Journal of Hydrology 590 (2020) 125427

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

Evaluating two multi-model simulation–optimization approaches for managing groundwater contaminant plumes

Ahmed S. Elshall[a], Ming Ye[a,*], Michael Finkel[b]

[a] Department of Earth, Ocean, and Atmospheric Science, Florida State University, Tallahassee, FL, USA
[b] Center of Applied Geoscience, University of Tübingen, Tübingen, Germany

ARTICLE INFO

This manuscript was handled by Huaming Guo, Editor-in-Chief

Keywords:
Groundwater remediation
Contaminant plume containment
Water quality control
Multi-objective optimization
Conceptual model uncertainty

ABSTRACT

Addressing conceptual model uncertainty using multi-model approaches has become an important topic in groundwater hydrology. Model uncertainty is of practical importance when designing a pump-and-treat system for groundwater remediation, as inappropriate consideration of model uncertainty may ultimately lead to ineffective system designs. This study introduces two multi-model approaches for addressing model uncertainty at a site contaminated by chlorinated hydrocarbons and BTEX compounds. We use a multi-objective simulation–optimization method to design a multiple-well pump-and-treat system to contain contaminant plumes within the site perimeter, while minimizing the pumping rate. The design parameters include the coordinates of the pumping well locations and pumping rates. We consider four groundwater models to address uncertainty in conceptualizing site geology, boundary condition, and recharge. Using two multi-model approaches, we evaluate different design concepts. The first approach is a model selection approach that identifies the critical models, which particularly govern the design process, within the multi-model ensemble. The second approach is a model aggregation approach that identifies system designs that meet the remediation objectives for all models of the ensemble. Results show that the model aggregation approach is a more conservative option, which results in reliable designs at the expense of higher pumping rates. This study evaluates these two approaches within the context of multi-model simulation–optimization, providing insights on multi-model simulation–optimization in groundwater management.

1. Introduction

The combination of simulation and optimization methods is a popular approach to tackle groundwater management problems (Gorelick et al., 2015; Yeh, 2015), namely, to design management solutions to the given problem. Simulation-optimization methods have been successfully applied to identify optimal well locations and withdrawal rates for sustainable groundwater management (Ashraf et al., 2017; Burnett et al., 2020; Delottier et al., 2017; Kicsiny and Varga, 2019; Luo et al., 2020; Shi et al., 2012; Wang et al., 2019). In coastal aquifer management, a specific focus is identifying optimal well settings that reduce saltwater intrusion, while satisfying the management objectives (Mondal et al., 2019; Rajabi and Ketabchi, 2017; Song et al., 2018). Simulation-optimization has also been used to assess hydro-economic aspects of groundwater management (Davidsen et al., 2015; Katic and Grafton, 2011; Peña-Haro et al., 2009; Zhang et al., 2018). In addition, simulation–optimization methods have been used to optimize data collection and monitoring networks (Ayvaz and Elçi, 2018; Christensen et al., 2009). Another core field of the application of simulation–optimization methods is the design of optimal systems to control and remediate groundwater contaminations. Applications include optimal designs for contamination-source identification, long-term contamination monitoring, water quality control, and clean-up of contaminated sites (Ayvaz, 2016; Chang et al., 2007; Guan and Aral, 1999; Lu et al., 2019; Luo et al., 2016, 2014; Raei et al., 2017; Saravanan et al., 2014). Parker et al. (2010) show that optimized remediation designs can yield cost savings of up to approximately 50% in comparison with a non-optimized design based on common engineering practices.

The description of groundwater systems in simulation models is inherently uncertain due to epistemic uncertainty in terms of lack of data and knowledge for characterizing hydrogeological properties, boundary and initial conditions, climate forcings, and some (residual) aleatory uncertainty (Beven, 2016). Model uncertainty needs to be considered within the simulation–optimization method to obtain reliable management or remediation designs (Freeze and Gorelick, 1999).

* Corresponding author.
E-mail address: mye@fsu.edu (M. Ye).

https://doi.org/10.1016/j.jhydrol.2020.125427
Received 13 April 2020; Received in revised form 11 August 2020; Accepted 13 August 2020
Available online 15 August 2020

Uncertainty analysis within the context of simulation–optimization for groundwater management and remediation generally focuses on parameter uncertainty by generating multiple parameter realizations using stochastic methods (Bayer et al., 2008; Warner et al., 2006), fuzzy set methods (Guan and Aral, 2004; Taravatrooy et al., 2019), chance constraint methods (He et al., 2008b; Medina et al., 1996), linear uncertainty analysis (Delottier et al., 2017), risk-based methods (De Barros et al., 2013), data worth analysis methods (Baú and Mayer, 2007; Cirpka et al., 2004), or combinations of these methods (He et al., 2008a; Lu et al., 2018, 2017). If uncertainty is attributed only to parameter errors, parameter estimates may be biased as they compensate for errors in model structure (Doherty and Welter, 2010; Rojas et al., 2008). Chitsazan et al. (2015) show that using a single conceptual model without accounting for conceptual model uncertainty results in overestimating the design reliability, which in turn increases the risk of failure of the groundwater remediation.

Conceptual uncertainty can be addressed using a total model error approach (Elshall et al., 2019; Pan et al., 2020; Xu et al., 2017) or multi-model analysis, where the latter is more common. In the multi-model approach, an ensemble of multiple alternative conceptual models is considered instead of a single model. The existence of alternative conceptualizations, methods, and assumptions during model development, results in conceptual model uncertainty. Conceptual uncertainties have been evaluated with respect to geological model structure (Mustafa et al., 2020; Ye et al., 2004), mathematical model structure (Elshall and Ye, 2019; Kopsiaftis et al., 2019; Maghrebi et al., 2015; Makropoulos et al., 2008; Vansteenkiste et al., 2014), recharge (Rojas et al., 2010a; Ye et al., 2016), boundary conditions (Elshall and Tsai, 2014), hydrogeochemical reactions (Lu et al., 2015), surface water-groundwater interaction (Enemark et al., 2019a), groundwater dependent ecosystems (Bianchi Janetti et al., 2019; Gondwe et al., 2011), and climate (Hartmann et al., 2012; Pholkern et al., 2019). For a detailed discussion, readers are referred to recent review articles about groundwater multi-model analysis (Enemark et al., 2019b; Heße et al., 2019; Höge et al., 2019; Refsgaard et al., 2012). In addition, Elshall et al. (2020) discuss several theoretical and practical advantages of multi-model analysis.

While multi-model analysis is a trending topic in groundwater hydrology (Ferré, 2017a, 2017b; Guillaume et al., 2016; Neuweiler and Helmig, 2017), multi-model analysis has received much less attention within the context of simulation–optimization for both groundwater management and remediation. This may be attributed to the fact that simulation–optimization methods are computationally expensive, which discourages using multiple conceptual models. This problem can be addressed through using surrogate models (Yin and Tsai, 2020, 2018; Zeng et al., 2018, 2016) and parallel computing (Chitsazan et al., 2015; Elshall et al., 2015). Apart from this practical concern, a more fundamental issue is how multiple conceptual models should exactly be incorporated for the evaluation of candidate management and remediation designs. Our systematic search in peer-reviewed groundwater literature shows that there are only few simulation–optimization studies that used multiple conceptual models. Table 1 summarizes the scopes of these studies, and the multi-model approaches used, including this study.

Given the studies shown in Table 1, we may distinguish four approaches that are used to incorporate multiple models in a simulation–optimization framework (Fig. 1). The simplest approach (Fig. 1a) uses multiple conceptual models independently, and compares the model predictions to draw conclusions. For example, Aksoy and Culver (2004) consider multiple conceptual models representing different levels of heterogeneity of the spatial distribution of both hydraulic conductivity and sorption coefficient. The study shows that the solution variability increases as the heterogeneity of the hydraulic conductivity field increases, and the time to compliance increases for systems with both chemical and physical heterogeneity, as compared to systems with only physical heterogeneity.

The second approach of cross-validating optimal designs (CVOD, Fig. 1b), which also uses multiple models independently, cross-validates the optimal design of well setting in each conceptual model using the other conceptual models. In this way, designs of multiple independent simulation–optimization runs are ranked to identify the most reliable designs. This approach is used in this study. By cross-validating optimal designs, CVOD can identify the critical model, which is the model that has the largest effect on the design reliability (Bayer et al., 2010, 2008; Kourakos and Mantoglou, 2008; Mantoglou and Kourakos, 2007).

A third approach (Fig. 1c) uses Bayesian model averaging (BMA) to account for groundwater model uncertainty in the simulation–optimization framework (Chitsazan et al., 2015; Chitsazan and Tsai, 2015; Mani et al., 2016b, 2016a; Tsai, 2010). BMA with chance-constrained programming (Chitsazan and Tsai, 2015; 2015) can additionally account for parameter uncertainty to improve the design reliability. The approach bases the design on the weighted averaged predictions of the candidate models. The weights are estimated based on each model's fit to the calibration data and the level of model complexity. While BMA makes better predictions than a single model (e.g., Liu et al., 2016a, 2016b; Elshall et al., 2018), the model that has the largest weight is not necessarily the model that makes the best prediction, and vice-versa (Seifert et al., 2012). In other words, the best model given a calibration principle (i.e., a better model is the model that can better reproduce the calibration data given model complexity) is not necessarily the best model given a prediction principle (i.e., a better model is the model that makes better prediction given the cross-validation data). Accordingly, if a critical model receives low or no weight in the multi-model ensemble (i.e., given a calibration principle) this can affect the reliability of the identified best design option. To overcome this issue, the fourth approach of concurring ensemble members (CEM, Fig. 1d) involves all individual conceptual models explicitly in the evaluation. This allows to check that certain requirements of a robust remediation design are met in each of the models. CEM is used by Timani and Peralta (2015) and in this study.

In this study, we develop a multi-model simulation–optimization framework for designing pump-and-treat systems for water quality control. We apply this framework to an industrial plant in Italy, which is contaminated with both CHC (chlorinated hydrocarbons) and BTEX (benzene, toluene, ethylbenzene and xylene). Finkel et al. (2008) show that disjoined capture and separate treatment of these two groups of contaminants result in cost savings for remediation, and the disjoined capture can be achieved by using a simulation–optimization framework with a particle-tracking-based advective control scheme to evaluate the performance of pump-and-treat remediation designs. This manuscript has several innovative components. First, we provide an improved formulation of the objective function by introducing a constraint to restrict the capture zone to the site perimeter (i.e., to avoid legal or financial liability), extending the objective function to account for multiple conceptual models, and updating the multi-objective weighting coefficients to improve the solution robustness. Second, we introduce the concepts of CVOD and CEM to simulation–optimization methods for addressing conceptual uncertainty in groundwater management and remediation. Third, we show the importance of conceptual uncertainty analysis to improve the design reliability. Fourth, we illustrate how CVOD can identify the critical model(s), and how CEM can identify robust design(s) under conceptual uncertainty. In the remainder of the manuscript, we introduce the study site to illustrate the need for disjoined capture and multiple conceptual models in Section 2. We present the simulation–optimization method of multi-model disjoined capture in Section 3. We discuss the simulation–optimization results and the performance of CVOD and CEM, and give an outlook of the multi-model simulation–optimization framework in Section 4. Finally, we summarize the main findings, and draw conclusions in Section 5.

Table 1
Summary of peer-reviewed literature on groundwater multi-model simulation–optimization.

Paper	Simulation-Optimization Problem	Conceptual Uncertainty[1]	Multi-model Approach[2]	No. of Conceptual Models
Aksoy and Culver (2004)	Aquifer cleanup, and water quality control using pump-and-treat system	GMS,HGC	IMR	12
Chitsazan and Tsai (2015)	Aquifer cleanup using saltwater extraction	GMS	HBMA-CC	36
Chitsazan et al. (2015)	Water quality control using hydraulic barrier	GMS	BMA-CC	3
Kopsiaftis et al. (2019)	Coastal aquifer management based on optimal withdrawal	MMS	IMR	4
Mani et al. (2016a), Mani et al. (2016b)	Conjunctive management of surface water – groundwater	HCM	BMA	11
Timani and Peralta (2015)	Evaluation of perennial aquifer yield strategies	GMS,BC	IMR, CEM	2
Tsai (2010)	Aquifer cleanup using saltwater extraction, and water quality control using hydraulic barrier	GMS,BC	BMA	15
Yang et al. (2017)	Aquifer cleanup using pump-and-treat system	GMS	IMR	2
Yin and Tsai (2018)	Aquifer cleanup using saltwater extraction	SM	BMA-CC	1
Yin and Tsai (2020)	Aquifer cleanup using saltwater extraction	SM	BMA	1
This study	Water quality control using pump-and-treat system	GMS,BC,R	CVOD, CEM	4

[1] Conceptual uncertainty: boundary conditions (BC), hydrogeochemistry (HGC), geological model structure (GMS), hydro-climate model (HCM), mathematical model structure (MMS), recharge (R), and surrogate model (SM).
[2] Multi-model approach: Bayesian model averaging (BMA), chance constrained (CC), concurring ensemble members (CEM), hierarchical Bayesian model averaging (HBMA), cross-validating optimal designs (CVOD), independent model run (IMR).

2. Study area, numerical modeling, and alternative models

2.1. Description of study area and contamination

The simulation–optimization approach in this study has been developed to support the design of groundwater remediation measures at a petrochemical plant in Italy. The site (Fig. 2) has an area of approximately 2 km^2. The plant is located on a coastal plain, surrounded by urban and industrial areas, agricultural land, areas used by the salt industry, and wetlands that are hydraulically connected to the Mediterranean Sea. The subsurface stratigraphic sequence in the study area has the following three main layers. The upper layer 1 consists of terrace alluvial sediments, and reaches a maximum thickness of approximately 20 m. The alluvium consists of highly heterogeneous

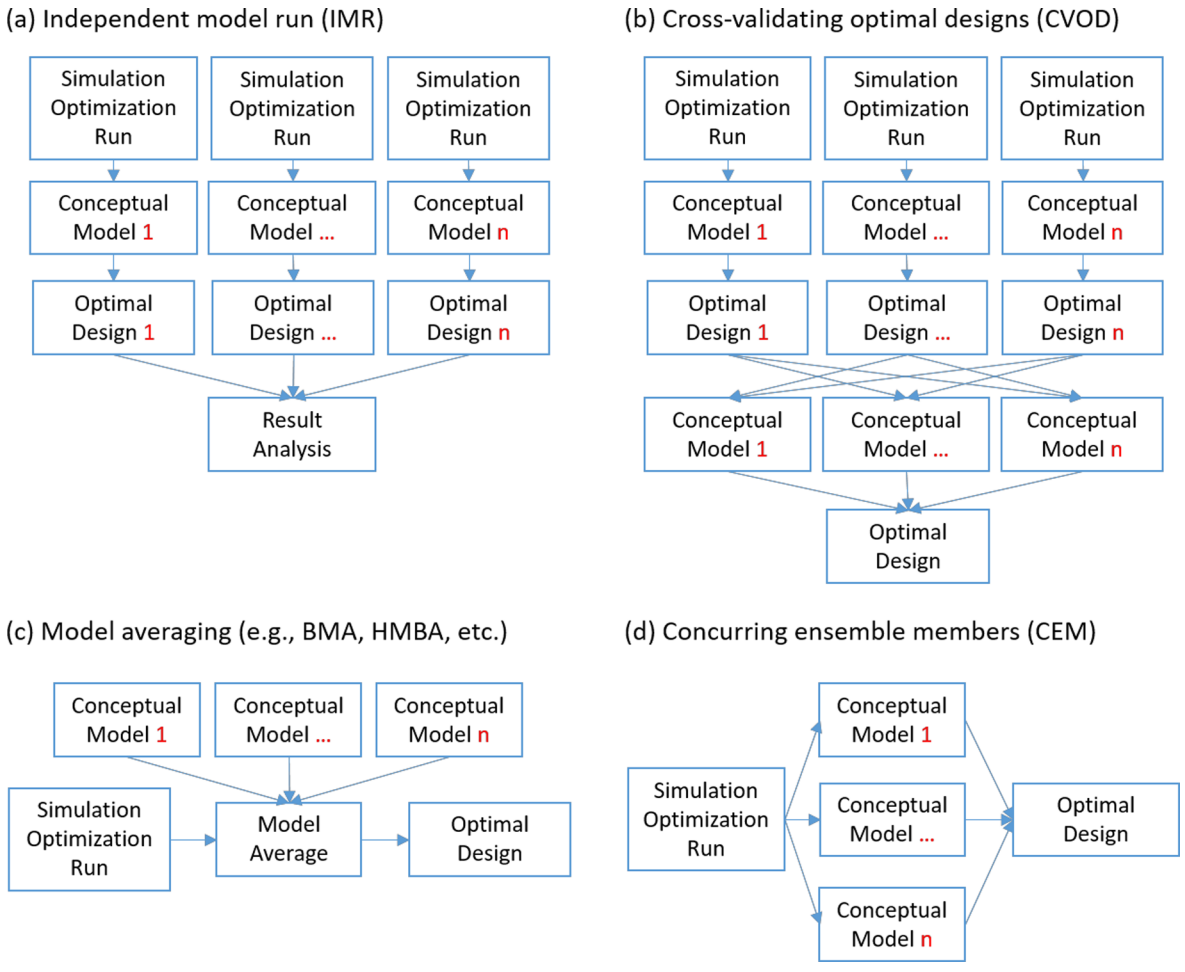

Fig. 1. Four approaches to incorporate multiple conceptual models in a simulation–optimization framework.

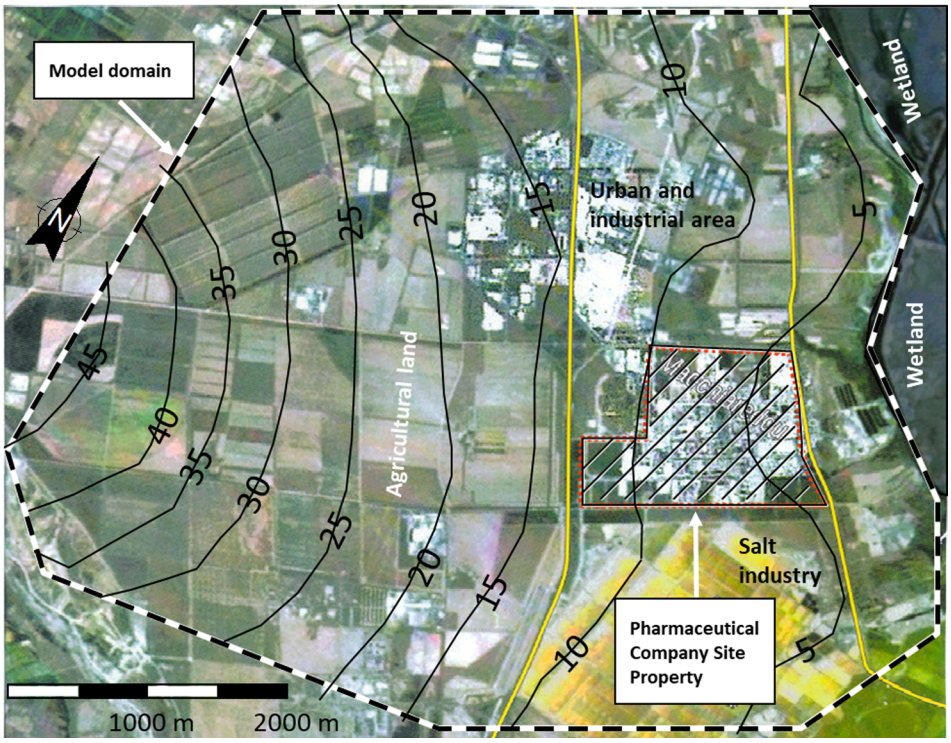

Fig. 2. Site map showing the model domain (thick dashed line), location of the petrochemical plant (red dotted line), and isolines of ground level (thin solid lines). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Pleistocene-age deposits ranging from conglomerates to sand and silty gravel with local lenses of silty clay. Layer 2 is an intermediate layer consisting of clay and silt, and is present in the whole study area with variable thickness. The lower stratigraphic layer 3 consists of alternating sub-layers of gravel and clay. The original morphology of the upper layer was changed by site leveling during construction of the petrochemical plant.

The hydrogeology in the study area is characterized by a phreatic aquifer (i.e. layer 1), an intermediate impervious layer (i.e. layer 2), and a confined aquifer (i.e. layer 3). This study only considers the phreatic aquifer where contaminants were detected. Hydraulic conductivity of the phreatic aquifer varies considerably in space with a mean value of approximately 3×10^{-4} m/s. The mean hydraulic gradient is 0.6%. Recharge of groundwater has several components, including infiltration from precipitation, runoff from the granite hills located in the west, leakage in water supply lines within the plant perimeter, and infiltration from deeper geological formations. The groundwater flow conditions are characterized by diverging flow directions (Fig. 3a).

The groundwater contamination consists of two major contaminant plumes and three major compounds: benzene, 1,2-DCA, and TCE (Fig. 3b–d). The BTEX plume is located in the northern part, with benzene being the major compound (Fig. 3b). Benzene concentrations observed in groundwater are up to 4.2 mg/l. In the area of this plume, also some minor occurrences of chlorinated hydrocarbons compounds (CHC) were found. The CHC contamination predominantly consists of 1,2-DCA and TCE, with concentrations of up to 240 mg/l and 3 mg/l, respectively (Fig. 3c and d). These plumes are in the southern part of the plant and largely overlap. Due to the large number of contaminant concentration measurements (Fig. 3), both the BTEX and the CHC plume were considerably well identified. They are in close proximity to one another.

2.2. Groundwater flow and particle tracking models

The MODFLOW code (Harbaugh et al., 2000) is used to develop steady-state groundwater flow models of the study area for each of the four conceptual models considered in this study. In this section, we describe all features of model 1, which is used as reference in the next section to describe how the other three conceptual models differ from model 1. The model grid covers an area of approximately 34.2 km^2. Variable cell dimensions range from 25 × 25 m in the main area of interest (e.g. the petrochemical plant) to 100 × 100 m in the model edges as shown in Fig. 4a. The western model boundary is specified as a general head boundary, which describes inflow into the model domain in relation to the hydrologic conditions beyond the model boundary. The head level has its maximum at the most western cell, and decreases with a linear gradient towards the south and the north. The eastern model boundary is assumed to be a constant head boundary (i.e. the sea level). The northern and the southern model boundaries approximately run parallel to the groundwater flow direction, and are therefore considered as no-flow boundaries. It is assumed that recharge within the model domain is limited to irrigated agricultural areas, and water losses of the industrial facility. Data from eight pumping tests and expert knowledge are used to delineate the hydraulic conductivity zones (Lantschner, 2006). Hydraulic conductivity in each conceptual model is calibrated based on site investigation data (i.e. pumping tests and borelogs of observation wells), and head measurements in 48 observation wells (Fig. 3). The spatial distribution and values of calibrated hydraulic conductivity are shown in Fig. 4a. The model comprises of two layers, an unconfined layer (i.e., layer 1) with a thickness ranging from 10 m to 20 m, and a confined layer (i.e., layer 2) with a thickness ranging from 5 m (east) to 45 m (west) (Fig. 4b). Hydraulic conductivity in the unconfined layer ranges from 0.43 to 17.3 m/d, and a constant hydraulic conductivity of 0.01 m/d is assigned to the confined layer of the model. Calibration of other alternative models is discussed in Section 2.3.

The MODPATH code (Pollock, 1994) is employed to simulate forward-tracking of particles from the boundary of the plumes. This forms the so-called advective control scheme for evaluating design options as has been used previously (Bayer and Finkel, 2007, 2004; Cousquer et al., 2019; Mulligan and Ahlfeld, 1999). In this work, the advective

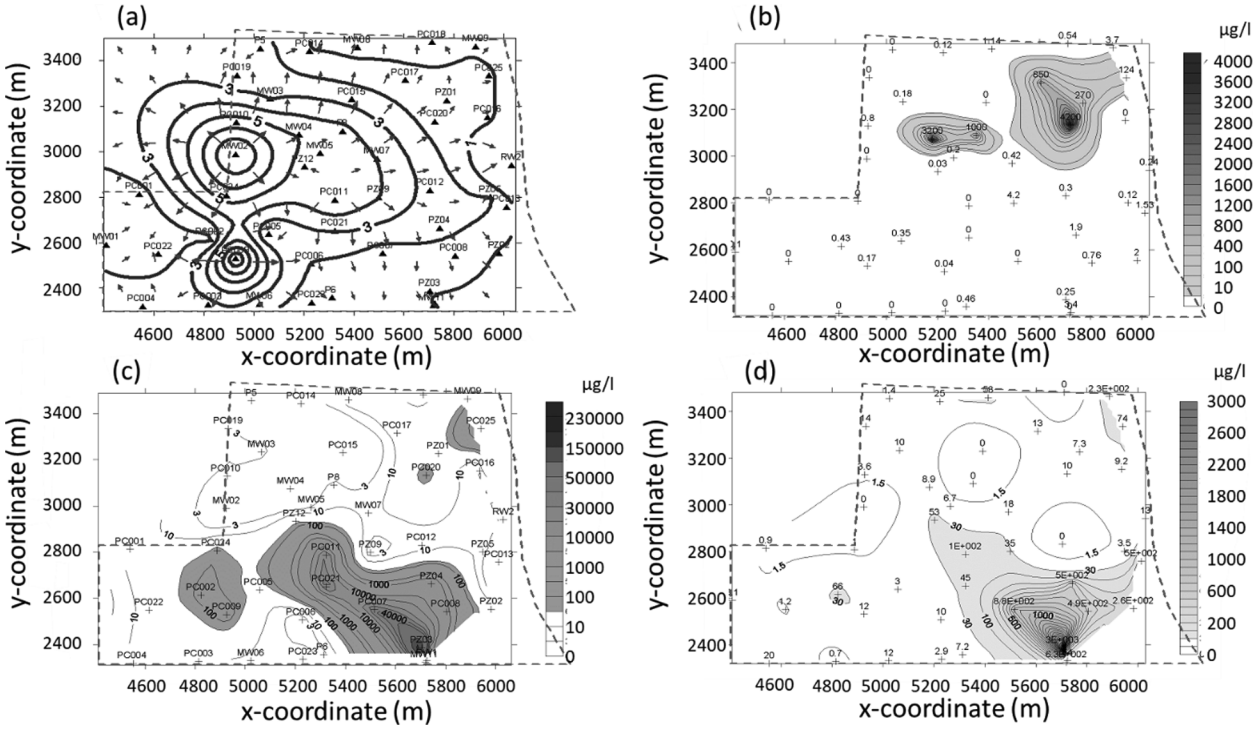

Fig. 3. (a) Isolines of hydraulic head and flow directions, and measurements, and isolines of concentration of (b) benzene, (c) 1,2-DCA, and (d) TCE in groundwater. The areas with high contaminant concentrations are shaded. The dashed line delineates the plant perimeter. Crosses marks are the measurement points. Spatial distribution of contaminant plumes was estimated using natural neighbor interpolation.

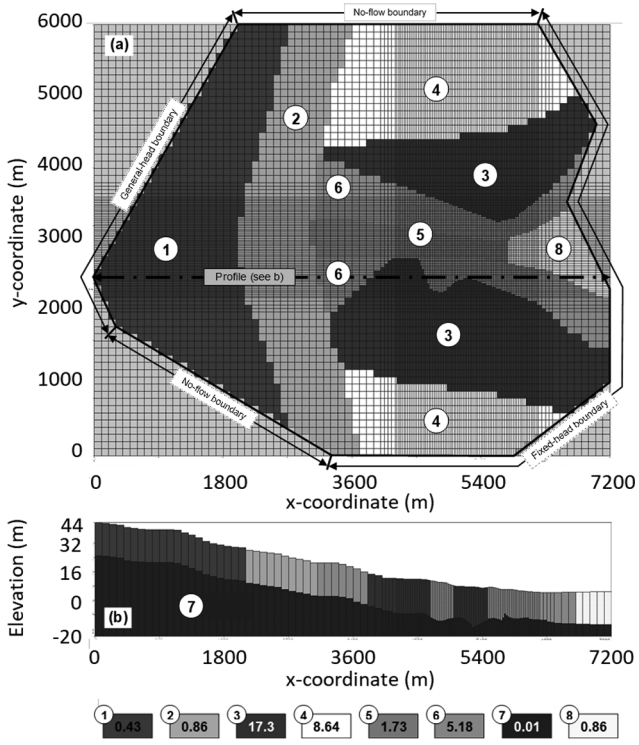

Fig. 4. Geometry, boundary conditions, and hydraulic conductivity distribution (m/d) of flow model.

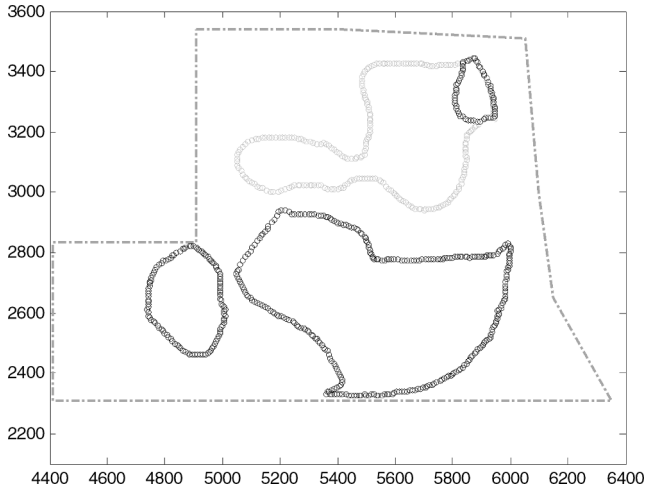

Fig. 5. Starting locations of particles for the advective control scheme. The CHC plumes and BTEX plume are delineated by 472 particles (black symbols), and 259 particles (gray symbols), respectively. The dashed line indicates the plant perimeter.

control scheme is enhanced to differentiate the plumes as well as the particles with respect to the contaminant group. The extension of the plumes is approximated based on the isolines of certain concentration levels, which are selected such that the main area of contamination is included (see shaded areas in Fig. 3b–d). The three CHC plumes are delineated by 472 particles, and the BTEX plume by 259 particles (Fig. 5). Due to the overlap of the CHC and the BTEX plumes in the northern part of the plant, a perfectly disjoined capture is not possible, and this is taken into account in the setup of the optimization framework (as discussed in Section 3.1). The result of the advective transport simulation of the particles from the plume boundary to proposed pumping wells is the key input for assessing the appropriateness of candidate remediation designs during the optimization, as discussed in Section 3.

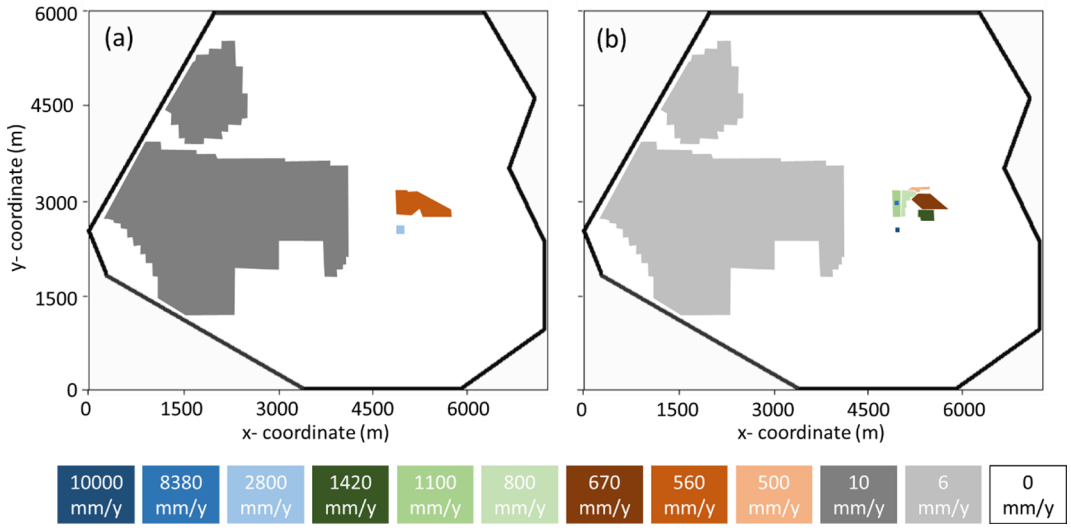

Fig. 6. Spatial distribution and values of recharge rate (mm/y) for (a) Models 1, 3, and 4, and (b) Model 2.

2.3. Setup and calibration of alternative flow models

When developing the flow model for the site, it has become evident that several elements of the model's conceptualization, such as hydraulic conductivity zones, recharge zones and lateral boundary conditions, as well as the thickness of aquifer layers, are considerably uncertain. Alternative models were developed to address this conceptual model uncertainty. Note that parametric uncertainty is related to model parameters that can be continuous or discrete variables, while conceptual uncertainty refers to different mathematical and geological structures, and alternative methods and conceptualizations (Elshall et al., 2020).

In Model 1, as shown in Fig. 6a, there are two recharge zones in the area of the plant with recharge rates of 560 mm/yr and 2800 mm/yr, respectively. The high recharge accounts for water input from artificial sources (e.g., from broken water pipes). In the rest of the model area, the recharge rate is set to 10 mm/yr for irrigated agricultural land, and set to zero for non-agricultural land. In Model 2, the water table mounds (Fig. 3a) are further explained by an increased heterogeneity of the recharge, by adding more recharge zones (Fig. 6b). This is accompanied by reducing the heterogeneity of hydraulic conductivity (i.e. reducing the number of hydraulic conductivity zones in comparison to Model 1 as shown in Fig. 7a and b). The calibrated values of recharge rate and hydraulic conductivity are shown in Figs. 6b and 7b, respectively.

Model 3 considers the uncertainty in geological structure with respect to the elevation of the top of layer 2, which is only known within the plant perimeter, as no drilling were made outside. The elevation is lowered at the western general head boundary by approximately 40% (the thickness of layer 1 is increased accordingly). The thickness of the layers in the model domain is modified proportionally as shown in Fig. 8. In Model 4 the hydraulic head of the general head boundary at the western boundary is reduced in Model 4 by 33%. In this model, the significance of the uncertainty associated with the water table depth is reflected. This uncertainty exists as there is no available data outside the plant. For Models 3 and 4, the hydraulic conductivity of each hydraulic conductivity zone is calibrated, and the calibrated values are shown in Fig. 7c and d.

Table 2 summarizes the differences between the four conceptual models, calibrated parameters, simulated inflow rates from recharge and across the general head boundary, and root mean square error (RMSE) normalized by average head. The simulated rates of inflow (via recharge and laterally) reflect the differences of the alternative models. For example, the recharge of Model 2 is larger than those in the other models, and the inflow from the general head boundary is the smallest for Model 4 among the four models. The increased conductivity in the north-western zone in Model 3 (25.1 m/d) compared to the other models might be attributed to the reduction of the layer thickness. The normalized RMSE indicates that the average error in the simulated heads is about 10%–15%. With respect to the description of hydraulic heads within the plant perimeter, Models 3 and 4 are the best models with respect to normalized RMSE, followed by Model 2 and finally Model 1.

3. Optimization scheme for multiple-well pump-and-treat designs

3.1. Objective function for a single model

The remediation design here is a well setting with multiple pumping wells. Each pumping well is characterized by three design variables: x and y coordinates and pumping rate. In its simplest version, the objective of optimizing multiple well settings for a pump-and-treat measure is to find a particular well setting that satisfies the constraints (e.g., ensuring capture of existing plume(s) at a minimum total pumping rate). This is a constraint optimization problem, which we transfer into an unconstraint problem using a penalty function. The corresponding objective function is

$$\min_{Q_i,x_i,y_i} f = \phi \sum_{i=1}^{n_{w,\max}} Q_i(x_i, y_i) \tag{1}$$

where ϕ is the penalty function that integrates one or more constraints into the objective function, $n_{w,\max}$ is the maximum number of the pumping wells, and x_i, y_i, and Q_i are the coordinates, and the pumping rate of each pumping well, respectively. The penalty function is formulated such that $\phi = 1$ if constraints are perfectly met, and $\phi > 1$ otherwise. In this section, we first present the basic penalty function that ensures capturing of existing plume(s), and extend this basic function to consider more constraints as follows: ensuring well placement within accessible placement areas, containing the capture zone within a given area, and disjoined capturing of plumes.

3.1.1. Capturing existing plumes

The basic penalty function for capture of existing plume(s) is

$$\phi = \left(1 + w_{cap}\left(1 - \sum_{i=1}^{n_{w,\max}} R_{cap,i}\right)\right) \tag{2}$$

wherew_{cap} is a weighting factor and $R_{cap,i}$ is the ratio of particles

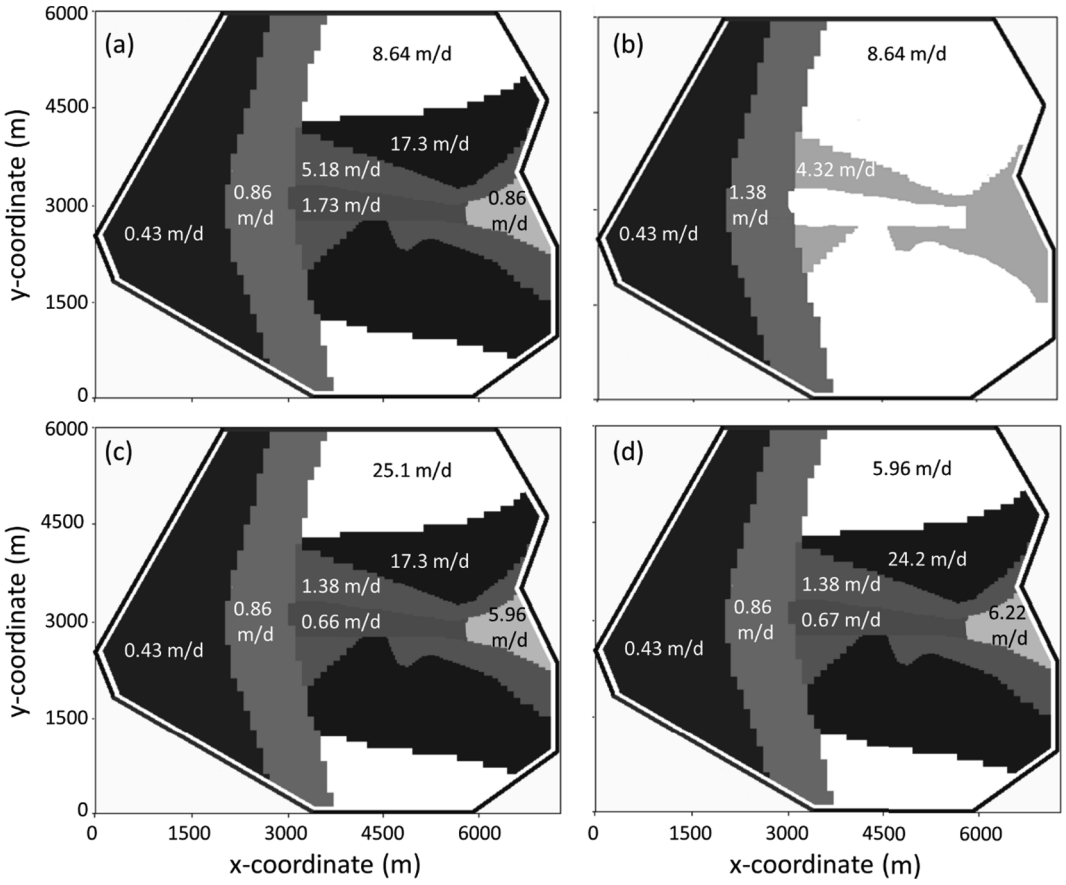

Fig. 7. Spatial distribution and values of hydraulic conductivity (m/d) for (a) Model 1, (b) Model 2, (c) Model 3, and (d) Model 4.

captured by well i. The capture ratio $R_{cap,i}$ is defined as

$$R_{cap,i} = \frac{n_{p,captured\ by\ well\ i}}{n_{p,tot}} \tag{3}$$

where $n_{p,captured\ by\ well\ i}$ is the number of particles captured by well i, and $n_{p,tot}$ is the total number of particles delineating the contaminant plume (s). The values of $R_{cap,i}$ are determined by running a forward particle tracking model based on a groundwater flow model, and a sufficient number of particles should be used to represent the spatial extent of existing contaminant plume(s). The summation of the ratios of all the captured particles $\sum_{i=1}^{n_{w,\max}} R_{cap,i}$ for the total number of wells $n_{w,\max}$ indicates the achieved degree of hydraulic plume containment. A full containment is achieved, if all particles delineating the contaminant plume(s) are captured, i.e. $\sum_{i=1}^{n_{w,\max}} R_{cap,i} = 1$ and $\phi = 1$.

3.1.2. Ensuring well placement within accessible placement areas

For organizational and legal reasons, wells are typically placed inside the plant perimeter. The restriction of the pumping well locations to certain placement areas extends the penalty function to

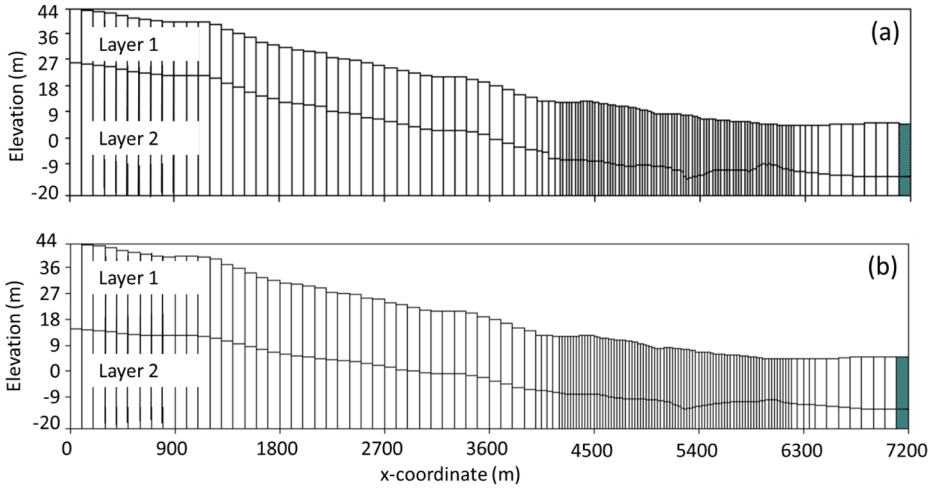

Fig. 8. Vertical cross-section comparing the layers' geometry of (a) Model 1 and (b) Model 3 along one row of the grid. In Model 3, the top of Layer 2 is lowered outside the plant perimeter.

Table 2
Summary of the conceptual differences, calibration parameters, simulated inflow rates from recharge and across the general head boundary, and root mean square error (RMSE) normalized by average hydraulic head.

	Model 1	Model 2	Model 3	Model 4
Conductivity zones (K-zone)	Reference	*Reduced*	Reference	Reference
Recharge zones (R-zone)	Reference	*Increased*	Reference	Reference
Thickness of model layers	Reference	Reference	*Reduced in layer 2, and increased in layer 1*	*Same as Model 3*
General head	Reference	Reference	Reference	*Reduced by 33%*
Calibration parameters	7 K-zones	4 K-zones, and 8 R-zones	7 K-zones	7 K-zones
Recharge (m^3/day)	747	807	755	755
Lateral inflow W' boundary (m^3/day)	80	82	181	23
Normalized RMSE [%]	14.5	11.3	9.8	9.8

$$\phi = \left(1 + w_{cap}\left(1 - \sum_{i=1}^{n_{w,\max}} R_{cap,i}\right)\right)(1 + w_{out}R_{out}) \tag{4}$$

where w_{out} is a weighting factor, andR_{out}

$$R_{out} = \frac{n_{w,outside}}{n_{w,\max}} \tag{5}$$

is the ratio of the number, $n_{w,outside}$, of wells located outside the targeted well placement area with the total number of pumping wells, $n_{w,\max}$. The R_{out} term can be calculated by comparing the indices of model cells (or nodes) representing the pumping wells with the list of cell indices representing the well placement area. We penalize those well settings that include wells located outside of the plant perimeter. We use a multiplicative penalty function because it is known to be more robust than an additive penalty function (Chan Hilton and Culver, 2005).

3.1.3. Containing the capture zone within a given area

Another necessary objective is to hydraulically capture the existing plumes such that contaminants propagate only within a given operation area, e.g. to avoid any legal or financial liability. For the advective control scheme applied in this study, this means that the pathlines of all particles from their starting positions to the pumping wells should be kept within the plant perimeter (used here as the operation area). The cell-wise travel time data in the MODPATH output is used to evaluate the pathline of each particle with respect to the travel time inside ($t_{in,i}$) and outside ($t_{out,i}$) of the plant perimeter. The goal is to achieve $t_{out,i} = 0$ for each particle, and that all the particles travel exclusively within the operation area. An overall time ratio R_{time} can be determined to quantitatively measure the proportion of the plume that undesirably evolves into outlying areas

$$R_{time} = \frac{\sum_{i=1}^{n_{p,tot}} t_{out,i}}{\sum_{i=1}^{n_{p,tot}} t_{in,i}} \tag{6}$$

This information can be incorporated into the penalty function as one further multiplier

$$\phi = \left(1 + w_{cap}\left(1 - \sum_{i=1}^{n_{w,\max}} R_{cap,i}\right)\right)(1 + w_{out}R_{out})(1 + R_{time}^{w_{time}}) \tag{7}$$

where the weighting factor w_{time} is a penalty term for R_{time}.

3.1.4. Disjoined capturing of plumes

A fourth constraint refers to the disjoined capture of existing contaminant plumes, such that plumes containing different groups of contaminants are not captured together. This allows for a separate and specific treatment of the contaminant groups, which may increase the treatment efficiency, thus reducing the overall treatment cost. Ideally, each of the wells shall pump water originating exclusively from one single type of plume (here either BTEX or CHC). To meet this objective, the capturing of water stemming from more than one type of plume must be penalized. This is done defining a cross-capture ratio, CC_i, for each pumping well as

$$CC_i = \frac{\min_j R_{cap,i,j}}{\max_j R_{cap,i,j}} \tag{8}$$

where $R_{cap,i,j}$is the plume-type-specific capture ratio of well i with respect to plume type j. It is defined as

$$R_{cap,i,j} = \frac{n_{p\ of\ type\ j,\ captured\ by\ well\ i}}{n_{p\ of\ type\ j,\ total}} \quad \begin{matrix} \forall\, i \in [1.\,..n_{well}] \\ \forall\, j \in [1.\,..n_{plume}] \end{matrix} \tag{9}$$

where $n_{p\ of\ type\ j,\ captured\ by\ well\ i}$ is the number of particles of plume type j captured at well i. The sum of the cross-capture ratio CC_i of all wells, multiplied by a weighting factor, w_{CC}, constitutes the penalty term for disjoined capture as follows

$$\phi = \left(1 + w_{cap}\left(1 - \sum_{i=1}^{n_{w,\max}} R_{cap,i}\right) + w_{CC}\sum_{i=1}^{n_{w,\max}} CC_i\right)(1 + w_{out}R_{out})(1 + R_{time}^{w_{time}}) \tag{10}$$

Small cross-capture ratios indicate that the respective well pumps water preferably from one contaminant group, and $\sum_{i=1}^{n_{w,\max}} CC_i = 0$ means that plumes of different contaminant groups are perfectly separated by the groundwater flow scheme produced by the pumping wells. Note that the formulation in Eq. (10) does not require any predefinition in terms of links between particular plumes and wells. Which and how many wells will be used to capture a certain plume will be determined as an inherent part of the search for optimal well settings. A mandatory requirement associated with Eq. (10) is that the particle tracking model must be able to distinguish between different groups of particles, to discriminate between plumes consisting of different contaminant groups (e.g., as shown in Fig. 5).

3.1.5. Weighting factors

In this study, the weighting factors in Eq. (10) are determined as $w_{cap} = 100$, $w_{CC} = 10$, $w_{out} = 2$, and $w_{time} = 1.2$. These values are estimated in a series of preliminary optimization runs to appropriately weigh in the sensitivity required to guide the search for the optimal solutions (i.e. well settings) that meet all constraints. Substituting these weighting factors into Eq. (10) and then into Eq. (1) leads to the objective function for any model k as

$$f_k = \left[\left(1 + 100\left(1 - \sum_{i=1}^{n_{w,\max}} R_{cap,i}\right) + 10\sum_{i=1}^{n_{w,\max}} CC_i\right)(1 + 2R_{out})(1 + R_{time}^{1.2})\right]\sum_{i=1}^{n_{w,\max}} Q_i(x_i, y_i) \tag{11}$$

The solution fitness in the objective function value.

3.2. Multi-model simulation–optimization approaches

To determine the optimal well setting for multiple models, both the cross-validating optimal designs (CVOD), and the concurring ensemble members (CEM) approach may be used.

3.2.1. Cross-validating optimal designs (CVOD)

CVOD investigates whether a certain well setting that has been identified as optimal in one model can achieve the remediation objectives in the other models. Since the optimization is conducted for individual models, the objective function of CVOD is the same as that defined in Eq. (11) for a single model, i.e.

$$f_{CVOD,k} = f_k \quad \forall \quad k \in [1, n_{model}] \tag{12}$$

where n_{model} is the number of alternative models.

The idea of CVOD is to cross-validate the feasibility of a well setting, which has been optimized for one model, with respect to its performance for the other models. This can identify well setting solutions with high reliability, and to determine if one or more critical models exist within the multi-model ensemble. A conceptual model is considered 'critical' if the optimal solution that has been found for the model does not fail to meet the solution objectives when it is cross-validated, i.e. being applied to other conceptual models, and if the optimal solutions based on the other conceptual models may fail when applied to the critical model. The CVOD approach follows these steps:

1. Cross-validate the solution of each of the repeated simulation–optimization runs of one model on the other models (e.g., that results in 20 fitness values for each model given four alternative models, and five repeated runs).
2. For each solution, sum-up the fitness values of the four models.
3. Rank the cumulative fitness values.
4. Use the ranking results to determine if one or more critical models exist.
5. While a critical model is identified here for the given multi-model ensemble, in other situations (i.e. for other sites) this may not be the case for the initial set of models. Optional next steps are:
   - develop further conceptual models to identify the critical conceptual model, and start again from step 1, or
   - proceed to step 6
6. Use the ranking results to select top solutions and analyze their technical qualification. Irrespective of whether a critical model could be found, this analysis is useful to identify the models that have high impact on the solution reliability.

Note that several studies use similar ranking procedures (Bayer et al., 2010, 2008; Kourakos and Mantoglou, 2008; Mantoglou and Kourakos, 2007; Paly et al., 2013) for identifying critical realizations in a single model ensembles generated by parameter perturbation. Kourakos and Mantoglou (2008) defines critical realizations as those realizations that have the most impact of the solution given the required reliability level.

3.2.2. Concurring ensemble members (CEM)

The goal of CEM is to identify the best well setting based on the performance in each of the alternative models. Accordingly, the aggregated objective function value of CEM is the weighted sum of all individual models' objective function values

$$f_{CEM} = n_{model} \sum_{k}^{n_{model}} w_k f_k \tag{13}$$

where w_k is the model weight.

Model weights may be set on the basis of model performance metrics using maximum likelihood estimation (e.g. Neuman, 2003; Ye et al., 2005a, 2005b), generalized likelihood uncertainty estimation (Rojas et al., 2010c, 2010b, 2010a, 2008), marginal likelihood estimation (Elshall and Ye, 2019; Liu et al., 2016a, 2016b), scoring rules (Elshall et al., 2018), among other methods as discussed by Rojas et al. (2008). The performance-based estimation of weights may either follow a calibration principle (e.g., BMA with MLE, GLUE or marginal likelihood estimation) such that models that reproduce calibration data better will have higher weights, or a prediction principle (e.g., model combination using scoring rules) such that models that make better predictions based on cross-validation data will have higher model weights. Multi-model aggregation using performance-based weights is not necessarily superior compared to the simple ensemble average using equal model weights (Rojas et al., 2012). Assuming equal model weights ($w_i = 1/n_{\text{mod } el} \; \forall \; i$) Eq. (13) simplifies to

$$f_{CEM} = \sum_{k}^{n_{model}} f_k \tag{14}$$

For the objective of optimization in this study, which is the search for a robust remediation design, assuming equal model weights is an appropriate and conservative choice. By applying equal weights, it is ensured that the critical model has the same weight as the other models. If we would apply some performance-based methods (such as BMA) the critical model might not necessarily have a considerable weight. Using equal model weights as a conservative choice is further discussed in previous studies, e.g., by Rojas et al. (2012).

3.3. Optimization method

The objective function (Eq. (12) or Eq. (14)) is minimized by using the covariance matrix adaptation evolution strategy (CMA-ES, Hansen et al., 2003; Hansen and Ostermeier, 2001). CMA-ES is an evolutionary algorithm that shows robust performance in terms of computational efficiency and search capacity for groundwater remediation problems (Bayer et al., 2010, 2008; Bayer and Finkel, 2007, 2004; Bürger et al., 2007; Chitsazan et al., 2015). Inspired by the evolutionary theory, CMA-ES works on the basis of populations. Each population is composed of a number λ of solutions each with n-dimensional decision variables. A solution in this context is a design (i.e. well setting). For example, a well setting with 9 pumping wells, in which the well coordinates (i.e., $\boldsymbol{x}$ and $\boldsymbol{y}$) and pumping rates (i.e., $\boldsymbol{Q}$) are the decision variables, is a 27-dimensional problem. This number of 9 pumping wells is a reasonable choice to satisfy the optimization objective, while maintaining a reasonably well-posed problem with an affordable computational cost.

To search for the optimal solution, CMA-ES randomly generates an initial population. Each solution in the population is evaluated in terms of its fitness (i.e., objective function value). In a so-called selection process, the best solutions (or, more generally, the information they carry) are retained. To explore the search space, a new population is created by a recombination of the information of different selected solutions in a so-called mutation step, which is a probabilistic change of the selected solutions or their combinations (Bayer et al., 2008). New populations are iteratively created and evaluated during the search for an optimal solution, until a stopping criterion is reached. The best solution (i.e., the most optimal solution) is the individual solution having the lowest fitness.

The default stopping criterion for the CMA-ES optimization is a fixed number of iterations. Multiplying the iteration size with the total number of iterations gives the total number of model runs. The default iteration size $\lambda = 4 + \lfloor 3\ln(n) \rfloor$ is a function of the number, n, of decision variables. In this case, the number of decision variables of $n = 27$ gives an iteration size of $\lambda = 13$. The number of iterations is not known a priori, and requires several trails to be initially performed, to avoid premature termination, and additional model runs with no significant variations in the objective function. For these purposes, we use 308 iterations with the default iteration size $\lambda = 13$ resulting in approximately 4000 model runs.

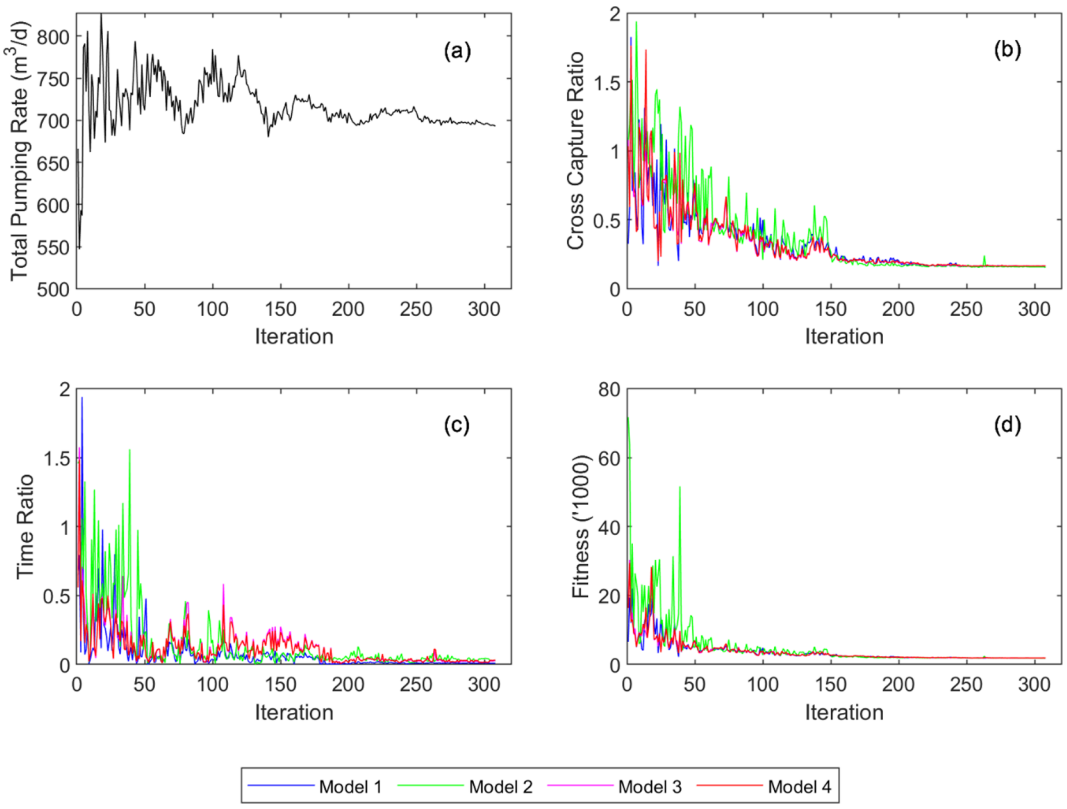

Fig. 9. Search progress in a CEM simulation–optimization run, showing the convergence of (a) total pumping rate, (b) cross-capture ratio, (c) time ratio, and (d) objective function value (i.e., fitness) with all the constraints included. Note that in CEM, the total pumping rate (a) is model-independent.

4. Results and discussion

4.1. Simulation-optimization performance

The progress of a search for the optimal well setting is shown in Fig. 9 for the case of a CEM run, which is more computationally demanding than a CVOD run to satisfy, as one well setting must successfully perform on all the four models. Fig. 9 shows the convergence of total pumping rate (Fig. 9a), the cross-capture ratio quantifying the degree of disjoined capture (Fig. 9b), the time ratio to evaluate the containment of the capture zone within the plant perimeter (Fig. 9c), and the fitness (Fig. 9d). A very similar search progress with convergence for the decision variables (e.g. the pumping rate), the penalty terms (e.g. cross-capture ratio and time ratio) of the objective function, and the fitness is observed for all other optimization runs carried out in this study (results not shown). To provide a reliable basis for analysis, five repeated runs were conducted because certain simulation–optimization runs (SOR) may converge to a better solution than other runs. This is due to the stochastic features of evolutionary algorithms, causing any randomly initialized optimization run to show a different search progress and results, i.e., values of decision variables and fitness (Bayer and Finkel, 2004).

Table 3 shows the simulation–optimization results of five repeated runs for each of the four models. The results indicate that the optimization gives well settings that succeeded in capturing all the particles, placing the pumping wells within the plant perimeter, and disjointly capturing the two plumes. The cross-capture ratio in most of the simulation–optimization runs is brought to its lower possible limit (i.e. the lowest limit ≈ 0.14 since the BTEX and CHC plumes overlap in the north east side as shown in Fig. 3).

To illustrate the impact of the constraint to limit the operation area of hydraulic capture (defined here along the plant perimeter), we compare the results of the optimization with and without this

Table 3
Performance parameter of optimal well settings in five repeated simulation–optimization runs (SOR) for each of the four conceptual models. The number of well placement outside the plant perimeter is zero for all the models given all the runs, and thus is not shown in Table 3.

Model	SOR	Total pumping rate (m^3/day)	Number of particles not captured	Cross-capture ratio (–)	Time ratio (–)	Mean particle travel time (year)	Fitness (–)
1	1	456	0	0.14	0.01	20	1111
	2	518	0	0.14	0.02	20	1232
	3	456	0	0.15	0.04	18	1143
	4	454	0	0.15	0.05	21	1184
	5	487	0	0.14	0.02	21	1159
2	1	619	0	0.28	0.02	11	2361
	2	614	0	0.17	0.02	17	1669
	3	654	0	0.17	0.01	25	1770
	4	635	0	0.18	0.00	17	1766
	5	572	0	0.17	0.03	12	1575
3	1	444	0	0.14	0.02	43	1070
	2	535	0	0.15	0.06	12	1399
	3	476	0	0.15	0.02	28	1181
	4	459	0	0.15	0.05	13	1169
	5	452	0	0.15	0.06	12	1173
4	1	437	0	0.14	0.01	19	1066
	2	447	0	0.14	0.04	15	1099
	3	452	0	0.15	0.03	13	1141
	4	489	0	0.15	0.04	19	1254
	5	451	0	0.14	0.02	14	1096
Mean		505	0	0.16	0.03	19	1331
Std		73	0	0.03	0.02	7	335
Minimum		437	0	0.14	0.00	11	1066
Maximum		654	0	0.28	0.06	43	2361

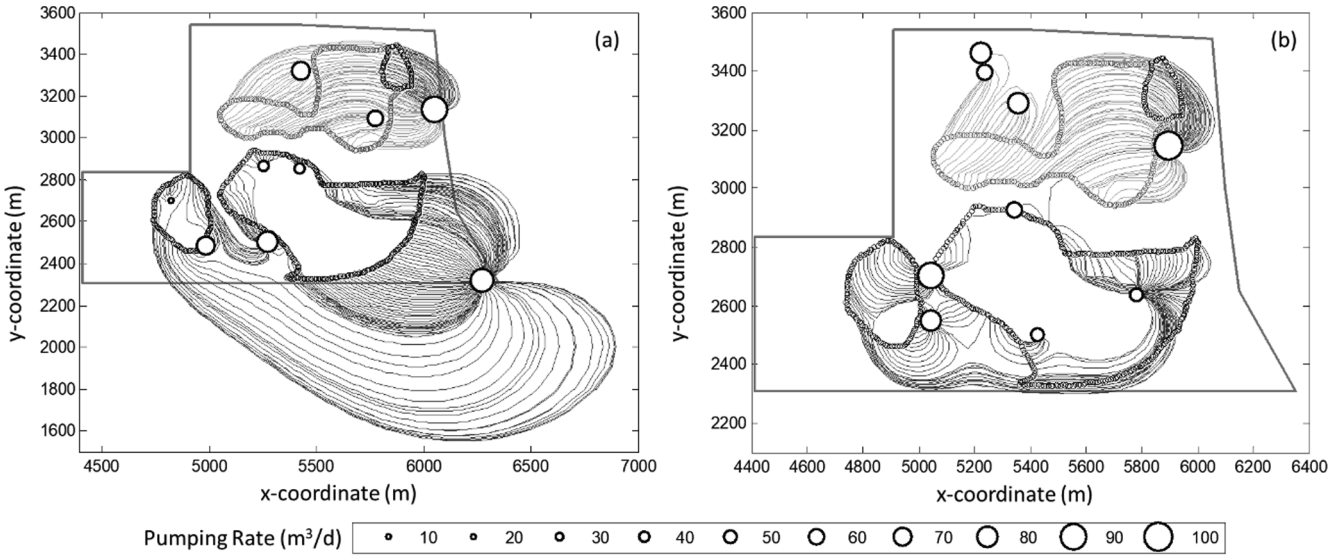

Fig. 10. Optimized well settings (a) not including and (b) including the operation area penalty term to constrain the capture zone to the plant perimeter (shown for Model 2). The well locations are represented by hollow circles with the circle area corresponding to pumping rate (m^3/day).

Table 4
Result of cross-validation: performance parameter of the optimal well settings identified using Model 2, Model 3, and Model 4 when being applied to Model 1. Similar results for cross-validation applied to Model 2, Model 3, and Model 4 are shown in the supplement.

	SOR	Total pumping rate (m^3/d)	Number of particles not captured	Cross-capture ratio (–)	Time ratio (–)	Mean particle travel time (year)	Fitness (–)
Well Setting of Model 1	1	456	0	0.14	0.01	20	1111
	2	518	0	0.14	0.02	20	1232
	3	456	0	0.15	0.04	18	1143
	4	454	0	0.15	0.05	21	1184
	5	487	0	0.14	0.02	21	1159
Apply Well Setting of Model 2 to Model 1	1	619	0	0.60	0.03	19	4402
	2	614	0	0.26	0.18	14	2496
	3	654	0	0.42	0.01	18	3418
	4	635	0	0.31	0.09	23	2729
	5	572	0	0.52	0.08	14	3710
Apply Well Setting of Model 3 to Model 1	1	444	162	0.59	2.22	96	46,508
	2	535	0	0.81	0.18	18	5500
	3	476	110	0.09	2.11	91	27,802
	4	459	63	0.12	1.94	44	15,895
	5	452	52	0.11	2.09	38	14,278
Apply Well Setting of Model 4 to Model 1	1	437	162	0.08	3.06	71	50,487
	2	447	56	0.11	0.96	39	8456
	3	452	20	0.14	0.57	27	3518
	4	489	39	0.13	1.03	51	7570
	5	451	61	0.10	1.16	40	10,235
Mean		505	36.25	0.25	0.79	35	10,642
Std		73	53	0.21	0.97	25	14,542
Minimum		437	0	0.08	0.01	14	1111
Maximum		654	162	0.81	3.06	96	50,487

constraint. This constraint is implemented by means of the time ratio of the particle travel time inside and outside the plant perimeter (Eq. (6)). For the purpose of comparison, we conducted five simulation–optimization runs for all the models without including the time ratio constraint in the objective function by setting $R_{time} = 0$ in Eq. (13); results are shown in the Supplement. For this case of no time ratio constraint, an optimal well setting (for the objectives of capturing all the particles, placing all individual pumping wells within the perimeter of the plant, and establishing a disjoined capture of the two types of plumes) is shown in Fig. 10a. By adding the penalty term for the time-ratio constraint in the objective function, the search for optimal solutions converges to well settings with capture zones entirely located within the plant perimeter (Fig. 10b). Accounting for time ratio increases pumping rate from 480 ± 61 m^3/d to 505 ± 73 m^3/d given all runs and models. Apart from increasing the pumping rate, adding the time ratio as an additional constraint does not result in a trade-off with respect to the other objectives. The cross-capture ratios for the cases of not accounting for and accounting for time ratio are similar with mean values for the two cases of 0.16 ± 0.05 and 0.16 ± 0.03, respectively, given all runs and models.

However, accounting for the time-ratio penalty term has significantly succeeded in containing the capture zone to the plant perimeter. One-way ANOVA is used to quantify the difference in time ratio between the two cases of not accounting and accounting for time ratio in the objective function. A P-value of 3×10^{-5} is obtained indicating a statistically significant difference between the two data sets. The results for the case of accounting for time ratio in the objective function (Table 3) indicates that the capture zone is successfully contained within the plant perimeter such that the mean time ratio is 0.03 ± 0.01 in comparison with a mean time ratio of 0.77 ± 0.33 for

Table 5
Summary of the cross-validation (CVOD): performance ranking of the 20 simulation–optimization runs (SOR) based on fitness values for all models.

Model-SOR	Fitness Model 1	Fitness Model 2	Fitness Model 3	Fitness Model 4	Σ Fitness (1000′)
2-3	3418	1770	4089	4065	13
2-5	3710	1575	4852	4777	15
2-1	4402	2361	5029	5067	17
2-2	2496	1669	13,440	5627	23
2-4	2729	1766	14,492	12,010	31
3-2	5500	23,427	1399	1486	32
1-4	1184	40,619	2402	3837	48
4-3	3518	45,245	1342	1141	51
1-3	1143	43,369	4698	3486	53
3-4	15,895	36,719	1169	1854	56
3-5	14,278	45,192	1173	1181	62
4-2	8456	55,832	1372	1099	67
4-5	10,235	63,047	1393	1096	76
1-5	1159	37,063	80,054	1472	120
4-4	7570	117,133	1511	1254	128
1-2	1232	115,036	6367	7587	130
4-1	50,487	152,773	1381	1066	206
3-3	27,802	200,765	1181	3437	233
3-1	46,508	207,582	1070	5982	261
1-1	1111	245,021	30,698	28,370	305
Σ Fitness (1000′)	213	1438	179	96	1926

the case of not including this penalty term. Table 3 further shows that some simulation–optimization runs (e.g. model 2 SOR 4) succeed in totally containing the capture zone within the plant perimeter such that the particle travel time outside the plant perimeter is almost zero. On the other hand, some simulation–optimization runs without time ratio penalty have failed to contain the capture zone within the plant perimeter such that the particle travel time outside the plant perimeter can be as three times higher than inside the plant perimeter. In conclusion, the objective function that incorporates time ratio as an additional constraint is successfully formulated. This shows that the concept of using the time ratio of the particles travel time outside and inside the plant perimeter is a valid approximation for the particle locations (i.e. travel path inside and outside the plant perimeter). For the comparison of CVOD and CEM, the objective function with time ratio is used.

In summary, Table 3 shows that almost all proposed 20 well settings are successful in terms of meeting the optimization objective (with the only exception being SOR1 for Model 2 that has relatively high cross-capture ratio). The fitness values of the 20 simulation–optimization runs have a wide range from 1070 to 2360. While the statistical mean for the fitness values for models 1, 3 and 4 is 1150 ± 59, Model 2 has a mean of 1828 ± 383. This suggests that model 2 is more challenging. Given the variability in the performance of the well settings given the four alternative models, it is an obvious next step to evaluate the well settings for their reliability. To this end, we use two different approaches, the cross-validating optimal designs (CVOD) and the concurring ensemble members (CEM) approach.

4.2. Cross-validating optimal designs (CVOD)

According to CVOD, each of the 20 well settings (Table 3) is evaluated for its performance on the other three models. Table 4 lists the performance results of the well settings optimized for Models 2–4 when applied to Model 1. Similar cross-validation results for the other three models are shown in the Supplement. The results shown in Table 4 indicate that the optimal well settings of Models 3 and 4 fail to perform on Model 1 with respect of capturing the plumes for all the SORs. The only exception is Model 3 – SOR2, which still have higher fitness than all the SORs of Model 2. The results show that Model 3 has high risk of failure since only one out of five repeated runs did not fail to capture the plumes. The performance of the optimal well setting for Model 2 when being applied to Model 1 is comparatively consistent in all SORs. The performance is particularly good with respect to capturing the particles, and in containing the capture zone within the plant perimeter. However, performance with respect to cross-capture ratio is relatively poor. In summary, results in Table 4 demonstrate that well settings that are optimized for one conceptual model can fail in other models. This might be due to the impact of model specific hydraulic and geological conditions that are difficult to satisfy, if the solution is not optimized for this model. This suggests that multiple models should be considered during optimization, as proposed in the CEM approach, to obtain robust remediation designs.

Table 6
Performance parameter of the five best well settings identified by CVOD using Model 2 in different simulation-optimization runs (SOR) when applied to the four alternative models.

Model-SOR	Post-processed on Model#	Total pumping rate (m^3/d)	Number of particles not captured	Cross-capture ratio (–)	Time ratio (–)	Mean particle travel time (year)	Fitness	Σ Fitness ('1000)
2-3	1	654	0	0.42	0.01	18	3418	13.3
2-3	2	654	0	0.17	0.01	25	1770	
2-3	3	654	0	0.52	0.01	12	4089	
2-3	4	654	0	0.52	0.01	12	4065	
2-5	1	572	0	0.52	0.08	14	3710	14.9
2-5	2	572	0	0.17	0.03	12	1575	
2-5	3	572	0	0.66	0.16	9	4852	
2-5	4	572	0	0.66	0.15	10	4777	
2-1	1	619	0	0.60	0.03	19	4402	16.9
2-1	2	619	0	0.28	0.02	11	2361	
2-1	3	619	0	0.67	0.09	11	5029	
2-1	4	619	0	0.67	0.09	11	5067	
2-2	1	614	0	0.26	0.18	14	2496	23.2
2-2	2	614	0	0.17	0.02	17	1669	
2-2	3	614	6	0.45	2.13	38	13,440	
2-2	4	614	0	0.46	0.69	20	5627	
2-4	1	635	0	0.31	0.09	23	2729	31.0
2-4	2	635	0	0.18	0.00	17	1766	
2-4	3	635	30	0.30	1.65	24	14,492	
2-4	4	635	25	0.31	1.42	22	12,010	
Mean	619	3.05	0.41	0.35	17	4967	19.9	
Std	28	9	0.18	0.63	7	3834	6.5	
Minimum	572	0	0.17	0.00	9	1575	13.3	
Maximum	654	30	0.67	2.13	38	14,492	31.0	

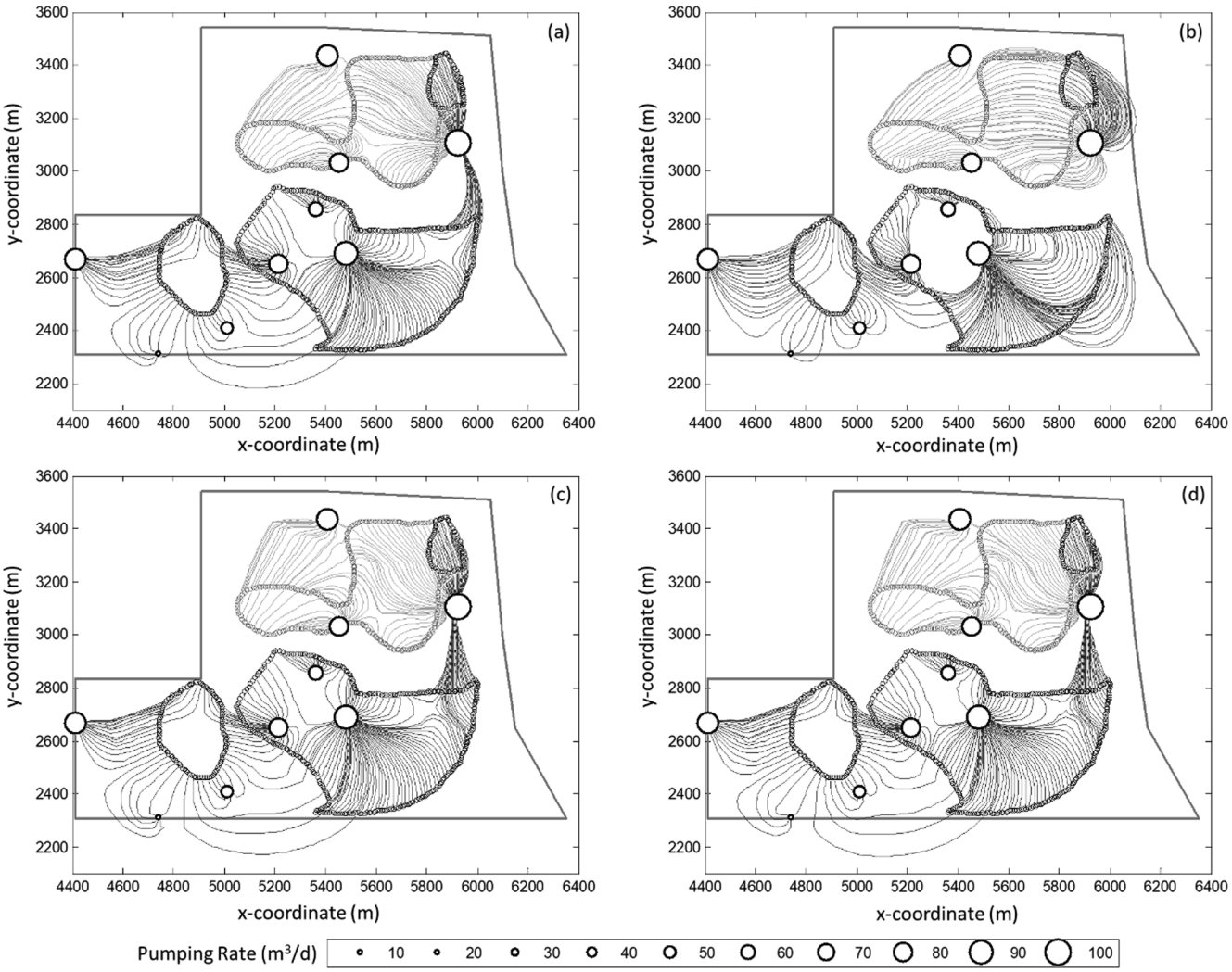

Fig. 11. Results of CVOD: performance of optimal well setting identified in SOR3 of Model 2 in (a) Model 1, (b) Model 2, (c) Model 3, and (d) Model 4. The well locations are represented by hollow circles with the circle area corresponding to pumping rate (m^3/d).

The next steps of the CVOD approach are to rank the well settings, and to determine the best ones, which perform most successfully on multiple models. The ranking is based on the fitness values of all cross-validations (Table 5). Top-ranked are the well settings of Model 2, which perform comparatively well on the other models resulting in relatively low total fitness value per single SOR, as shown in the last column. This is due to the rather challenging character of Model 2 as has been observed previously (i.e. fitness values for models 1, 3 and 4 is 1150 ± 59 and for Model 2 is 1828 ± 383 as shown in Table 3), which renders the performance of well settings optimized for the other models very poor on Model 2. Following the same analogy, since Model 4 is the least challenging, only one optimal well setting of its five simulation–optimization runs is among the top ten well settings. These results show that the optimization for Model 2 yields well settings with high reliability in comparison to the other three models.

As a final step, the technical feasibility of the top ranked well settings needs to be assessed. The results for the five best well settings are shown in Table 6. The best well setting (i.e., outcome of SOR3 for Model2) succeeds in containing the capture zone within the plant perimeter as indicated from the time ratio being as low as 0.01 for all models. Yet, as illustrated in Fig. 11, this well setting fails to achieve a disjoined capture of the two plumes, as indicated by the cross-capture ratio, which is as high as 0.5 for some models. The second best well setting (Model 2 – SOR5) does not only fail to disjointly capture the two plumes, but also in effectively containing the capture zone within the plant perimeter (see the high time ratio in Table 7). These two examples demonstrate that CVOD does produce well settings that meet the objectives set in this case study only for the model used in the optimization, but not for other alternative models. This might be attributed not only to the large differences in the alternative models, but also to the complexity and ambitiousness of the objectives. In other studies, CVOD might be the approach of choice, depending on the particular purpose and the given situation.

4.3. Concurring ensemble members (CEM)

In CEM, the simulation–optimization is based on a performance evaluation that considers all four models. Therefore, in contrast to CVOD, CEM succeeds in finding optimal well settings that are capable of hydraulically containing the contaminant plumes in each of the four models (Table 7). The five SORs show rather consistent results. The identified best well settings succeed in containing the capture zone within the plant perimeter for each of the four models. The average time ratio is 0.04 ± 0.04 (Table 7), which is distinctly lower than the average time ratio of the five top-ranked well settings identified using the CVOD (0.35 ± 0.63, see Table 6). The comparison of the results with respect to the cross-capture ratio shows a similar outcome; CEM (the cross-capture ratio of five best well settings being 0.2 ± 0.08) outperforms CVOD (0.41 ± 0.18) with respect to the well settings' capability of disjointly capturing the two plumes.

Table 7
Performance parameter of the five best well settings identified by CEM using all models in different simulation–optimization runs (SOR) when applied to the four alternative models.

SOR	Model	Total pumping rate (m^3/d)	Number of particles not captured	Cross-capture ratio (–)	Time ratio (–)	Particle mean travel time (year)	Fitness	Σ Fitness ('1000)
1	1	616	0	0.18	0.07	7	1828	7.6
	2	616	0	0.25	0.16	14	2415	
	3	616	0	0.16	0.08	11	1672	
	4	616	0	0.16	0.07	11	1657	
2	1	608	0	0.17	0.00	13	1625	8.4
	2	608	0	0.43	0.10	10	3450	
	3	608	0	0.17	0.00	9	1671	
	4	608	0	0.17	0.00	9	1669	
3	1	639	0	0.19	0.00	6	1829	7.9
	2	639	0	0.31	0.08	12	2730	
	3	639	0	0.16	0.03	9	1669	
	4	639	0	0.16	0.02	9	1657	
4	1	678	0	0.17	0.02	13	1827	8.8
	2	678	0	0.37	0.10	12	3354	
	3	678	0	0.17	0.01	12	1830	
	4	678	0	0.17	0.01	12	1832	
5	1	693	0	0.16	0.01	13	1837	7.3
	2	693	0	0.15	0.03	23	1781	
	3	693	0	0.16	0.03	9	1854	
	4	693	0	0.16	0.03	9	1864	
	Mean	647	0	0.20	0.04	11	2003	8.0
	Std	34	0	0.08	0.04	3	548	0.6
	Minimum	608	0	0.15	0.00	6	1625	7.3
	Maximum	693	0	0.43	0.16	23	3450	8.8

The good performance of the best well settings identified with CEM in each of the alternative models (see an illustration in Fig. 12 for the example of SOR5) demonstrates their robustness and reliability. Yet this robustness is at the expense of a higher pumping rate according to the mean values of best well settings; the mean total pumping rates of CVOD and CEM are 619 ± 28 m^3/d (Table 6) and 647 ± 34 m^3/d (Table 7), respectively. Fig. 13 gives a summary of the performance statistics of the best well settings obtained with CVOD and CEM.

4.4. Discussion

Both CVOD and CEM are useful approaches for multi-model simulation–optimization. CVOD has several practical implications. First, it provides multiple storylines about the problem of interest (Ferré, 2017a). Enemark et al. (2019b) note that current multi-model approaches require a shift in mindset towards acknowledging that multiple-narratives and unknown unknowns exist, rather than assuming that there is a single outcome of the modelling process. Enemark et al. (2019b) define the unknown unknowns as "the conceptual models that current data has not yet uncovered and will lead to conceptual surprises if they are". CVOD provides multiple narratives about the solution by assessing the design performance given "a reality" that is different from the conceptual model that is used to develop the design. Second, this cross-validation approach further facilitates a reliability analysis (Russell and Rabideau, 2000) that defines the probability and cost of failure to guide the selection between the different candidate solutions. Third, the concept of critical model adds a new alternative for model selection – in addition to the available model performance-based approaches (e.g. BMA with MLE, or marginal likelihood estimation) as previously discussed. Fourth, CVOD identifies the critical model in case there is a need to only use one model for a specific research or management purpose. For examples about the application of the concept of critical model in groundwater management, the reader is referred to Elshall et al. (2020). Fifth, the optimal design in any newly developed conceptual model, or BMA result can be cross-validated using the critical model to evaluate the reliability of the proposed remediation design. Finally, a point that is worth to be further investigated is the value of the critical model in data-worth analysis, which depends on the conditions that most significantly impact the final outcome.

On the other hand, CEM is a risk-averse approach for simulation–optimization-based groundwater management under model uncertainty. As Ranjithan et al. (1993) note, the design that performs well for many realizations (or conceptual models) is considered to be more reliable than one that performs well for only a few. However, the CEM-based designs require higher pumping rates, implying higher remediation cost. Similarly, using a chance-constrained BMA approach, Chitsazan et al. (2015) show that an increased injection rate is needed to avoid overestimating the design reliability when using a single model.

As no specific model weights are used, the presented approaches of CVOD and CEM do not necessarily require the working assumption that the developed conceptual models need to be mutually exclusive and collectively exhaustive (that is required by BMA for example). The multi-model ensemble considered in this study covers plausible conceptualizations of the site's hydrogeology and is − though being possibly under-sampled − suitable for demonstrating the multi-model-optimization approaches. It is beyond the scope of this work to evaluate the appropriateness of this ensemble of four alternative flow models with regard to the quality of uncertainty assessment (see, e.g., Rojas et al., 2010a, 2010b).

4.5. Outlook

The alternative flow models considered here represent the heterogeneity of the subsurface approximately at depositional environment scale (100–1000 m); see Fig. 2.9 of Rubin (2003) for a classification. Heterogeneities below that scale have not been accounted for in this study. Therefore, based on the results presented, we cannot ultimately rate the robustness of the demonstrated multi-model optimization approaches. Further studies are warranted to quantify the sensitivity of the approaches' performance to multimodal and unimodal heterogeneity at channel and stratigraphic features scale, respectively. These studies may also involve the consideration of synthetic test cases, and the evaluation of the remediation designs with systematic multi-model ensemble development (Enemark et al., 2019a; Tsai and Elshall, 2013) to account for other sources of uncertainty.

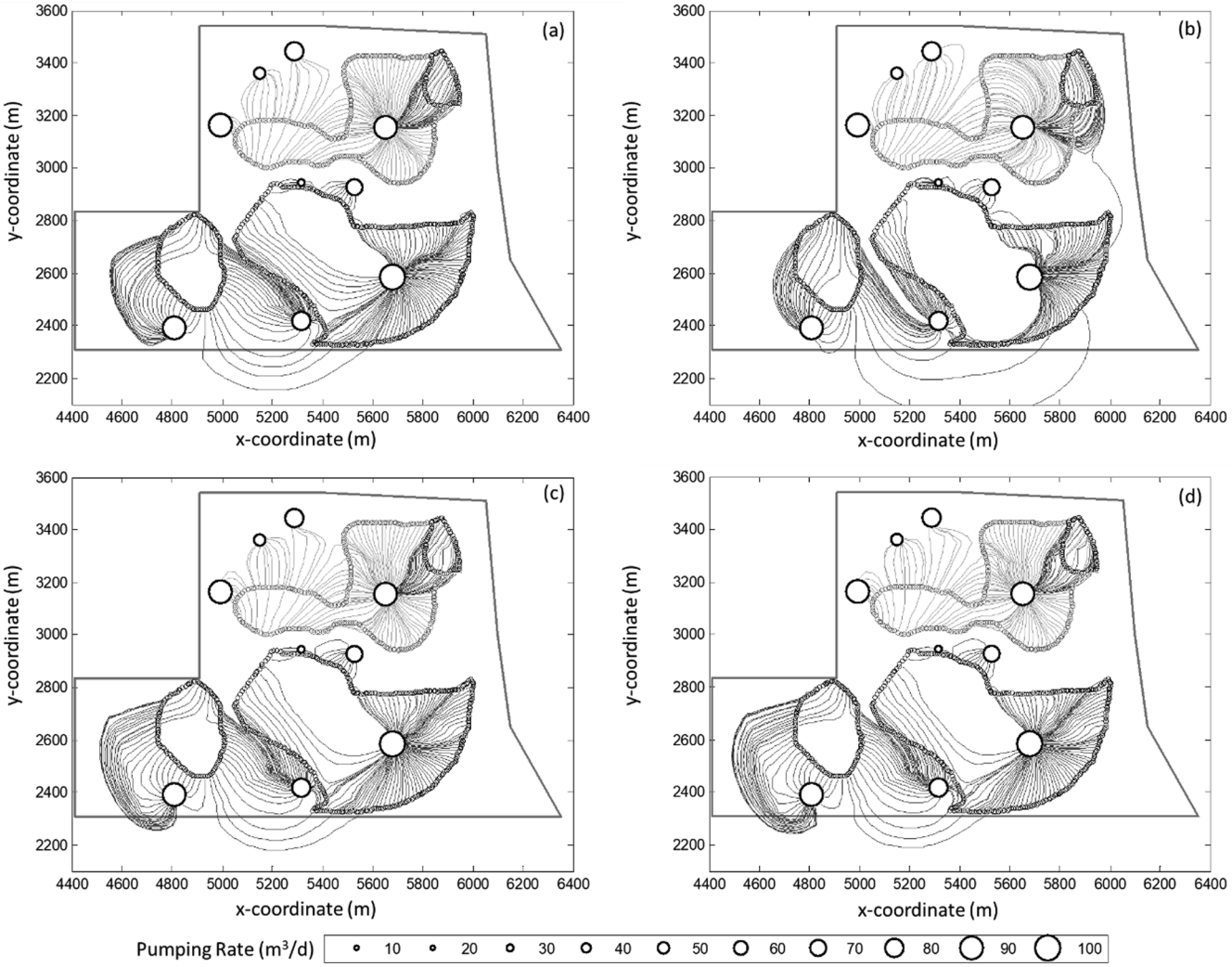

Fig. 12. Results of CEM: performance of optimal well setting identified in SOR5 in (a) Model 1, (b) Model 2, (c) Model 3, and (d) Model 4. The well locations are represented by hollow circles with the circle area corresponding to pumping rate (m^3/d).

The optimization presented here is focused on the hydraulic objectives of the pump-and-treat measure only. The objectives are defined to optimally control the core area of two contaminant plumes with multiple wells. To fully assess the technological and economical aspects of the remediation measure, both a solute transport model (for the prediction of contaminant concentrations in the pumped water) and an economical model (for calculation of pumping and treatment costs) need to be integrated in the simulation–optimization framework. The objective function can be modified accordingly (e.g. by replacing the pumping rate with the remediation costs). The integration of costs into the objective function will also allow to better assess the costs of reliability, including data-worth analysis that incorporates the cost of collecting new data.

5. Conclusions

This paper discusses two alternative approaches of identifying optimal remediation designs (i.e., well settings) to groundwater water quality control for contaminated sites under conceptual model uncertainty. Both approaches involve the evaluation of candidate remediation designs in multiple models. Following the cross-validating optimal designs (CVOD) approach, remediation designs are optimized in individual models and then cross-validated in other models. In the concurring ensemble members (CEM) approach, all models are considered for evaluation during optimization. We have implemented these concepts within a simulation–optimization framework to design a multiple-well pump-and-treat system for the disjoined capture of a contaminant plumes, using an advective control scheme based on particle tracking. The framework does not set any specific requirements to the models to be considered (e.g., whether and in what kind of detail and scale the models consider the heterogeneity of the subsurface, and the related parameter field of hydraulic conductivity), and to the comprehensiveness of the ensemble members. When new models emerge with new data/knowledge, they can be easily incorporated into the presented framework. Note that concerns related to the quality of the model ensemble are similar to other frameworks; by not developing all plausible models, we may be in for a conceptual surprise (Bredehoeft, 2005; Enemark et al., 2019b).

Conceptual model uncertainty is addressed by four alternative models representing different conceptual site models for the study area with respect to layer thickness, boundary conditions, recharge zones, and aquifer conductivity zones. The results for CVOD show that the remediation design based on one model may fail on other models. Also, a critical conceptual model is identified that – if used in the optimization – results in rather reliable well settings compared to the other models. The well settings that are optimized employing the CEM

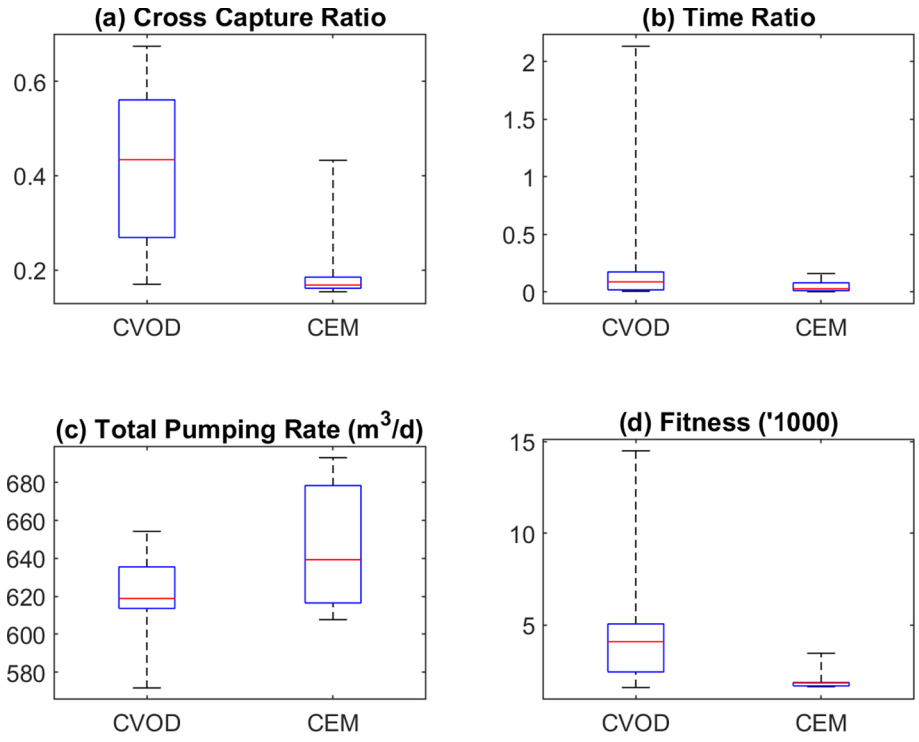

Fig. 13. The performance of CVOD runs (cmp. Table 6) and CEM runs (cmp. Table 7) with respect to (a) cross-capture ratio ($P = 3 \times 10^{-5}$), (b) time ratio ($P = 0.04$), (c) total pumping rate ($P = 0.007$), and (d) fitness ($P = 0.002$). Small P-values indicate that there is a significant difference between the means of the two data sets (Tables 6 and 7), with significance level of 0.05 for the P-values.

approach clearly outperform those obtained by CVOD in terms of robustness and reliability. The design objectives, namely the disjoined capture of the plumes and the restriction of the capture zone to the plant perimeter, are met in all models with the CEM approach, yet at a higher pumping rate.

CRediT authorship contribution statement

Ahmed S. Elshall: Methodology, Software, Investigation, Visualization, Validation, Writing - original draft, Writing - review & editing. **Ming Ye:** Validation, Writing - original draft, Writing - review & editing, Resources, Supervision, Project administration, Funding acquisition. **Michael Finkel:** Conceptualization, Data curation, Methodology, Software, Investigation, Visualization, Validation, Writing - original draft, Writing - review & editing, Resources, Supervision, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We thank two anonymous reviewers for their constructive comments that helped to improve the manuscript. This research was supported by NSF Grant EAR-1552329.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhydrol.2020.125427.

References

Aksoy, A., Culver, T.B., 2004. Impacts of physical and chemical heterogeneities on aquifer remediation design. J. Water Resour. Plann. Manage. 130, 311–320. https://doi.org/10.1061/(ASCE)0733-9496(2004)130:4(311).

Ashraf, B., AghaKouchak, A., Alizadeh, A., Baygi, M.M., Moftakhari, H.R., Mirchi, A., Anjileli, H., Madani, K., 2017. Quantifying anthropogenic stress on groundwater resources. Sci. Rep. 7. https://doi.org/10.1038/s41598-017-12877-4.

Ayvaz, M.T., 2016. A hybrid simulation-optimization approach for solving the areal groundwater pollution source identification problems. J. Hydrol. 538, 161–176. https://doi.org/10.1016/j.jhydrol.2016.04.008.

Ayvaz, M.T., Elçi, A., 2018. Identification of the optimum groundwater quality monitoring network using a genetic algorithm based optimization approach. J. Hydrol. 563, 1078–1091. https://doi.org/10.1016/j.jhydrol.2018.06.006.

Baú, D.A., Mayer, A.S., 2007. Data-worth analysis for multiobjective optimal design of pump-and-treat remediation systems. Adv. Water Resour. 30, 1815–1830. https://doi.org/10.1016/j.advwatres.2007.02.008.

Bayer, P., Bürger, C.M., Finkel, M., 2008. Computationally efficient stochastic optimization using multiple realizations. Adv. Water Resour. 31, 399–417. https://doi.org/10.1016/j.advwatres.2007.09.004.

Bayer, P., De Paly, M., Bürger, C.M., 2010. Optimization of high-reliability-based hydrological design problems by robust automatic sampling of critical model realizations. Water Resour. Res. 46. https://doi.org/10.1029/2009WR008081.

Bayer, P., Finkel, M., 2007. Optimization of concentration control by evolution strategies: formulation, application, and assessment of remedial solutions. Water Resour. Res. 43. https://doi.org/10.1029/2005WR004753.

Bayer, P., Finkel, M., 2004. Evolutionary algorithms for the optimization of advective control of contaminated aquifer zones. Water Resour. Res. 40. https://doi.org/10.1029/2003WR002675.

Beven, K.J., 2016. Facets of uncertainty: epistemic uncertainty, non-stationarity, likelihood, hypothesis testing, and communication. Hydrol. Sci. J. 61, 1652–1665. https://doi.org/10.1080/02626667.2015.1031761.

Bianchi Janetti, E., Guadagnini, L., Riva, M., Guadagnini, A., 2019. Global sensitivity analyses of multiple conceptual models with uncertain parameters driving groundwater flow in a regional-scale sedimentary aquifer. J. Hydrol. 574, 544–556. https://doi.org/10.1016/j.jhydrol.2019.04.035.

Bredehoeft, J., 2005. The conceptualization model problem – surprise. Hydrogeol. J. 13, 37–46. https://doi.org/10.1007/s10040-004-0430-5.

Bürger, C.M., Bayer, P., Finkel, M., 2007. Algorithmic funnel-and-gate system design optimization. Water Resour. Res. 43. https://doi.org/10.1029/2006WR005058.

Burnett, K.M., Elshall, A.S., Wada, C.A., Arik, A., El-Kadi, A., Voss, C.I., Delevaux, J.M.S., Bremer, L.L., 2020. Incorporating historical spring discharge protection into sustainable groundwater management: a case study from pearl Harbor Aquifer, Hawai'i. Front. Water 2, 1–12. https://doi.org/10.3389/frwa.2020.00014.

Chan Hilton, A.B., Culver, T.B., 2005. Groundwater remediation design under uncertainty using genetic algorithms. J. Water Resour. Plann. Manage. 131, 25–34. https://doi.org/10.1061/(ASCE)0733-9496(2005)131:1(25).

Chang, L.C., Chu, H.J., Hsiao, C.T., 2007. Optimal planning of a dynamic pump-treat-inject groundwater remediation system. J. Hydrol. 342, 295–304. https://doi.org/10.1016/j.jhydrol.2007.05.030.
Chitsazan, N., Pham, H.V., Tsai, F.T.C., 2015. Bayesian chance-constrained hydraulic barrier design under geological structure uncertainty. Groundwater 53, 908–919. https://doi.org/10.1111/gwat.12304.
Chitsazan, N., Tsai, F.T.C., 2015. Uncertainty segregation and comparative evaluation in groundwater remediation designs: a chance-constrained hierarchical Bayesian model averaging approach. J. Water Resour. Plann. Manage. 141. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000461.
Christensen, S., Zlotnik, V.A., Tartakovsky, D.M., 2009. Optimal design of pumping tests in leaky aquifers for stream depletion analysis. J. Hydrol. 375, 554–565. https://doi.org/10.1016/j.jhydrol.2009.07.006.
Cirpka, O.A., Bürger, C.M., Nowak, W., Finkel, M., 2004. Uncertainty and data worth analysis for the hydraulic design of funnel-and-gate systems in heterogeneous aquifers. Water Resour. Res. 40. https://doi.org/10.1029/2004WR003352.
Cousquer, Y., Pryet, A., Delbart, C., Valois, R., Dupuy, A., 2019. Adaptive optimization of a vulnerable well field. Hydrogeol. J. 27, 1673–1681. https://doi.org/10.1007/s10040-019-01963-8.
Davidsen, C., Liu, S., Mo, X., Holm, P.E., Trapp, S., Rosbjerg, D., Bauer-Gottwein, P., 2015. Hydroeconomic optimization of reservoir management under downstream water quality constraints. J. Hydrol. 529, 1679–1689. https://doi.org/10.1016/j.jhydrol.2015.08.018.
De Barros, F.P.J., Fernàndez-Garcia, D., Bolster, D., Sanchez-Vila, X., 2013. A risk-based probabilistic framework to estimate the endpoint of remediation: concentration rebound by rate-limited mass transfer. Water Resour. Res. 49, 1929–1942. https://doi.org/10.1002/wrcr.20171.
Delottier, H., Pryet, A., Dupuy, A., 2017. Why should practitioners be concerned about predictive uncertainty of groundwater management models? Water Resour. Manage. 31, 61–73. https://doi.org/10.1007/s11269-016-1508-2.
Doherty, J., Welter, D., 2010. A short exploration of structural noise. Water Resour. Res. 46. https://doi.org/10.1029/2009WR008377.
Elshall, A.S., Arik, A.D., El-Kadi, A.I., Pierce, S., Ye, M., Burnett, K.K., Wada, C., Bremer, L.L., Chun, G., 2020. Groundwater sustainability: a review of the interactions between science and policy. Environ. Res. Lett. https://doi.org/10.1088/1748-9326/ab8e8c.
Elshall, A.S., Pham, H.V., Tsai, F.-T.-C., Yan, L., Ye, M., 2015. Parallel inverse modeling and uncertainty quantification for computationally demanding groundwater-flow models using covariance matrix adaptation. J. Hydrol. Eng. 20, 04014087. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001126.
Elshall, A.S., Tsai, F.-T.-C., 2014. Constructive epistemic modeling of groundwater flow with geological structure and boundary condition uncertainty under the Bayesian paradigm. J. Hydrol. 517. https://doi.org/10.1016/j.jhydrol.2014.05.027.
Elshall, A.S., Ye, M., Niu, G.-Y., Barron-Gafford, G.A., 2019. Bayesian inference and predictive performance of soil respiration models in the presence of model discrepancy. Geosci. Model Dev. 12, 2009–2032. https://doi.org/10.5194/gmd-12-2009-2019.
Elshall, A.S., Ye, M., Pei, Y., Zhang, F., Niu, G.-Y.-G.-Y., Barron-Gafford, G.A., 2018. Relative model score: a scoring rule for evaluating ensemble simulations with application to microbial soil respiration modeling. Stoch. Environ. Res. Risk Assess. 32, 2809–2819. https://doi.org/10.1007/s00477-018-1592-3.
Elshall, Ye, 2019. Making steppingstones out of stumbling blocks: a Bayesian model evidence estimator with application to groundwater transport model selection. Water 11, 1579. https://doi.org/10.3390/w11081579.
Enemark, T., Peeters, L.J., Mallants, D., Batelaan, O., Valentine, A.P., Sambridge, M., Enemark, T., Peeters, L.J., Mallants, D., Batelaan, O., Valentine, A.P., Sambridge, M., 2019a. Hydrogeological Bayesian hypothesis testing through trans-dimensional sampling of a stochastic water balance model. Water 11, 1463. https://doi.org/10.3390/w11071463.
Enemark, T., Peeters, L.J.M.M., Mallants, D., Batelaan, O., 2019b. Hydrogeological conceptual model building and testing: a review. J. Hydrol. 569, 310–329. https://doi.org/10.1016/j.jhydrol.2018.12.007.
Ferré, T.P.A., 2017a. Revisiting the relationship between data, models, and decision-making. Groundwater 55, 604–614. https://doi.org/10.1111/gwat.12574.
Ferré, T.P.A., 2017b. Modelers: is objectivity overrated? Groundwater 55(5), 603–603. https://doi.org/10.1111/gwat.12575.
Finkel, M., Schad, H., Bayer, P., Lantschner, L., 2008. Disjoined capture and treatment of multiple contaminant plumes in groundwater to improve the cost-efficiency of remediation. In: Trefry, M. (Ed.), International Conference on Groundwater Quality: Securing Groundwater Quality in Urban and Industrial Environments. IAHS-AISH publication, vol. 324, Fremantle, Australia, pp. 94–101.
Freeze, R.A., Gorelick, S.M., 1999. Convergence of stochastic optimization and decision analysis in the engineering design of aquifer remediation. Ground Water 37, 934–954. https://doi.org/10.1111/j.1745-6584.1999.tb01193.x.
Gondwe, B.R.N., Merediz-Alonso, G., Bauer-Gottwein, P., 2011. The influence of conceptual model uncertainty on management decisions for a groundwater-dependent ecosystem in karst. J. Hydrol. 400, 24–40. https://doi.org/10.1016/j.jhydrol.2011.01.023.
Gorelick, S., Zheng, C., Section, S., 2015. Global change and the groundwatermanagement challenge Steven. Water Resour. Res. 3031–3051. https://doi.org/10.1002/2014WR016825.Received.
Guan, J., Aral, M.M., 2004. Optimal design of groundwater remediation systems using fuzzy set theory. Water Resour. Res. 40. https://doi.org/10.1029/2003WR002121.
Guan, J., Aral, M.M., 1999. Optimal remediation with well locations and pumping rates selected as continuous decision variables. J. Hydrol. 221, 20–42. https://doi.org/10.1016/S0022-1694(99)00079-7.
Guillaume, J.H.A., Hunt, R.J., Comunian, A., Blakers, R.S., Fu, B., 2016. Methods for exploring uncertainty in groundwater management predictions. In: Integrated Groundwater Management: Concepts, Approaches and Challenges. Springer International Publishing, pp. 711–737. https://doi.org/10.1007/978-3-319-23576-9_28.
Hansen, N., Müller, S.D., Koumoutsakos, P., 2003. Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES). Evol. Comput. 11, 1–18. https://doi.org/10.1162/106365603321828970.
Hansen, N., Ostermeier, A., 2001. Completely derandomized self-adaptation in evolution strategies. Evol. Comput. https://doi.org/10.1162/106365601750190398.
Harbaugh, A.W., Banta, E.R., Hill, M.C., McDonald, M.G., 2000. MODFLOW-2000, The U. S. Geological Survey Modular Ground-Water Model – User Guide to Modularization Concepts and the Ground-Water Flow Process, Open-File Report. https://doi.org/10.3133/OFR200092.
Hartmann, A., Lange, J., Aguado, A.V., Mizyed, N., Smiatek, G., Kunstmann, H., 2012. A multi-model approach for improved simulations of future water availability at a large Eastern Mediterranean karst spring. J. Hydrol. 468, 130–138. https://doi.org/10.1016/j.jhydrol.2012.08.024.
He, L., Huang, G.H., Lu, H.W., 2008a. A simulation-based fuzzy chance-constrained programming model for optimal groundwater remediation under uncertainty. Adv. Water Resour. 31, 1622–1635. https://doi.org/10.1016/j.advwatres.2008.07.009.
He, L., Huang, G.H., Lu, H.W., Zeng, G.M., 2008b. Optimization of surfactant-enhanced aquifer remediation for a laboratory BTEX system under parameter uncertainty. Environ. Sci. Technol. 42, 2009–2014. https://doi.org/10.1021/es071106y.
Heße, F., Comunian, A., Attinger, S., 2019. What we talk about when we talk about uncertainty. Toward a unified, data-driven framework for uncertainty characterization in hydrogeology. Front. Earth Sci. https://doi.org/10.3389/feart.2019.00118.
Höge, M., Guthke, A., Nowak, W., 2019. The hydrologist's guide to Bayesian model selection, averaging and combination. J. Hydrol. 572, 96–107. https://doi.org/10.1016/j.jhydrol.2019.01.072.
Katic, P., Grafton, R.Q., 2011. Optimal groundwater extraction under uncertainty: resilience versus economic payoffs. J. Hydrol. 406, 215–224. https://doi.org/10.1016/j.jhydrol.2011.06.016.
Kicsiny, R., Varga, Z., 2019. Differential game model with discretized solution for the use of limited water resources. J. Hydrol. 569, 637–646. https://doi.org/10.1016/j.jhydrol.2018.12.029.
Kopsiaftis, G., Christelis, V., Mantoglou, A., 2019. Comparison of sharp interface to variable density models in pumping optimisation of coastal aquifers. Water Resour. Manage. 33, 1397–1409. https://doi.org/10.1007/s11269-019-2194-7.
Kourakos, G., Mantoglou, A., 2008. Remediation of heterogeneous aquifers based on multiobjective optimization and adaptive determination of critical realizations. Water Resour. Res. 44. https://doi.org/10.1029/2008WR007108.
Lantschner, L., 2006. Cost-efficient remediation of multiple contaminant plumes using differentiated capture approach (Master's Thesis). Center for Applied Geoscience, University of Tuebingen.
Liu, M., Nie, Z.L., Wang, J.Z., Wang, L.F., Tian, Y.L., 2016a. An assessment of the carrying capacity of groundwater resources in North China Plain region-Analysis of potential for development. J. Groundw. Sci. Eng. 4, 174–187.
Liu, P., Elshall, A.S.A.S., Ye, M., Beerli, P., Zeng, X., Lu, D., Tao, Y., 2016b. Evaluating marginal likelihood with thermodynamic integration method and comparison with several other numerical methods. Water Resour. Res. 52, 734–758. https://doi.org/10.1002/2014WR016718.
Lu, D., Ye, M., Curtis, G.P., 2015. Maximum likelihood Bayesian model averaging and its predictive analysis for groundwater reactive transport models. J. Hydrol. 529, 1859–1873. https://doi.org/10.1016/j.jhydrol.2015.07.029.
Lu, H., Li, J., Chen, Y., Lu, J., 2019. A multi-level method for groundwater remediation management accommodating non-competitive objectives. J. Hydrol. 570, 531–543. https://doi.org/10.1016/j.jhydrol.2019.01.018.
Lu, H., Li, J., Ren, L., Chen, Y., 2018. Optimal groundwater security management policies by control of inexact health risks under dual uncertainty in slope factors. Chemosphere 198, 161–173. https://doi.org/10.1016/j.chemosphere.2018.01.121.
Lu, H., Ren, L., Chen, Y., Tian, P., Liu, J., 2017. A cloud model based multi-attribute decision making approach for selection and evaluation of groundwater management schemes. J. Hydrol. 555, 881–893. https://doi.org/10.1016/j.jhydrol.2017.10.009.
Luo, Q., Wu, J., Yang, Y., Qian, J., Wu, J., 2016. Multi-objective optimization of long-term groundwater monitoring network design using a probabilistic Pareto genetic algorithm under uncertainty. J. Hydrol. 534, 352–363. https://doi.org/10.1016/j.jhydrol.2016.01.009.
Luo, Q., Wu, J., Yang, Y., Qian, J., Wu, J., 2014. Optimal design of groundwater remediation system using a probabilistic multi-objective fast harmony search algorithm under uncertainty. J. Hydrol. 519, 3305–3315. https://doi.org/10.1016/j.jhydrol.2014.10.023.
Luo, Q., Yang, Y., Qian, J., Wang, X., Chang, X., Ma, L., Li, F., Wu, J., 2020. Spring protection and sustainable management of groundwater resources in a spring field. J. Hydrol. 582, 124498. https://doi.org/10.1016/j.jhydrol.2019.124498.
Maghrebi, M., Jankovic, I., Weissmann, G.S., Matott, L.S., Allen-King, R.M., Rabideau, A.J., 2015. Contaminant tailing in highly heterogeneous porous formations: Sensitivity on model selection and material properties. J. Hydrol. 531, 149–160. https://doi.org/10.1016/j.jhydrol.2015.07.015.
Makropoulos, C., Koutsoyiannis, D., Stanić, M., Djordjević, S., Prodanović, D., Dašić, T., Prohaska, S., Maksimović, Č., Wheater, H., 2008. A multi-model approach to the simulation of large scale karst flows. J. Hydrol. 348, 412–424. https://doi.org/10.1016/j.jhydrol.2007.10.011.
Mani, A., Tsai, F.T.C., Kao, S.C., Naz, B.S., Ashfaq, M., Rastogi, D., 2016a. Conjunctive management of surface and groundwater resources under projected future climate change scenarios. J. Hydrol. 540, 397–411. https://doi.org/10.1016/j.jhydrol.2016.

06.021.
Mani, A., Tsai, F.T.C., Paudel, K.P., 2016b. Mixed integer linear fractional programming for conjunctive use of surface water and groundwater. J. Water Resour. Plann. Manage. 142. https://doi.org/10.1061/(asce)wr.1943-5452.0000676.
Mantoglou, A., Kourakos, G., 2007. Optimal groundwater remediation under uncertainty using multi-objective optimization. Water Resour. Manage. 21, 835–847. https://doi.org/10.1007/s11269-006-9109-0.
Medina, M.A., Jacobs, T.L., Lin, W., Lin, K.-C., 1996. Ground water solute transport, optimal remediation planning, and decision making under uncertainty. J. Am. Water Resour. Assoc. 32, 1–12. https://doi.org/10.1111/j.1752-1688.1996.tb03429.x.
Mondal, R., Benham, G., Mondal, S., Christodoulides, P., Neokleous, N., Kaouri, K., 2019. Modelling and optimisation of water management in sloping coastal aquifers with seepage, extraction and recharge. J. Hydrol. 571, 471–484. https://doi.org/10.1016/j.jhydrol.2019.01.060.
Mulligan, A.E., Ahlfeld, D.P., 1999. Advective control of groundwater contaminant plumes: model development and comparison to hydraulic control. Water Resour. Res. 35, 2285–2294. https://doi.org/10.1029/1999WR900106.
Mustafa, S.M.T., Nossent, J., Ghysels, G., Huysmans, M., 2020. Integrated Bayesian Multi-model approach to quantify input, parameter and conceptual model structure uncertainty in groundwater modeling. Environ. Model. Softw. 126, 104654. https://doi.org/10.1016/j.envsoft.2020.104654.
Neuman, S.P., 2003. Maximum likelihood Bayesian averaging of uncertain model predictions. Stoch. Environ. Res. Risk Assess. 17, 291–305. https://doi.org/10.1007/s00477-003-0151-7.
Neuweiler, I., Helmig, R., 2017. Debates—Hypothesis testing in hydrology: a subsurface perspective. Water Resour. Res. https://doi.org/10.1002/2016WR020047.
Paly, M.D., Bürger, C.M., Bayer, P., 2013. Optimization under worst case constraints – a new global multimodel search procedure. Struct. Multidiscip. Optim. 48, 1153–1172. https://doi.org/10.1007/s00158-013-0950-5.
Pan, Y., Zeng, X., Xu, H., Sun, Y., Wang, D., Wu, J., 2020. Assessing human health risk of groundwater DNAPL contamination by quantifying the model structure uncertainty. J. Hydrol. 584, 124690. https://doi.org/10.1016/j.jhydrol.2020.124690.
Parker, J., Kim, U., Kitanidis, P.K., Cardiff, M., Liu, X., 2010. Stochastic cost optimization of multistrategy DNAPL site remediation. Gr. Water Monit. Remediat. 30, 65–78. https://doi.org/10.1111/j.1745-6592.2010.01287.x.
Peña-Haro, S., Pulido-Velazquez, M., Sahuquillo, A., 2009. A hydro-economic modelling framework for optimal management of groundwater nitrate pollution from agriculture. J. Hydrol. 373, 193–203. https://doi.org/10.1016/j.jhydrol.2009.04.024.
Pholkern, K., Saraphirom, P., Cloutier, V., Srisuk, K., 2019. Use of alternative hydrogeological conceptual models to assess the potential impact of climate change on groundwater sustainable yield in central Huai Luang Basin, Northeast Thailand. Water 11, 241. https://doi.org/10.3390/w11020241.
Pollock, D.W., 1994. User's guide for MODPATH/MODPATH-PLOT, Version 3; a particle tracking post-processing package for MODFLOW, the U.S. Geological Survey finite-difference ground-water flow model, Open-File Report. https://doi.org/10.3133/OFR94464.
Raei, E., Nikoo, M.R., Pourshahabi, S., 2017. A multi-objective simulation-optimization model for in situ bioremediation of groundwater contamination: application of bargaining theory. J. Hydrol. 551, 407–422. https://doi.org/10.1016/j.jhydrol.2017.06.010.
Rajabi, M.M., Ketabchi, H., 2017. Uncertainty-based simulation-optimization using Gaussian process emulation: application to coastal groundwater management. J. Hydrol. 555, 518–534. https://doi.org/10.1016/j.jhydrol.2017.10.041.
Ranjithan, S., Eheart, J.W., Garrett, J.H., 1993. Neural network-based screening for groundwater reclamation under uncertainty. Water Resour. Res. 29, 563–574. https://doi.org/10.1029/92WR02129.
Refsgaard, J.C., Christensen, S., Sonnenborg, T.O., Seifert, D., Højberg, A.L., Troldborg, L., 2012. Review of strategies for handling geological uncertainty in groundwater flow and transport modeling. Adv. Water Resour. 36, 36–50. https://doi.org/10.1016/j.advwatres.2011.04.006.
Rojas, R., Batelaan, O., Feyen, L., Dassargues, A., 2010a. Assessment of conceptual model uncertainty for the regional aquifer Pampa del Tamarugal – North Chile. Hydrol. Earth Syst. Sci. 14, 171–192. https://doi.org/10.5194/hess-14-171-2010.
Rojas, R., Feyen, L., Batelaan, O., Dassargues, A., 2010b. On the value of conditioning data to reduce conceptual model uncertainty in groundwater modeling. Water Resour. Res. 46. https://doi.org/10.1029/2009WR008822.
Rojas, R., Feyen, L., Bianchi, A., Dosio, A., 2012. Assessment of future flood hazard in Europe using a large ensemble of bias-corrected regional climate simulations. J. Geophys. Res. Atmos. 117. https://doi.org/10.1029/2012JD017461.
Rojas, R., Feyen, L., Dassargues, A., 2008. Conceptual model uncertainty in groundwater modeling: combining generalized likelihood uncertainty estimation and Bayesian model averaging. Water Resour. Res. 44. https://doi.org/10.1029/2008WR006908.
Rojas, R., Kahunde, S., Peeters, L., Batelaan, O., Feyen, L., Dassargues, A., 2010c. Application of a multimodel approach to account for conceptual model and scenario uncertainties in groundwater modelling. J. Hydrol. 394, 416–435. https://doi.org/10.1016/j.jhydrol.2010.09.016.
Rubin, Y., 2003. Applied Stochastic Hydrogeology. Oxford University Press, Oxford.
Russell, K.T., Rabideau, A.J., 2000. Decision analysis for pump-and-treat design. Groundw. Monit. Remediat. 20, 159–168. https://doi.org/10.1111/j.1745-6592.2000.tb00281.x.
Saravanan, K., Kashyap, D., Sharma, A., 2014. Model assisted design of scavenger well system. J. Hydrol. 510, 313–324. https://doi.org/10.1016/j.jhydrol.2013.12.031.
Seifert, D., Sonnenborg, T.O., Refsgaard, J.C., Højberg, A.L., Troldborg, L., 2012. Assessment of hydrological model predictive ability given multiple conceptual geological models. Water Resour. Res. 48. https://doi.org/10.1029/2011WR011149.
Shi, F.Z., Chi, B.M., Zhao, C.Y., Yang, T., de la Paix, M.J., Lu, Y., Gao, S.Q., de la Paix, M.J., Lu, Y., Gao, S.Q., 2012. Identifying the sustainable groundwater yield in a Chinese semi-humid basin. J. Hydrol. 452, 14–24. https://doi.org/10.1016/j.jhydrol.2012.05.017.
Song, J., Yang, Y., Wu, J., Wu, J., Sun, X., Lin, J., 2018. Adaptive surrogate model based multiobjective optimization for coastal aquifer management. J. Hydrol. 561, 98–111. https://doi.org/10.1016/j.jhydrol.2018.03.063.
Taravatrooy, N., Nikoo, M.R., Adamowski, J.F., Khoramshokooh, N., 2019. Fuzzy-based conflict resolution management of groundwater in-situ bioremediation under hydrogeological uncertainty. J. Hydrol. 571, 376–389. https://doi.org/10.1016/j.jhydrol.2019.01.063.
Timani, B., Peralta, R., 2015. Multi-model groundwater-management optimization: reconciling disparate conceptual models. Hydrogeol. J. 23, 1067–1087. https://doi.org/10.1007/s10040-015-1259-9.
Tsai, F.-T.-C., Elshall, A.S., 2013. Hierarchical Bayesian model averaging for hydrostratigraphic modeling: uncertainty segregation and comparative evaluation. Water Resour. Res. 49. https://doi.org/10.1002/wrcr.20428.
Tsai, F.T.C., 2010. Bayesian model averaging assessment on groundwater management under model structure uncertainty. Stoch. Environ. Res. Risk Assess. 24, 845–861. https://doi.org/10.1007/s00477-010-0382-3.
Vansteenkiste, T., Tavakoli, M., Ntegeka, V., De Smedt, F., Batelaan, O., Pereira, F., Willems, P., 2014. Intercomparison of hydrological model structures and calibration approaches in climate scenario impact projections. J. Hydrol. 519, 743–755. https://doi.org/10.1016/j.jhydrol.2014.07.062.
Wang, H., Asefa, T., Bracciano, D., Adams, A., Wanakule, N., 2019. Proactive water shortage mitigation integrating system optimization and input uncertainty. J. Hydrol. 571, 711–722. https://doi.org/10.1016/j.jhydrol.2019.01.071.
Warner, J.W., Tamayo-Lara, C., Khazaei, E., Manghi, F., 2006. Stochastic management modeling of a pump and treat system at the Rocky Mountain Arsenal near Denver, Colorado. J. Hydrol. 328, 523–537. https://doi.org/10.1016/j.jhydrol.2005.12.007.
Xu, T., Valocchi, A.J., Ye, M., Liang, F., 2017. Quantifying model structural error: efficient Bayesian calibration of a regional groundwater flow model using surrogates and a data-driven error model. Water Resour. Res. 53, 4084–4105. https://doi.org/10.1002/2016WR019831.
Yang, Y., Wu, J., Luo, Q., Zhang, T., Wu, J., Wang, J., 2017. Effects of stochastic simulations on multiobjective optimization of groundwater remediation design under uncertainty. J. Hydrol. Eng. 22. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001510.
Ye, M., Neuman, S.P., Meyer, P.D., 2004. Maximum likelihood Bayesian averaging of spatial variability models in unsaturated fractured tuff. Water Resour. Res. 40, 1–17. https://doi.org/10.1029/2003WR002557.
Ye, M., Neuman, S.P., Meyer, P.D., Pohlmann, K., 2005ab. Sensitivity analysis and assessment of prior model probabilities in MLBMA with application to unsaturated fractured tuff. Water Resour. Res. 41, 1–14. https://doi.org/10.1029/2005WR004260.
Ye, M., Wang, L., Pohlmann, K.F., Chapman, J.B., 2016. Evaluating groundwater interbasin flow using multiple models and multiple types of data. Groundwater 54, 805–817. https://doi.org/10.1111/gwat.12422.
Yeh, W.W.G., 2015. Revue: Méthodes d'optimisation pour la modélisation et la gestion des eaux souterraines. Hydrogeol. J. 23, 1051–1065. https://doi.org/10.1007/s10040-015-1260-3.
Yin, J., Tsai, F.T.C., 2020. Bayesian set pair analysis and machine learning based ensemble surrogates for optimal multi-aquifer system remediation design. J. Hydrol. 580. https://doi.org/10.1016/j.jhydrol.2019.124280.
Yin, J., Tsai, F.T.C., 2018. Saltwater scavenging optimization under surrogate uncertainty for a multi-aquifer system. J. Hydrol. 565, 698–710. https://doi.org/10.1016/j.jhydrol.2018.08.021.
Zeng, X., Ye, M., Burkardt, J., Wu, J., Wang, D., Zhu, X., 2016. Evaluating two sparse grid surrogates and two adaptation criteria for groundwater Bayesian uncertainty quantification. J. Hydrol. 535, 120–134. https://doi.org/10.1016/j.jhydrol.2016.01.058.
Zeng, X., Ye, M., Wu, J., Wang, D., Zhu, X., 2018. Improved nested sampling and surrogate-enabled comparison with other marginal likelihood estimators. Water Resour. Res. 54, 797–826. https://doi.org/10.1002/2017WR020782.
Zhang, X., Ren, L., Wan, L., 2018. Assessing the trade-off between shallow groundwater conservation and crop production under limited exploitation in a well-irrigated plain of the Haihe River basin using the SWAT model. J. Hydrol. 567, 253–266. https://doi.org/10.1016/j.jhydrol.2018.09.041.