Title
Model Reduction of Unconfined Groundwater Flow Using Proper Orthogonal Decomposition and the Discrete Empirical Interpolation Method

Author
Stanko, Zachary Paul

Publication Date
2016

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Model Reduction of Unconfined Groundwater Flow Using Proper Orthogonal Decomposition and the Discrete Empirical Interpolation Method

A thesis submitted in partial satisfaction

of the requirements for the degree Master of Science

in Civil Engineering

by

Zachary Stanko

2016

ABSTRACT OF THE THESIS

Model Reduction of Unconfined Groundwater Flow Using Proper Orthogonal Decomposition and the Discrete Empirical Interpolation Method

by

Zachary Stanko

Master of Science in Civil Engineering

University of California, Los Angeles, 2016

Professor William W. Yeh, Chair

Nonlinear groundwater flow models have the propensity to be overly complex leading to burdensome computational demands. Reduced modelling techniques are used to develop an approximation of the original model that has smaller dimensionality and faster run times. The reduced model proposed is a combination of proper orthogonal decomposition (POD) and the discrete empirical interpolation method (DEIM). Solutions of the full model (snapshots) are collected to represent the physical dynamics of the system and Galerkin projection allows the formulation of a reduced model that lies in a subspace of the full model. Interpolation points are added through DEIM to eliminate the reduced model's dependence on the dimension of the full model. POD is shown to effectively reduce the dimension of the full model and DEIM is shown to speed up the solution by further reducing the dimension of the nonlinear calculations. To show

the concept can work for unconfined groundwater flow model, with added nonlinear forcings, one-dimensional and two-dimensional test cases are constructed in MODFLOW-OWHM. POD and DEIM are added to MODFLOW as a modular package. Comparing the POD and the POD-DEIM reduced models, the experimental results indicate similar reduction in dimension size with additional computation speed up for the added interpolation. The hyper-reduction method presented is effective for models that have fine discretization in space and/or time as well as nonlinearities with respect to the state variable. The dual reduction approach ensures that, once constructed, the reduced model can be solved in an equation system that depends only on reduced dimensions.

The thesis of Zachary Stanko is approved.

Steven Adam Margulis

Mekonnen Gebremichael

William W. Yeh, Committee Chair

University of California, Los Angeles

2016

Contents

List Of Figures

List of Tables

Table 1: Variable symbols and definitions along with appropriate units and initial default values.

Symbol	Dimension	[units]	Description
$\boldsymbol{h}$	$n \times 1$	[m]	Vector of groundwater head
$\boldsymbol{A}_h$	$n \times n$	[-]	Matrix containing nonlinear coefficients for the groundwater flow equation (elements that are functions of head)
$\boldsymbol{B}$	$n \times n$		Matrix containing linear coefficients for the groundwater flow equation (elements that are not a function of head)
$\boldsymbol{b}$	$n \times 1$	[-]	Right-hand side vector of the groundwater flow equation
K	scalar	[m/day]	Hydraulic conductivity in the x and y principal directions
$\boldsymbol{q}$	$n \times 1$	[m/day]	Sources/sinks of water in flow per unit volume
S_y	scalar	[-]	Specific yield coefficient
ε_{max}	scalar	[m]	Maximum difference in head in two consecutive iterations.
i, j			Indices for the model cells in the x and y direction
k			Index for the iteration of the nonlinear solver
t			Index for the model time steps
n	scalar	[-]	Number of finite difference cells of the model domain
r	scalar	[-]	Number of singular values retained from POD to create the reduced model
s	scalar	[-]	Number of snapshots taken of the full model
d	scalar	[-]	Number of interpolation points used for the POD-DEIM reduced model
Δ_t	scalar	[days]	Uniform time step length
$\boldsymbol{\Phi}_h$	$s \times n$		Snapshot set of h.
$\boldsymbol{\Phi}_b$	$s \times n$		Snapshot set of b.
σ_i	scalar		The i^{th} singular value
$\mathbf{P}$	$n \times r$		Projection matrix formed from POD on Φ_h
$\mathbf{D}$	$n \times d$		Projection matrix formed from POD on Φ_d
$\mathbf{Z}$	$n \times d$		Permutation matrix to select the dominant rows of the system
z_j	scalar		The j^{th} interpolation index
$\boldsymbol{c}$	$d \times 1$		Vector of reduced variables in DEIM formulation
$\boldsymbol{h_r}$	$r \times 1$		Vector of state variables in the reduced space
$\boldsymbol{z_m}$	$d \times 1$		Vector of interpolation indices for model m.

Introduction

Reduced modeling has become a necessary field of research given the near-complete scientific understanding of many physical processes and the ensuing complexity of mathematical models. For groundwater modeling specifically, reduced modeling techniques are being applied at many stages and to models of many forms. Traditionally, the proper orthogonal decomposition (POD) method is used to formulate a low dimension basis for high-dimension dynamical systems (Vermeulen et al. 2004; Antoulas et al. 2001). The key advantage of using POD for model reduction is that the reduced model maintains the physics of the full model and captures the dominating characteristics of the full model. POD is also known as Empirical Orthogonal Functions (EOF) (von Storch & Hannoschöck 1985; McPhee & Yeh 2008), Coherent Structures (CS) (Sirovich 1987), Principal Component Analysis (PCA), or Common Factor Analysis (CFA) (Reyment & Joreskog 1993).

For groundwater flow specifically, most previous applications of POD to groundwater models have utilized the confined flow equation, which is linear with respect to the state variable of interest, namely, the hydraulic head (Boyce & Yeh 2014). Alternatively, linearization techniques, such as quasilinearization (Siade et al. 2012), can be utilized to facilitate POD reduced model construction for the case of nonlinear parameter estimation. To briefly describe this reduced model construction: one selects a set of model simulation results at specific instances of simulation time, which is called a snapshot set. POD is then applied to identify the singular values of the matrix composed of the snapshots. Only a selected few of the singular values are chosen such that most of the variance of the original system is retained. A subspace basis is then constructed and Galerkin projection is applied to form the reduced model

(Vermeulen et al. 2004). Extensions to this method for nonlinear problems include modifications to the Galerkin projection—such as using a Petrov-Galerkin projection for stabilization (Stef et al. 2015) or adaptive Ritz vectors (Nigro et al. 2015)—or alternate strategies such as minimization of the L^1 norm representing the reduced model approximation error (Abgrall & Amsallem 2015).

While confined flow models have nice linear properties that allow for flexible manipulation and superposition, nonlinearities are unavoidable in many groundwater modeling projects. Requiring unconfined flow, creates nonlinear equations that are harder to solve and hence more difficult to reduce successfully. Nonlinear model reduction has been addressed thoroughly in Cardoso et al. (2009), where a Markov chain Monte Carlo simulation was performed for an inverse problem utilizing Bayesian inference. Boyce et at (2015) also successfully reduced an unconfined groundwater model using the Newton formulation of MODFLOW, MODFLOW-NWT (Niswonger et al. 2011). These examples of successful nonlinear groundwater model reduction illustrate the added difficulty and present methods that are applicable in unique contexts. That is, the solution scheme in both studies involves Newton's method which requires approximation of a Jacobian that may not be easy to obtain. Also, approximating the Jacobian and formulating Newton's method effectively linearizes the system since the Jacobian can be evaluated at prior values of head. Lastly, there are more inherent memory requirements for a Newton solution than the traditional MODFLOW (exactly twice as much), which may restrict large-scale applications (Niswonger et al. 2011).

The discrete empirical interpolation method (DEIM) is an effective approach to nonlinear approximations. Originally developed as the empirical interpolation procedure (EIP) (Barrault et al. 2004), with the discrete form introduced in (Chaturantabut & Sorensen 2010), It has been

used in conjunction with POD for reducing FitzHugh-Nagumo equations (Chaturantabut & Sorensen 2010), shallow water equations (Ştefănescu & Navon 2013), and an advection-diffusion-reaction system (Cardoso et al. 2009). These model reduction procedures are also called reduced basis (RB) methods and an EIP has been developed in this context by (Drohmann et al. 2012). Successful applications of RB methods to various forms of the Navier-Stokes equations (both steady and unsteady) are demonstrated to have significant computational advantages (Quarteroni & Rozza 2007).

The joint application of POD and DEIM for nonlinear model reduction in the literature of other fields is also gaining popularity. In electrical engineering, the methods have been successfully applied to a magnetostatic problem coupled to an electric circuit (Henneron & Clenet 2014) and in mechanical engineering, the methods were used for a solid mechanics problem involving nonlinear elasticity (Radermacher & Reese 2015). However, unconfined groundwater flow models have not yet been reduced in this manner. In this study, we propose a combined model reduction approach that: 1) performs POD on an unconfined groundwater flow model; 2) applies DEIM to the nonlinear component of the governing equation; and 3) implements the procedure within MODFLOW. By enabling model reduction within MODFLOW, a very large assortment of existing MODFLOW models—many have single run times on the order of hours or even days—can be reduced to increase computational efficiency. This would permit large-scale optimization or uncertainty analysis that can require hundreds of thousands of model runs.

More robust reduced modeling techniques are needed for nonlinear dynamics in groundwater flow. Methods developed previously have resorted to using strategies from the linear world of modeling and do not address the additional time required to solve a system of nonlinear

equations. In large-scale simulations, having nonlinear calculations that still have a computational complexity of the full model greatly inhibits the value of developing the reduced model, which can have significant overhead. Application of DEIM to the unconfined groundwater flow equations presents an opportunity to develop a hyper reduced model utilizing controllable accuracy and computer runtimes that scale with reduced dimensions only.

Methods

The following methodology expands upon recent developments in POD and DEIM reduced modeling. The implementation of these methods within the commonly used MODFLOW software will be described. All variable definitions are compiled in Table 1 with the following convention: uppercase letters for scalar variables; bold lowercase for one-dimensional vectors; bold uppercase for two-dimensional matrices. The general three-dimensional governing equation for constant-density groundwater flow in an unconfined aquifer is given by (Keating & Zyvoloski):

$$\nabla \cdot (K\boldsymbol{h}\nabla\boldsymbol{h}) \pm \boldsymbol{q} = S_y \frac{\partial \boldsymbol{h}}{\partial t} \quad (1)$$

where $\nabla\cdot$ is the divergence operator, ∇ is the gradient operator, K is the isotropic hydraulic conductivity tensor [L/T], $\boldsymbol{h}$ is the hydraulic head [L], $\boldsymbol{q}$ is a volumetric flux per unit volume in or out of the system [T^{-1}], t is the time [T], and S_y is the specific yield [-]. To obtain a two-dimensional equation (assuming horizontal flow only), vertical integration can be performed in the z direction using Liebniz's rule. The associated two-dimensional governing equation for an unconfined aquifer then becomes the Boussinesq equation.

$$\frac{\partial}{\partial x}\left[K_{xx}h\frac{dh}{dx}\right]+\frac{\partial}{\partial y}\left[K_{yy}h\frac{dh}{dy}\right]+W=S_y\frac{\partial h}{\partial t} \quad (2)$$

Where K_{xx} and K_{yy} are the hydraulic conductivity parameters assumed to align with the x and y, and coordinates, respectively. In this case, W is the net source/sink into the aquifer (including areal recharge and point source wells) [LT^{-1}]. Here it is also assumed that the specific storage, S_s, resulting from compressed porous media, is negligible compared to the specific yield, S_y.

In our study, we consider two-dimensional unconfined flow as defined by MODFLOW. The PDE used for the MODFLOW governing equation is shown in Equation 3. MODFLOW has a Quasi-3D approach to vertical discretization which simply stacks 2D layers together and optionally simulates a confining layer between them with vertical leakage only (Harbaugh, Arlen 2005). From the continuity equation, assuming constant density, a finite-difference representation of Equation 3 is solved within MODFLOW. For strictly two-dimensional flow in MODFLOW, the specification of only one layer translates to a zero for all terms involving vertical flow parameters and the discretized PDE becomes Equation 4.

$$\frac{\partial}{\partial x}\left[K_{xx}\frac{dh}{dx}\right]+\frac{\partial}{\partial y}\left[K_{yy}\frac{dh}{dy}\right]+\frac{\partial}{\partial z}\left[K_{zz}\frac{dh}{dz}\right]+W=S_y\frac{\partial h}{\partial t} \quad (3)$$

$$C_{-j}\left(\Delta h_{-j}\right)+C_{+j}\left(\Delta h_{+j}\right)+C_{-i}(\Delta h_{-i})+C_{+i}(\Delta h_{+i})+p_{ij}h_{ij}+q_{ij}=S_y\Delta i\Delta j\frac{dh}{dt} \quad (4)$$

Here, MODFLOW introduces a new parameter called conductance (C), which is defined as the hydraulic conductivity ($K_x\ or\ K_y$) times the cross-sectional area of flow ($\Delta y * h\ or\ \Delta x * h$) divided by the length of the flow path ($\Delta x\ or\ \Delta y$). For unconfined conditions, the cross-sectional area depends on the height of the water table (h) at that location, making Equation 4 nonlinear in

each term involving C. The notation $C_{\pm i,j}$ refers to the direction in which the conductance is specified; $+j$ being from the direction of increasing j and $-i$ being from the direction of decreasing i. The conductance is also related to Transmissivity (T) via the equation $C = TW/L$, where $T = K * h$. Boundary conditions contained in W are separated into head-dependent (p) and head-independent (q) fluxes. The block centered finite-difference scheme produces a set of nonlinear ordinary differential equations, represented in matrix form as Equation 5.

$$\boldsymbol{Ah} + \boldsymbol{f} = \boldsymbol{B}\frac{dh}{dt} \quad (5)$$

Where $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{h}, \boldsymbol{f} \in \mathbb{R}^{n\times 1}$, n being the number of finite difference nodes on the model domain. $\boldsymbol{A}$ contains all coefficients of head that are internally calculated as functions of head at each time step, making the term nonlinear. $\boldsymbol{B}$ contains constant coefficients for the temporal head change and constant spatial discretization values and $\boldsymbol{f}$ contains all head-independent sources or sinks of water and head-independent boundary conditions. The head at any time is then calculated using a backward-difference approach to ensure stability. After reordering some terms and multiplying through by -1, the following matrix equation (Equation 6) is calculated to represent the flow system at each time step:

$$\left[\frac{\boldsymbol{B}}{\Delta_t} - \boldsymbol{A}\right]\boldsymbol{h}^{t+1} = \frac{\boldsymbol{B}}{\Delta_t}\boldsymbol{h}^{t} + \boldsymbol{f} \quad (6)$$

or

$$\boldsymbol{A}_h\boldsymbol{h}^{t+1} = \boldsymbol{b}$$

Defining $\left[\frac{\boldsymbol{B}}{\Delta_t} - \boldsymbol{A}\right]$ as the nonlinear system matrix, $\boldsymbol{A}_h$, and $\left[\frac{B}{\Delta_t}\boldsymbol{h}^{t} + \boldsymbol{f}\right]$ as the right hand side vector of constants, $\boldsymbol{b}$. A variety of indirect methods can be used to solve for $\boldsymbol{h}^{t+1}$ at each time step. for the full model, the current study utilizes the Preconditioned Conjugate-Gradient Solver (PCG), documented in (Hill 1990) and Picard iteration for every time step, stopping when $\max_{1<i<n}\left(h_i^k - h_i^{k+1}\right) \leq \varepsilon_{max}$ (the maximum head difference from iteration k to $k+1$ is

sufficiently small) or $\max_{1<i<n}\left(inflow_i^{k+1} - outflow_i^{k+1}\right) \leq r_{max}$ (the maximum flow residual is sufficiently small). These methods are already programmed into MODFLOW and the recently updated MODFLOW-OWHM (Hanson et al. 2014), which served as the base code for the POD-DEIM development and includes the Newton Formulation of MODFLOW. For the reduced model, the LAPACK *LU* decomposition routine GESV with partial pivoting and row interchanges (documented: https://software.intel.com/en-us/node/520973). When implementing solvers for reduced models, additional factors ought to be considered yet this is beyond the scope of this project; see (Forstall 2015) for more thorough analysis of linear and nonlinear reduced model solvers.

POD

Equation 6, which is called the full model, lies in the dimension $n \times n$, since there are n equations and n unknowns. A brief derivation of POD begins by approximating $\boldsymbol{h}^{t+1}$ with $\boldsymbol{P}\boldsymbol{h}_r^{t+1}$. $P \in \mathbb{R}^{n \times r}$ is generated by applying singular value decomposition (SVD) on $\boldsymbol{\Phi}_h \in \mathbb{R}^{n \times s}$, a matrix composed of a set of simulated values of $\boldsymbol{h}$, called solution snapshots, for s selected time steps. $\boldsymbol{h}_r \in \mathbb{R}^{r \times 1}$ is then a reduced vector of dependent variables, where $r \ll n$ is chosen based in the r largest singular values (σ) of $\boldsymbol{\Phi}_h$ that account for nearly all (e.g., 99.99% or 99.999%) of the matrix's embedded information, called percent energy (ε_{energy}).

$$Percent\ Energy = \frac{\sum_{i=1}^{r} \sigma_i}{\sum_{i=1}^{rank(\boldsymbol{\Phi}_h)} \sigma_i} \times 100 \geq \varepsilon_{energy}$$

Galerken Projection is then used to project the full model space onto a reduced subspace by pre-multiplying both sided by $\boldsymbol{P}^T$, resulting in Equation 7. This process is described more thoroughly in Vermeulen et al. (2004) and as applied to a newton formulation of MODFLOW in

Boyce et al. (2015) (Vermeulen et al. 2004; Boyce et al. 2015). Equation 7 is called the POD-reduced model and still requires matrix multiplication with a dimension of n when $\boldsymbol{P}^T\boldsymbol{A}_h\boldsymbol{P}$ is computed at each time step.

$$\boldsymbol{P}^T\boldsymbol{A}_h\,\boldsymbol{P}\boldsymbol{h}_r^{t+1} = \widehat{\boldsymbol{A}_h}\,\boldsymbol{h}_r^{t+1} = \boldsymbol{P}^T\boldsymbol{b} \tag{7}$$

DEIM

The discrete empirical interpolation method is implemented specifically to reduce the nonlinear term's dependency on the full dimension of the original model. At each time step, it may be computationally burdensome and unnecessary to compute any nonlinear approximations in the POD formulation using the full system's state space. To approximate nonlinearities in a reduced space, the projection of nonlinear term is evaluated at selected interpolation points. The nonlinear projection basis is obtained via POD on snapshots of the nonlinear components only. The initial DEIM approximation (Equation 8) is made by approximating the nonlinear on the left-hand side of Equation 7 with a linear interpolation. The selection algorithm for the interpolation indices chooses points that have the largest residual error when iteratively approximating the nonlinear basis vectors (columns of $\boldsymbol{D}$) from vector of new reduced variables, c, times the basis. The algorithm for this selection is shown in Figure 1.

$$\widehat{\boldsymbol{A}_h}\,\boldsymbol{h}_r^{t+1} \cong \boldsymbol{D}\boldsymbol{c} \tag{8}$$

Where $\boldsymbol{D} \in \mathbb{R}^{n\times d}$ is generated by performing POD on snapshots of $\boldsymbol{b}$ from Equation 6 (taken at the same s time steps as snapshots of $\boldsymbol{h}$) and $\boldsymbol{c} \in \mathbb{R}^{d\times 1}$ is a vector of coefficients still to be determined. Since Equation 8 is overdetermined, only d equations are required to solve for $\boldsymbol{c}$. The retained rows of the system are the interpolation indices that are selected

through the process described in Stefanescu & Navon (2013) and illustrated in Figure 1(Ştefănescu & Navon 2013). Using $\boldsymbol{D_d} \stackrel{\text{def}}{=} \boldsymbol{Z}^T\boldsymbol{D}$ and $\boldsymbol{Z} \in \mathbb{R}^{n\times d}$ as a permutation matrix that selects the rows of $\widehat{\boldsymbol{A}_h}$ and $\boldsymbol{D}$ that correspond to d interpolation points, Equation 8 becomes:

$$\boldsymbol{Z}^T\widehat{\boldsymbol{A}_h}\,\boldsymbol{h}_r^{t+1} = \boldsymbol{Z}^T\boldsymbol{D}\boldsymbol{c} \tag{9}$$

Equation 9 can be inverted to solve for $\boldsymbol{c}$ and subsequently substituted back into Equation 8 to obtain the final approximation of the nonlinear term (Equation 10).

$$\boldsymbol{c} = [\boldsymbol{Z}^T\boldsymbol{D}]^{-1}\boldsymbol{Z}^T\widehat{\boldsymbol{A}_h}\,\boldsymbol{h}_r^{t+1} \tag{10}$$

$$\widehat{\boldsymbol{A}_h}\,\boldsymbol{h}_r^{t+1} \cong \boldsymbol{D}[\boldsymbol{Z}^T\boldsymbol{D}]^{-1}\boldsymbol{Z}^T\widehat{\boldsymbol{A}_h}\,\boldsymbol{h}_r^{t+1} = \widetilde{\boldsymbol{A}_h}\boldsymbol{h}_r^{t+1} \tag{11}$$

The final reduced model arises when the approximation in Equation 11 replaces the nonlinear term of Equation 7, which in this case is the only left-hand side term. Equation 11 is now considered the POD-DEIM reduced model, which is solved for $\boldsymbol{h}_r^{t+1}$at each time step entirely within the reduced dimension. In other words, there are no nonlinear operations that must be carried out in the original n-dimension allowing the reduced model to be solved at each iteration depending only on dimensions r and d.

The DEIM procedure involves the following steps:

1. Construct a basis $D \in \mathbb{R}^{n \times d}$ from POD on snapshots of a nonlinear term
2. Select z_1 as the index of the largest element of $\boldsymbol{d}_1$, the first column of D
3. For the remaining $j = 2, \ldots, d$ columns
    a. Calculate $\boldsymbol{c} = \left(Z^T [\boldsymbol{d}_1, \ldots, \boldsymbol{d}_j]\right)^{-1} \left(Z^T \boldsymbol{d}_j\right)$
        Where $Z = \left[\boldsymbol{e}_{z_1}, \ldots, \boldsymbol{e}_{z_j}\right]$
    b. Compute the residual $\boldsymbol{r} = \boldsymbol{d}_j - D\boldsymbol{c}$
    c. Take z_j to be the index of $\|\boldsymbol{r}\|_\infty$

Figure 1: Algorithm for selecting the interpolation indices z_j for $j = 1, \ldots, d$.

MODFLOW

The implementation within MODFLOW is contained within the subroutine GWF2MRED1AR1 and a new reduced solver is added. The structure of the POD-DEIM algorithm is demonstrated in the flow chart of Figure 2. The LAPACK LU-decomposition routine is used as the reduced solver and the MODFLOW implementation of PCG is used as the full model solver.

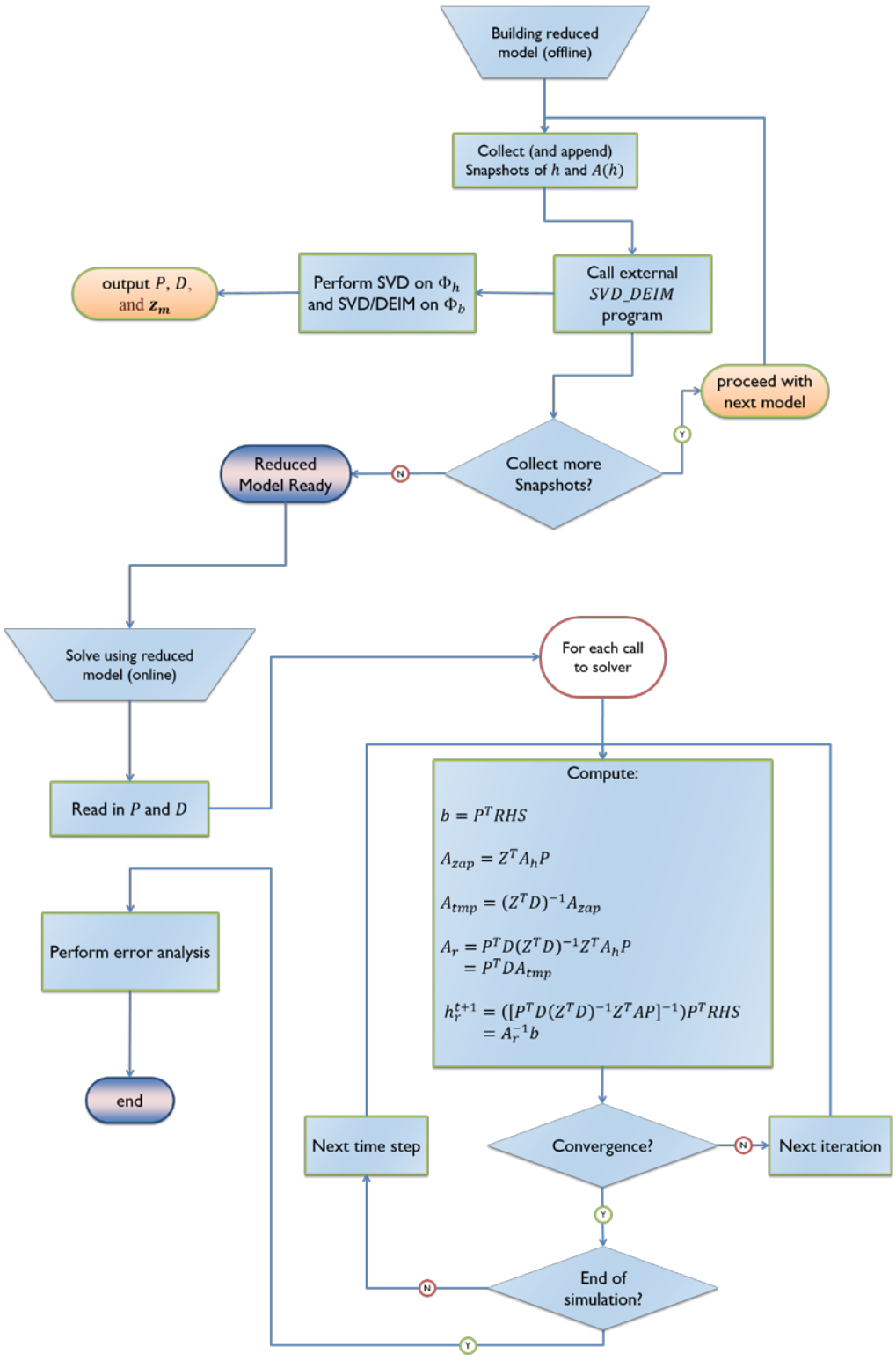

Figure 2: The flow chart describes the process of collecting snapshots for both traditional POD and DEIM, constructing the snapshot set, obtaining the basis, and conducting the interpolation.

Model Application

First, a simple one-dimensional (1D) groundwater model was developed to test the proposed methodology. 200 finite difference cells, with a discretization of $\Delta x = 10$ m, compose the model domain. The model simulates a pumping well at node 107, with constant zero head boundaries at nodes 1 and 200, an initial condition of zero head everywhere, and a saturated thickness of 50 m. Two hydraulic conductivity zones were used ($K_{x1} = 0.4$ and $K_{x2} = 1.9$ m/day) to introduce minimal heterogeneity. The value of K_{x2} was varied to account for an uncertain parameter. The well begins extracting at a rate of $Q = 150$ m^3/day on day 30, pumps for 30 days, and then shuts off for 30 days. The pumping rate was also varied to account for uncertain pumping. There are a total of 90 1-day time steps in the transient model. Figure 3 displays the one-dimensional model domain with two zones and the water table contour after 30 days of pumping (i.e. at day 60) for pumping rates of 100 and 200 m^3/day. The reduced model was generated with 90 snapshots each of four different model runs: $(K_{x2}, Q) = (1.2,100); (1.2,200); (2.2,100), (2.2,200)$ for a total of 360 snapshots. The reduced model was then tested with the simulation: $(K_{x2}, Q) = (1.9,150)$, to evaluate its effectiveness at parameter and pumping values not used to generate snapshots.

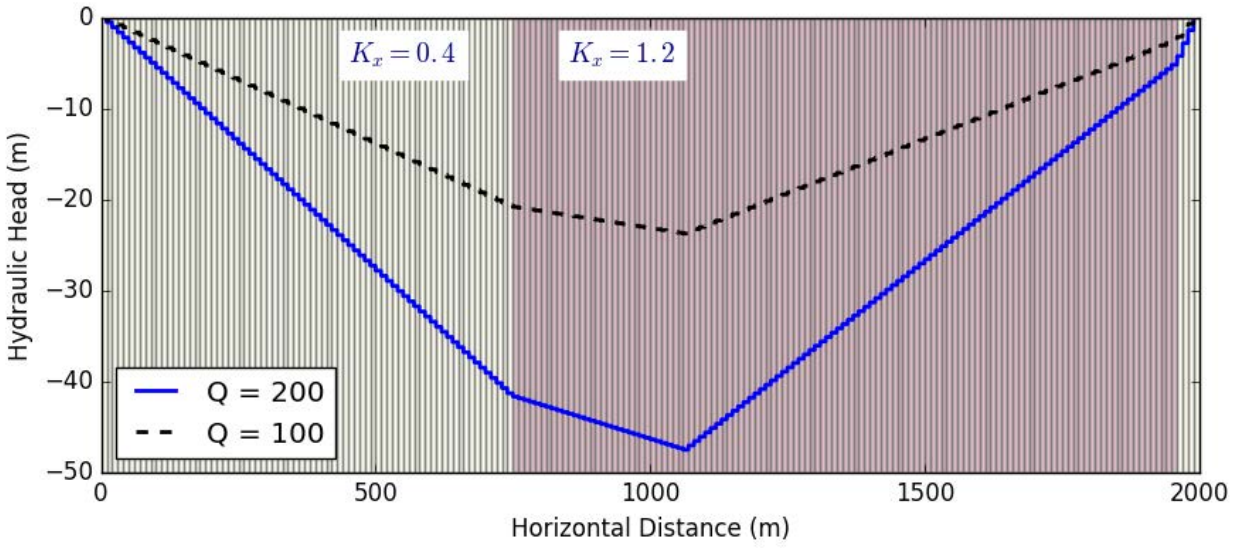

Figure 3: The 1-dimensional model shows the conductivity zones and the water table after 30 days of pumping, creating the unconfined conditions, for pumping rates of 100 and 200 m^3/day.

After the 1D case has been successfully verified, we then extend the methodology to a more realistic 2D case. Six zones of hydraulic conductivity were used to span four orders of magnitude. For the 2D case, the head starts at -5.0 m everywhere, creating unconfined conditions in the single layer with a thickness of 100 m, and pumping at five production wells begins to drawdown the water table. Three transient, 30-day stress periods (uniform time steps of 0.5 day) with various pumping rates are used to create a dynamic head distribution. A final stress period is added to allow the aquifer to recover without any pumping. A mix of constant head, constant flux, and no flow boundaries are maintained. Three additional head-dependent boundary conditions—a river, drain, and evapotranspiration zone—were added with the RIV, DRN, and EVT packages, respectively. The model is shown in Figure 4 with its zonation pattern and well locations.

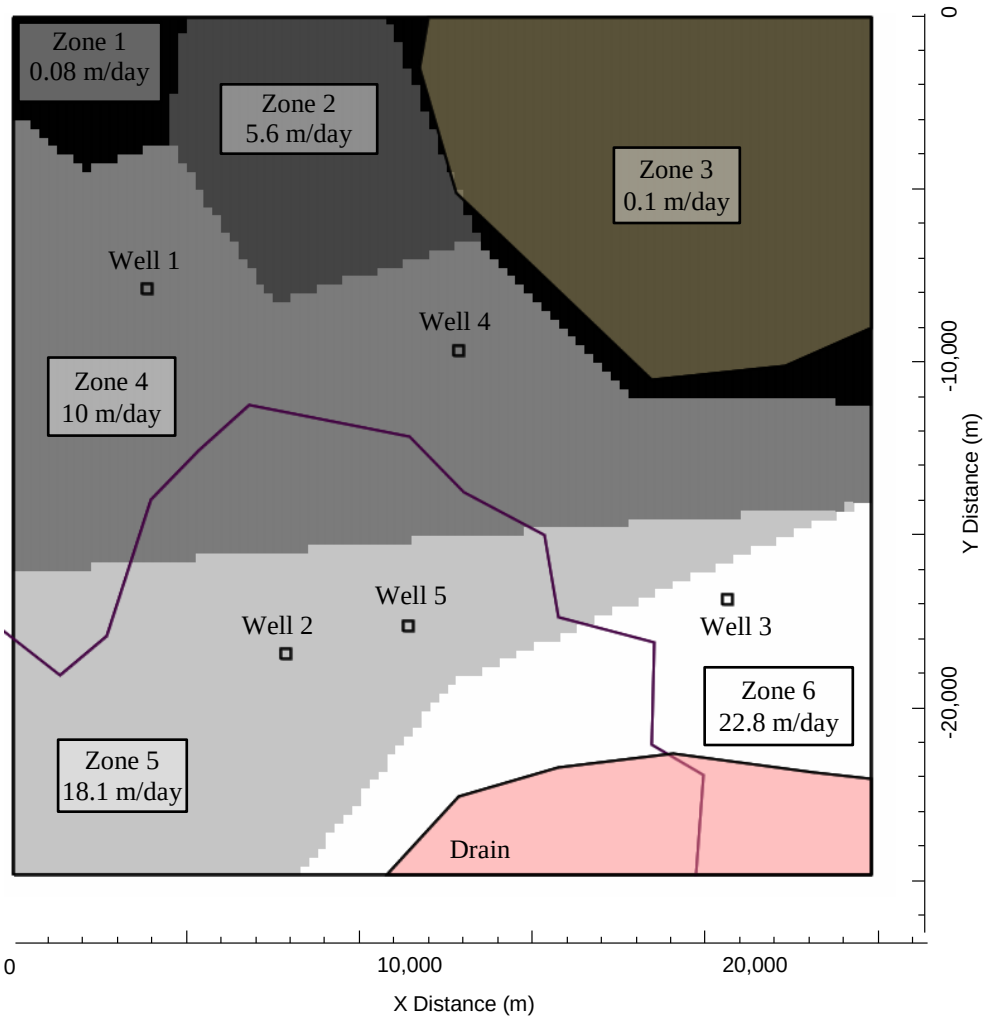

Figure 4: The domain of the 2D test case is shown with a model grid of 99 rows and 99 columns and six conductivity zones that span several orders of magnitude. There are six wells that pump at various rates and head-dependent features (river, drain, and ET) are included.

Error Analysis

The error for any reduced model is henceforth defined as the difference between the original full-model solution and the reduced model solution. This error is calculate for both the final head solution in the full model dimension as well as the nonlinear result of the operation $\boldsymbol{A}_h * \boldsymbol{h}$, which is the same as the vector $\boldsymbol{b}$ for the governing equation under consideration. Absolute error is presented in meters of hydraulic head. Since, the significance of the absolute error depends on

the precision of the model, a normalized root mean squared error (NRMSE) is also calculated for each case so the error is weighted by the span of head values (Equation 12). The location of the maximum error is also show in terms of the model cell number where it occurs. This information allows for quick identification of model features that may not be sufficiently captured with the current set of snapshots or interpolation points. While it is not necessary for the reduced dimensions, r and d, to be the same for the POD-DEIM model, they are initially chosen to be nearly equal for simplicity. The different errors between the POD and POD-DEIM model is therefore more comparable. The dimension d, however, experimentally appears to be most effective for solver convergence when $d \geq r + 1$.

$$NRMSE = \frac{\|full - reduced\|_2}{\sqrt{n} * (h_{max} - h_{min})} = \frac{RMSE}{(h_{max} - h_{min})} \quad 12$$

Results

Two to three orders of dimension reduction is achieved with both the 1D and 2D test problems. In Table 2, a summary of the reduced model's performance is compared to the full model using several metrics. The minimum head value, which is also the maximum drawdown, is recorded to show the head range magnitude (used to normalize the RMSE). The maximum error, maximum error location, and NRMSE are presented for both the head results and the values of the nonlinear operation $\boldsymbol{A}_h * \boldsymbol{h}$. The location of the maximum error informs of the time step and cell location that would be the next choice for an added snapshot or interpolation index, respectively. For the 1D model, eight interpolation indices were chosen: $\boldsymbol{z_{1D}} = [107, 104, 53, 196, 76, 141, 168, 29]$; and for the 2D model, 40 indices were selected. The indices indicate a diverse spread across the well's capture zone in the domain's interior. Points

near wells, which capture drawdown information, are often selected. Points near the boundaries are only selected if there is difficulty in resolving head-dependent boundary conditions.

Table 2: Model reduction results are displayed in comparison to the full, unreduced model.

			$\boldsymbol{h}$				$\boldsymbol{A * h}$		
		Dim	**h_{min} [m]**	**max error [m]**	**max loc $(t, x, [y])$**	**NRMSE [m]**	**max error**	**max loc $(t, x, [y])$**	**NRMSE**
1D	Full Model	200	-26.23	-	-	-	-	-	-
1D	POD	4	-26.23	5.66E-03	$(61, 69)$	9.70E-02	-	-	-
1D	POD-DEIM	5	-26.23	5.66E-03	$(61, 69)$	9.70E-02	3.31E-01	$(61, 106)$	1.64E-02
2D	Full Model	9801	-32.453	-	-	-	-	-	-
2D	POD	20	-32.451	2.10E-02	$(64, 67, 82)$	4.844E-02	-	-	-
2D	POD-DEIM	40	-32.451	2.12E-02	$(66, 67, 82)$	4.990E-02	343.6	$(65, 66, 82)$	3.95E-02

The full model is compared to both the POD (Equation 7) and the POD-DEIM model (Equation 11) to investigate errors in the 1D test case. First, the absolute residual errors ($|\boldsymbol{h} - \boldsymbol{h}_r|$) between the simulated head of the full model and the POD-DEIM model are illustrated with the simulation time on the y-axis (Figure 5). The beginning of each stress period (day 30 and day 60) can be seen to have a sharp increase in error. The maximum errors occur in the first time steps of a new stress period near the interface between the two hydraulic conductivity zones. However, the maximum error of 0.566 cm is less than 0.03% of the minimum simulated head value of -26.23 m and can be deemed insignificant. The errors diminish as the stress periods progress and the head values stabilize to a smooth gradient. If needed, more accurate reduced model results could be obtained by adding more snapshots (specifically at the beginning of each stress period), adding more interpolation points near the zone of changing conductivity, or slightly increasing the dimension of the reduced basis, r.

The performance of the nonlinear reduction is analyzed by comparing the result of the nonlinear operation. Figure 6 shows the residual error ($|b - b_r|$) between $\boldsymbol{A}_h * \boldsymbol{h}$ for the full and POD-DEIM reduced models. While minor errors still occur near the onset of a new stress period, the maximum errors now appear at the pumping cell and persist through the entire model horizon. Again, the errors are small with respect to the values being compared and more precision could be obtained by constructing another reduced model under the guidance of the maximum error locations.

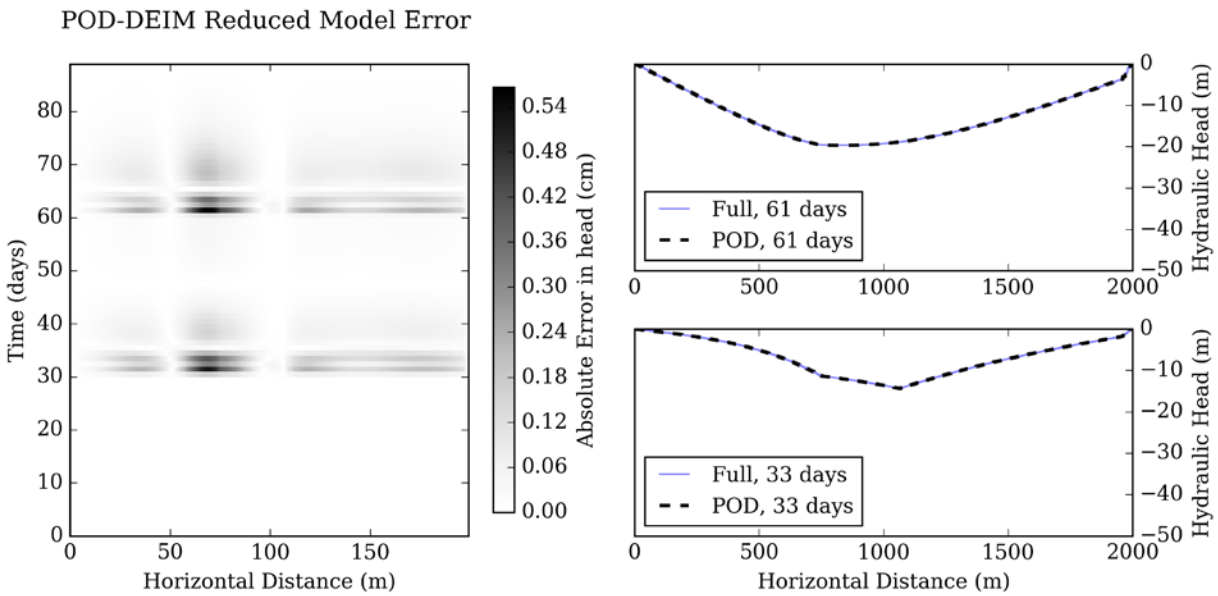

Figure 5: The results from the 1-D test are shown as a residual between full model head and reduced model head, for each of the 90 time steps in the A) POD reduced model and B) POD-DEIM reduced model

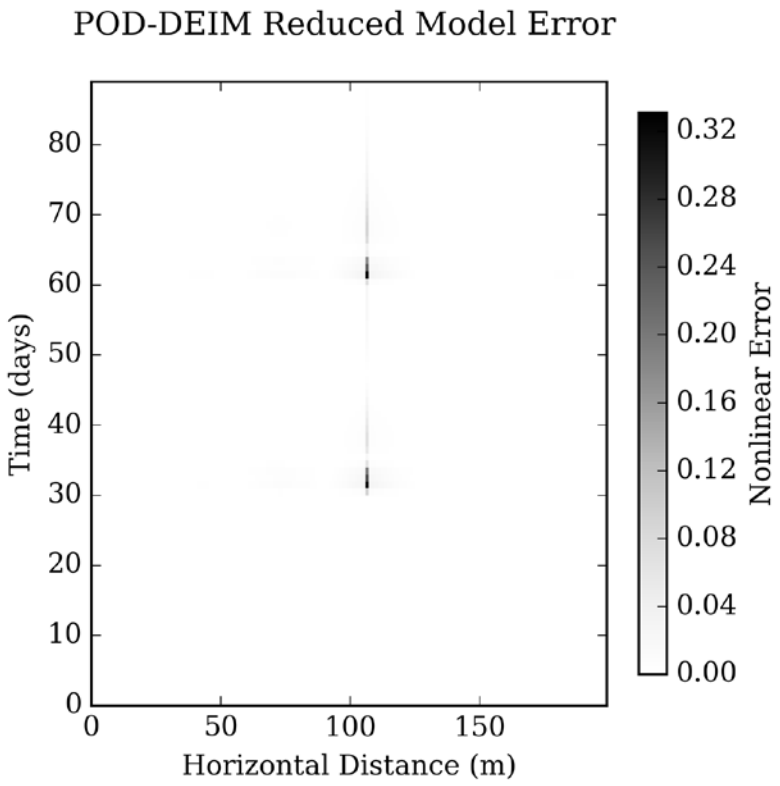

Figure 6: The error in the nonlinear approximation is shown to be at a maximum where the well is located, cell 107, and at the beginning of each time step.

For the two-dimensional case, the NRMSE is calculated for each model cell over all time steps to obtain an overall assessment of the reduced model's performance. Figure 7 displays the error in head for the 2D model. The maximum NRMSE is less than 0.5% and occurs at the location of Well 3. The ripple patterns emanating from some regions are a typical oscillatory behavior of POD errors. Red dots mark the interpolation points identified by the DEIM algorithm. Zones with larger regions of darker shades may indicate that additional snapshots or interpolation points are needed to capture more of a response from a particular model feature.

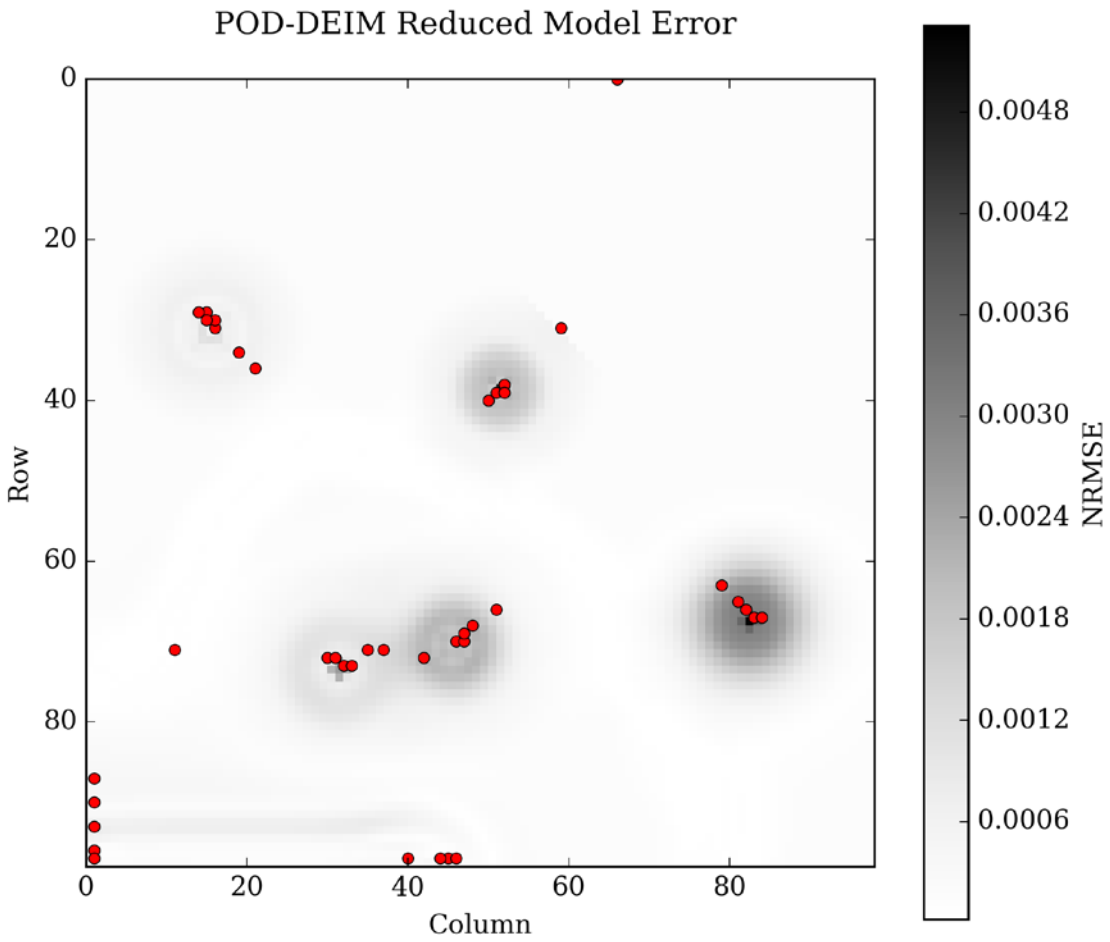

Figure 7: The NRMSE in head for each model cell is calculated over all time steps and shown over the model domain with DEIM interpolation points shown as red dots.

Errors are most likely to occur where the head changes significantly from one time step to another or one model cell to another. Large head gradients are present in the vicinity of the wells and in the time step following a change in pumping rate. To test the model with these situations, Figure 8 shows the water table plotted for two regions of the model. Even with significant drawdown inducing a steep head gradient toward the wells, the reduced model produces a head indistinguishable from the full model. In time step 66, immediately following a change in pumping, the temporal change in head does not produce discernable error.

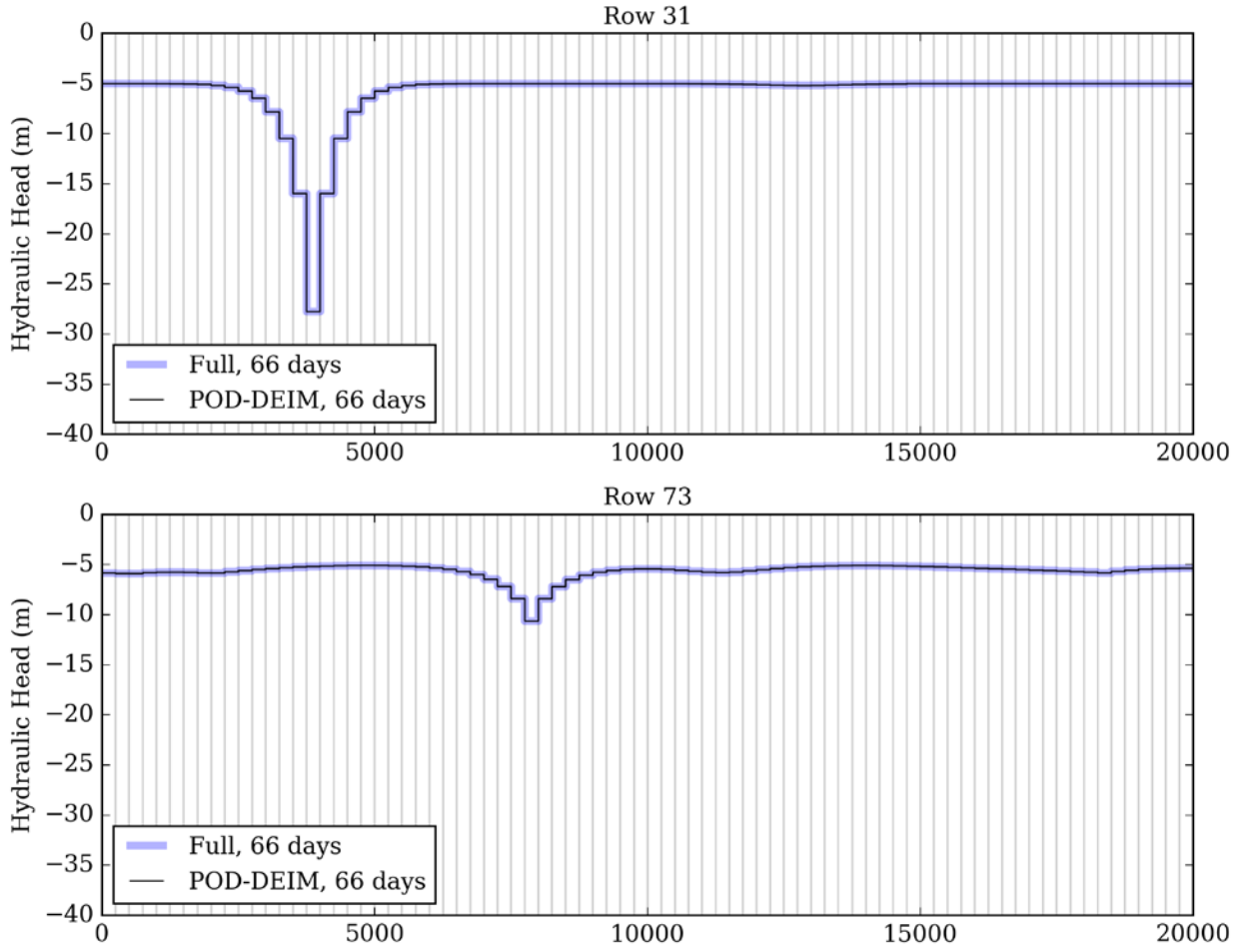

Figure 8: The water table drawdown in two regions is shown for the Full and POD-DEIM reduced model after 66 days, corresponding to the beginning of the third pumping period.

Further analysis of the error at specific time steps reveal different spatial distribution patterns. Examples of error in head and in the nonlinear term are illustrated in Figure 9. At day 65, the errors in head are concentrated at Well 1 and Well 3, yet remain on the order of millimeters. At 66 days, where a change in pumping rate occurs, the error pattern around the wells shifts, with maximum errors occurring in the cells adjacent to the wells. A head error of about 1 cm occurs in rings around the wells and fluctuates slightly between 0.5 and 1.5 cm. For

the nonlinear error, similar spatial patterns emerge at the corresponding time steps, even though interpolation points are clustered around the wells. The magnitude of the error in the nonlinear term has no physical interpretation but still identifies areas where the nonlinear approximation is relatively better than others. With enough interpolation points, the error pattern in the nonlinear operation closely follows that of the error in the head.

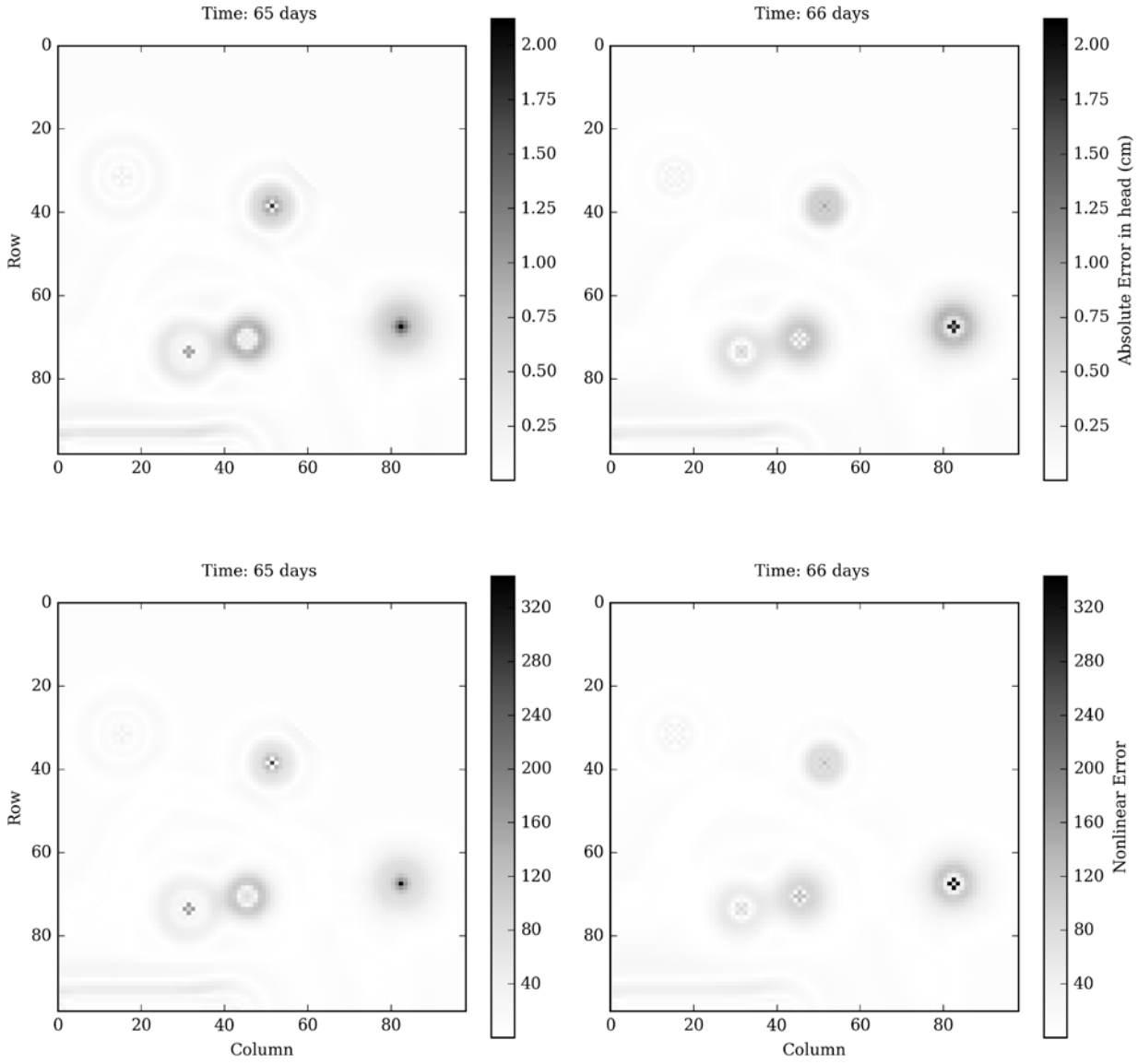

Figure 9: The absolute error in head, as well as the error in the nonlinear approximation, is shown for day 65 and 66, times when the largest errors were observed.

The nature of the errors change dramatically over time. An oscillatory pattern is observed in the time series of errors shown in Figure 10. This phenomena is commonly observed in POD-based model reduction. The maximum absolute error spikes to an amplitude that is still within an acceptable error tolerance but the MAE (mean average error) tends to grow initially to a peak of 2.1 cm but oscillations are attenuated as time progresses.

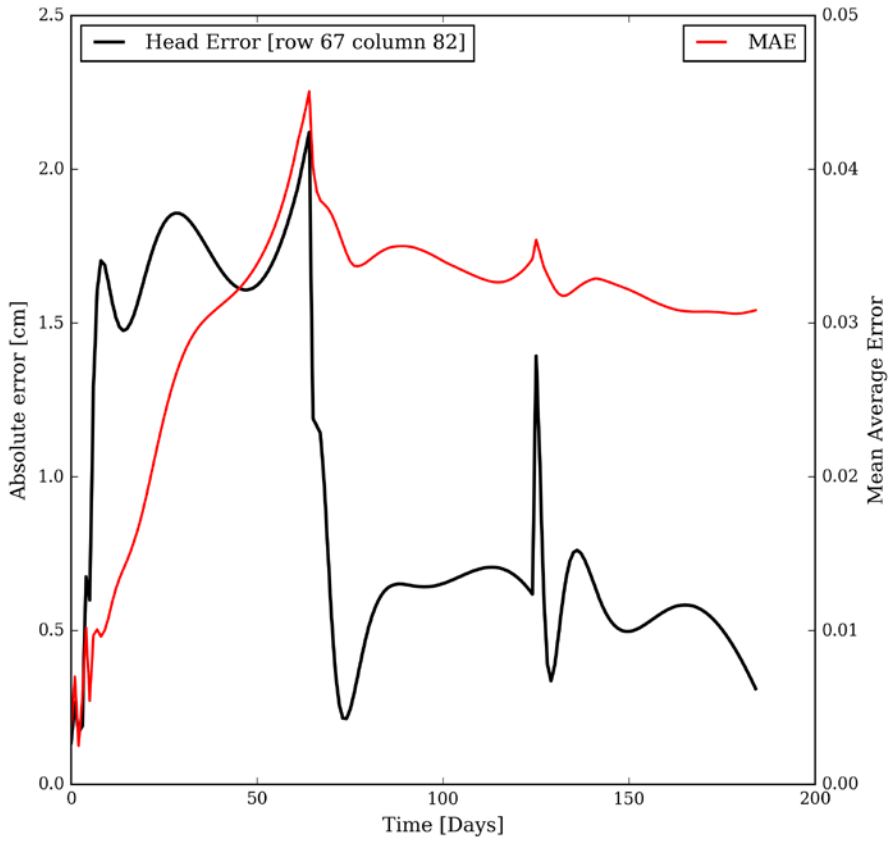

Figure 10: Absolute error at the location found to have the largest error (row 67, column 82) and MAE for the entire domain are computed for each time step and shown as a time series.

For a slightly different perspective on the reduced model error, the RMSE is calculated for the head solution at each time step. Figure 11 illustrates the exceedance curve for the RMSE for each of the reduced models. The POD-DEIM$_{30}$ model (the dual-reduced model using 30 interpolation points) is shown to reach a maximum RMSE of around 0.032 cm a small percentage of the time and the majority of the head RMSE (> 50%) are below 0.03 cm. For the POD-DEIM$_{40}$ model (40 interpolation points with the same dual-reduced approach), the RMSE approaches that of the traditional POD approach, yet has superior speed.

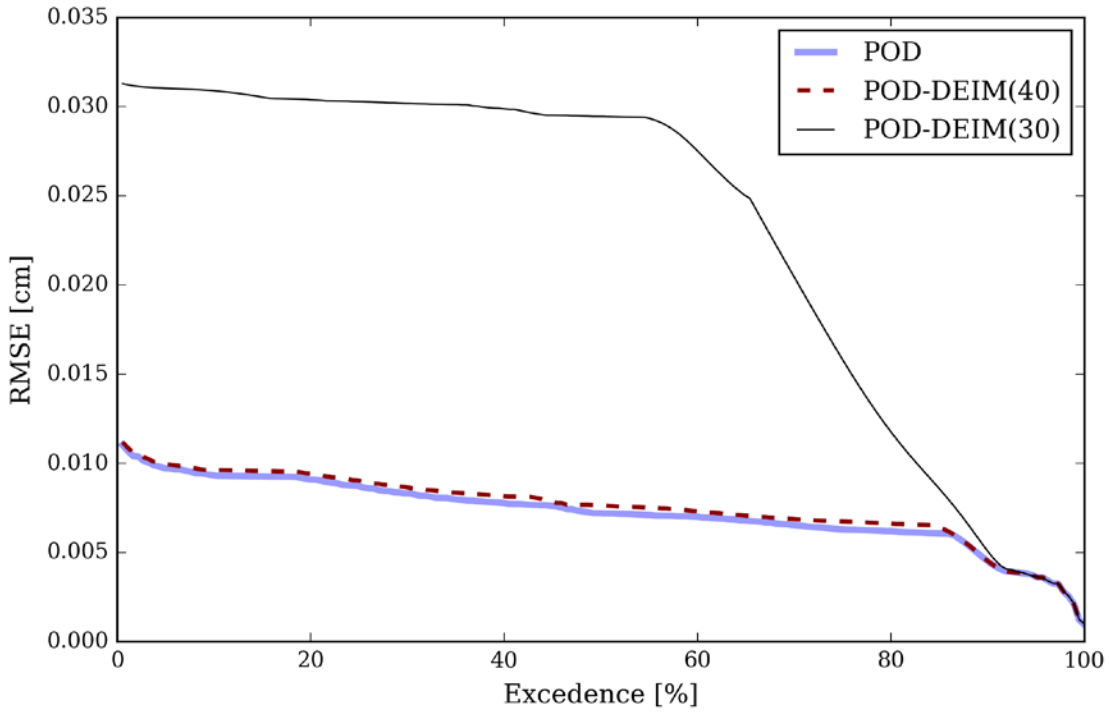

Figure 11: A comparison of the RMSE for the residual error in head ($|\boldsymbol{h} - \boldsymbol{h}_r|$) for each of the reduced models, POD, POD-DEIM$_{30}$ and POD-DEIM$_{40}$, when measured against the full model.

Timing experiments show significant speed improvements, even for small test problems. The speed up for the POD-DEIM$_{40}$ model over the POD model is 0.6 seconds when nearly equal error is obtained. This result may seem insignificant on this relatively small 2D application but when systems approach millions of nodes, this speed improvement will be highly beneficial. Additionally, if slightly more error is accepted with the POD-DEIM$_{30}$ model, a speed up of 1.51 seconds is achieved. The error is controllable with the number of interpolation points above r, depending on the accuracy desired.

Table 3: Timed results compared for each of the reduced models for the 2D test case.

Model	Reduced dimension (r)	Interpolation points (d)	NRMSE	Time in Solver (seconds)
POD	20	0	0.0484	3.61
POD-DEIM$_{40}$	20	40	0.0499	3.01
POD-DEIM$_{30}$	20	30	0.1585	2.10

Discussion

Both POD and POD-DEIM models perform well and have relatively insignificant errors. The nonlinear reduction with DEIM is obtained with only a small loss in accuracy. With both the 1D and 2D numerical experiments, the slight increase in reduced dimension and approximation error is minimal when adding DEIM. Thus, performing the nonlinear operation in the reduced space is preferred. The simplicity of the examples allows for satisfactory proof-of-concept in the 1D and 2D unconfined groundwater flow applications. Obtaining even smaller error would be feasible with additional snapshots, larger reduced dimension, or with added interpolation indices.

More than the value of the error, the structure and distribution of the error are of interest. The results presented analyze times and zones of maximum error, which frequently occur at the beginning of stress periods. However, it is not common to require simulations to be accurate in the first few time steps of a stress period. Since it takes a few iterations for the solution to smoothly adjust heads when new forcings are introduced, it is toward the end of a stress period where results are trusted most, even in the full model. This fact allows further confidence in the reduced model's adequacy. Structurally, the oscillatory appearance of very minor errors appears as ripples in Figure 7 and Figure 9. Irregular waveforms are visible in the error time series shown in Figure 10. This effect may be produced by the nature of the basis functions generated from

POD and the Galerkin projection. Alternative methods may reduce this behavior if the errors are approaching levels that prohibit the application of the reduced model. With careful error assessment, a new reduced model could be constructed to allow for accuracy at a specified radial distance from the well. Therefore, the error introduced from the POD-DEIM reduction is quantifiable and controllable, giving the modeler choices according to the tradeoff between reduced model size, time to construct, and relative importance of some model results (time and location of head observations, for instance) more than others.

Additional strategies could be implemented to construct a more robust reduced model using the POD-DEIM method. The DEIM indices that were selected by the algorithm were unique for each specified reduced dimension. Though the specific DEIM indices were not modified, using a priori knowledge of the system could allow specification of desired indices. Changes to the mesh refinement, snapshot selection, and temporal discretization could also contribute to a more accurate reduced model. Optimizing these variables is outside the scope of this study, though it is prudent to consider the amount of flexibility one would have when constructing a reduced model for more complex projects. If parameter uncertainty is a concern, systematic variation of the parameters when collecting snapshots can generate a parameter independent reduced model.

For models with additional nonlinear processes, the same approach can be used. Fortunately for the case of unconfined groundwater flow, there is only one nonlinear term that is easily resolved with other approximation/linearization techniques. Other applications of the POD-DEIM method have shown success when approximating many different nonlinear operations, each with their own basis and set of interpolation points. For more complex unconfined flow modeling in MODFLOW, these additional nonlinearities might come from any of the head-dependent boundary conditions (MNW, RCH, GHB, DRN, RIV). As more of these features are

modeled, the simulation requires more nonlinear computations at each time step, leading to run times that make any form of optimization (parameter estimation, sensitivity analysis, resource allocation, etc.) quite computationally expensive. With only one POD projection, errors might grow too large when certain values of parameters or boundary conditions were not incorporated into the basis. Therefore, the POD-DEIM method adds an element of structural independence from these changes by approximating the nonlinear dynamics of each additional term with its own interpolation. The dependence on snapshot selection is still an issue, but different sets of snapshots may be used for each nonlinear term, allowing the capture of a variety of possible inputs while maintaining the smallest possible reduced model. To perform this level of parameter independence in the traditional POD, a very large set of snapshots might be necessary, which could lead to only a moderate dimension reduction, and sometimes even an increase in dimension without limits on the size of the snapshot set.

While construction of the reduced model can be time intensive, the overhead investment is easily overcome by the savings once the reduced model is used to solve a variety of problems. Possible applications that would now be feasible include the effects of changing formulations within an optimization problem. Different methods of aggregating decision variables and objectives—necessary because the variable dimensions may be too large for a chosen algorithm or a multi-objective framework may limit the number of objectives or be dependent on weighting factors—can lead to Arrow's Paradox(Kasprzyk et al. 2015). With a sufficiently robust reduced model, suites of optimizations can be performed so any algorithmic tuning mechanism can also be optimized. Additionally, optimization and uncertainty analysis are still underutilized in the development and application of groundwater models for real-world decision making. Mostly, these techniques are not fully implemented due to the immense computational requirements each

method requires for a full solution. Using a reduced model to save even a few minutes or seconds on each model run can add up to time savings on the order of days, weeks, or even months.

Conclusion

A traditional linear model reductions technique for groundwater flow has been combined with an interpolation scheme to further reduce nonlinear components. The result is a reduced model of an unconfined flow equation that can be solved entirely in the reduced dimension with no dependence on the original, full model complexity. This additional approximation allows for faster calculations of nonlinear operations at each time step while sacrificing a tractably small amount of accuracy. As simulation models get more complex, with finer discretization, larger domains, and more nonlinear processes, faster calculations become more important. The combined model reduction approach with POD and DEIM greatly improves a modeler's ability to obtain solutions quickly. The results from the two test problems show a two to three orders of dimension reduction. A key advantage of the POD-DEIM model is that nonlinear operations are carried in the reduced space. The faster overall simulation times are critical when embedding within or linking the model to any form of optimization (e.g., parameter estimation, experimental design, resource allocation) or extensive uncertainty analysis (e.g., Monte Carlo). Additionally, more and more optimization algorithms are taking advantage of parallel computing power, yet long simulation runtimes still inhibit the attainment of optimal solutions in reasonable amounts of time. Therefore, reduced models such as those developed with POD-DEIM can be used within parallel architectures to facilitate searching very large feasible regions—regions with dimensions so large that they would otherwise be impossible to explore.

Acknowledgements

This material is based on co-authored work that has been submitted to *Advances in Water Resources* by Zachary P. Stanko, Scott E. Boyce, and William W.-G. Yeh

The author would like to thank Scott Boyce for his contribution to the computer code with a singular value decomposition routine and the implementation of compressed row storage for all matrices.

This material is based on work supported by NSF under award EAR-1314422. Partial support also was provided by an AECOM endowment.

References

Abgrall, R. & Amsallem, D., 2015. Robust Model Reduction by L 1 -norm Minimization and Approximation via Dictionaries : Application to Linear and Nonlinear Hyperbolic Problems. , 0(1).

Antoulas, A., Sorensen, D. & Gugercin, S., 2001. A survey of model reduction methods for large-scale systems. *Contemporary Mathematics*, 280, pp.193–220.

Barrault, M. et al., 2004. An "empirical interpolation" method: application to efficient reduced-basis discretization of partial differential equations. *Comptes Rendus Mathematique*, 339, pp.667–672.

Boyce, S.E., Nishikawa, T. & Yeh, W.W.-G., 2015. Reduced order modeling of the Newton formulation of MODFLOW to solve unconfined groundwater flow. *Advances in Water Resources*, 83, pp.250–262. Available at: http://www.sciencedirect.com/science/article/pii/S0309170815001311.

Boyce, S.E. & Yeh, W.W.-G., 2014. Parameter-independent model reduction of transient groundwater flow models: Application to inverse problems. *Advances in Water Resources*, 69, pp.168–180. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0309170814000761.

Cardoso, M.A., Durlofsky, L.J. & Sarma, P., 2009. Development and application of reduced-order modeling procedures for subsurface flow simulation. *International Journal for Numerical Methods in Engineering*, 77(9), pp.1322–1350. Available at: http://doi.wiley.com/10.1002/nme.2453.

Chaturantabut, S. & Sorensen, D.C., 2010. Nonlinear Model Reduction via Discrete Empirical Interpolation. *SIAM Journal on Scientific Computing*, 32(5), pp.2737–2764.

Drohmann, M., Haasdonk, B. & Ohlberger, M., 2012. Reduced Basis Approximation for Nonlinear Parametrized Evolution Equations based on Empirical Operator Interpolation. *SIAM Journal on Scientific Computing*, 34(2), pp.A937–A969.

Forstall, V.H., 2015. *Iterative Solution Methods for Reduced-Order Models of Parameterized Partial Differential Equations*.

Hanson, R.T. et al., 2014. *One-Water Hydrologic Flow Model (MODFLOW-OWHM)*,

Harbaugh, Arlen, W., 2005. MODFLOW-2005 , The U . S . Geological Survey Modular Ground-Water Model — the Ground-Water Flow Process. *U.S. Geological Survey Techniques and Methods*, p.253.

Henneron, T. & Clenet, S., 2014. Model Order Reduction of Non-Linear Magnetostatic Problems Based on POD and DEI Methods. *IEEE Transactions on Magnetics*, 50(2), pp.33–36.

Hill, M.C., 1990. Preconditioned conjugate-gradient 2 (PCG2), a computer program for solving ground-water flow equations. *U.S. Geological Survey Water-Resources Investigations Report*, (90–4048), p.43.

Kasprzyk, J.R., Reed, P.M. & Hadka, D.M., 2015. Battling Arrow's Paradox to Discover Robust Water Management Alternatives. *Journal of Water Resources Planning and Management*, p.04015053. Available at: http://ascelibrary.org/doi/10.1061/(ASCE)WR.1943-5452.0000572.

Keating, E. & Zyvoloski, G., A stable and efficient numerical algorithm for unconfined aquifer analysis. *Ground water*, 47(4), pp.569–79. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19341374 [Accessed February 8, 2016].

McPhee, J. & Yeh, W.W.-G., 2008. Groundwater Management Using Model Reduction via Empirical Orthogonal Functions. *Journal of Water Resources Planning and Management*, 134(2), pp.161–170.

Nigro, P.S.B. et al., 2015. An adaptive model order reduction with Quasi-Newton method for nonlinear dynamical problems. *International Journal for Numerical Methods in Engineering*, p.n/a–n/a.

Niswonger, R.G., Panday, S. & Ibaraki, M., 2011. MODFLOW-NWT, a Newton formulation for MODFLOW-2005. *U.S. Geological Survey Techniques and Methods*, 6–A37, p.44.

Quarteroni, A. & Rozza, G., 2007. Numerical solution of parametrized Navier–Stokes equations by reduced basis methods. *Numerical Methods for Partial Differential Equations*, 23(4), pp.923–948. Available at: http://doi.wiley.com/10.1002/num.20249 [Accessed May 22, 2016].

Radermacher, A. & Reese, S., 2015. POD-based model reduction with empirical interpolation applied to nonlinear elasticity. *International Journal for Numerical Methods in Engineering*, p.n/a–n/a.

Reyment, R.A. & Joreskog, K.G., 1993. *Applied factor analysis in the natural sciences*, Cambridge University Press.

Siade, A.J., Putti, M. & Yeh, W.W.-G., 2012. Reduced order parameter estimation using quasilinearization and quadratic programming. *Water Resources Research*, 48(September 2011).

Sirovich, L., 1987. Turbulence and the dynamics of coherent structures Part I: Coherent structures. *Quart Appl Math*, 45(3), pp.561–71.

Stef, R., Sandu, A. & Navon, I.M., 2015. POD / DEIM reduced-order strategies for efficient four dimensional variational data assimilation. , 295, pp.569–595.

Ştefănescu, R. & Navon, I.M., 2013. POD/DEIM nonlinear model order reduction of an ADI implicit shallow water equations model. *Journal of Computational Physics*, 237, pp.95–114. Available at: http://www.sciencedirect.com/science/article/pii/S0021999112007152.

von Storch, H. & Hannoschöck, G., 1985. Statistical aspects of estimated principal vectors (EOFs) based on small sample sizes. *Journal of Climate and Applied Meteorology*, 24(7), pp.716–724.

Vermeulen, P.T.M., Heemink, a. W. & Te Stroet, C.B.M., 2004. Reduced models for linear groundwater flow models using empirical orthogonal functions. *Advances in Water Resources*, 27, pp.57–69.

Vermeulen, P.T.M., Heemink, A.W. & te Stroet, C.B.M., 2004. Low-dimensional modelling of numerical groundwater flow. *Hydrological Processes*, 18(8), pp.1487–1504. Available at: http://dx.doi.org/10.1002/hyp.1424.