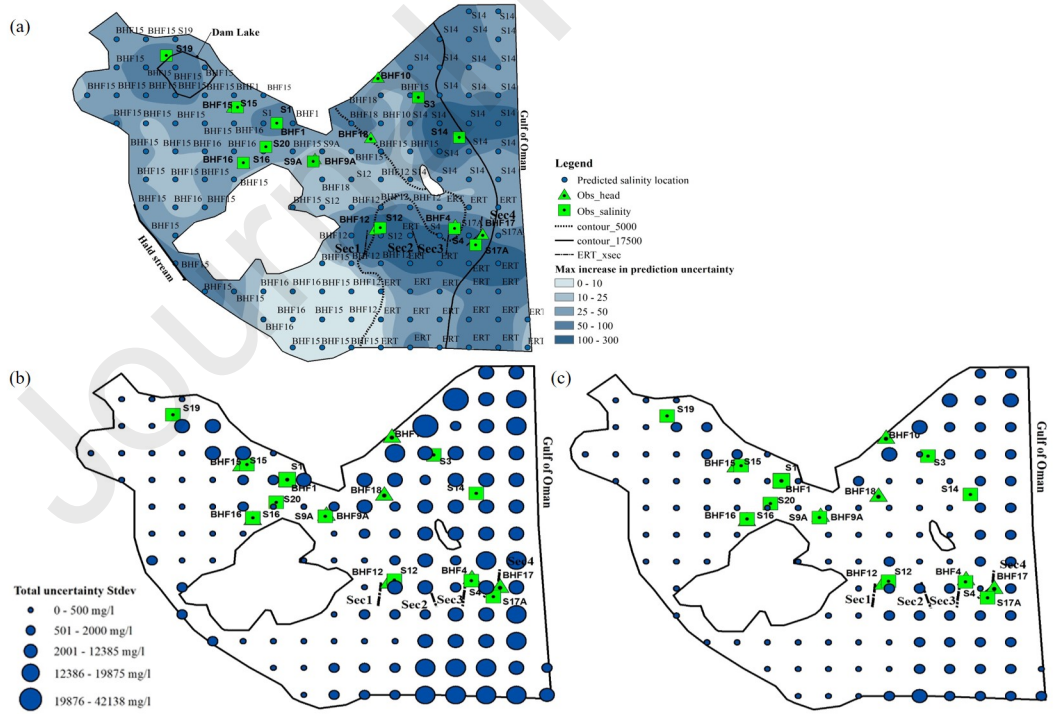
Zhao, Z. et al., 2016. A hybrid sampling method for the fuzzy stochastic uncertainty analysis of seawater intrusion simulations. Journal of Coastal Research, 32(3): 725-734.

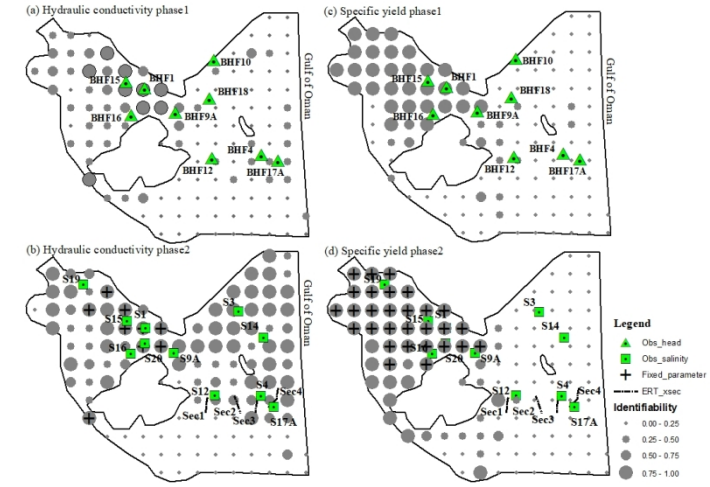


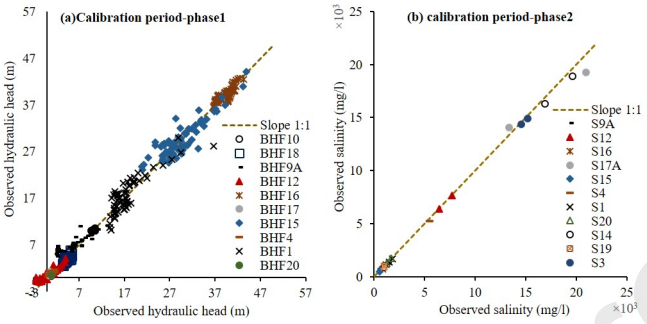


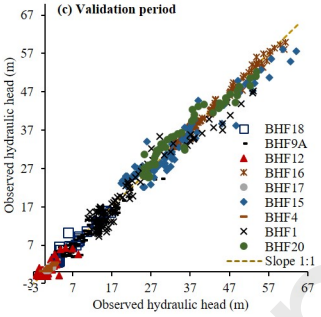


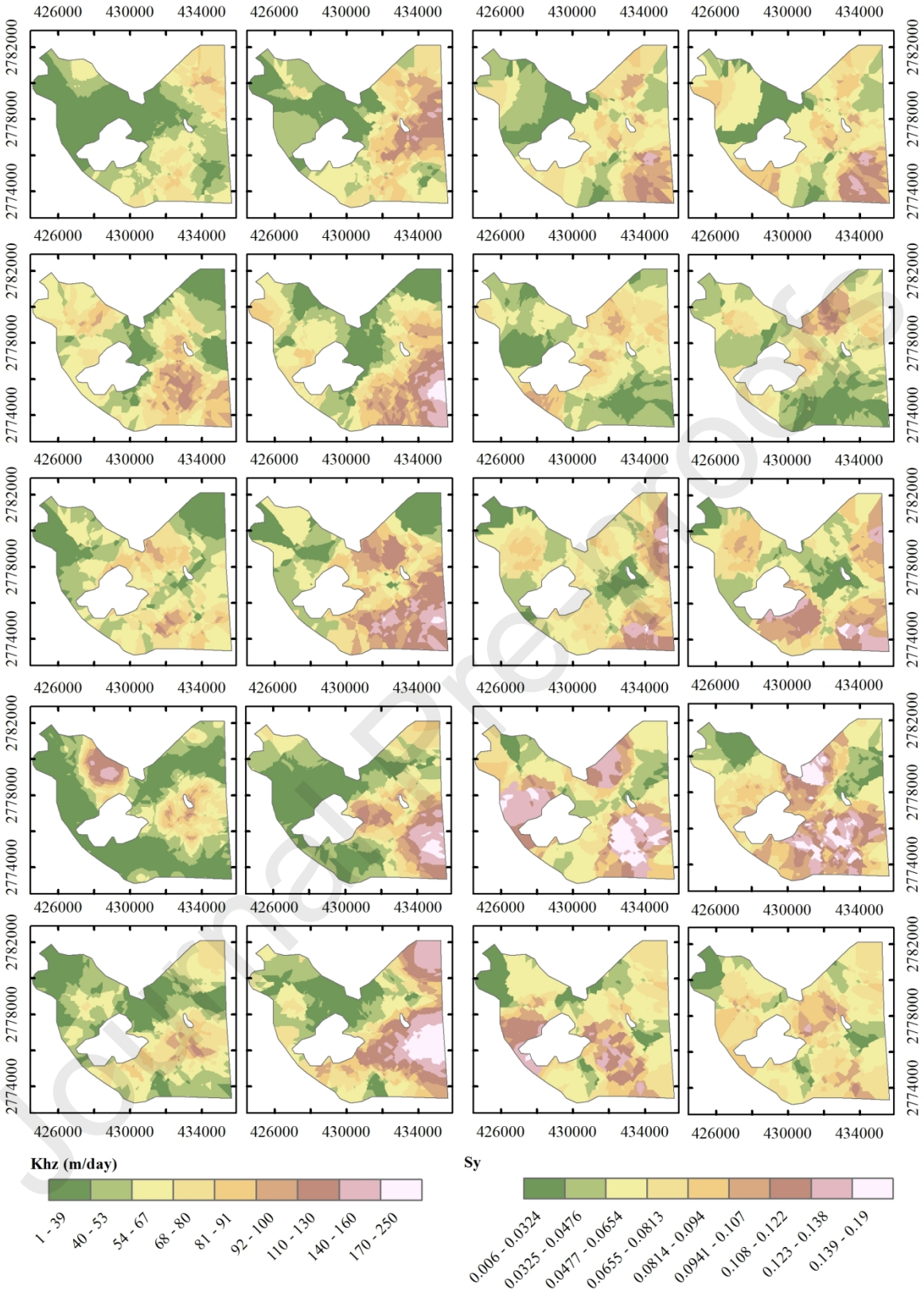


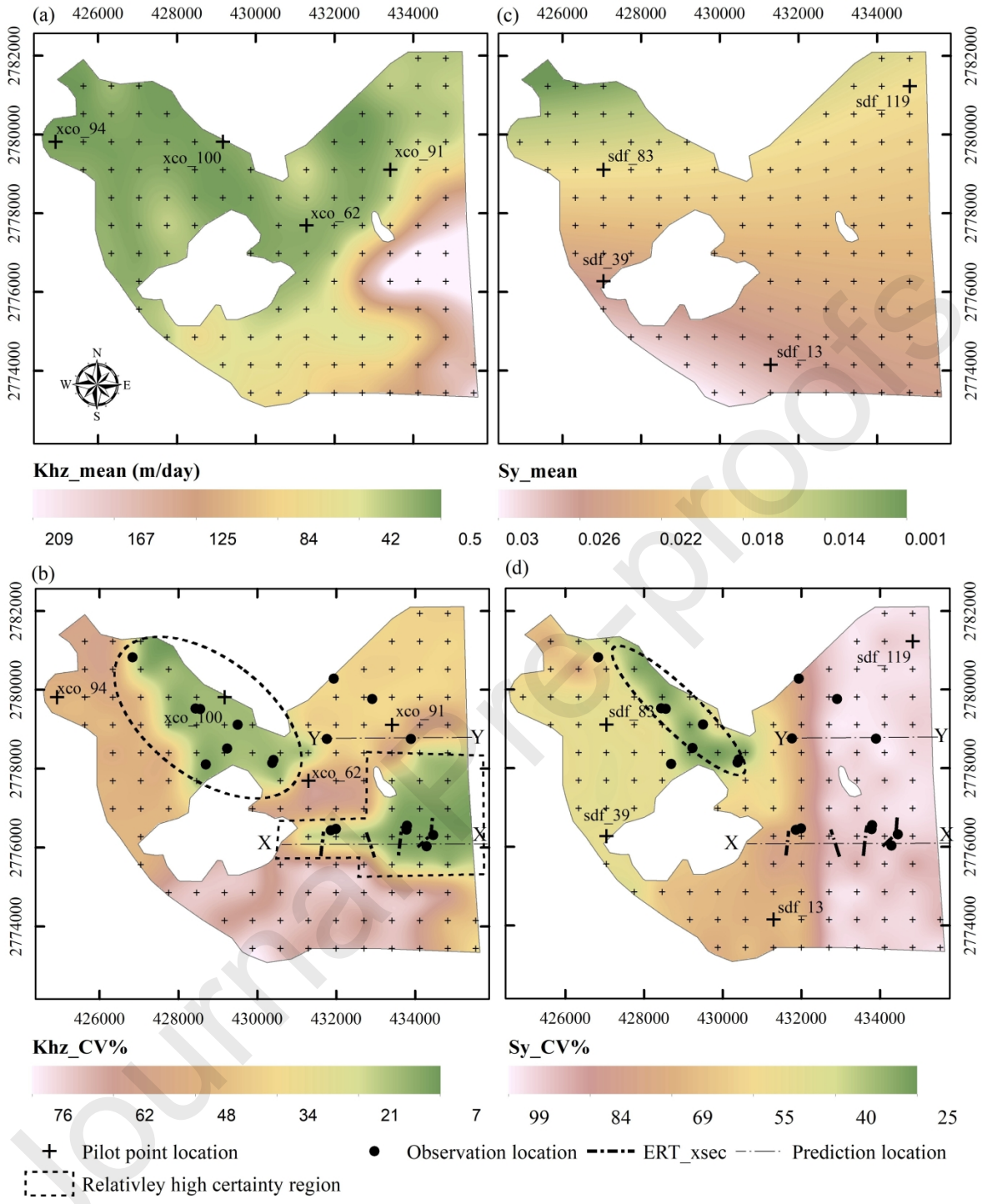












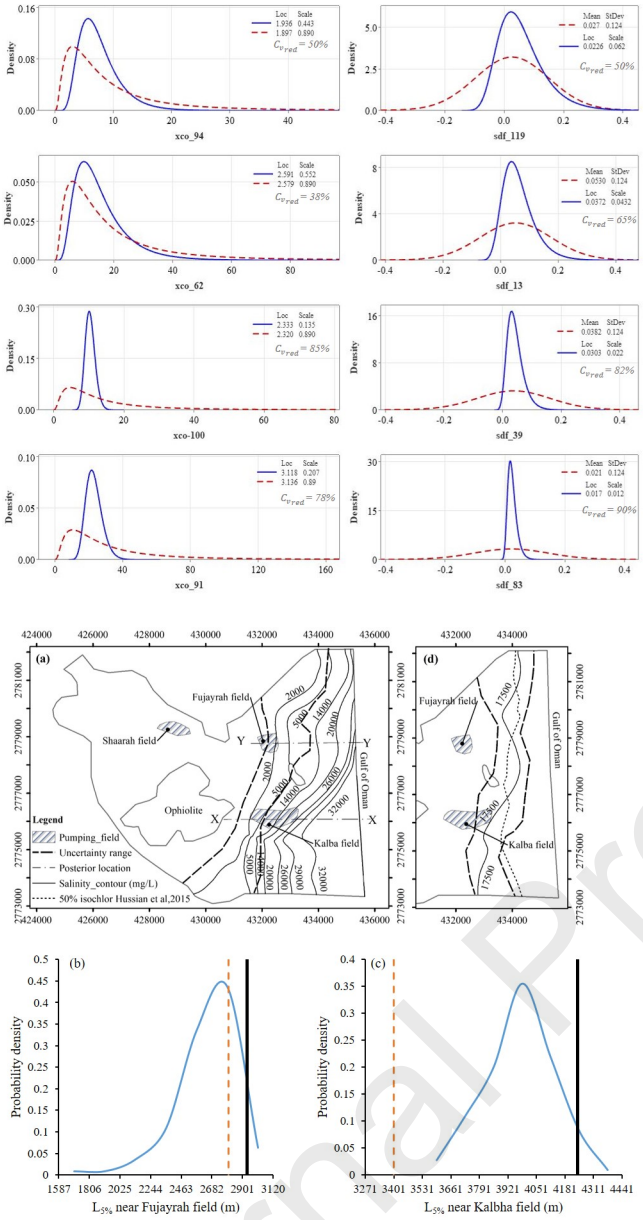


Figure 1 Flowchart of the proposed framework to quantify the prediction uncertainty. Numbers in circles represent the framework steps, and rectangles for the NSMC method based on OLHS, and a filtering technique.

Figure 2 Flow chart of the filtering technique to screen out less likely parameter sets from undergoing the recalibration process.

Figure 3 (a) Study area and the aquifer geological map, (b) Initial concentration distribution in 1988, (c) Initial hydraulic head in the main aquifer in 1988. Vertical exaggeration is 5.

Figure 4 Spatial distribution of hydrogeological parameters:(a,b) horizontal hydraulic conductivity and (c,d) specific yield for pre-calibration and post-calibration respectively. Circles represent the pumping test locations.

Figure 5 (a) Maximum increase in the predictive uncertainty variance of the salinity distribution, incurred by the removal of each observation group successively. The locations of predicted salinity concentrations are shown in circles, their labels indicate the observation group, which increased the prediction uncertainty the most when was left out. (b) & (c) Pre- calibration & post-calibration total predictive uncertainty for the salinity concentration respectively.

Figure 6 ParID of the hydraulic conductivity and the specific yield at the pilot points in phases 1&2 (a,b) for the hydraulic conductivity and (c,d) for the specific yield.

Figure 7 Goodness of fit of model outputs and their measured counterparts. Calibration results of (a) heads in phase1, (b) Salinity in phase2. (c) Validation results of heads in phase2.

Figure 8 Prior and posterior fields of the logarithm of hydraulic conductivity and specific yield fields. Prior fields were generated using OLHS. 1st and 3rd columns are for log Khz and Sy priors, 2nd and 4th columns are for their posteriors. Kriging interpolation method was used to generate the spatial distribution of the fields.

Figure 9 (a) mean and (b) C_v% of the Khz posterior distribution, (c) mean and (d) C_v% of the Sy posterior distribution. + represents the pilot points locations. The larger symbol size shows the location of randomly selected parameters to display their distributions.

Figure 10 Prior and posterior parameter probability distributions of eight randomly selected parameters, where dashed lines are for the priors and solid lines for the posteriors. The left and right columns are for Log Khz and Sy respectively. The plots were arranged in ascending order of their corresponding ParID values. The locations of those parameters were shown in Figure 9.

Figure 11 (a) Isochlor distribution at the end of 2050 using the calibrated parameter set of the minimum error variance, (b & c)Posterior predictive uncertainty distributions of $L_{5\%}$ isochlor (2000 mg/l isoline) near Fujayrah (Y-Y sec) and Kalba (X-X sec) respectively, thick solid lines represent 95% percentile of $L_{5\%}$ extent, and dashed lines indicate the distance from coastline to the center of the pumping field, (d) 50% Ischlor at the end of 2015 compared to other study results, dashed lines represent uncertainty intervals.

Table 1. Summary of previous studies on quantifying the prediction uncertainty in SWI

Ref.	Application	Number/name of uncertain inputs	Prediction Uncertainty analysis method	Alleviate computational burden	Key features
Rajabi and Ataie-Ashtiani, 2014; Rajabi et al., 2015	Two cases: Conceptual, 2D, Henry problem of SWI in coastal aquifers, and circular island	2 and 6/ hydraulic conductivity and inflow, and for the other case, porosity, horizonal and vertical hydraulic conductivities, and disperivity in the three dimensions	Monte Carlo simulations (MCS)	Effiicent sampling (OLHS)- Data-driven surrogate model	MCS method generates many independent realizations of random parameter sets, and subsequently employing each set to run simulation models to assess the prediction uncertainty. The uncertainty range of the parameter sets is identified based on prior knowledge range (e.g pumping tests, step-drawdown tests, laboratory tests, etc.). It is straightforward, and efficient for various probability distributions, but
Zhao et al. (2016)	Conceptual, 2D.	2/hydraulic conductivity and longitudianl dispersivity	Monte Carlo simulations (MCS)	Effiicent sampling (LHS and restricted stratified sampling)	
Miao et al. (2019)	Longkou City, China, 3D	2/ hydraulic conductivity and porosity	Monte Carlo simulations (MCS)	Data-driven surrogate model	
Mostafaei-Avandari and	Coastal aquifer of Ajabshir, Iran, 3D	10/hydraulic conductivity of 10 layers	Monte Carlo simulations (MCS)	Effiicent sampling (OLHS) - Parallel computing	

Ketabchi (2020)					may result in significant differences between real-world outcomes and expected results if parameter sets are not conditioned on measured state variables (e.g. piezometric heads, concentration data, etc).
Yang et al. (2021)	Baldwin County, USA,3D	22/hydraulic conductivity at 22 zones	Monte Carlo simulations (MCS)	N/A	
Ketabchi and Jahangir (2021)	Conceptual, 2D.	3/hydraulic conductivity, sea water level, land surface inundation	Monte Carlo simulations	N/A	
Rajabi and Ataie-Ashtiani (2016)	Two cases: Conceptual, 2D, Henry problem of SWI in coastal aquifers, and Kish Island, Iran, 3D	1 and 2/ hydraulic conductivity , and for the other case, hydraulic conductivity, and longitudinal disperivity	Conditional Monte Carlo simulations/ Markov Chain Monte Carlo (MCMC) simulation	Effiicent sampling (OLHS)- Data-driven surrogate model	MCMC propagates the prior parameter probabilities through the model to generate posterior parameter and prediction probabilities considering the model nonlinearity. Non-uniqueness becomes no longer a problem but an advantage, however it is computationally expensive and unaffordable for variable-density aquifer models (Carrera et al., 2010; Llopis-Albert et al., 2016), especially for highly parameterized models (Jiang et al., 2021).
Zeng et al. (2016)	Laizhou Bay,China,3D	7/hydraulic conductivities at four zones, inflow , recharge ratio, and longitudinal dispersivity	Conditional Monte Carlo simulations/ Markov Chain Monte Carlo (MCMC) simulation	N/A (30000 simulations)	
Pool et al. (2015)	Mar del Plata, Argentina,2D	1/hydraulic conductibvity field (conductivity distribution)	Conditional Monte Carlo simulations/ Stochastic Self-calibrating method	Sharp interface approach	Self-calibrating and gradual conditioning methods deform prior stochastic parameter sets to achieve a desired objective function and assess prediction uncertainty. Non-uniqueness becomes an advantage, but is
Llopis-Albert et al. (2016)	Conceptual, 2D.	1/hydraulic conductibvity field (conductivity distribution)	Conditional Monte Carlo simulations/ Gradual conditioning	Sharp interface approach	

			stochastic inverse model		computationally expensive and could be unaffordable for variable-density aquifer models (Carrera et al., 2010; Llopis-Albert et al., 2016), especially for highly parameterized models (Jiang et al., 2021).
Hugman and Doherty (2022)	Vale do Lobo subsystem of the Campina de Faro aquifer,Portugal,3D	565/hydraulic conductivity, specific storage and conductance of the aquitard which separates the deep aquifer from the upper aquifer	Conditional Monte Carlo simulations/ Iterative ensemble smoother (iES)	a fast-running model based on non-physical parameters, and a complex model to estimate equivalent values for the non-physical parameters	iES is a batch form of the Ensemble Kalman Filter, which conditions the ensemble of realizations obtained through Monte-Carlo simulations and assimilates all available data at once in a single analysis step (Rajabi et al., 2018). The prediction stage involves the model simulation for each realization in the ensemble. It is computationally efficient for highly parameterized models (White, 2018), but can handle only modestly nonlinear models (Rajabi et al., 2018).
Herckenrath et al. (2011)	Conceptual, 2D, Henry problem of SWI in coastal aquifers.	50/hydraulic conductivity	Conditional Monte Carlo simulations/ Null space Monte Carlo (NSMC)	Parallel computing	NSMC method generates multiple realizations that have the same solution space projection as the calibrated parameter set, satisfy an acceptable model-to-measurement fit (Tonkin and Doherty, 2009),

					and assess prediction uncertainty. It is a computationally efficient method for highly parameterized models (Safi et al., 2019; Tonkin and Doherty, 2009), but the multiple realizations may be biased towards the calibrated solution space (Yoon et al., 2013).
Coulon et al. (2021)	Magdalen Islands (Quebec, Canada), 3D	52/hydraulic conductivity	Linear prediction uncertainty	Sharp interface approximation	Linear analysis quantifies the prediction uncertainty by comparing its uncertainty variance before calibration (considering the prior knowledge of parameters values) and after calibration (considering the ability of the calibration dataset to constrain the parameters). It is efficient for highly parameterized models. However, it may provide inaccurate results for nonlinear models such as SWI models (Brunner et al., 2012), and the parameter set which corresponds to a prediction at a certain confidence level cannot be identified (Tonkin et al., 2007).

Present study	Wadi Ham aquifer, 3D	242/hydraulic conductivity and specific yield	Conditional Monte Carlo simulations/ Null space Monte Carlo (NSMC)	Parallel computing, Effiicent sampling (OLHS)	Combining the benefits of NSMC and OLHS to provide a practical, computationally efficient framework for quantifying uncertainty in SWI prediction.

Table 2. Sources of uncertainty in the calibration dataset, and their standard deviation

Source of uncertainty	Description	Standard deviation (stdev)	units	Remarks
Head measurement ($\boldsymbol{\sigma_{h_{measured}}}$)	Originated from human errors in execution, and the inaccuracies in the water depth measuring devices, borehole verticality, screen location determination, and determination of the geospatial position of the boreholes.	0.57	m	Adapted from Rau et al. (2019), Their estimation was based on best possible estimate or collected from the literature.
Ground elevation at the observation wells (σ_z)	Originated from inaccurate determination of ground elevations.	5	m	Ground elevation data were acquired by SRTM (Version 3) (NASA Shuttle Radar Topography Mission, 2013). The vertical accuracy differed according to the land cover. Elkhrachy (2018) assessed the accuracy in Najran city, KSA, which has similar site characteristics.
Fluid density of the water column inside the observation well (σ_ρ)	Errors propagated from estimating the fluid density of the water column inside the observation well.	1.5	Kg/m^3	Based on the mean of values considered in the studies (Coulon et al., 2021; Post et al., 2018).

Reproduction of the observed heads close to the pumping fields ($\boldsymbol{\sigma_{pump}}$)	Errors in reproducing the observed heads near pumping wells due to inaccurate identification of the cyclic operation of the pumping well.	0.40	m	The uncertainty was assigned by averaging the differences between the records at the observation wells near the pumping wells and the smoothed data using the moving average technique during the dry season period when the change was considered to be minimum due to the natural causes.
Salinity concentration measurement at the observation wells ($\boldsymbol{\sigma_{sal\_samples}}$)	The uncertainty in hydrochemical data from observation wells, including errors from the conductivity measurement device and human errors.	3%	-	Adapted from Sherif et al. (2011a) by comparing the change in two successive salinity measurements at one observation well within a short period of less than one month, then the average was considered.
Salinity concentration measurement deduced from a calibration raster ($\boldsymbol{\sigma_{sal\_raster}}$)	The uncertainty stemmed from interpolating the point salinity measurements, in addition to the measurement uncertainty.	6%	-	Crude estimate
ERT intrusion length observation ($\boldsymbol{\sigma_{ERT}}$)	Errors in geophysical and hydrochemical measurements, pertophysical model uncertainties (relationship between salinity and water resistivity), non-uniqueness in resistivity interpretation, geophysical inversion approach, and inversion parameters.	9%	m	Adapted from Zeynolabedin et al. (2021)

Table 3 Characteristics of the study area and its model discretization information

Hydraulic and transport parameters

Feature	Description
Horizontal Hydraulic conductivity (Khz) in X-direction (m/d)	Varied spatially, their values are based on Short-duration pumping tests (Sherif et al., 2012a; Sherif et al., 2014). The pumping test sites are shown in Figure 4
Specific yield (Sy)	Varied spatially, based on Sherif et al., 2012a study
Anisotropy ratio of Khz in X-direction to that in Y direction	1
Anisotropy ratio of vertical hydraulic conductivity to Khz	10
Molecular diffusion coefficient ($\boldsymbol{D_d}$) (m^2/s)	1×10^{-9}
Longitudianl dispersivity ($\boldsymbol{\beta_l}$) (m)	20
Transverse dispersivity ($\boldsymbol{\beta_T}$) (m)	2
Porosity ($\boldsymbol{\varepsilon}$)	0.3
Fluid dynamic viscosity ($\boldsymbol{\mu_0}$) (kg/m.s)	0.0011

Based on Sherif et al., 2012a study

Hydrogeological & Geological information

Feature	*Description*
Hydrogeological & Geological infromation	The subsystem consists of two primary aquifers: a surficial unconfined alluvial shallow aquifer composed of Quaternary sediments ranging from unconsolidated sand and gravels to consolidated gravels (Figure 3) with a thickness ranging from 24 to 100 m from the upstream side of the Wadi Ham dam to the coast. Underlying the shallow aquifer is a deep aquifer composed of consolidated rocks of Semail formation with a thickness ranging from 10 to 150m. The alluvial aquifer serves as the primary source of water for agricultural use (Sherif et al., 2017; Sherif et al., 2014).

Model discretization	
Feature	***Description***
Horizontal discretization	Using 17470 linear triangular prisms, which became finer close to the study area boundaries, pumping wells, and the dam's reservoir where the change in the hydraulic gradient was abrupt.
Vertical discretization	Using three layers, the main aquifer (Wadi gravel) split into two layers, while the third layer was for the deep aquifer.
Temporal discretization	Using Automatic time-stepping method with predictor-corrector scheme, adjusting time step based on convergence behavior with error tolerance of 1×10^{-3}.

Table 4 Boundary conditions used in the flow and transport model

Flow boundary conditions			
Location	**Type**	**Value**	**Remarks**
Sea-side boundary along Gulf of Oman	Dirichlet	Constant zero level above the mean sea level (m)	The freshwater and saltwater densities are 1000, and1025 kg/m^3 respectively.
Wadi Ham stream	Neumann	0.103 m/day	Adopted from Sherif et al., 2014.
Wadi Hald stream	Neumann	0.025 m/day	Adopted from Sherif et al., 2014.
Wadi AlHayal stream	Dirichlet	Constant head 10.75 m	The examination of BHF10 hydrograph showed that there was no significant change in the water level, with seasonal fluctuations ranging from 0.25 m to 0.5 m.
Shaarah pumping field	Well-type	5500 m^3/day from1988 to 1994, gradually decreased to 3000 m^3/day	Fully penetrating the main aquifer. The abstraction reduction was due to the deterioration in the water quality. Data adopted from Sherif et al., 2014.

Fujayrah pumping field	Well-type	1500 m^3/day from 1988 to 1994, then scaled down to 750 m^3/day	
Kalba pumping field	Well-type	4000 m^3/day from 1988 to 1994, then significantly increased to 20000 m^3/day	Fully penetrating the main aquifer. Data adopted from Sherif et al., 2014.
Dam reservoir recharge	Cauchy	Varied elevations, ranged from 73m (considered empty) to the maximum water level of 88m. The transfer rate was considered 11.7/day	The storage elevation was calculated based on the relationship between the rainfall depth and the storage volume behind Wadi Ham dam developed by Sherif et al. (2011b) study, and the reservoir stage-storage curve. The reservoir stage-storage curve was estimated using ArcGIS based on the SRTM digital elevation model.
Rainfall Recharge (Top surface of the model)	Neumann	Ranged from 0.2 to 172 mm/month. The median was 12 mm/month	The rainfall data from Dec 1988 till Oct 2004 were deduced from Sherif et al. (2006), from Jan 2009 till 2020, the data were derived from the open-access of remotely sensed-derived data, the Water Productivity Open Access Portal (WaPOR) database (FAO, 2018). The missed and forecasted events followed the same trend. The recharge factor was 20% of the measured rainfall (Sherif et al., 2017)
Solute transport boundary conditions			
Sea-side boundary	Dirichlet	constant salinity concentration 35,000 mg/l	Combined with a constraint allowing the salinity BC to be active only when saltwater enters, while outflowing parts are to be switched to an open BC, allowing the free movement of fresh and mixed water through the upper part of the boundary.
Sea-side boundary	Outflowing boundary	-	$-(D.\nabla C).n=0$, where n corresponds to the normal unit vector. The boundary is impervious for dispersive fluxes, while it is convectively pervious.
Wadi Ham, Wadi Hald, and Wadi AlHayal streams	Dirichlet	250 mg/l	-
Dam reservoir saline mass flux (Top surface of the dam reservoir)	Dirichlet	250 mg/l	-

Rainfall recharge saline mass flux (Top surface of the model)	Neumann	100mg/l multiplied by rainfall recharge rate ($gm/m^2/day$)	Adopted from Hussain et al. (2015). A dummy layer of 3 m thickness was assigned on the top of the aquifer to reduce the adverse effect of the mass imbalance. The imbalance results from applying the convective form of the mass transport equation (Saaltink et al., 2004).

Table 5. Observation dataset for calibration and validation, estimated parameters, and their prior parameter distributions

Observation data					
Type	**No. and form of available measurements**	**Measurements for calibration**	**Measurements for validation**	**Source**	**Remarks**
Heads	1324 monthly measurements of depth to water at 9 observation wells from December 1988 to April 2005	Using 546 monthly head measurements from December 1988 to September 1994	778 monthly head measurements from October 1994 to April 2005	The observation dataset was extracted from previous studies (Sherif et al., 2012a; Sherif et al., 2006; Sherif et al., 2014)	The observation wells' locations are presented in Figure 3
Salinity concentrations	21 salinity concentrations at 11 observation wells collected in Jan 2004	Used all available data	-		
	A salinity calibrated raster in Sep 1994	Used all available data	-	Extarcted from Sherif et al., 2012a study	-
Geophysical survey data (ERT survey data)	Four profiles of the true resistivity,parallel to the shoreline, were utilized to estimate the average SWI length in Jan 2004	Used all available data	-	Produced in Sherif et al. (2006) study	Even transects run perpendicular transects can delineate the seawate

					r interface more accurately, however, parallel transects can still provide useful information to determine the possible location of the seawater wedge (Costall et al., 2018).

Parameters

Hydrogeological property	No. of pilot points	Geostatistical model	Initial sill and range	Source	Remarks
Logarithm of Khz	121 pilot points	Gaussian variogram	0.21 & 4825	The geostatistical parameters were obtained by cross-validation using ArcGIS 10.1 Geostatistical Analyst.	- The ordinary Kriging interpolation method was used to obtain the parameters value. - The varigorams describe the mean of the prior
Sy	121 pilot points	Spherical variogram	5.82×10-5 & 4721		

							parameter distributions at the pilot point locations.

Prior parameter distribution

Parameter	Distribution	Mean range	95% Confidence interval		Standard deviation	Source	Remarks
			Lower limit	Upper limit			
Khz (m/day)	Log-normal	2.3 - 211	0.086	300	0.89	Nine short-duration pumping tests found in the literature (Sherif et al., 2012a; Sherif et al., 2014)	- The 95% confidence interval was large due to limited data, post-processed data, and to account for model structural noise.
Sy	Normal	0.0057 – 0.04	0.007	0.3	0.075	Estimated Sy from the pumping tests was low 1.3×10^{-20} to 0.002. These tests give unreliable values (Nwankwor et al. 1992). Hence it was based on the calibration of Sherif et al. (2012a)	- The standard deviation was estimated by dividing the confidence interval by 4.

Table 6 Statistical descriptors of the residuals after model calibration and model validation

Observation well	Calibration				Validation				*Stdev	measurement noise- 95% CI (4× Stdev)	Units
	No. of readings	RMSE	NSE	RQS	No. of readings	RMSE	NSE	RQS			
BHF1	64	1.262	0.73	0.74	123	**3.695**	0.93	0.94	0.699	2.796	m
BHF9A	62	1.113	0.77	0.78	122	1.020	0.75	0.61	0.571	2.284	
BHF18	48	0.832	0.72	0.74	28	1.250	0.94	0.92	0.573	2.292	
BHF4	63	0.893	0.37	0.70	121	0.209	0.83	0.83	0.575	2.3	
BHF16	64	0.937	0.71	0.71	105	0.852	0.99	0.98	0.576	2.304	
BHF12	62	1.181	0.72	0.80	121	0.912	0.71	0.71	0.699	2.796	
BHF17	66	0.836	0.26	0.78	112	0.131	0.83	0.84	0.578	2.312	
BHF15	66	**3.516**	0.81	0.81	79	**3.250**	0.92	0.94	0.700	2.8	
BHF10	51	0.438	0.94	**0.14**					0.577	2.308	
BHF20					67	1.208	0.94	0.95	0.633	2.532	
S20	2	68.99	0.97	1.00					92.83	371.32	Mg/l
S4	1	685.82							352.0	1408	
S9A	2	129.46	0.11	1.00					96.75	387	
S1	2	168.12	0.23	1.00					120.49	481.96	
S12	2	298.40	0.99	1.00					519.40	2077.6	

S14	2	548.50	0.75	1.00		1183.17	4732.68	
S15	2	57.14	0.89	1.00		41.00	164	
S16	2	48.42	0.70	1.00		76.12	304.48	
S17A	2	658.76	0.75	1.00		1025.84	4103.36	
S19	2	92.23	0.27	1.00		66.83	267.32	
S3	2	771.34	0.07	1.00		1044.09	4176.3	
ERT	1	228.90				227.88	911.52	

*Stdev is calculated based on Eq.1, Eq.3, and Table 2, and is equal to the inverse of the weights assigned to the various observations (Supplementary materials S(C3)).

Highlights

- Observed heads, salinities, & retrieved SWI length from ERT data used in calibration
- Parameter estimation reduced the uncertainty in salinity predictions
- Quantifying uncertainty in SWI extent was performed using OLHS linked with NSMC
- Fixing identified parameters, OLHS & filtering technique reduced computational cost
- Identifying pumping fields threatened by SWI in 2050, Wadi Ham aquifer, UAE

CRediT authorship contribution statement

Samia Saad: Conceptualization, Methodology, Software, Formal analysis, Investigation, Validation, Resources, Visualization, Writing - original draft, Writing - review & editing, Funding acquisition. **Akbar Javadi:** Supervision, Conceptualization, Writing - review & editing. **Raziyeh Farmani:** Supervision, Writing - review & editing. **Mohsen Sherif:** Writing - review & editing