

A robust experimental design for conceptual model discrimination based on information theory

Hai V. Pham^a, Frank T.-C. Tsai^{b,*}

^a INTERA Incorporated, Austin, TX 78759, USA

^b Department of Civil and Environmental Engineering, Louisiana State University, Baton Rouge, LA 70803, USA

ARTICLE INFO

Keywords:

Robust experimental design
Entropy
Information theory
Model discrimination
Uncertainty

ABSTRACT

This study introduces firm information gain for model discrimination based on Shannon entropy and worst-case scenario experimental design. Firm information gain is the minimal additional information gained by an experimental design with respect to existing information. Robust experimental design aims to maximize the firm information gain by searching for the least number of new pumping wells and observation wells. Robust experimental design includes a Bayes factor threshold to ensure that new data provide strong evidence for model discrimination. To maximize the firm information gain, a framework is proposed that combines the parallel-sequential genetic algorithm (GA) for parallel computing and the nested quadrature rule for efficiently solving multidimensional integrals. The numerical experiment involves the true model for the purpose of verification. The results show that using a full covariance matrix is imperative to avoid exaggerating firm information gain. Collecting new groundwater data is prioritized over exploring additional pumping wells. Maximizing firm information gain is able to identify the same and true model.

1. Introduction

Groundwater is a crucial source of freshwater throughout the world for both hydrologic and human systems (Alley et al., 2002; Giordano, 2009; Siebert et al., 2010). Groundwater modeling has been widely used for decades as essential tools for the planning and management of groundwater resources (Gleeson et al., 2012; Wada et al., 2010). However, developing a groundwater model has never been an easy task as groundwater data is always sparse and uncertainty always exists. Multiple conceptualizations of a groundwater system are often investigated. Yet considering too many conceptual models indicates high model prediction uncertainty and may lose the purpose of model development (Bredehoeft, 2005; Højberg and Refsgaard, 2005). Collecting and incorporating new data into groundwater models helps advance conceptual understanding and management of groundwater resources (Kikuchi, 2017) and in turn, reduces the number of models. Nevertheless, collecting groundwater data is usually costly, and optimal experimental design techniques are often conducted before data collection to gain the maximum amount of information given a pre-defined monitoring objective.

According to Sun (1994), experimental design in groundwater

modeling generally falls into two parts: the observation part (e.g., state variables to be observed, the number and locations of observation wells, and observation frequency) and the excitation part (e.g., the number and locations of extraction and injection wells, pumping and injection rates, and periods of extraction and injection). If the excitation part is pre-determined and only the observation part is considered, the experimental design is referred to as an observation network design.

Observation network designs have been studied extensively in the literature. A variety of methodologies have been introduced to design a groundwater observation network (Kollat et al., 2011; Loaiciga et al., 1992; Mogheir et al., 2006). Among these methods, physically-based simulation approaches (Cieniawski et al., 1995; Cleveland and Yeh, 1990; Dhar and Datta, 2007; Hudak and Loaiciga, 1992; McKinney and Loucks, 1992; Meyer et al., 1994; Reed et al., 2000; Storck et al., 1997) and information theory (entropy-based method) (Alfonso et al., 2010; Mogheir et al., 2006; Mogheir and Singh, 2002; Nowak and Guthke, 2016; Poeter and Anderson, 2005) are commonly employed owing to their flexibility in examining design scenarios and design constraints.

The objectives of observation network designs are usually to: (1) improve parameter estimation (Altmann-Dieses et al., 2002; Chang et al., 2005; Cleveland and Yeh, 1990; Herrera and Pinder, 2005; Hsu

* Corresponding author.

E-mail address: ftsai@lsu.edu (F.T.-C. Tsai).

<https://doi.org/10.1016/j.advwatres.2023.104485>

Received 10 August 2022; Received in revised form 30 April 2023; Accepted 7 June 2023

Available online 8 June 2023

0309-1708/© 2023 Elsevier Ltd. All rights reserved.

and Yeh, 1989; Sciortino et al., 2002; Siade et al., 2017; Sun and Yeh, 2007; Ushijima and Yeh, 2015), (2) minimize prediction uncertainty (Chadalavada and Datta, 2008; Janssen et al., 2008; McKinney and Loucks, 1992; Nowak et al., 2010; Wagner, 1995; Wöhling et al., 2016), (3) detect plumes (Bode et al., 2019; Dhar and Datta, 2007; Dokou and Pinder, 2009; Kim and Lee, 2007; Leube et al., 2012; Meyer and Brill, 1988; Storck et al., 1997), and (4) to discriminate among candidate models and identify the most probable model (Kikuchi et al., 2015; Knopman and Voss, 1988; Pham and Tsai, 2016, 2015; Usunoff et al., 1992; Yakirevich et al., 2013). Readers are referred to several in-depth review articles (Hassan, 2003; Kollat et al., 2011; Loaiciga et al., 1992; Minsker, 2003).

To achieve the objective of model discrimination, observation networks aim to provide the most useful information with respect to model discrimination. Several criteria have been developed for model discrimination in optimal observation network designs based on the maximum differences between model predictions (Knopman et al., 1991; Knopman and Voss, 1988; Nordqvist and Voss, 1996; Usunoff et al., 1992), the maximum Kullback-Leibler information (Kikuchi et al., 2015; Nowak and Guthke, 2016; Yakirevich et al., 2013), the maximum change in entropy (Box and Hill, 1967; Alfonso et al., 2010), and the maximization of posterior model probability (Pham and Tsai, 2016, 2015). The basic concept underlying all these criteria is to sample the state variable(s) at spatiotemporal locations (i.e., predicted data) where the variance among the ensemble of proposed competing model's predictions is maximized.

The worth of new data has been analyzed in various water-related problems such as prediction uncertainty reduction (Dausman et al., 2010; Feyen and Gorelick, 2005; Freer et al., 1996; Gates and Kisiel, 1974; Rojas et al., 2010; Sohn and Small, 2000; Tiedeman et al., 2004, 2003; Yokota and Thompson, 2004), model selection (Wöhling et al., 2015), decision making (Ben-Zvi et al., 1988; Davis and Dvoranchik, 1971; James et al., 1996; Reichard and Evans, 1989), and cost-effectiveness (James and Gorelick, 1994; Neuman et al., 2012; Norberg and Rosén, 2006; Wagner, 1999). Though optimal observation network designs have been studied extensively in the past, there is still a lack of clear understanding of the amount and the worth of new data required to justify a certain level of model discrimination and identify the most probable model. Besides, none of these studies could guarantee identifying the same, most probable model. Moreover, all these studies only considered the observation part of the experimental design.

In this study, we introduce a robust experimental design for model discrimination based on the Shannon entropy (Shannon, 1948) and the worst-case scenario experimental design (Sun and Yeh, 2007). First, we introduce a "firm information gain" concept and derive a new model discrimination criterion based on the Shannon entropy and the Bayes factor. The "firm information" is defined to be the minimum information guaranteed from an experimental design. The crux of experimental design based on firm information gain is that the design objective can be achieved with the least information. As a result, any other experimental designs under the same experimental conditions will result in higher information gain and therefore guarantee the same design outcome. According to Sun and Yeh (2007), an experimental design is considered "robust" if it accounts for both the excitation part (pumping activities) and observation part (observation activities) to maximize the firm information gain. This is achieved through a max-min optimization problem to improve model discrimination while using the fewest possible pumping and observation wells. In the context of robust experimental design, "robust" means that the optimized pumping well network performs well across all possible observation well networks, as it performs well even with the worst-case observation well network. We hypothesize that the same most probable model can be identified from a pool of competing models by maximizing the firm information gain for a system (e.g., a set of conceptual groundwater models that differ in boundary conditions, geological structures, etc.) and satisfying a Bayes factor threshold. Any other experimental design solutions having

information (negative Shannon entropy) higher than the firm information will result in the same most probable model. Second, we introduce a parallel computing framework that combines the parallel-sequential GA (Carroll, 1996) and the nested quadrature rule (Genz and Keister, 1996) to efficiently solve the time-consuming max-min optimization problem. Finally, we test the proposed framework and conduct the robust experimental design on a hypothetical numerical example where nine competing groundwater models were generated and a robust experimental design is needed to discriminate among the models and identify the same most probable model. The robust experimental design in this study is different from that in Box and Hill (1967) and Pham and Tsai (2016). First, this study considers measurement errors and data correlation in the experimental design. Second, a new model discrimination criterion is introduced that maximizes firm information gain to obtain the robust experimental design, instead of finding an upper bound of the expected information gain (usually referred to the Box-Hill discrimination function).

2. Methodology

2.1. Shannon entropy and expected information gain

Shannon entropy (Shannon, 1948) provides a measure of the information value of a system using probabilities of the occurrence of events in the system. Consider that a set of m candidate models, $\mathbf{M} = \{M_1, M_2, \dots, M_m\}$, represents the events of the system (e.g., candidate models are groundwater models that differ in model conceptualizations such as boundary conditions, geological structures, and parameter structures). Their posterior model probabilities are $\Pr(M_i|\Delta^{\text{obs}})$ given existing observation data Δ^{obs} . The Shannon entropy of the system is

$$S(\mathbf{M}|\Delta^{\text{obs}}) = -\sum_{i=1}^m \Pr(M_i|\Delta^{\text{obs}}) \ln \Pr(M_i|\Delta^{\text{obs}}) \quad (1)$$

where $\ln \Pr(M_i|\Delta^{\text{obs}})$ is the information of the model M_i . Negative entropy ($-S$) represents the average amount of information (I) provided by all candidate models:

$$I(\mathbf{M}|\Delta^{\text{obs}}) = -S(\mathbf{M}|\Delta^{\text{obs}}) \quad (2)$$

The least information corresponds to the maximum entropy when all models have an equal posterior model probability. The maximum information from the system corresponds to the minimum entropy when one model has a 100% posterior model probability and other models have zero posterior model probability.

The main purpose of an experimental design (D) for model discrimination is to maximize information gain through acquiring new data such that the most probable model can be identified from a pool of candidate models of the system. The information gain is defined as follows:

$$I_G = I(\mathbf{M}|\Delta_D^{\text{new}}) - I(\mathbf{M}|\Delta^{\text{obs}}) \quad (3)$$

where I_G is the information gain after an experimental design, $\Delta_D^{\text{new}} \in R^N$ is a vector of N new data, and $I(\mathbf{M}|\Delta_D^{\text{new}})$ represents the combined information obtained from both the new and existing data.

New data are unknown and uncertain before sampling. This study proposes an expected information gain of the new data for the experimental design:

$$\bar{I}_G = E[I(\mathbf{M}|\Delta_D^{\text{new}})] - I(\mathbf{M}|\Delta^{\text{obs}}) \quad (4)$$

where $\bar{I}_G$ is the expected information gain and E is the expectation operator. The expected new information under a probability distribution function of new data Δ_D^{new} is

$$E[I(\mathbf{M}|\Delta_D^{\text{new}})] = \int_{-\infty}^{\infty} \sum_{i=1}^m \Pr(\mathbf{M}_i|\Delta_D^{\text{new}}) \ln[\Pr(\mathbf{M}_i|\Delta_D^{\text{new}})] q(\Delta_D^{\text{new}}) d\Delta_D^{\text{new}} \quad (5)$$

where $\Pr(\mathbf{M}_i|\Delta_D^{\text{new}})$ are the posterior model probabilities given new data Δ_D^{new} ; and $q(\Delta_D^{\text{new}})$ is the averaged probability density function of new data Δ_D^{new} via Bayesian model averaging:

$$q(\Delta_D^{\text{new}}) = \sum_{j=1}^m \Pr(\mathbf{M}_j|\Delta^{\text{obs}}) p(\Delta_D^{\text{new}}|\mathbf{M}_j) \quad (6)$$

where $p(\Delta_D^{\text{new}}|\mathbf{M}_j)$ is the probability density function of predicted new data Δ_D^{new} using the model $\mathbf{M}_j$.

Inserting Eqs. (2) and (5) into the Eq. (4), Appendix A shows the expected information gain as follows:

$$\bar{I}_G = \sum_{i=1}^m \Pr(\mathbf{M}_i|\Delta^{\text{obs}}) \int_{-\infty}^{\infty} p(\Delta_D^{\text{new}}|\mathbf{M}_i) \ln \frac{p(\Delta_D^{\text{new}}|\mathbf{M}_i)}{q(\Delta_D^{\text{new}})} d\Delta_D^{\text{new}} \quad (7)$$

The integral in the Eq. (7) is the Kullback-Leibler (KL) divergence that measures the difference between the BMA weighted probability distribution $q(\Delta_D^{\text{new}})$ and the probability distribution $p(\Delta_D^{\text{new}}|\mathbf{M}_i)$. Maximizing $\bar{I}_G$ enhances the “diversity” of probability distributions of each model’s prediction compared to the BMA weighted probability distribution. The expected information gain is the averaged KL divergence weighted by $\Pr(\mathbf{M}_i|\Delta^{\text{obs}})$. Because the KL divergence is always non-negative, the expected information gain is always non-negative.

To solve Eq. (7), we need to know the probability density function $p(\Delta_D^{\text{new}}|\mathbf{M}_j)$ of predicted new data Δ_D^{new} . Considering that new data are correlated and multivariate Gaussian, the probability density function $p(\Delta_D^{\text{new}}|\mathbf{M}_i)$ is:

$$p(\Delta_D^{\text{new}}|\mathbf{M}_i) = (2\pi)^{-\frac{N}{2}} |\hat{\Sigma}_i|^{-\frac{1}{2}} e^{-\frac{1}{2}(\Delta_D^{\text{new}} - \Delta_i)^T \hat{\Sigma}_i^{-1} (\Delta_D^{\text{new}} - \Delta_i)} \quad (8)$$

where Δ_i are the expected values of new data estimated by the model $\mathbf{M}_i$, $\hat{\Sigma}_i = \Sigma + \Sigma_e$ is the total covariance matrix of new data involving the use of the model $\mathbf{M}_i$, which is the sum of the covariance matrix of the estimated new data (representing parameter and model structure uncertainties) and the covariance matrix of measurement errors in new data. This study considers correlated data, which results in a full covariance matrix.

Appendix B further expands the Eq. (7) with the multivariate Gaussian distribution as follows:

$$\bar{I}_G = \sum_{i=1}^m \Pr(\mathbf{M}_i|\Delta^{\text{obs}}) \left\{ \ln \left[(2\pi)^{-\frac{N}{2}} |\hat{\Sigma}_i|^{-\frac{1}{2}} \right] - \frac{N}{2} - E_{\Delta_D^{\text{new}}|\mathbf{M}_i} [\ln q(\Delta_D^{\text{new}})] \right\} \quad (9)$$

$E_{\Delta_D^{\text{new}}|\mathbf{M}_i} [\ln q(\Delta_D^{\text{new}})]$ in the Eq. (9) is the expectation of $\ln q(\Delta_D^{\text{new}})$ under randomness of Δ_D^{new} given model $\mathbf{M}_i$, which is

$$E_{\Delta_D^{\text{new}}|\mathbf{M}_i} [\ln q(\Delta_D^{\text{new}})] = \int_{-\infty}^{\infty} p(\Delta_D^{\text{new}}|\mathbf{M}_i) \ln q(\Delta_D^{\text{new}}) d\Delta_D^{\text{new}} \quad (10)$$

For one-dimensional integral (i.e., only one new observation is collected), Gaussian quadrature rules and Monte Carlo methods are powerful. However, when the new data are in high dimensions and correlated, these approaches become impractical due to prohibitive computing costs (rising exponentially with the number of dimensions) and there is no analytical solution for the Eq. (10) as far as authors’ knowledge. It is noted that the method presented in this study is not limited to the Gaussian distribution of the predicted new data. The general form of the expected information gain is in Eq. (7). As soon as one knows the probability density function $p(\Delta_D^{\text{new}}|\mathbf{M}_j)$ Eq. (7) can be solved. Assuming the Gaussian distribution is for the convenience purpose that the general form of $\bar{I}_G$ in Eq. (7) is reduced to a simpler form in Eqs. (9) and (10).

Genz and Keister (1996) presented a nested quadrature rule to efficiently calculate high-dimensional integrals for the multivariate normal

distribution with zero means and an identity matrix. Unfortunately, the integral in Eq. (10) was for multivariate Gaussian with a non-zero mean and a full covariance matrix. Therefore, this study adopted the Cholesky decomposition to transform $E_{\Delta_D^{\text{new}}|\mathbf{M}_i} [\ln q(\Delta_D^{\text{new}})]$ into a multivariate normal distribution with zero means and an identity matrix (i.e., the covariance matrix is an identity matrix in which all the diagonal elements are ones and all off-diagonal elements are zeros), and used (Genz and Keister, 1996) approach to calculate $E_{\Delta_D^{\text{new}}|\mathbf{M}_i} [\ln q(\Delta_D^{\text{new}})]$ numerically as shown in Appendix C.

2.2. Max–min information gain criterion for model discrimination

Consider an experimental design (D) that includes a pumping design and an observation design to collect new groundwater level data using the least number of pumping wells and observation wells. The new head observation locations serve to obtain firm information gain while the new pumping test locations serve to maximize the firm information gain. Data from new head observation wells stimulated by new pumping test locations will serve to discriminate groundwater models such that the same most probable groundwater model can be identified.

This study adopts the Bayes factor as a model discrimination function. To achieve the design objective, the max–min optimization problem to maximize firm information gain is introduced for the robust experimental design as follows:

$$\max_{D_Q} \min_{D_\Delta} \bar{I}_G(D) \quad (11)$$

where D_Q are the pumping design and D_Δ are the observation design.

Eq. (11) is subject to

$$\min \left\{ \text{BF}_{ki} = \frac{p(\Delta_D^{\text{new}}|\mathbf{M}_k)}{p(\Delta_D^{\text{new}}|\mathbf{M}_i)}, i = 1, 2, \dots, \text{mandi} \neq k \right\} \gamma \quad (12)$$

where $\min \bar{I}_G(D)$ is the firm information gain from experimental design, BF_{ki} is the Bayes factor, which is the likelihood ratio of the most probable model $\mathbf{M}_k$ (having the highest posterior model probability) against other models $\mathbf{M}_i$, and γ is a Bayes factor threshold. $p(\Delta_D^{\text{new}}|\mathbf{M}_i)$ is the likelihood that new data are predicted using the model $\mathbf{M}_i$. $p(\Delta_D^{\text{new}}|\mathbf{M}_k)$ is the highest likelihood among m models. Given pumping locations in an experimental design D , $\min \bar{I}_G(D)$ can be obtained by minimizing $\bar{I}_G$ given new observation data as dependent variables. The maximum of $\min \bar{I}_G(D)$ in Eq. (11) can be solved by solving the maximization optimization problem where the dependent variables are pumping locations.

Eq. (12) ensures that the same most probable model has sufficient evidence to be discriminated from all other models. The classification of Harold Jeffreys (Jeffreys, 1998) presents how strong the new data evidence supports one model over other models. The higher the γ value, the stronger the data evidence that supports one model over the other competing models. For example, when the Bayes factor is between 5 and 10, the data evidence is classified as substantial. When Bayes factor is greater than 10, the data evidence is classified as strong (Jeffreys, 1998).

2.3. Total covariance matrix $\hat{\Sigma}_i$ for new observation data

The total covariance matrix of new data includes the covariance matrix of measurement errors in new data and the covariance matrix of the estimated new data. Random measurement errors are usually modeled by uncorrelated Gaussian noise with zero means. Therefore, the covariance matrix of measurement errors in new data can be $\Sigma_e = \sigma_e^2 \mathbf{I}$, where σ_e^2 is a constant error variance and $\mathbf{I}$ is an identity matrix.

Monte Carlo simulation on model parameters is adopted to calculate the covariance matrix of the expected values of new data estimated by model $\mathbf{M}_i$:

$$\sum_i = \frac{1}{Q-1} \sum_{q=1}^Q (\Delta(\beta_i^{(q)}) - \bar{\Delta}_i) (\Delta(\beta_i^{(q)}) - \bar{\Delta}_i)^T \quad (13)$$

where Q is the number of realizations of model parameters $\beta_i^{(q)}$ of the model M_i . These parameter realizations are sampled from the posterior distribution upon history matching for each model. $\bar{\Delta}_i$ is the mean of the new data simulated by model M_i :

$$\bar{\Delta}_i = \frac{1}{Q} \sum_{q=1}^Q \Delta(\beta_i^{(q)}) \quad (14)$$

The BMA method (Hoeting et al., 1999; Draper, 1995) is used to calculate the covariance matrix Σ of the estimated new data as follows:

$$\sum = \sum_{\text{BMA}} = \sum_{i=1}^m \left[\sum_i + (\Delta(\hat{\beta}_i) - \bar{\Delta}_D^{\text{new}}) (\Delta(\hat{\beta}_i) - \bar{\Delta}_D^{\text{new}})^T \right] \Pr(M_i | \Delta^{\text{obs}}) \quad (15)$$

where $\Delta(\hat{\beta}_i)$ is the predicted new data using the model M_i and the estimated model parameters $\hat{\beta}_i$ of the model M_i . $\bar{\Delta}_D^{\text{new}}$ is the BMA mean of the predicted new data. The total covariance matrix for Δ_D^{new} is $\hat{\Sigma}_i = \Sigma + \Sigma_\epsilon = \Sigma_{\text{BMA}} + \Sigma_\epsilon$.

2.4. Model calibration and posterior model probability

The covariance matrix adaptation-evolution strategy (CMA-ES) (Hansen and Ostermeier, 2001; Hansen et al., 2003) is employed to estimate model parameters and to obtain a covariance matrix for the estimated model parameters. Model parameters are estimated by minimizing the root mean square error (RMSE) between calculated and observed heads. The CMA-ES is a global-local stochastic derivative-free algorithm, that was parallelized for time-consuming groundwater model calibration and uncertainty analysis (Elshali et al., 2013).

Once the estimated parameters and their covariance matrix are obtained by the CMA-ES, the marginal likelihood function is calculated as follows for existing observation data Δ^{obs} , which is similar to the Eq. (8):

$$p(\Delta^{\text{obs}} | M_i) = (2\pi)^{-\frac{N_1}{2}} |\Sigma_\epsilon + \Sigma'_i|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} (\Delta^{\text{obs}} - \Delta'_i)^T (\Sigma_\epsilon + \Sigma'_i)^{-1} (\Delta^{\text{obs}} - \Delta'_i) \right] \quad (16)$$

where N_1 is the number of existing observation data; Σ_ϵ is the covariance matrix of measurement errors; Δ'_i is the simulated observation data using the model M_i with the estimated model parameters $\hat{\beta}_i$ obtained by the CMA-ES; and Σ'_i is the covariance matrix of simulated observation data, which is calculated by the Monte Carlo simulation based on the estimated model parameters and their covariance matrix obtained by the CMA-ES. It is important to acknowledge that the covariance matrix generated by the CMA-ES is merely an estimate. It is crucial to confirm the accuracy of the covariance matrix obtained through this method. This can be achieved by running simulations using realizations of model parameters and verifying that the resulting root mean square errors (RMSEs) are comparable to those obtained using the estimated model parameters. The posterior model probability for each groundwater model is commonly calculated the same as the likelihood given the assumption that all models have the same prior model probability. Other than the CMA-ES, the Null-Space Monte Carlo method (Siade et al., 2017) and the iterative ensemble smoother method (White, 2018) can also quantify model output uncertainty.

3. Numerical example

This study uses a steady-state groundwater flow condition in a 5-layer synthetic anisotropic confined aquifer to illustrate the robust

experimental design based on the firm information gain. The size of the aquifer is 5 km by 5 km and is discretized into 5 layers, 25 rows, and 25 columns. See Fig. 1. The cell size is 200 m by 200 m with variable thickness. There are two pumping wells (PWs) screened at layer 1 and layer 5 and one injection well (IW) screened at layer 3. Fig. 1 shows the well locations and pump rates. The true constant-head boundary condition of 50 m is assigned to the boundary cells of all layers at the south boundary. No-flow boundary condition is assigned to the north, east and west boundaries to allow better variations (higher sensitivities) in simulated heads upon pumping. Fig. 2a shows the true aquifer structure. Table 1 lists the true model parameters. The USGS MODFLOW-2005 (Harbaugh, 2005) is adopted to simulate true steady-state groundwater levels at the 5 existing observation wells in the model domain (see Fig. 1). Gaussian noises of a zero mean and a standard deviation of 0.1 m are added to the groundwater level data to simulate measurement errors.

We pretend that we do not know the true aquifer structure (i.e., the geometry of the aquifer system and lithology), the true constant-head boundary value, and the true horizontal hydraulic conductivity. Three-dimensional geometry views of three aquifer structures are given in Fig. 2, denoted as G_1 , G_2 , and G_3 , respectively. These aquifer structures were extracted from the real-world case study of the Baton Rouge aquifer system using three different geostatistical methods (the generalized parameterization, the indicator zonation methods, and the indicator kriging, respectively) (Pham and Tsai, 2016, 2015). G_3 is a highly connected aquifer system following up with G_1 and G_2 . The number of active model cells is 2018, 1566, and 2021 for GP, IZ, and IK, respectively. Three head values 49, 50, and 51 m for the south boundary are considered and denoted as B_1 , B_2 , and B_3 , respectively. The number of boundary cells at the south boundary for G_1 , G_2 , and G_3 are 30, 25, and 29, respectively. These boundary cells are only in layers from 3 to 5 for all three aquifer structures. The connections between the south boundary cells and the aquifer are weaker in the IZ structure in comparison to the GP and IK structures. Combinations of three aquifer structures and three head boundary values result in nine conceptual groundwater models.

4. Solving the max-min optimization problem to obtain robust experimental design

This study solved Eqs. (11) and (12) to identify the robust experimental design for model discrimination to identify the most probable groundwater model. The Bayes factor threshold was set to be 10 such that the most probable model will be at least strongly discriminated from the other eight competing models. Decision variables were the number of new pumping wells and the number of new observation wells. A pumping rate of 200 m³/day was assigned for all new pumping wells. Experimental designs were conducted by gradually increasing the number of pumping wells and the number of observation wells of the system. The robust experimental design was the one that optimizes Eq. (11) using the least number of new pumping wells and new observation wells until Eq. (12) is satisfied.

To maximize the firm information gain in Eq. (11) (the max-min optimization problem), this study utilized a parallel-sequential genetic algorithm (GA) optimization scheme. Given a number of pumping wells and observation wells, a parallel GA was employed to optimize pumping locations (the outer loop of the max-min optimization problem); and under the parallel GA, a sequential GA (the inner loop) was employed to optimize observation locations. The GA code of Carroll (1996) was employed to solve the max-min optimization problem and was parallelized to be run in SuperMIC, a supercomputer at Louisiana State University using an embarrassingly parallel technique. A population size of 80 (i.e., used 80 cores) was assigned to the parallel GA and a population size of five was assigned to the sequential GA (micro-GA). The number of generations was 50 for the parallel GA and 500 for the sequential GA. Other default settings were set the same in the GA code.

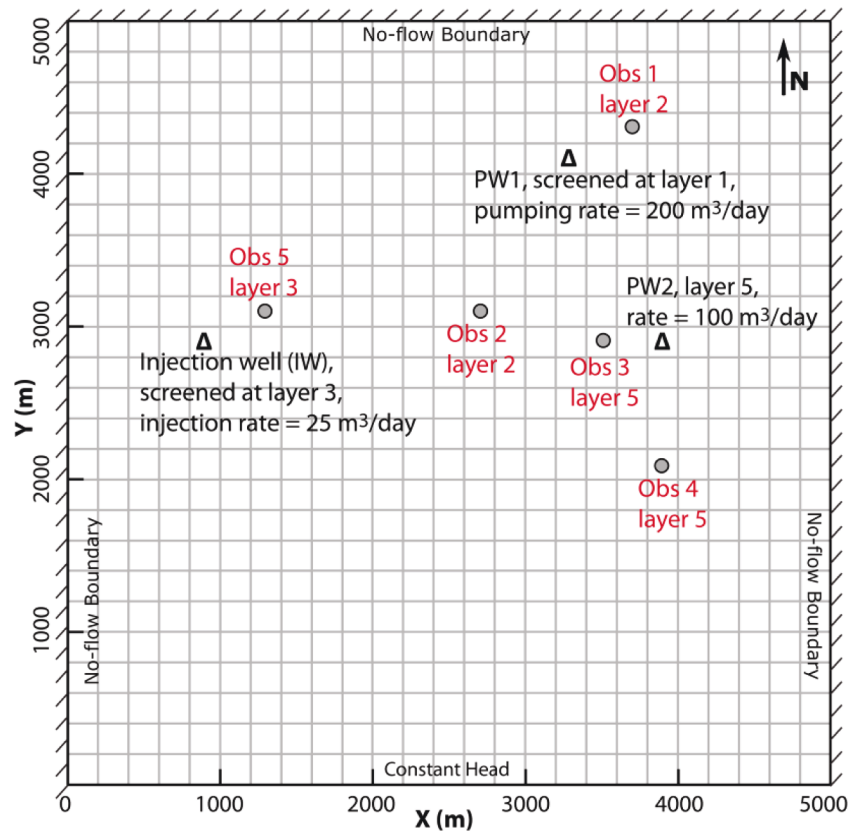


Fig. 1. Plan view of the model domain and the existing pumping wells (PW), the injection well (IW), and the observation wells (Obs).

5. Results

5.1. Model calibration, posterior model probability, and entropy of the current system

Table 2 shows model calibration results, the posterior model probability, and Bayes factor for each groundwater model. The parallel CMA-ES estimated horizontal hydraulic conductivity for all layers using the five “noisy” head observation data. The top five models (G_1B_1 , G_1B_2 , G_1B_3 , G_3B_1 , and G_3B_2) showed comparably small RMSEs. The three conceptual models with G_2 aquifer structure resulted in a much larger RMSE. It indicates that the aquifer structure significantly affected the model calibration results. The models with the G_1 and G_3 aquifer structures better represented the aquifer than the G_2 aquifer structure. The top five models had posterior model probabilities greater than 17%. G_3B_2 model had the highest posterior model probability but did not have the lowest RMSE because of the impact of the covariance matrix Σ'_i in Eq. (16). Insufficient observation data used in the model calibration prevented the true model (G_1B_2) from having the highest posterior model probability and outperforming the other models. Nevertheless, the Bayes factor suggested that the current data did not discriminate G_3B_2 model from the other top-four models.

The entropy of the system was 1.748 nat calculated using Eq. (1) and the posterior model probabilities in Table 2. The nat (the natural unit of information) is the natural unit for information entropy. Given a system of nine models, the entropy of the system is between zero (highest information) and 2.197 nat (lowest information). Therefore, 1.748 nat (79.6% of the maximal entropy) of the system was a high value. This indicates that more data are needed to reduce the entropy (increase information) of the system and to identify the most probable model.

5.2. Information gain and data correlation evaluation using the current system

In this section, we intend to study the changes in expected information gain $\bar{I}_G$ and firm information gain $\min \bar{I}_G(D)$ by systematically adding new observation data before conducting an exhaustive robust experimental design. No new pumping and injection wells were added. We only draw new head data out of active cells that are in common in three aquifer structures (i.e., 1024 possible locations). Additionally, we investigate the impacts of data correlation on $\min \bar{I}_G(D)$.

Fig. 3 shows the spatial distributions of expected information gain $\bar{I}_G$ from drawing one new head data in layers 2 to 5. $\bar{I}_G$ was found varied between 0.723 nat and 1.472 nat. High expected information gain occurred in the areas near the constant-head boundary and near the injection well, where heads predicted by the candidate models were quite different. Drawing one additional head data for either layer 3, 4, or 5 gained higher $\bar{I}_G$ than that from layer 2. Data collected from different locations provided different $\bar{I}_G$. Robust experimental designs are needed to identify optimal locations.

Fig. 4 compares firm information gain $\min \bar{I}_G(D)$ calculated by using an experimental design D (i.e., using the existing pumping and injection wells, and adding one to five new head data) for both cases of uncorrelated and correlated heads. The result indicates that $\min \bar{I}_G(D)$ increased as the size of new head data increased. Experimental designs considering uncorrelated new data overestimated $\min \bar{I}_G(D)$. The degree of overestimation increased dramatically with the size of uncorrelated data. Data correlation significantly impacted on $\min \bar{I}_G(D)$; therefore, this study will only focus on experimental designs utilizing correlated data in the later sections.

Experimental designs using one to five new head data and the existing pumping and injection wells were unable to reach the highest possible information gain $\bar{I}_G$ of 1.748 nat. The Bayes factor threshold

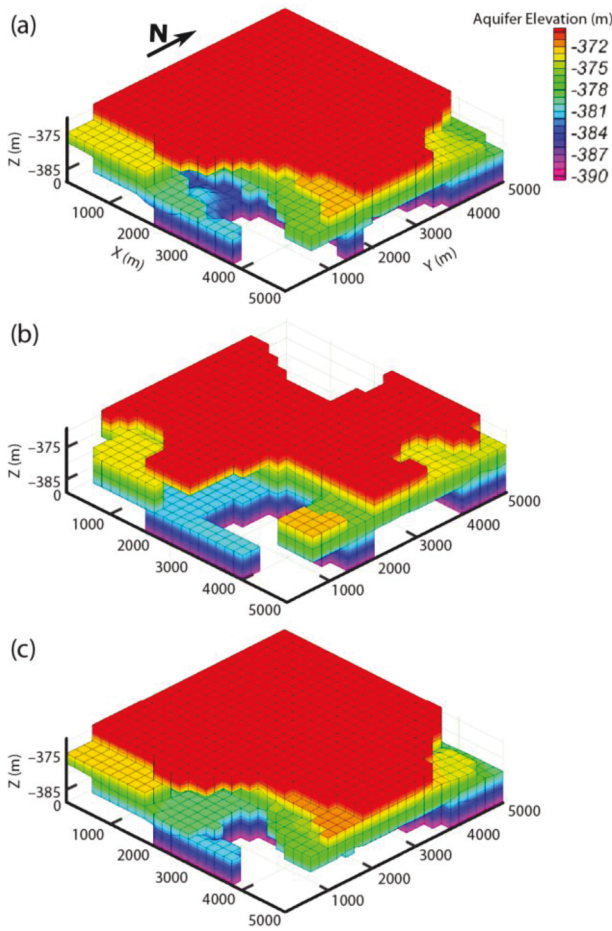


Fig. 2. Aquifer structures generated using: (a) the generalized parameterization (G_1), (b) the indicator zonation (G_2), and (c) the indicator kriging (G_3).

Table 1

True model parameter values of the confined aquifer.

Layer	Layer 1, 2	Layer 3	Layer 4, 5
Horizontal hydraulic conductivity (m/d)	25	1.00×10^{-4}	25
Vertical hydraulic conductivity (m/d)	2.5	1.00×10^{-5}	2.5

Table 2

Root mean square error (RMSE), estimated horizontal hydraulic conductivity, posterior model probability (Pr), and Bayes factor.

Model	RMSE (m)	Horizontal hydraulic conductivity (m/d)			Pr (%)	Bayes factor
		Layer 1, 2	Layer 3	Layer 4, 5		
G ₁ B ₁	0.081	30.05	4.21×10^{-2}	45.53	17.72	0.82
G ₁ B ₂	0.073	22.77	2.54×10^{-5}	25.47	18.96	0.48
G ₁ B ₃	0.095	21.32	3.14×10^{-6}	16.77	18.75	2.74
G ₂ B ₁	0.900	4768	2.88×10^{-1}	64.52	1.71×10^{-11}	1.62×10^{12}
G ₂ B ₂	1.110	1252	2.50×10^{-1}	32.95	4.26×10^{-10}	2.44×10^{13}
G ₂ B ₃	1.331	755.8	2.17×10^{-1}	22.25	6.21×10^{-8}	3.01×10^{14}
G ₃ B ₁	0.086	24.37	3.25×10^{-2}	38.31	17.14	1.07
G ₃ B ₂	0.081	20.28	2.65×10^{-6}	21.04	20.49	1.00
G ₃ B ₃	0.154	19.09	5.58×10^{-6}	13.93	6.94	66.10

(Eq. (12)) was also not satisfied (the Bayes factor values will be presented in Section 5.5). The experimental designs using one to five new head data showed that new pumping wells are needed.

5.3. Information gain using one new well

Detailed maximum firm information gain was illustrated by the case of searching for one optimal pumping location and one optimal observation location. Although there are 1024 possible locations (model cells) available for installing new pumping and observation wells, only 256 locations (every other model cell) were considered for potential pumping locations to ensure the experimental design remains tractable. The potential pumping locations were indexed from 1 to 256. The potential head observation locations are indexed from 1 to 1024.

Fig. 5(a) shows the firm information gain $\min \bar{I}_G(D)$ given by each of the 256 potential pumping well locations. For each new pumping location, 1024 alternatives of new observation wells were investigated and the observation location that resulted in $\min \bar{I}_G(D)$ was recorded. The result showed that $\min \bar{I}_G(D)$ were varied from 0.532 nat to 0.684 nat. Not much changes in $\min \bar{I}_G(D)$ were detected if placing a new pumping well in layer 1, layer 2, layer 4, or layer 5. However, large changes in $\min \bar{I}_G(D)$ were found if placing a new pumping well in layer 3. The maximum change of $\min \bar{I}_G(D)$ was found to be 0.684 nat, occurring at pumping location index 141 in layer 3, which is denoted as circle A in Fig. 5(a) and (c). The horizontal coordinates of the optimal pumping location are ($x = 4300$ m, $y = 3100$ m).

Given the optimal pumping location, Fig. 5(b) shows the expected information gain $\bar{I}_G$ for each of the 1024 potential observation locations. The result showed that $\bar{I}_G$ were varied from 0.684 nat to 1.110 nat. Drawing a new head data from layer 3 (low conductivity) generally provided higher $\bar{I}_G$ than other layers. Some observation locations in layers 4 and 5 also provided higher $\bar{I}_G$. The firm information gain of 0.684 nat was obtained at the observation location index 548 in layer 3 and denoted as circle B in Fig. 5(b) and (d). The horizontal coordinates of the optimal observation location are also ($x = 500$ m, $y = 2100$ m). The result verified that all experimental designs using one new head observation and one new pumping well (at pumping location index 141 in layer 3) resulted in higher $\bar{I}_G$ than 0.684 nat. Similar to Section 5.2, experimental designs using one new pumping well and one new head data were unable to reach the highest possible information gain $\bar{I}_G$ of 1.748 nat and failed to meet the Bayes factor threshold. Two or more pumping wells are needed to achieve the design objective.

5.4. Data worth of adding new pumping wells versus new observation wells

Adding more pumping wells or adding more observation wells showed different maximize firm information gain $\max \min \bar{I}_G(D)$ as illustrated in Fig. 6. Red circles show $\max \min \bar{I}_G(D)$ by increasing the number of new pumping wells up to five while keeping the number of new observation wells to be one. Yellow squares show $\max \min \bar{I}_G(D)$ by increasing the number of new observation wells up to five while keeping the number of new pumping wells to be one. Given the same number of new wells (e.g., one new pumping well or one new observation well), adding new observation wells always resulted in higher $\max \min \bar{I}_G(D)$ than adding new pumping wells. Fig. 6 suggested that experimental designs should emphasize new head data collection before exploring new pumping wells. We acknowledge that this observation may vary depending on the specific case. In this numerical example, where the model domain is relatively small, the addition of a single pumping well can potentially influence the entire model domain. Therefore, incorporating additional observation wells would be a more effective strategy than adding more pumping wells.

5.5. Robust experimental designs

Fig. 7 shows whether the first rank model can or cannot be discriminated from others, considering only adding up to five new head

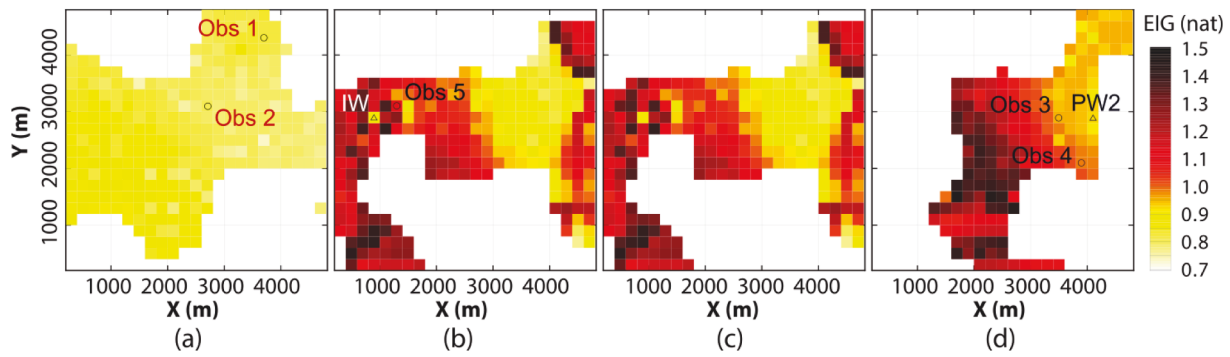


Fig. 3. Expected information gain (EIG) at each cell for (a) layer 2, (b) layer 3, (c) layer 4, and (d) layer 5. The existing pumping and injection wells remain the same. Symbol (Δ) refers to an existing pumping or injection well, and ($\circ$) refers to an existing head observation location.

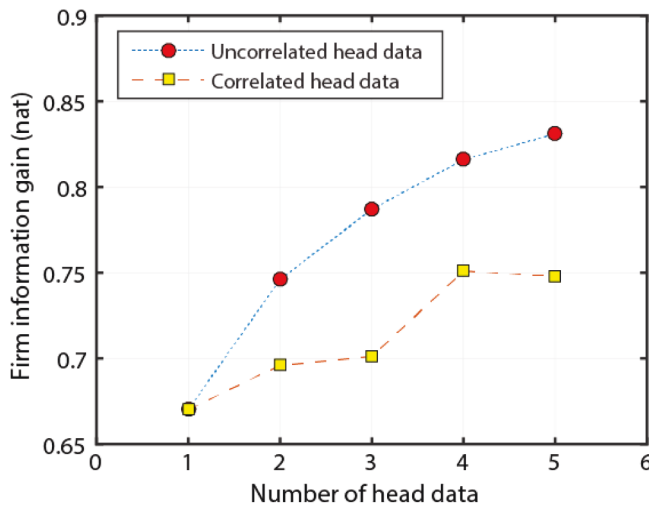


Fig. 4. Firm information gain by adding new head measurements. The existing pumping and injection wells remain the same.

data. BF_{12} is the Bayes factor of the first rank model to the second rank model, BF_{13} is the Bayes factor of the first rank model to the third rank model, and so forth. The model rank is determined by the likelihood after new data are acquired. The rank may change for different scenarios.

Fig. 7(a) presents the Bayes factors for the status quo (no new pumping well and no new head data). The result showed that adding a new head data discriminated the first rank model from the last three models. By adding two new head data, the first rank model was discriminated from the last two models. Adding up to five new head data only discriminated the first rank model from the last five models. Experimental designs using up to 5 new observations and the current system were unable to discriminate the most probable model from the other eight competing models.

If one new pumping well was added, Fig. 7(b) showed that adding a new head data discriminated the first rank model from six other models. Adding two new head data discriminated the first rank model from five other models. The results also indicate that increasing new head data increased the maximum firm information gain (see Fig. 6b), but might not increase the number of models to be discriminated against. Similar to Fig. 7(a), experimental designs using one new pumping well and up to five new observations were unable to discriminate the most probable model from the other eight competing models.

If two new pumping wells were added, the same most probable model can be identified by using two to five new head data as shown in Fig. 7(c), where the first rank model dominated all other models. The most probable model was the G_1B_2 model, which was the true model.

The robust experimental design found that two new pumping wells and two new head observation wells sufficed with firm information gain of 1.707 nat and reduce the entropy of the system to 0.041 nat. The minimum Bayes factor of 152.98 exceeded the selected threshold of 10. Given the optimal locations of the two pumping wells, we verified all possible locations of two new head observation wells produced entropy of the system less than 0.041 nat. All identified most probable models, which met the Bayes factor threshold, were the G_1B_2 model, the true model. This verification indicates that the same most probable model can be consistently identified regardless of sample locations.

6. Discussion

The presence of head data correlation, attributable to several factors like spatiotemporal location, model domain size, boundary conditions, and model parameterizations, showed significant impacts on firm information gain $\min_{\bar{G}}(D)$ and should be considered in experimental designs. This is because the most probable model tends to receive overwhelming posterior model probability (close to 100%) when the data size is large, and the data are assumed uncorrelated. This finding is consistent with Lu et al. (2013) that suggests accounting for the correlation of model data errors in the covariance matrix to avoid deriving unrealistic posterior model probabilities. For this study, it poses a serious concern that exaggerated $\min_{\bar{G}}(D)$ by assuming data uncorrelated may eventually fail the experimental designs due to low information gain in actual data collection.

To gain maximum firm information, this study found that the best locations to draw new pumping wells are in low hydraulic conductivity zones (i.e., layer 3 in this case study, see Fig. 5a and 5c) and the best locations to draw new observation wells are the areas that are far from the pumping wells (See Fig. 5d). This is because pumping in these areas tends to generate high variation in groundwater levels (in these low conductivity zones) and thereafter, provides higher expected information gain (in comparison with pumping in high conductivity zones). Therefore, this numerical example suggests drawing new pumping wells in the low conductivity zone and observe at a “far-enough distance” from the pumping wells to obtain firm information for model discrimination and identification. It is noted that different aquifer settings (e.g., boundary conditions) will result in different design outcomes.

Groundwater systems are highly heterogeneous and nonlinear. Different locations of pumping wells and observation wells yield different information given a design objective. Determining the best pumping and observation locations is an important step before any field data collection as pumping tests are costly and time-consuming. Considering only the observation part (e.g., adding new observation wells) was not a good strategy for this case study potentially because many new observation locations might yield similar information (i.e., did not help to increase firm information gain). Simultaneously accounting for both the observation part and the excitation part (e.g.,

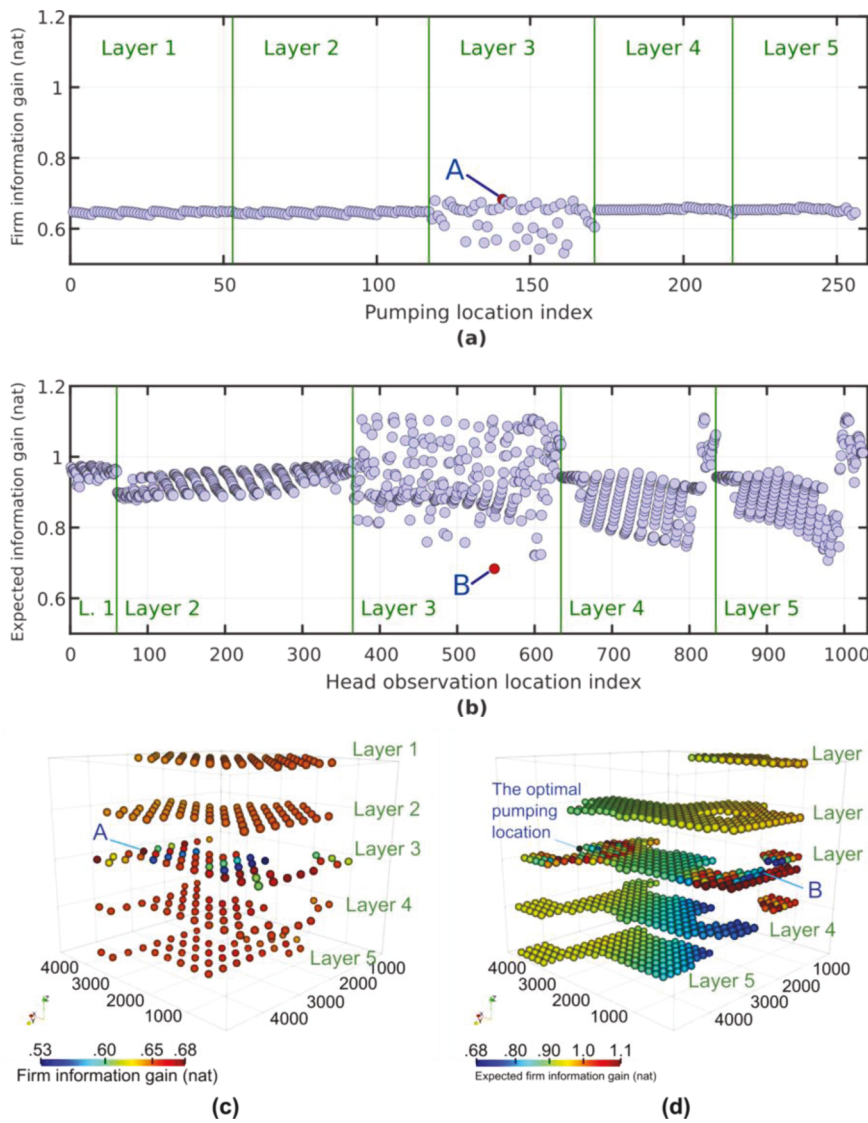


Fig. 5. Potential solutions for an experimental design seeking one new pumping well and one new head data. (a) Firm information gain for 256 potential locations to install a new pumping well. Letter A denotes the pumping location that gains maximum firm information gain; (b) Expected information gain for 1024 potential locations to draw a new groundwater head data given the optimal pumping well at circle A in (a). Letter B denotes the observation location that gives firm (minimal) information gain; (c) and (d) are three-dimensional views of (a) and (b), respectively.

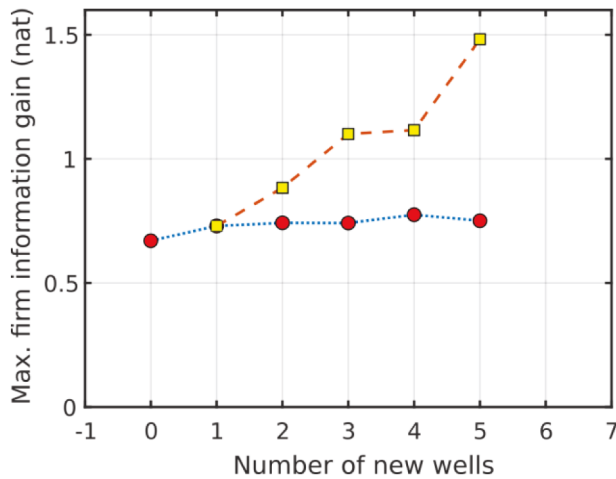


Fig. 6. Maximum firm information gain from experimental designs for seeking zero to five new pumping locations (red circles) and one new head data and seeking one new pumping location and one to five new head data (yellow squares).

adding new pumping wells) was found a more efficient way to obtain new additional information. The robust experimental design was succeeded in determining the optimal locations to draw new pumping wells and measure groundwater levels to achieve the design objective using the least number of wells. After the robust experimental design was succeeded, all other designs (using the same number of wells such as 2 pumping wells and two observation wells) identified the same most probable groundwater model which differs from the author's previous model discrimination criterion based on posterior model probability (Pham and Tsai, 2016, 2015) where the most probable model was varied by design alternatives. It is important to recognize that drilling a new pumping well is generally more expensive than drilling a new observation well, and groundwater managers are not typically interested in drilling a new pumping well solely for model discrimination purposes. Consequently, in a real-world application, it is more feasible to apply the method to the existing pumping network and concentrate on drilling new observation wells only.

The nested quadrature rule (Genz and Keister, 1996) was found an efficient approach to calculate high-dimensional integrals such as $E_{\Delta_D^{new}|M_i}[\ln q(\Delta_D^{new})]$ for deriving the expected information gain $\bar{I}_G$ in this study. Using 5 nodes and searching for one new observation at a time (the dimension of Δ_D^{new} is one), calculating $E_{\Delta_D^{new}|M_i}[\ln q(\Delta_D^{new})]$ required nine samples of Δ_D^{new} . Calculating $E_{\Delta_D^{new}|M_i}[\ln q(\Delta_D^{new})]$ required sample

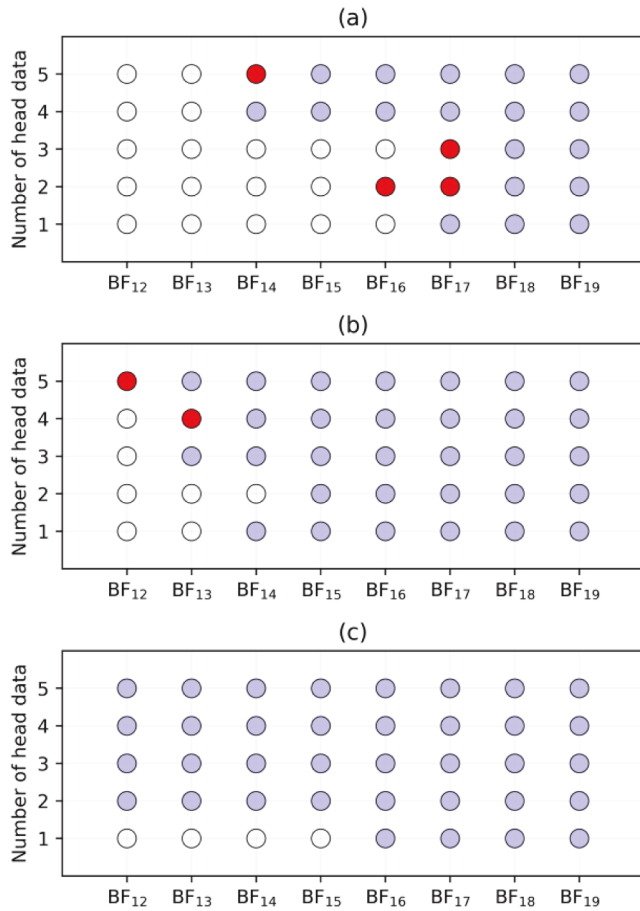


Fig. 7. Solutions (circles) of experimental designs that have Bayes factor (BF) greater than 10 for seeking one to five new head data with (a) no new pumping well and no new head data (status quo), (b) one new pumping location, and (c) two new pumping locations. Open circles for $BF < 5$. Red circles for $5 \leq BF < 10$. Blue circles for $BF \geq 10$.

sizes of 37, 93, 201, and 401 when the dimension Δ_D^{new} increased from 2 to 5. The computation time dramatically increased with Δ_D^{new} dimension. Yet, the number of samples required was small in comparison with the traditional Monte Carlo simulation approaches that usually require thousands of samples.

Solving the max-min programming problem to maximize the firm information gain was extremely time-consuming even with the hypothetical case study where a single model run was less than one minute. For example, using the parallel GA to search for two new pumping wells, the computation time for solving the max-min programming problem was 6.68, 7.86, 8.87, 11.5, and 16.4 h for Δ_D^{new} size to be 1, 2, 3, 4, and 5 using 80 cores. Computing time grew substantially by just increasing a few new observation data. The most time-consuming part came from $E_{\Delta_D^{new} | M_i} [\ln q(\Delta_D^{new})]$ calculations. Combining the parallel-sequential GA and the nested quadrature rule efficiently solved the time-consuming max-min optimization problem.

The presented methodology assumed that the probability distribution function of observable states (i.e., groundwater level) given a realization of model events follows a multivariate Gaussian distribution. This assumption may not hold for the nonlinear groundwater problem (e.g., the reactive transport model Shi et al., 2014) and may have an impact on the results of the robust experimental design. However, assuming the multivariate Gaussian allows transferring the complicated, multiple integrals of Eq. (7) into an easier form of Eqs. (9) and (10). These equations can accurately and efficiently be solved by utilizing the Cholesky decomposition and the nested quadrature rule. This Gaussian

assumption can be resolved by using Monte Carlo approaches such as the DREAM package (Vrugt, 2016) however, this approach requires high computational cost (e.g., requires thousands of sample sizes) and is not suitable for solving the max-min optimization program, even with the hypothetical numerical example in this study.

The robust experimental design may not guarantee a global optimal solution when the search dimension increases (e.g., greater than five) because solving the nonlinear and non-convex max-min problem is challenging and the computation time increases exponentially with the increase in search dimension. To increase the global search capability in finding the global optimal solution, one can increase the population size in the GA. However, this will significantly increase the computation time. Therefore, we limited our search dimensions to less than five (i.e., less than five new wells) to avoid potential numerical issues in the numerical calculation of $E_{\Delta_D^{new} | M_i} [\ln q(\Delta_D^{new})]$ and make our optimization problem trackable. The computational burden of the robust experimental design may be reduced by using surrogate modeling approaches (also known as reduced-order model) where a complex model is replaced with an approximate, but computationally-efficient model (Ushijima and Yeh, 2013; Asher et al., 2015; Jefferson et al., 2015; Yin and Tsai, 2020).

For the numerical example in this study, the true model was added to a pool of 9 competing models for verification purposes. When the true model was removed from the robust experimental design presented in Section 5.5, the minimum Bayes factor decreased to 44.73, which was almost 3.42 times less than when the true model was included. Consequently, the most probable model became G1B3. It is worth noting that a true model is typically unknown, and including it does not reflect a realistic scenario. However, whether or not the true model is included has no impact on the methodology, but it could affect the optimal locations for pumping and observation wells and the number of wells used in the robust experimental design. If a model that is close to the true model is included among the competing models, there is a higher likelihood of obtaining a robust experimental design with lower costs (i.e., using fewer pumping and observation wells).

While achieving the minimum Bayes factor of 10 was possible in the numerical example, expensive experimental designs may be resulted for real-world applications. It is analyst's discretion in setting the Bayes factor threshold to achieve a certain level of model discrimination (Jeffreys, 1998).

7. Conclusions

Incorporating the concept of firm information gain (nominally, the minimum expected information gain) in the robust experimental design reveals the minimum information required while acquiring new data to identify the most probable model. This is a robust approach and places the experimental design in the context of information theory. The Bayes factor threshold of 10 in the robust experimental design ensures that new data provides strong evidence to discriminate the most probable model from other candidate models.

Considering a full covariance matrix of data substantially affects the calculation of firm information gain. The full covariance matrix in this study is comprehensive, which accounts for measurement errors and errors from model conceptualization and model parameters. The Bayesian model averaging method and the Monte Carlo approach are suitable to quantify covariances due to conceptual uncertainty and parametric uncertainty, respectively. Neglecting covariances between data tends to exaggerate true firm information gain and results in unrealistic Bayes factor values.

Maximizing the firm information gain in the robust experimental design is a unique choice and results in more direct solutions than those from maximizing the value of the Box-Hill discrimination function (an upper bound of the expected information gain). However, calculating the firm information gain is not straightforward. This study found that

the Genz-Keister (Genz and Keister, 1996) method can efficiently calculate the multi-dimensional integral in the expected information gain when data size is small. This study also found that the parallel-sequential genetic algorithm scheme is an efficient scheme to maximize the firm information gain, which is posed as a max–min programming problem.

Through the numerical groundwater example, this study found that (1) maximum firm information gain grows faster with the size of new head data than with the number of new pumping wells. In other words, this study suggests that experimental designs should emphasize new head data collection before exploring new pumping wells for this specific numerical example; and (2) the same most probable groundwater model could be identified as long as solutions of experimental designs result in higher than firm information gain and satisfy a Bayes factor threshold.

Future research should focus on evaluating the impacts of Gaussian assumption on the robust experimental design and comparing the informatics metric proposed in this study with available metrics presented in the introduction section.

CRediT authorship contribution statement

Hai V. Pham: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Visualization, Writing – original

draft. **Frank T.-C. Tsai:** Conceptualization, Methodology, Writing – original draft, Resources, Supervision, Project administration, Funding acquisition.

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.

Data availability

Data will be made available on request.

Acknowledgments

This study was supported in part by the U.S. Geological Survey under Grant/Cooperative Agreement No. G21AP10577 and the U.S. National Science Foundation (Award No. 2019561). LSU High Performance Computing and LSU Center for Computation & Technology are acknowledged for providing a supercomputer for this study. All of the numerical data are provided in the tables and figures produced by solving the equations in the paper.

Appendix A: Derivations for expected information gain

The expected information gain $\bar{I}_G$ in the Eq. (4) can be further expanded as:

$$\bar{I}_G = \int_{-\infty}^{+\infty} \left[\sum_{i=1}^m \Pr(M_i | \Delta^{\text{new}}) \ln \Pr(M_i | \Delta^{\text{new}}) \right] q(\Delta_D^{\text{new}}) d\Delta_D^{\text{new}} - \sum_{i=1}^m \Pr(M_i | \Delta^{\text{obs}}) \ln \Pr(M_i | \Delta^{\text{obs}}). \quad (\text{A1})$$

where

$$q(\Delta_D^{\text{new}}) = \sum_{j=1}^m \Pr(M_j | \Delta^{\text{obs}}) p(\Delta_D^{\text{new}} | M_j), \quad (\text{A2})$$

and

$$\Pr(M_i | \Delta^{\text{new}}) = \frac{p(\Delta_D^{\text{new}} | M_i) \Pr(M_i | \Delta^{\text{obs}})}{q(\Delta_D^{\text{new}})}. \quad (\text{A3})$$

Inserting Eqs. (A3) and (A2) into (A1), we have:

$$\begin{aligned} \bar{I}_G &= \int_{-\infty}^{+\infty} \sum_{i=1}^m p(\Delta_D^{\text{new}} | M_i) \Pr(M_i | \Delta^{\text{obs}}) \ln \Pr(M_i | \Delta^{\text{obs}}) d\Delta_D^{\text{new}} \\ &+ \int_{-\infty}^{+\infty} \left[\sum_{i=1}^m p(\Delta_D^{\text{new}} | M_i) \Pr(M_i | \Delta^{\text{obs}}) \ln \frac{p(\Delta_D^{\text{new}} | M_i)}{q(\Delta_D^{\text{new}})} \right] d\Delta_D^{\text{new}} \\ &- \sum_{i=1}^m \Pr(M_i | \Delta^{\text{obs}}) \ln \Pr(M_i | \Delta^{\text{obs}}) \end{aligned} \quad (\text{A4})$$

Then

$$\bar{I}_G = \sum_{i=1}^m \Pr(M_i | \Delta^{\text{obs}}) \int_{-\infty}^{\infty} p(\Delta_D^{\text{new}} | M_i) \ln \frac{p(\Delta_D^{\text{new}} | M_i)}{q(\Delta_D^{\text{new}})} d\Delta_D^{\text{new}}. \quad (\text{A5})$$

Appendix B: Expected information gain for correlated Multivariate-Gaussian data

The expected information gain is

$$\begin{aligned} \bar{I}_G = & \sum_{i=1}^m \Pr(M_i | \Delta^{\text{obs}}) \int_{-\infty}^{\infty} p(\Delta_D^{\text{new}} | M_i) \ln p(\Delta_D^{\text{new}} | M_i) d\Delta_D^{\text{new}} \\ & - \sum_{i=1}^m \Pr(M_i | \Delta^{\text{obs}}) \int_{-\infty}^{\infty} p(\Delta_D^{\text{new}} | M_i) \ln q(\Delta_D^{\text{new}}) d\Delta_D^{\text{new}} \end{aligned} \quad (\text{B1})$$

The probability density function is

$$p(\Delta_D^{\text{new}} | M_i) = (2\pi)^{-\frac{N}{2}} |\hat{\Sigma}_i|^{-\frac{1}{2}} e^{-\frac{1}{2}(\Delta_D^{\text{new}} - \Delta_i)^T \hat{\Sigma}_i^{-1} (\Delta_D^{\text{new}} - \Delta_i)}, \quad (\text{B2})$$

where $\hat{\Sigma}_i = \Sigma + \Sigma_\epsilon$ is the total covariance matrix for Δ_D^{new} when model M_i is used

Substituting $p(\Delta_D^{\text{new}} | M_i)$ in (B1) with (B2), the first integral in (B1) is

$$\begin{aligned} & \int_{-\infty}^{\infty} p(\Delta_D^{\text{new}} | M_i) \ln p(\Delta_D^{\text{new}} | M_i) d\Delta_D^{\text{new}} \\ &= E[\ln p(\Delta_D^{\text{new}} | M_i)] \\ &= E\left[\ln\left[(2\pi)^{-\frac{N}{2}} |\hat{\Sigma}_i|^{-\frac{1}{2}} e^{-\frac{1}{2}(\Delta_D^{\text{new}} - \Delta_i)^T \hat{\Sigma}_i^{-1} (\Delta_D^{\text{new}} - \Delta_i)}\right]\right] \\ &= \ln\left[(2\pi)^{-\frac{N}{2}} |\hat{\Sigma}_i|^{-\frac{1}{2}}\right] - \frac{1}{2} E\left[(\Delta_D^{\text{new}} - \Delta_i)^T \hat{\Sigma}_i^{-1} (\Delta_D^{\text{new}} - \Delta_i)\right] \\ &= \ln\left[(2\pi)^{-\frac{N}{2}} |\hat{\Sigma}_i|^{-\frac{1}{2}}\right] - \frac{N}{2} \end{aligned} \quad (\text{B3})$$

where E is the expectation operator and

$$\begin{aligned} & E\left[(\Delta_D^{\text{new}} - \Delta_i)^T \hat{\Sigma}_i^{-1} (\Delta_D^{\text{new}} - \Delta_i)\right] \\ &= E\left[\text{tr}\left((\Delta_D^{\text{new}} - \Delta_i)^T \hat{\Sigma}_i^{-1} (\Delta_D^{\text{new}} - \Delta_i)\right)\right] \\ &= E\left[\text{tr}\left(\hat{\Sigma}_i^{-1} (\Delta_D^{\text{new}} - \Delta_i) (\Delta_D^{\text{new}} - \Delta_i)^T\right)\right] \\ &= \text{tr}\left(E\left[\hat{\Sigma}_i^{-1} (\Delta_D^{\text{new}} - \Delta_i) (\Delta_D^{\text{new}} - \Delta_i)^T\right]\right) \\ &= \text{tr}\left(\hat{\Sigma}_i^{-1} E\left[(\Delta_D^{\text{new}} - \Delta_i) (\Delta_D^{\text{new}} - \Delta_i)^T\right]\right) \\ &= \text{tr}(\hat{\Sigma}_i^{-1} \hat{\Sigma}_i) = \text{tr}(\mathbf{I}) = N \end{aligned} \quad (\text{B4})$$

where tr is the trace of a square matrix and $\mathbf{I}$ is the identity matrix.

The second integral in (B1) is

$$E_{\Delta_D^{\text{new}} | M_i} [\ln q(\Delta_D^{\text{new}})] = \int_{-\infty}^{\infty} p(\Delta_D^{\text{new}} | M_i) \ln q(\Delta_D^{\text{new}}) d\Delta_D^{\text{new}} \quad (\text{B5})$$

Appendix C: Calculate $E_{\Delta_D^{\text{new}} | M_i} [\ln q(\Delta_D^{\text{new}})]$

The correlated new data Δ_D^{new} are transformed into uncorrelated random variables, $\mathbf{x}$ by the Cholesky decomposition since the covariance matrices $\hat{\Sigma}_i$ are positive definite and symmetric. Let $\hat{\Sigma}_i = \mathbf{L}_i \mathbf{L}_i^T$ and $\Delta_D^{\text{new}} = \mathbf{L}_i \mathbf{x} + \Delta_i$, where $\mathbf{L}_i$ is a lower triangular matrix with real and positive diagonal entries. The random variables $\mathbf{x}$ have zero means and an identity matrix for the covariance matrix. Δ_D^{new} obtained through $\mathbf{x}$ include measurement errors. The probability density function $p(\Delta_D^{\text{new}} | M_i)$ in terms of $\mathbf{x}$ is

$$\begin{aligned} p(\Delta_D^{\text{new}} | M_i) &= (2\pi)^{-\frac{N}{2}} |\hat{\Sigma}_i|^{-\frac{1}{2}} e^{-\frac{1}{2}(\Delta_D^{\text{new}} - \Delta_i)^T \hat{\Sigma}_i^{-1} (\Delta_D^{\text{new}} - \Delta_i)} \\ &= (2\pi)^{-\frac{N}{2}} |\mathbf{L}_i \mathbf{L}_i^T|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{L}_i \mathbf{x} + \Delta_i - \Delta_i)^T (\mathbf{L}_i \mathbf{L}_i^T)^{-1} (\mathbf{L}_i \mathbf{x})} \\ &= (2\pi)^{-\frac{N}{2}} |\mathbf{L}_i|^{-1} e^{-\frac{1}{2} \mathbf{x}^T \mathbf{x}} \end{aligned} \quad (\text{C1})$$

Therefore, $q(\Delta_D^{\text{new}})$ can be calculated in terms of $\mathbf{x}$:

$$\begin{aligned} q(\Delta_D^{\text{new}}) &= \sum_{j=1}^m \Pr(M_j | \Delta^{\text{obs}}) p(\Delta_D^{\text{new}} | M_j) \\ &= \sum_{j=1}^m \Pr(M_j | \Delta^{\text{obs}}) (2\pi)^{-\frac{N}{2}} |\hat{\Sigma}_j|^{-\frac{1}{2}} e^{-\frac{1}{2}(\Delta_D^{\text{new}} - \Delta_j)^T \hat{\Sigma}_j^{-1} (\Delta_D^{\text{new}} - \Delta_j)} \\ &= \sum_{j=1}^m \Pr(M_j | \Delta^{\text{obs}}) (2\pi)^{-\frac{N}{2}} |\hat{\Sigma}_j|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{L}_j \mathbf{x} + \Delta_j - \Delta_j)^T \hat{\Sigma}_j^{-1} (\mathbf{L}_j \mathbf{x} + \Delta_j - \Delta_j)} \\ &= q(\mathbf{x}) \end{aligned} \quad (\text{C2})$$

The transformation from $d\Delta_D^{\text{new}}$ to $d\mathbf{x}$ needs the Jacobian, which is the determinant of $\mathbf{L}_i: d\Delta_D^{\text{new}} = \text{abs}(|\mathbf{L}_i|)d\mathbf{x}$, where $\text{abs}(|\mathbf{L}_i|)$ is the absolute value of the determinant of $\mathbf{L}_i$. Since positive diagonal entries in $\mathbf{L}_i$, it becomes $\Delta_D^{\text{new}} = |\mathbf{L}_i|d\mathbf{x}$. Therefore,

$$\begin{aligned} E_{\Delta_D^{\text{new}}|\mathbf{M}_i}[\ln q(\Delta_D^{\text{new}})] &= \int_{-\infty}^{\infty} p(\Delta_D^{\text{new}}|\mathbf{M}_i) \ln q(\Delta_D^{\text{new}}) d\Delta_D^{\text{new}} \\ &= \int_{-\infty}^{\infty} (2\pi)^{-\frac{N}{2}} |\mathbf{L}_i|^{-1} e^{-\frac{1}{2}\mathbf{x}^T \mathbf{x}} \ln q(\mathbf{x}) |\mathbf{L}_i| d\mathbf{x} \\ &= (2\pi)^{-\frac{N}{2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\mathbf{x}^T \mathbf{x}} \ln q(\mathbf{x}) d\mathbf{x} \end{aligned} \quad (\text{C3})$$

The nested quadrature rule for N-dimensional numerical integration (Genz and Keister, 1996) is:

$$E_{\Delta_D^{\text{new}}|\mathbf{M}_i}[\ln q(\Delta_D^{\text{new}})] = \sum_{r=1}^R \ln [q(x_{r,1}, \dots, x_{r,N})] w_r, \quad (\text{C4})$$

where N is the number of dimensions of Δ_D^{new} , $r = 1, \dots, R$ where R is the number of nodes after removing duplicates (Heiss and Winschel, 2008), and x_r is a set of nodes and w_r is a set of weights. Sampling $x_{r,1}, \dots, x_{r,N}$ from a sparse grid we get Δ_D^{new} .

References

- Alfonso, L., Lobbrecht, A., Price, R., 2010. Information theory-based approach for location of monitoring water level gauges in polders. *Water Resour. Res.* 46, W03528. <https://doi.org/10.1029/2009WR008101>.
- Alley, W.M., Healy, R.W., LaBaugh, J.W., Reilly, T.E., 2002. Flow and storage in groundwater systems. *Science* 296, 1985–1990. <https://doi.org/10.1126/science.1067123>.
- Altmann-Dieses, A.E., Schlöder, J.P., Bock, H.G., Richter, O., 2002. Optimal experimental design for parameter estimation in column outflow experiments. *Water Resour. Res.* 38, 1186. <https://doi.org/10.1029/2001WR000358>.
- Asher, M.J., Croke, B.F.W., Jakeman, A.J., Peeters, L.J.M., 2015. A review of surrogate models and their application to groundwater modeling. *Water Resour. Res.* 51, 5957–5973. <https://doi.org/10.1002/2015WR016967>.
- Ben-Zvi, M., Berkowitz, B., Kesler, S., 1988. Pre-posterior analysis as a tool for data evaluation: application to aquifer contamination. *Water Resour. Manag.* 2, 11–20. <https://doi.org/10.1007/BF00421927>.
- Bode, F., Reed, P., Reuschen, S., Nowak, W., 2019. Search space representation and reduction methods to enhance multiobjective water supply monitoring design. *Water Resour. Res.* 55, 2257–2278. <https://doi.org/10.1029/2018WR023133>.
- Box, G.E.P., Hill, W.J., 1967. Discrimination among mechanistic models. *Technometrics* 9 (1), 57–71. <https://doi.org/10.1080/00401706.1967.10490441>.
- Bredheoef, J., 2005. The conceptualization model problem-surprise. *Hydrogeol. J.* 13, 37–46. <https://doi.org/10.1007/s10040-004-0430-5>.
- Carroll, D.L., 1996. Chemical laser modeling with genetic algorithms. *AIAA J.* 34, 338–346. <https://doi.org/10.2514/3.13069>.
- Chadalavada, S., Datta, B., 2008. Dynamic optimal monitoring network design for transient transport of pollutants in groundwater aquifers. *Water Resour. Manag.* 22, 651–670. <https://doi.org/10.1007/s11269-007-9184-x>.
- Chang, L.F., Sun, N.Z., Yeh, W.W.G., 2005. Optimal observation network design for parameter structure identification in groundwater modeling. *Water Resour. Res.* 41, W03002. <https://doi.org/10.1029/2004WR003514>.
- Cieniawski, S.E., Eheart, J.W., Ranjithan, S., 1995. Using genetic algorithms to solve a multiobjective groundwater monitoring problem. *Water Resour. Res.* 31, 399–409. <https://doi.org/10.1029/94WR02039>.
- Cleveland, T., Yeh, W., 1990. Sampling network design for transport parameter identification. *J. Water Resour. Plan. Manag.* 116, 764–783. [https://doi.org/10.1061/\(ASCE\)0733-9496\(1990\)116:6\(764\)](https://doi.org/10.1061/(ASCE)0733-9496(1990)116:6(764)).
- Dausman, A.M., Doherty, J., Langevin, C.D., Sukop, M.C., 2010. Quantifying data worth toward reducing predictive uncertainty. *Ground Water* 48, 729–740. <https://doi.org/10.1111/j.1745-6584.2010.00679.x>.
- Davis, D.R., Dvoranchik, W.M., 1971. Evaluation of the worth of additional data. *J. Am. Water Resour. Assoc.* 7, 700–707. <https://doi.org/10.1111/j.1752-1688.1971.tb04981.x>.
- Dhar, A., Datta, B., 2007. Multiobjective design of dynamic monitoring networks for detection of groundwater pollution. *J. Water Resour. Plan. Manag.* 133, 329–338. [https://doi.org/10.1061/\(ASCE\)0733-9496\(2007\)133:4\(329\)](https://doi.org/10.1061/(ASCE)0733-9496(2007)133:4(329)).
- Dokou, Z., Pinder, G.F., 2009. Optimal search strategy for the definition of a DNAPL source. *J. Hydrol.* 376, 542–556. <https://doi.org/10.1016/j.jhydrol.2009.07.062>.
- Draper, D., 1995. Assessment and propagation of model uncertainty. *J. R. Stat. Soc. Ser. B Methodol.* 45–97.
- Elshall, A.S., Tsai, F.T.C., Hanor, J.S., 2013. Indicator geostatistics for reconstructing Baton Rouge aquifer-fault hydrostratigraphy, Louisiana, USA. *Hydrogeol. J.* 21, 1731–1747. <https://doi.org/10.1007/s10040-013-1037-5>.
- Feyen, L., Gorelick, S.M., 2005. Framework to evaluate the worth of hydraulic conductivity data for optimal groundwater resources management in ecologically sensitive areas. *Water Resour. Res.* 41, W03019. <https://doi.org/10.1029/2003WR002901>.
- Freer, J., Beven, K., Ambrose, B., 1996. Bayesian estimation of uncertainty in runoff prediction and the value of data: an application of the GLUE approach. *Water Resour. Res.* 32, 2161–2173. <https://doi.org/10.1029/95WR03723>.
- Gates, J.S., Kisiel, C.C., 1974. Worth of additional data to a digital computer model of a groundwater basin. *Water Resour. Res.* 10, 1031–1038. <https://doi.org/10.1029/WR010i005p01031>.
- Genz, A., Keister, B.D., 1996. Fully symmetric interpolatory rules for multiple integrals over infinite regions with Gaussian weight. *J. Comput. Appl. Math.* 71, 299–309. [https://doi.org/10.1016/0377-0427\(95\)00232-4](https://doi.org/10.1016/0377-0427(95)00232-4).
- Giordano, M., 2009. Global groundwater? Issues and solutions. *Annu. Rev. Environ. Resour.* 34, 153–178. <https://doi.org/10.1146/annurev.enviro.030308.100251>.
- Gleeson, T., Wada, Y., Bierkens, M.F.P., van Beek, L.P.H., 2012. Water balance of global aquifers revealed by groundwater footprint. *Nature* 488, 197–200. <https://doi.org/10.1038/nature11295>.
- 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.
- Hansen, N., Ostermeier, A., 2001. Completely derandomized self-adaptation in evolution strategies. *Evol. Comput.* 9, 159–195. <https://doi.org/10.1162/106365601750190398>.
- Harbaugh, A.W., 2005. MODFLOW-2005, The U.S. Geological Survey Modular Groundwater Model - the Ground-Water Flow process. Book 6. Modeling Techniques, Section A. Ground Water. US Dept. of the Interior, US Geological Survey, Virginia, p. 253.
- Hassan, A., 2003. Long-term Monitoring Plan for the Central Nevada Test Area. U.S. Department of Energy, Office of Scientific and Technical Information, Oak Ridge, TN.
- Heiss, F., Winschel, V., 2008. Likelihood approximation by numerical integration on sparse grids. *J. Econ.* 144, 62–80. <https://doi.org/10.1016/j.jeconom.2007.12.004>.
- Herrera, G.S., Pinder, G.F., 2005. Space-time optimization of groundwater quality sampling networks. *Water Resour. Res.* 41, W12407. <https://doi.org/10.1029/2004WR003626>.
- Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T., 1999. Bayesian model averaging: a tutorial. *Stat. Sci.* 14, 382–401.
- Højberg, A., Refsgaard, J., 2005. Model uncertainty-parameter uncertainty versus conceptual models. *Water Sci. Technol. J. Int. Assoc. Water Pollut. Res.* 52, 177–186.
- Hsu, N.S., Yeh, W.W.G., 1989. Optimum experimental design for parameter identification in groundwater hydrology. *Water Resour. Res.* 25, 1025–1040. <https://doi.org/10.1029/WR025i005p01025>.
- Hudak, P.F., Loaiciga, H.A., 1992. A location modeling approach for groundwater monitoring network augmentation. *Water Resour. Res.* 28, 643–649. <https://doi.org/10.1029/91WR02851>.
- James, B., Gwo, J., Toran, L., 1996. Risk-cost decision framework for aquifer remediation design. *J. Water Resour. Plan. Manag.* 122, 414–420. [https://doi.org/10.1061/\(ASCE\)0733-9496\(1996\)122:6\(414\)](https://doi.org/10.1061/(ASCE)0733-9496(1996)122:6(414)).
- James, B.R., Gorelick, S.M., 1994. When enough is enough: the worth of monitoring data in aquifer remediation design. *Water Resour. Res.* 30, 3499–3513. <https://doi.org/10.1029/94WR01972>.
- Janssen, G.M.C.M., Valstar, J.R., van der Zee, S.E.A.T.M., 2008. Measurement network design including traveltime determinations to minimize model prediction uncertainty. *Water Resour. Res.* 44, W02405. <https://doi.org/10.1029/2006WR005462>.
- Jefferson, J.L., Gilbert, J.M., Constantine, P.G., Maxwell, R.M., 2015. Active subspaces for sensitivity analysis and dimension reduction of an integrated hydrologic model. *Comput. Geosci.* 83, 127–138. <https://doi.org/10.1016/j.cageo.2015.07.001>.
- Jeffreys, H., 1998. *Theory of Probability*, 3rd ed. Oxford University Press, USA.
- Kikuchi, C., 2017. Toward increased use of data worth analyses in groundwater studies. *Groundwater* 55, 670–673. <https://doi.org/10.1111/gwat.12562>.
- Kikuchi, C.P., Ferré, T.P.A., Vrugt, J.A., 2015. On the optimal design of experiments for conceptual and predictive discrimination of hydrologic system models. *Water Resour. Res.* 51, 4454–4481. <https://doi.org/10.1002/2014WR016795>.
- Kim, K.H., Lee, K.K., 2007. Optimization of groundwater-monitoring networks for identification of the distribution of a contaminant plume. *Stoch. Environ. Res. Risk Assess.* 21, 785–794. <https://doi.org/10.1007/s00477-006-0094-x>.

- Knopman, D.S., Voss, C.I., 1988. Discrimination among one-dimensional models of solute transport in porous media: implications for sampling design. *Water Resour. Res.* 24, 1859–1876. <https://doi.org/10.1029/WR024i01p01859>.
- Knopman, D.S., Voss, C.I., Garabedian, S.P., 1991. Sampling design for groundwater solute transport: tests of methods and analysis of Cape Cod tracer test data. *Water Resour. Res.* 27, 925–949. <https://doi.org/10.1029/90WR02657>.
- Kollat, J.B., Reed, P.M., Maxwell, R.M., 2011. Many-objective groundwater monitoring network design using bias-aware ensemble Kalman filtering, evolutionary optimization, and visual analytics. *Water Resour. Res.* 47, W02529. <https://doi.org/10.1029/2010WR009194>.
- Leube, P.C., Geiges, A., Nowak, W., 2012. Bayesian assessment of the expected data impact on prediction confidence in optimal sampling design. *Water Resour. Res.* 48. <https://doi.org/10.1029/2010WR010137> n/a-n/a.
- Loaiciga, H., Charbeneau, R., Everett, L., Fogg, G., Hobbs, B., Rouhani, S., 1992. Review of ground-water quality monitoring network design. *J. Hydraul. Eng.* 118, 11–37. [https://doi.org/10.1061/\(ASCE\)0733-9429\(1992\)118:1\(11\)](https://doi.org/10.1061/(ASCE)0733-9429(1992)118:1(11)).
- Lu, D., Ye, M., Meyer, P.D., Curtis, G.P., Shi, X., Niu, X.F., Yabusaki, S.B., 2013. Effects of error covariance structure on estimation of model averaging weights and predictive performance. *Water Resour. Res.* 49, 6029–6047. <https://doi.org/10.1002/wrcr.20441>.
- McKinney, D.C., Loucks, D.P., 1992. Network design for predicting groundwater contamination. *Water Resour. Res.* 28, 133–147. <https://doi.org/10.1029/91WR02397>.
- Meyer, P.D., Brill, E.D., 1988. A method for locating wells in a groundwater monitoring network under conditions of uncertainty. *Water Resour. Res.* 24, 1277–1282. <https://doi.org/10.1029/WR024i008p01277>.
- Meyer, P.D., Valocchi, A.J., Eheart, J.W., 1994. Monitoring network design to provide initial detection of groundwater contamination. *Water Resour. Res.* 30, 2647–2659. <https://doi.org/10.1029/94WR00872>.
- Minsker, B., 2003. *Long-Term Groundwater Monitoring Design: The State of the Art*. American Society of Civil Engineers, Reston, VA, p. 103.
- Mogheir, Y., Singh, V.P., 2002. Application of information theory to groundwater quality monitoring networks. *Water Resour. Manag.* 16, 37–49. <https://doi.org/10.1023/A:1015511811686>.
- Mogheir, Y., Singh, V.P., de Lima, J.L.M.P., 2006. Spatial assessment and redesign of a groundwater quality monitoring network using entropy theory, Gaza Strip, Palestine. *Hydrogeol. J.* 14, 700–712. <https://doi.org/10.1007/s10040-005-0464-3>.
- Neuman, S.P., Xue, L., Ye, M., Lu, D., 2012. Bayesian analysis of data-worth considering model and parameter uncertainties. *Adv. Water Resour.* 36, 75–85. <https://doi.org/10.1016/j.advwatres.2011.02.007>. Special Issue on Uncertainty Quantification and Risk Assessment.
- Norberg, T., Rosén, L., 2006. Calculating the optimal number of contaminant samples by means of data worth analysis. *Environmetrics* 17, 705–719. <https://doi.org/10.1002/env.787>.
- Nordqvist, R., Voss, C.I., 1996. A simulation-based approach for designing effective field-sampling programs to evaluate contamination risk of groundwater supplies. *Hydrogeol. J.* 4, 23–39. <https://doi.org/10.1007/s100400050081>.
- Nowak, W., de Barros, F.P.J., Rubin, Y., 2010. Bayesian geostatistical design: task-driven optimal site investigation when the geostatistical model is uncertain. *Water Resour. Res.* 46, W03535. <https://doi.org/10.1029/2009WR008312>.
- Nowak, W., Guthke, A., 2016. Entropy-based experimental design for optimal model discrimination in the geosciences. *Entropy* 18, 409. <https://doi.org/10.3390/e18110409>.
- Pham, H.V., Tsai, F.T.C., 2016. Optimal observation network design for conceptual model discrimination and uncertainty reduction. *Water Resour. Res.* 52, 1245–1264. <https://doi.org/10.1002/2015WR017474>.
- Pham, H.V., Tsai, F.T.C., 2015. Bayesian experimental design for identification of model propositions and conceptual model uncertainty reduction. *Adv. Water Resour.* 83, 148–159. <https://doi.org/10.1016/j.advwatres.2015.05.024>.
- Poeter, E., Anderson, D., 2005. Multimodel ranking and inference in ground water modeling. *Ground Water* 43, 597–605. <https://doi.org/10.1111/j.1745-6584.2005.0061.x>.
- Reed, P., Minsker, B., Valocchi, A.J., 2000. Cost-effective long-term groundwater monitoring design using a genetic algorithm and global mass interpolation. *Water Resour. Res.* 36, 3731–3741. <https://doi.org/10.1029/2000WR900232>.
- Reichard, E.G., Evans, J.S., 1989. Assessing the value of hydrogeologic information for risk-based remedial action decisions. *Water Resour. Res.* 25, 1451–1460. <https://doi.org/10.1029/WR025i007p01451>.
- Rojas, R., Feyen, L., Batelaan, O., Dassargues, A., 2010. On the value of conditioning data to reduce conceptual model uncertainty in groundwater modeling. *Water Resour. Res.* 46, W08520. <https://doi.org/10.1029/2009WR008822>.
- Sciortino, A., Harmon, T.C., Yeh, W.W.G., 2002. Experimental design and model parameter estimation for locating a dissolving dense nonaqueous phase liquid pool in groundwater. *Water Resour. Res.* 38. <https://doi.org/10.1029/2000WR000134>, 15–1–15–9.
- Shannon, C., 1948. A mathematical theory of communication. *Bell Syst. Tech. J.* 27 (3), 379–423.
- Shi, X., Ye, M., Curtis, G.P., Miller, G.L., Meyer, P.D., Kohler, M., Yabusaki, S., Wu, J., 2014. Assessment of parametric uncertainty for groundwater reactive transport modeling. *Water Resour. Res.* 50, 4416–4439. <https://doi.org/10.1002/2013WR013755>.
- Siade, A.J., Hall, J., Karelse, R.N., 2017. A practical, robust methodology for acquiring new observation data using computationally expensive groundwater models: robust experimental design for groundwater models. *Water Resour. Res.* 53, 9860–9882. <https://doi.org/10.1002/2017WR020814>.
- Siebert, S., Burke, J., Faures, J.M., Frenken, K., Hoogeveen, J., Döll, P., Portmann, F.T., 2010. Groundwater use for irrigation - a global inventory. *Hydrol. Earth Syst. Sci.* 14, 1863–1880. <https://doi.org/10.5194/hess-14-1863-2010>.
- Sohn, M.D., Small, M.J., 2000. Reducing uncertainty in site characterization using Bayes Monte Carlo methods. *J. Environ. Eng.* 126, 893.
- Storck, P., Eheart, J.W., Valocchi, A.J., 1997. A method for the optimal location of monitoring wells for detection of groundwater contamination in three-dimensional heterogeneous aquifers. *Water Resour. Res.* 33, 2081–2088. <https://doi.org/10.1029/97WR01704>.
- Sun, N.Z., 1994. *Inverse Problems in Groundwater Modeling*, 1994th ed. Springer.
- Sun, N.Z., Yeh, W.W.G., 2007. Development of objective-oriented groundwater models: 2. Robust experimental design. *Water Resour. Res.* 43, W02421. <https://doi.org/10.1029/2006WR004888>.
- Tiedeman, C.R., Ely, D.M., Hill, M.C., O'Brien, G.M., 2004. A method for evaluating the importance of system state observations to model predictions, with application to the Death Valley regional groundwater flow system. *Water Resour. Res.* 40, W12411. <https://doi.org/10.1029/2004WR003313>.
- Tiedeman, C.R., Hill, M.C., D'Agnese, F.A., Faunt, C.C., 2003. Methods for using groundwater model predictions to guide hydrogeologic data collection, with application to the Death Valley regional groundwater flow system. *Water Resour. Res.* 39, 1010. <https://doi.org/10.1029/2001WR001255>.
- Ushijima, T.T., Yeh, W.W.G., 2015. Experimental design for estimating unknown hydraulic conductivity in an aquifer using a genetic algorithm and reduced order model. *Adv. Water Resour.* 86, 193–208. <https://doi.org/10.1016/j.advwatres.2015.09.029>.
- Ushijima, T.T., Yeh, W.W.G., 2013. Experimental design for estimating unknown groundwater pumping using genetic algorithm and reduced order model. *Water Resour. Res.* 49, 6688–6699. <https://doi.org/10.1002/wrcr.20513>.
- Usunoff, E., Carrera, J., Mousavi, S.F., 1992. An approach to the design of experiments for discriminating among alternative conceptual models. *Adv. Water Resour.* 15, 199–214. [https://doi.org/10.1016/0309-1708\(92\)90024-V](https://doi.org/10.1016/0309-1708(92)90024-V).
- Vrugt, J.A., 2016. Markov chain Monte Carlo simulation using the DREAM software package: theory, concepts, and MATLAB implementation. *Environ. Model. Softw.* 75, 273–316. <https://doi.org/10.1016/j.envsoft.2015.08.013>.
- Wada, Y., van Beek, L.P.H., van Kempen, C.M., Reckman, J.W.T.M., Vasak, S., Bierkens, M.F.P., 2010. Global depletion of groundwater resources. *Geophys. Res. Lett.* 37, L20402. <https://doi.org/10.1029/2010GL044571>.
- Wagner, B., 1999. Evaluating data worth for ground-water management under uncertainty. *J. Water Resour. Plan. Manag.* 125, 281–288. [https://doi.org/10.1061/\(ASCE\)0733-9496\(1999\)125:5\(281\)](https://doi.org/10.1061/(ASCE)0733-9496(1999)125:5(281)).
- Wagner, B.J., 1995. Sampling design methods for groundwater modeling under uncertainty. *Water Resour. Res.* 31, 2581–2591. <https://doi.org/10.1029/95WR02107>.
- White, J.T., 2018. A model-independent iterative ensemble smoother for efficient history-matching and uncertainty quantification in very high dimensions. *Environ. Model. Softw.* 109, 191–201. <https://doi.org/10.1016/j.envsoft.2018.06.009>.
- Wöhling, T., Geiges, A., Nowak, W., 2016. Optimal design of multitype groundwater monitoring networks using easily accessible tools. *Groundwater* 54, 861–870. <https://doi.org/10.1111/gwat.12430>.
- Wöhling, T., Schöninger, A., Gayler, S., Nowak, W., 2015. Bayesian model averaging to explore the worth of data for soil-plant model selection and prediction. *Water Resour. Res.* <https://doi.org/10.1002/2014WR016292> n/a-n/a.
- Yakirevich, A., Pachepsky, Y.A., Gish, T.J., Guber, A.K., Kuznetsov, M.Y., Cady, R.E., Nicholson, T.J., 2013. Augmentation of groundwater monitoring networks using information theory and ensemble modeling with pedotransfer functions. *J. Hydrol.* 501, 13–24. <https://doi.org/10.1016/j.jhydrol.2013.07.032>.
- 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, 124280. <https://doi.org/10.1016/j.jhydrol.2019.124280>, 124280.
- Yokota, F., Thompson, K.M., 2004. Value of information literature analysis: a review of applications in health risk management. *Med. Decis. Mak.* 24, 287–298. <https://doi.org/10.1177/0272989X04263157>.