Advances in Water Resources 86 (2015) 193–208

Contents lists available at ScienceDirect

Advances in Water Resources

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

# Experimental design for estimating unknown hydraulic conductivity in an aquifer using a genetic algorithm and reduced order model

Timothy T. Ushijima, William W-G. Yeh*

*Department of Civil and Environmental Engineering, University of California, Los Angeles, CA 90095, USA*

ARTICLE INFO

*Article history:*
Received 18 August 2014
Revised 25 September 2015
Accepted 30 September 2015
Available online 23 October 2015

*Keywords:*
Model reduction
Proper orthogonal decomposition
Robust experimental design
Max–min optimization
Genetic algorithm

ABSTRACT

We develop an experimental design algorithm to select locations for a network of observation wells that provide the maximum robust information about unknown hydraulic conductivity in a confined, anisotropic aquifer. Since the information that a design provides is dependent on an aquifer's hydraulic conductivity, a robust design is one that provides the maximum information in the worst-case scenario. The design can be formulated as a max–min optimization problem. The problem is generally non-convex, non-differentiable, and contains integer variables. We use a Genetic Algorithm (GA) to perform the combinatorial search. We employ proper orthogonal decomposition (POD) to reduce the dimension of the groundwater model, thereby reducing the computational burden posed by employing a GA. The GA algorithm exhaustively searches for the robust design across a set of hydraulic conductivities and finds an approximate design (called the High Frequency Observation Well Design) through a Monte Carlo-type search. The results from a small-scale 1-D test case validate the proposed methodology. We then apply the methodology to a realistically-scaled 2-D test case.



## 1. Introduction

### 1.1. Background

Hydraulic conductivity is one of the most important parameters to take into account when modeling an aquifer. Unknown hydraulic conductivities can have drastic effects on the reliability of groundwater model results. Estimating these unknown hydraulic conductivities through inverse modeling is an important area of research in groundwater modeling. Solving the inverse problem requires field observations of head. However, collecting good observations is expensive, time consuming, and difficult, particularly if aquifer parameters vary spatially. As a result, inverse modeling always faces an observation scarcity problem. Given an infinite budget, we could take observations at infinitesimally fine spatial and temporal resolutions; however, in the real world there are budget constraints on the number of experiments (observations) that can be conducted. In general, the goal of optimal experimental design for parameter estimation is to select the observation locations and sampling frequency such that a specified criterion is optimized subject to a set of constraints. The constraints frequently encountered are cost, reliability of the estimated parameters, and time and duration of the experiments. In groundwater modeling, if one assumes that observations are taken from the beginning of the pumping test to the end, the experimental design problem is simplified to the determination of a network of observation locations. The most commonly used criterion for optimal experimental design is the maximization of a measure of the information matrix. This formulation generally lends itself to a combinatorial optimization problem that is nonlinear and non-convex. We then can determine the tradeoff between number of experiments and amount of useful information collected by varying the number of experiments and evaluating the amount of useful information received for each experiment. While the concept of experimental design for parameter estimation in the context of groundwater modeling has been investigated extensively, the combinatorial search required to solve for the optimal observation network of observations has challenged many studies. A realistic, highly discretized, large-scale groundwater model, referred to in this paper as the full model, may have tens or hundreds of thousands of nodes and a resulting equivalent number of equations. As a consequence, the dimension of the combinatorial search quickly becomes so large that it is infeasible to solve through mathematical programming techniques such as mixed integer nonlinear programming. Thus other methods are required to solve this optimization problem. Genetic Algorithms (GAs) are one of a number of methods that have been developed over the years to solve large-scale optimization problems that are difficult or impossible to solve through traditional mathematical programming techniques. GAs do not require the calculation of derivatives and also

* Corresponding author. Tel.: +1 310 825 2300; fax: +1 310 825 7581.
*E-mail addresses:* tim.ushijima@gmail.com (T.T. Ushijima), williamy@seas.ucla.edu (W.W-G. Yeh).

http://dx.doi.org/10.1016/j.advwatres.2015.09.029
0309-1708/

## Nomenclature

| | |
|---|---|
| $\boldsymbol{A} \in R^{N_n \times N_n}$ | stiffness matrix for the full model |
| $\tilde{\boldsymbol{A}} \in R^{N_p \times N_p}$ | stiffness matrix for the reduced model |
| $\boldsymbol{A}(\boldsymbol{k}) \in R^{N_n \times N_n}$ | matrix representing the system of linear equations in the full model |
| $\tilde{\boldsymbol{A}}(k) \in R^{n_p \times n_p}$ | matrix representing the system of linear equations in the reduced model |
| $\boldsymbol{B} \in R^{N_n \times N_n}$ | mass matrix for the full model |
| $\tilde{\boldsymbol{B}} \in R^{n_p \times n_p}$ | mass matrix for the reduced model |
| $\boldsymbol{b} \in R^{N_n}$ | vector describing the forcing of the full model |
| $\tilde{\boldsymbol{b}} \in R^{N_n}$ | vector describing the forcing of the reduced model |
| $\chi$ | matrix of snapshots collected at some hydraulic conductivity $\boldsymbol{k}$ |
| $d$ | scaling factor for the residual |
| $F$ | specific volumetric pumping rate |
| $\Delta K_j$ | perturbation of the $j$th hydraulic conductivity |
| $\boldsymbol{\Delta k}_j \in R^{n_z}$ | vector perturbing the $j$th hydraulic conductivity |
| $\Delta o_i$ | change in the $i$th observation |
| $\Delta\theta_j$ | perturbation of the $j$th parameter |
| $\Delta\boldsymbol{\theta}_j \in R^N$ | vector perturbing the $j$th parameter |
| $f_1, f_2, f_3$ | known functions describing initial and boundary conditions of an aquifer |
| $\boldsymbol{e}$ | error between the full and reduced model |
| $\Gamma_1$ | fixed head boundary |
| $\Gamma_2$ | flux boundary |
| $\boldsymbol{H} \in R^{N_n \times n_{sp}}$ | vector of initial head values |
| $\boldsymbol{h} \in R^{N_n \times n_{sp}}$ | vector of head values |
| HFO | high frequency observation well design |
| $\boldsymbol{I} \in R^{n_q \times n_q}$ | information matrix calculated by $\boldsymbol{J}_d{}^T * \boldsymbol{J}_d * \boldsymbol{W}$ |
| $\boldsymbol{J}_d \in R^{n_{obs} \times n_q}$ | Jacobian matrix containing only the observations of interest to all wells |
| $J_{d_{i,j}}$ | $i,j$th element in $\boldsymbol{J}_d$ |
| $\boldsymbol{K} \in R^{n_z \times n_l^{n_z}}$ | matrix of columns storing all parameter combinations to be searched |
| $K_i$ | hydraulic conductivity in the $i$th direction |
| $\hat{\boldsymbol{K}}$ | the space of feasible hydraulic conductivities |
| $\boldsymbol{k} \in R^{n_z}$ | vector of hydraulic conductivities |
| $\boldsymbol{M} \in R^{n_{obswell} \times N_n}$ | observation matrix isolating the rows of $\boldsymbol{P}$ corresponding to some $\boldsymbol{\omega}$ |
| $\boldsymbol{\Lambda} \in R^{N_n \times N_n}$ | matrix containing the eigenvalues of $\boldsymbol{X}$ |
| L | length unit (meters, feet, etc.) |
| $N$ | total number of parameters of interest |
| $n_l$ | number of parameter increments |
| $N_n$ | number of nodes in the full model |
| $n_{obs}$ | total number of observations taken |
| $n_{obswell}$ | maximum number of allowable observation wells |
| $n_p$ | number of principal components used in the reduced model |
| $n_{sp}$ | number of snapshots taken for each pumping well to build the reduced model |
| $n_z$ | number of hydrologic zones with unknown hydraulic conductivity |
| $\boldsymbol{P} \in R^{N_n \times n_p}$ | full projection matrix |
| $\hat{\boldsymbol{P}} \in R^{N_n \times n_{sp}}$ | principle vectors of a particular set of snapshots |
| PDE | partial differential equation |
| $\boldsymbol{\Omega}$ | the space of all feasible sets of observation well locations |
| $\boldsymbol{\omega} \in R^{n_{obswell}}$ | vector contacting some feasible set of observation well locations ($\boldsymbol{\omega} \subset \boldsymbol{\Omega}$) |
| $\hat{\boldsymbol{\omega}} \in R^{n_{obs} \times n_l^{n_z}}$ | the set of all $\tilde{\boldsymbol{\omega}}_i$ |
| $\hat{\boldsymbol{\omega}}_{+1} \in R^{n_{obs} \times n_l^{n_z}+1}$ | $\hat{\boldsymbol{\omega}}$ plus the HFO design |
| $\boldsymbol{\omega}^* \in R^{n_{obswell}}$ | the optimal experimental design solution given the true aquifer parameters |
| $\tilde{\boldsymbol{\omega}}_i \in R^{n_{obswell}}$ | $i$th solution output by the GA |
| $\tilde{\boldsymbol{\omega}}_i^{(j)}$ | $j$th location in $\tilde{\boldsymbol{\omega}}_i$ |
| ODE | ordinary differential equation |
| $o_i$ | $i$th observation |
| $o_i(\boldsymbol{\theta})$ | $i$th model simulated value using the parameter values in $\boldsymbol{\theta}$ |
| $\boldsymbol{Q}$ | matrix containing the eigenvectors of $\boldsymbol{X}$ |
| $\boldsymbol{q} \in R^{N_n}$ | vector of sinks for the full model |
| $q_n$ | specific discharge normal to the flux boundary ($\Gamma_2$) |
| $\boldsymbol{R}(k) \in R^{n_p \times n_{obs}}$ | matrix containing the reduced solution at the observation times |
| $\boldsymbol{r} \in R^{n_p}$ | vector of the reduced solution at time $t$ |
| $\boldsymbol{r_e}$ | the residual between the full and reduced model |
| $\boldsymbol{s} \in R^{N_n \times n_{sp}}$ | vector of drawdown values |
| $s_i$ | $i$th drawdown value |
| $\boldsymbol{\Sigma} \in R^{n_p \times n_p}$ | diagonal matrix containing the singular values of $\boldsymbol{X}$ |
| $S_s$ | specific storage |
| $\hat{\boldsymbol{s}}_t \in R^{N_n}$ | vector of the approximation of $\boldsymbol{s}_t$ |
| SVD | singular vale decomposition |
| T | time unit (days, hours, etc.) |
| $\tau$ | error tolerance |
| $\boldsymbol{\theta} \in R^N$ | vector of nominal parameter values |
| $\theta_j$ | $j$th parameter |
| $\tilde{V} \subset R^{N_n}$ | subspace spanned by the columns of $\boldsymbol{P}$ |
| $\boldsymbol{V} \in R^{n_p \times n_p}$ | matrix containing the right singular vectors of $\boldsymbol{X}$ |
| $\boldsymbol{W} \in R^{n_q \times n_q}$ | some user specified weighting matrix used in calculating $\boldsymbol{I}$ |
| $\boldsymbol{X} \in R^{N_n \times n_p}$ | matrix containing all collected snapshots |
| $\boldsymbol{x} \in R^{N_n}$ | vector of binary variables indicating if a node has an observation well |
| $x_i$ | $i$th element in $\boldsymbol{x}$ |
| *zone j* | all the nodes in the $j$th observation zone |

can deal with discontinuous functions. While GAs have been used by a number of studies to design optimal observation networks using various optimality criteria [1–3], many of these studies were challenged by the fact that GAs require a large number of model calls. Thus for a highly discretized groundwater model, using a GA to design an optimal observation network may be computationally inefficient, even infeasible. To overcome this, we apply a proper orthogonal decomposition (POD) model reduction to the groundwater model to reduce its spatial dimension. POD is an effective model reduction technique that maintains the physics of the groundwater model. In many instances it has been shown that the reduced model is orders of magnitudes smaller than the original full model and runs 1,000 times faster. Since the parameters themselves are to be estimated, we develop a parameter-independent reduced model for the optimal experimental design. By doing this, we are able to reduce an inordinately large-scale, combinatorial optimization problem to a manageable size.