Advances in Water Resources 97 (2016) 130–143

Contents lists available at ScienceDirect

Advances in Water Resources

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

Nonlinear model reduction of unconfined groundwater flow using POD and DEIM

Zachary P. Stanko[a], Scott E. Boyce[b,1], William W.-G. Yeh[a,*]

[a] Department of Civil and Environmental Engineering, University of California, Los Angeles, CA 90095, USA
[b] California Water Science Center, U.S. Geological Survey, 4165 Spruance Rd., Suite 200, San Diego, CA 92101-0812, USA

ARTICLE INFO

Article history:
Received 14 February 2016
Revised 29 July 2016
Accepted 12 September 2016
Available online 13 September 2016

Keywords:
Model reduction
Unconfined flow
Proper orthogonal decomposition
Discrete empirical interpolation method
Nonlinear differential equations

ABSTRACT

Nonlinear groundwater flow models have the propensity to be overly complex leading to burdensome computational demands. Reduced modeling 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.

1. 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. Reduced modeling techniques are commonly applied to decrease the computational burden associated with high dimensionality. 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. Applications of POD is also known in the literature as Empirical Orthogonal Functions (EOF) (von Storch and Hannoschöck, 1985; McPhee and Yeh, 2008), Coherent Structures (CS) (Sirovich, 1987), Principal Component Analysis (PCA), Karhunen-Loéve (KL) methods (Graham and Kevrekidis, 1996), or Common Factor Analysis (CFA) (Reyment and Joreskog, 1993).

For groundwater flow specifically, most previous applications of POD have utilized the confined flow equation, which is linear with respect to the state variable of interest, namely, the hydraulic head (Boyce and Yeh, 2014). 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 (Ştefănescu 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 and Amsallem, 2016). Alternatively, linearization techniques such as quasilinearization, can be utilized to

* Corresponding author.
E-mail address: williamy@seas.ucla.edu (W.W.-G. Yeh).
[1] Formerly at University of California, Los Angeles, USA.

http://dx.doi.org/10.1016/j.advwatres.2016.09.005
0309-1708/

facilitate POD reduced model construction for the case of nonlinear parameter estimation (Siade et al., 2012).

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 al. (2015) also successfully reduced an unconfined groundwater model using the Newton formulation of MODFLOW, MODFLOW-NWT (Niswonger et al., 2011). These examples of successful groundwater model reduction with reduced basis methods illustrate the added difficulty with nonlinearities and present methods that are only 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 can 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 and Sorensen, 2010). It has been used in conjunction with POD for reducing FitzHugh–Nagumo equations (Chaturantabut and Sorensen, 2010), shallow water equations (Ştefănescu and 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 and Rozza, 2007).

The joint application of POD and DEIM is a form of hyper-reduction for nonlinear equations. This type of 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 and Clenet, 2014) and in mechanical engineering, the methods were used for a solid mechanics problem involving nonlinear elasticity (Radermacher and 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. Reducing the dimension of these models would permit large-scale applications—such as multi-objective optimization (Reed et al., 2013) or Monte Carlo uncertainty analysis (Kasprzyk et al., 2009; Tonkin and Doherty, 2009; Siade et al., 2015)—that can require hundreds of thousands, or even millions of groundwater model runs.

More robust reduced modeling techniques are needed for nonlinear dynamics in groundwater flow. Methods developed previously have resorted to using strategies designed for linear equations 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 the computational complexity of the full model dimension greatly inhibits the value of developing the reduced model, which can have significant overhead itself. Application of POD-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.

2. Methods

The following methodology introduces unconfined groundwater flow equations and 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.

2.1. Unconfined groundwater flow

The general three-dimensional governing equation for constant-density groundwater flow in an unconfined aquifer is given by (Keating and Zyvoloski, 2009):

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

where $\nabla \bullet$ 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 [L/T], t is the time [T], and S_y is the specific yield [–]. For the case of a two-dimensional flow and under the Dupuit assumptions (essentially horizontal flow), the governing equation for an unconfined aquifer then becomes the Boussinesq equation (Willis and Yeh, 1987):

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

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) [$\mathrm{L\,T^{-1}}$]. Note that Eq. (2) is nonlinear as it involves the product terms of the state variable h.

Applying a finite difference approximation scheme to the spatial variables (x, y) yields the following system of nonlinear ordinary differential equations, represented in matrix form as Eq. (3):

$$\boldsymbol{A}\boldsymbol{h} + \boldsymbol{f} = \boldsymbol{B}\frac{d\boldsymbol{h}}{dt} \tag{3}$$

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 product term ($\boldsymbol{A}\boldsymbol{h}$) 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 (Eq. (4)) is calculated to represent the flow system at each time step:

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

Or

$$\boldsymbol{A}_h\boldsymbol{h}^{(t+1)} = b \tag{4}$$

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
$\varepsilon_{\max}$	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 ith 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 jth 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
$\boldsymbol{z}$	$d \times 1$		Vector of interpolation indices

Defining $[\frac{\boldsymbol{B}}{\Delta_t} - \boldsymbol{A}]$ as the nonlinear system matrix, $\boldsymbol{A}_h$, and $[\frac{\boldsymbol{B}}{\Delta_t}\boldsymbol{h}^t + \boldsymbol{f}]$ 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 $\max_{1<i<n}(h_i^k - h_i^{k+1}) \leq \varepsilon_{max}$ (the maximum head difference from iteration k to $k+1$ is sufficiently small) or $\max_{1<i<n}(inflow_i^{k+1} - outflow_i^{k+1}) \leq r_{max}$ (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. Eq. (4), which is called the full model, lies in the dimension $n \times n$, since there are n equations and n unknowns.

2.2. POD

A brief derivation of POD begins by approximating $\boldsymbol{h}^{t+1}$ with $\boldsymbol{P}\boldsymbol{h}_r^{t+1}$. The matrix $\boldsymbol{P} \in \mathbb{R}^{n \times r}$ 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. An error threshold of percent energy ($\varepsilon_{\text{energy}}$) is used to specify the amount of information retained in the reduced basis

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

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 Eq. (5). 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). Eq. (5) is called the POD-reduced model and still requires matrix multiplication with a dimension of n when $\boldsymbol{P}^T\boldsymbol{A}_h$ is computed at each time step. Due to nonlinearities, each element of $\boldsymbol{A}_h$ must be recomputed each time step; an operation that also scales with n.

$$\boldsymbol{P}^T\boldsymbol{A}_h\ \boldsymbol{P}\boldsymbol{h}_r^{t+1} = \boldsymbol{P}^T\boldsymbol{b} \tag{5}$$

2.3. 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 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 (Eq. (6)) is made by approximating the nonlinear operation ($\boldsymbol{A}_h\boldsymbol{h}^{t+1}$) with a linear interpolation. The selection algorithm for the interpolation indices chooses points that have the largest residual error when iteratively approximating the nonlinear term with basis vectors (selected columns of $\boldsymbol{D}$) times a vector of new reduced variables, $\boldsymbol{c}$.

$$\boldsymbol{A}_h\boldsymbol{h}^{t+1} \cong \boldsymbol{D}\boldsymbol{c} \tag{6}$$

where $\boldsymbol{D} \in \mathbb{R}^{n \times d}$ is generated by performing POD on snapshots of $\boldsymbol{b}$ from Eq. (4) (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 Eq. (6) 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 Ştefănescu and Navon (2013). Defining $\boldsymbol{Z} \in \mathbb{R}^{n \times d}$ as a diminished permutation matrix that retains only the columns corresponding to the interpolation indices. Premultiplying by $\boldsymbol{Z}^T$effectively selects only the rows of $\boldsymbol{A}_h$ and $\boldsymbol{D}$ that correspond to interpolation points and Eq. (6) becomes:

$$\boldsymbol{Z}^T\boldsymbol{A}_h\ \boldsymbol{h}^{t+1} = \boldsymbol{Z}^T\boldsymbol{D}\boldsymbol{c} \tag{7}$$

Eq. (7) can be inverted to solve for $\boldsymbol{c}$ (Eq. (8)) and subsequently substituted back into Eq. (6) to obtain the final approximation of

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 $\|\boldsymbol{r}\|_\infty$

Fig. 1. Algorithm for selecting the interpolation indices $\boldsymbol{z}_j$ for $j = 1, \ldots, d$.

the nonlinear term (Eq. (9)). Finally, POD and Galerkin projection can be applied, as in Eq. (5), to achieve a new reduced model (Eq. (10))

$$\boldsymbol{c} = \left[\boldsymbol{Z}^T\boldsymbol{D}\right]^{-1}\boldsymbol{Z}^T\boldsymbol{A}_h\ \boldsymbol{h}^{t+1} \tag{8}$$

$$\boldsymbol{A}_h\boldsymbol{h}^{t+1} \cong \boldsymbol{D}\left[\boldsymbol{Z}^T\boldsymbol{D}\right]^{-1}\boldsymbol{Z}^T\boldsymbol{A}_h\boldsymbol{h}^{t+1} \tag{9}$$

$$\boldsymbol{P}^T\boldsymbol{D}\left[\boldsymbol{Z}^T\boldsymbol{D}\right]^{-1}\boldsymbol{Z}^T\boldsymbol{A}_h\boldsymbol{P}\boldsymbol{h}_r^{t+1} = \boldsymbol{P}^T\boldsymbol{b} \tag{10}$$

All calculations dependent on the dimension n can be precomputed and the final POD-DEIM reduced model (Eq. (10)) 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 full n-dimension allowing the reduced model to be solved at each iteration depending only on dimensions r and d. Note that since $\boldsymbol{A}_h$ is a sparse matrix that results from finite difference discretization, the complexity of $\boldsymbol{A}_h\boldsymbol{P}$ scales with the dimension r, not n.

Fig. 1 shows the algorithm for selecting the interpolation indices and the steps involved in the DEIM procedure.

2.4. MODFLOW framework

The hyper-reduction within MODFLOW is performed via traditional online-offline paradigm. An additional MODFLOW package, MRED, has been created for MODFLOW-OWHM (Hanson et al., 2014) and it contains all model reduction subroutines. To further enhance the unconfined reduction capabilities, the DEIM algorithm was added to this package. The structure of the MRED package mirrors that of other MODFLOW packages. The POD-DEIM algorithm is demonstrated in the flow chart of Fig. 2. The offline portion need only be executed once for the entire simulation and the online portion is required for each iteration of the model's solver. The LAPACK LU-decomposition routine is used as the reduced solver (documented: https://software.intel.com/en-us/node/520973) and the MODFLOW implementation of the preconditioned conjugate gradient (PCG) method is used as the full model solver. 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.

2.5. Model development

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. There are a total of 90 1-day time steps in the transient model, which is divided into three 30-day uniform pumping periods (referred to as stress periods). The well begins extracting on day 30, pumps for 30 days, and then shuts off for 30 days. Constant-head boundaries are set to zero units of head at nodes 1 and 200 and a pumping well is placed at node 107. The initial condition is set to 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 pumping rate was varied to account for uncertain pumping. The reduced model was generated with 180 snapshots: 90 snapshots with $Q = 100\ \mathrm{m}^3$/day and 90 snapshots with $Q = 200\ \mathrm{m}^3$/day. Fig. 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 then tested with $Q = 150\ \mathrm{m}^3$/day to evaluate its effectiveness at pumping values not used to generate snapshots.

After the 1D case has been successfully verified, we then extend the methodology to a more realistic 2D case. The square model domain has sides of length 24,750 m and a discretization of 198 rows and 198 columns, resulting in a model grid with 39,204 uniform square cells (125 m by 125 m). The 395-day simulation has 14 stress periods: an initial 5-day period with zero pumping; 12 30-day periods with variable extractions; and a final 30-day period with zero pumping to simulate the aquifer recovery. The 5-day period has 19 time steps while other stress periods each have 13 time steps for a total of 188 time steps per simulation. The head starts at −5.0 m everywhere, creating unconfined conditions in the single layer with a thickness of 100 m. Constant head boundaries compose the top and bottom (via CHD package) while general head (GHB package) and no-flow boundaries compose the left and right boundaries. Two additional head-dependent boundary conditions—a river and a drain—were added with the RIV and DRN packages, respectively. Pumping at five production wells begins to drawdown the water table after five days. Six zones of hydraulic conductivity, all assumed to be isotropic, span four orders of magnitude. The model domain is shown in Fig. 4 with its zonation pattern and well locations. Parameter values and pumping rates for the tested model are shown in Table 2.

2.6. Snapshot selection

To mitigate for the reduced model's dependence on snapshot selection, the snapshots were taken at each time step of a simulation. Overall six simulations were used to develop the 2D snapshot set: one with all pumping rates at their respective maximums (see Table 2) and one for each of the five wells pumping at quarterly (three month) rates while all other wells are shut off. Quarterly rates for each well are 50, 75, 100, and 75% of the maximum rate for quarter 1, 2, 3, and 4, respectively. The total snapshot count is then 1128, making the snapshot sets ($\boldsymbol{\Phi}_h$ and $\boldsymbol{\Phi}_b$) matrices with dimension 39,204 × 1128. Truncated SVD on this matrix with $\varepsilon_{\text{energy}} = 99.999\%$ produces the projection matrix $\boldsymbol{P}$ and the same

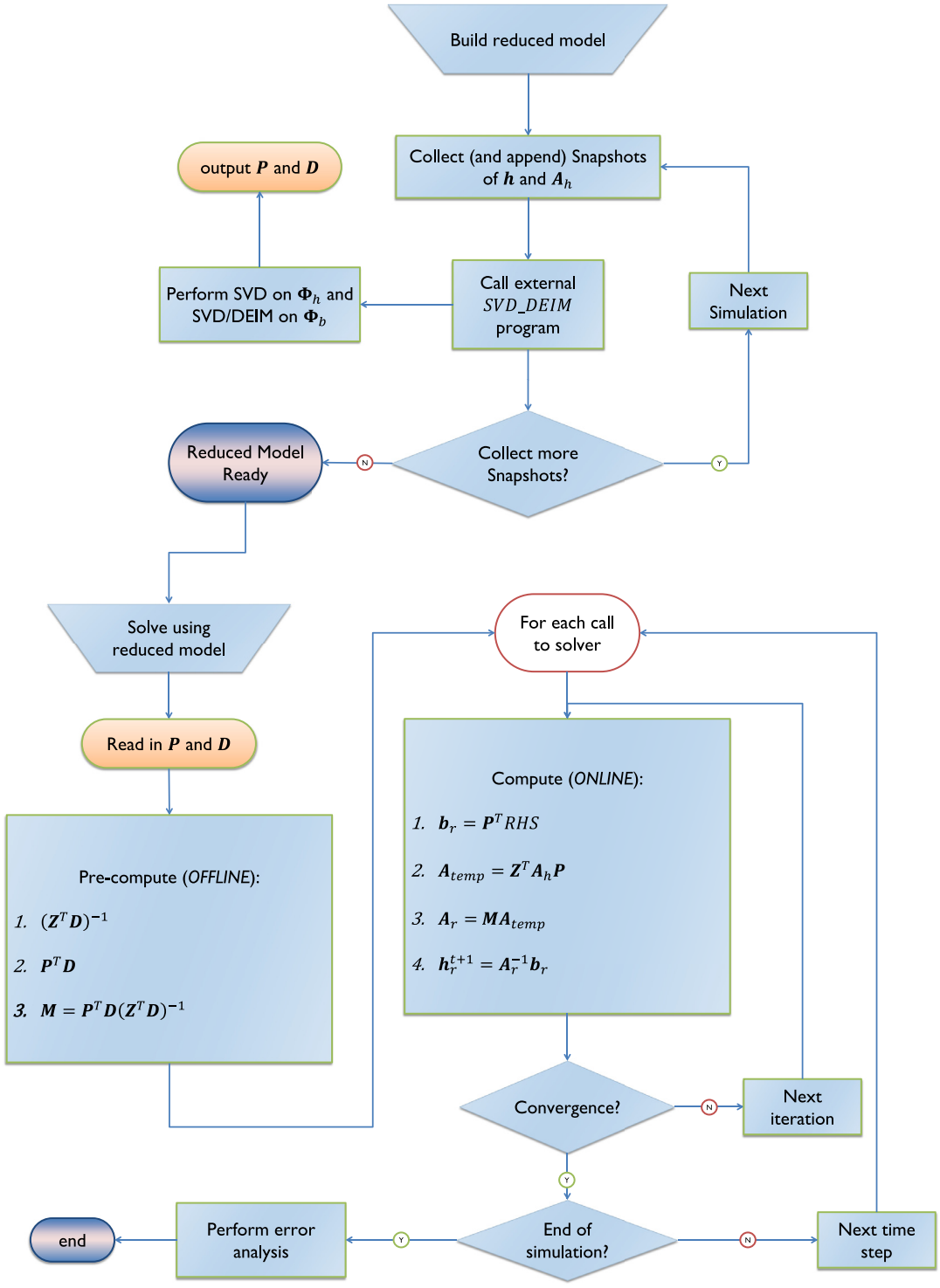

Fig. 2. The flow chart describes the process of collecting snapshots for both traditional POD and DEIM; constructing the snapshot sets and obtaining the bases; and solving with the POD-DEIM reduced model.

Table 2
Parameter and pumping values for the 2D test case.

Parameter	Value [m/day]	Pumping well	Location [row, col]	Max extraction [m^3/day]	Quarterly rates [1000 m^3/day]			
K_{x1}	0.08	1	(63, 31)	10,000	5	7.5	10	7.5
K_{x2}	5.60	2	(147, 63)	16,000	8	12	16	12
K_{x3}	8.20	3	(135, 165)	28,000	14	21	28	21
K_{x4}	10.00	4	(78, 103)	18,000	9	13.5	18	13.5
K_{x5}	18.10	5	(141, 92)	24,000	12	18	24	18
K_{x6}	22.80							

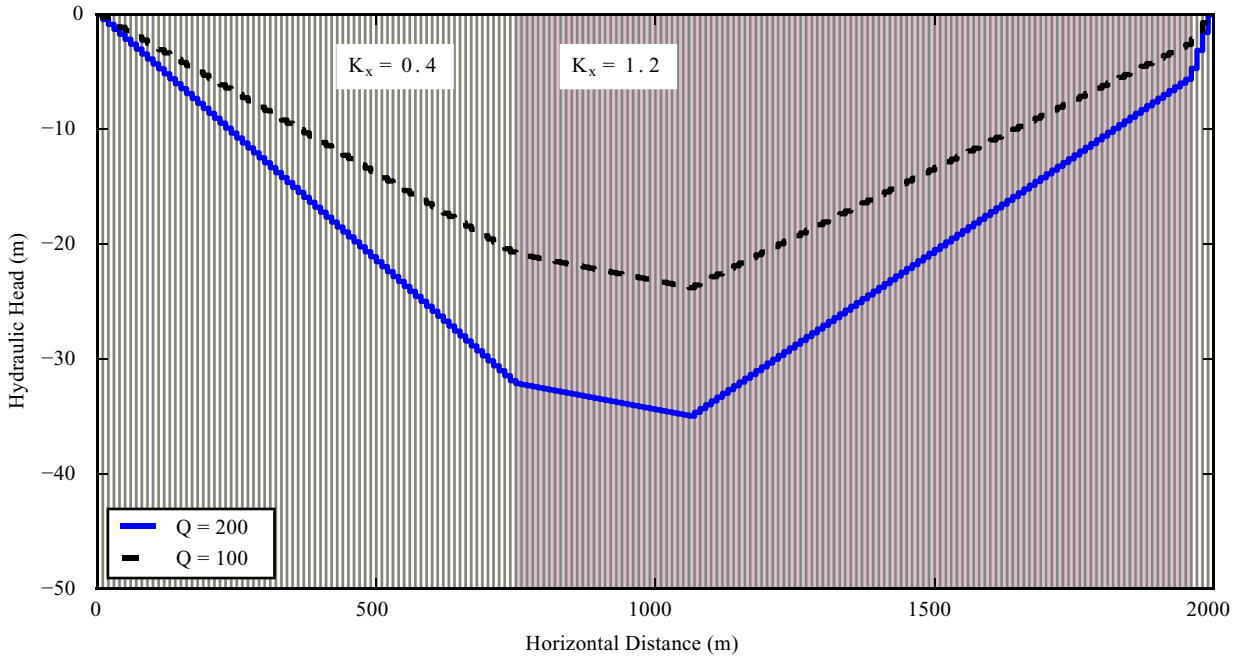

Fig. 3. The 1-dimensional model shows the conductivity zones and the water table after 30 days of pumping, creating the unconfined conditions, for pumping rates of 100 and 200 m^3/day.

method on $\boldsymbol{\Phi}_b$ produces $\boldsymbol{D}$ to which DEIM is applied to obtain $\boldsymbol{Z}$. Once built, the reduced model is first tested with pumping rates at 62.5% of the maximum. Then, 50 random samples were generated to allow the pumping rates to vary between 0 and 100% for each pumping quarter.

2.7. 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, the location of the maximum error is also shown 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. It is not necessary for the reduced dimensions, r and d, to be the same for the POD and POD-DEIM models, they are chosen experimentally based on what is required for acceptable error. For the purposes of this paper, errors less than 1 cm are deemed insignificant. Root mean squared errors (RMSE), and normalized root mean squared error (NRMSE) are calculated for each case so the errors can be weighted by the span of head values (Eq. (11))

$$NRMSE = \frac{\| full - reduced \|_2}{\sqrt{n} * (h_{max} - h_{min})} = \frac{RMSE}{(h_{max} - h_{min})}. \quad (11)$$

3. Results

The full model (Eq. (4)) is compared to both the POD (Eq. (5)) and the POD-DEIM model (Eq. (10)) to investigate errors in the 1D test case. 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 (Fig. 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. A cross-section of head is also shown for the two time steps with the largest error. The reduced model head profile is indistinguishable from the full model head. Fig. 5 also shows that the full model's head distribution at the end of the first pumping period and at the beginning of the recovery period is well matched by the reduced model.

The performance of the nonlinear reduction is analyzed by comparing the result of the nonlinear operation. Fig. 6 shows the residual error ($|\boldsymbol{b} - \boldsymbol{b}_r|$) 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 only at the pumping cell and persist through the entire model horizon. Distance is shown on the x-axis and time on the y-axis with shading to represent residual error. Some space and time results are omitted to focus on the only noticeable errors. Again, the errors are small with respect to the values being compared. Fig. 6 also shows a time series of head at the well and a time series of nonlinear error at the well. A very slight increase in error is observed in the aquifer recovery period over the pumping period.

Two to three orders of dimension reduction is achieved with both the 1D and 2D test problems. This reduction is quite significant, particularly for a highly discretized model. In Table 3, a summary of the reduced model's performance is compared to the full model using several metrics. The minimum head value, which is also corresponds to the maximum drawdown, is recorded to show the head range magnitude that is 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. The DEIM algorithm selected 5 interpolation indices for the 1D model with

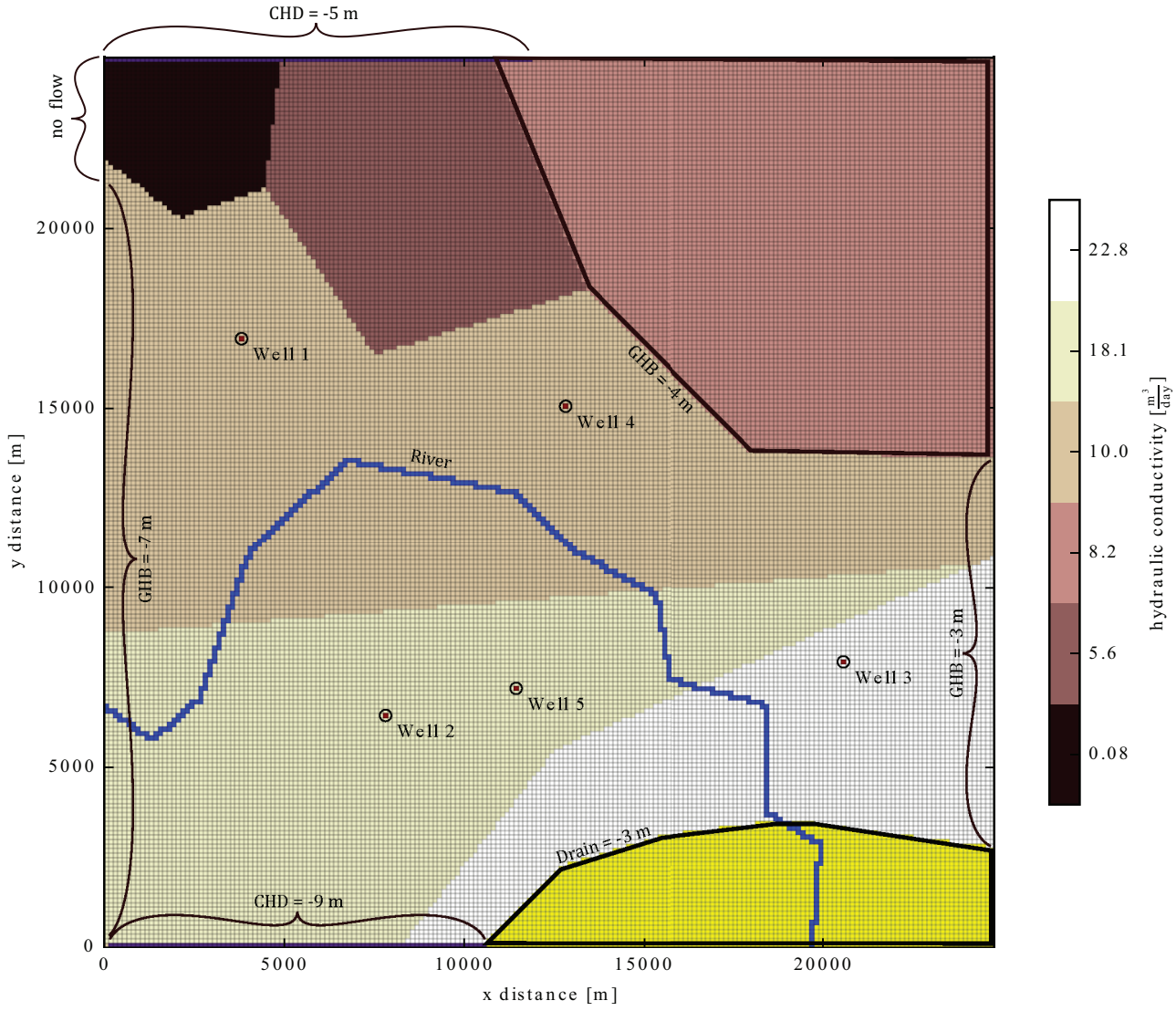

Fig. 4. The domain of the 2D test case is shown with a model grid of 198 rows and 198 columns and six conductivity zones that span several orders of magnitude. There are five wells that pump at various rates and head-dependent features (river and drain) are included.

Table 3
Model reduction results are displayed in comparison to the full, unreduced model.

				h			$A_h * h$		
		Dim	h_{min} [m]	max error [m]	max loc (t, row, [col])	NRMSE [m]	max error	max loc (t, row, [col])	NRMSE
1D	Full model	200	−26.23	–	–	–	–	–	–
1D	POD	4	−26.23	5.66E−3	(61, 69)	9.70E−2	–	–	–
1D	POD-DEIM	5	−26.23	5.66E−3	(61, 69)	9.70E−2	0.331	(61, 106)	1.64E−2
2D	Full model	39,204	−13.95	–	–	–	–	–	–
2D	POD	65	−13.95	9.3E−4	(188, 143, 153)	5.6E−6	–	–	–
2D	POD-DEIM	200	−13.95	3.2E−3	(188, 62, 90)	4.4E−5	62.9	(163, 93, 117)	1.1E−6
2D	POD-DEIM	250	−13.95	5.6E−3	(178, 30, 12)	7.1E−5	111.4	(176, 30, 12)	1.8E−6

$\varepsilon_{energy} = 99.99\%$. For the 2D model, two sets of indices were selected experimentally: one with 200 interpolation points and one with 250. 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, such as near the GHB.

For the two-dimensional case, the RMSE is calculated for each model cell over all time steps to obtain an overall assessment of the reduced model's performance. Fig. 7 displays the error in head for two versions of the 2D model: one with 200 interpolation points ($d = 200$) and one with 250 interpolation points ($d = 250$). The maximum RMSE is less than 0.02% of the span in head values (NRMSE = 0.0002) for $d = 200$ and occurs at between Well 3 and Well 4. The RMSE for $d = 250$ is distributed differently yet still extremely low throughout the entire domain. The ripple patterns emanating from some regions are a typical oscillatory behavior of POD errors. 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

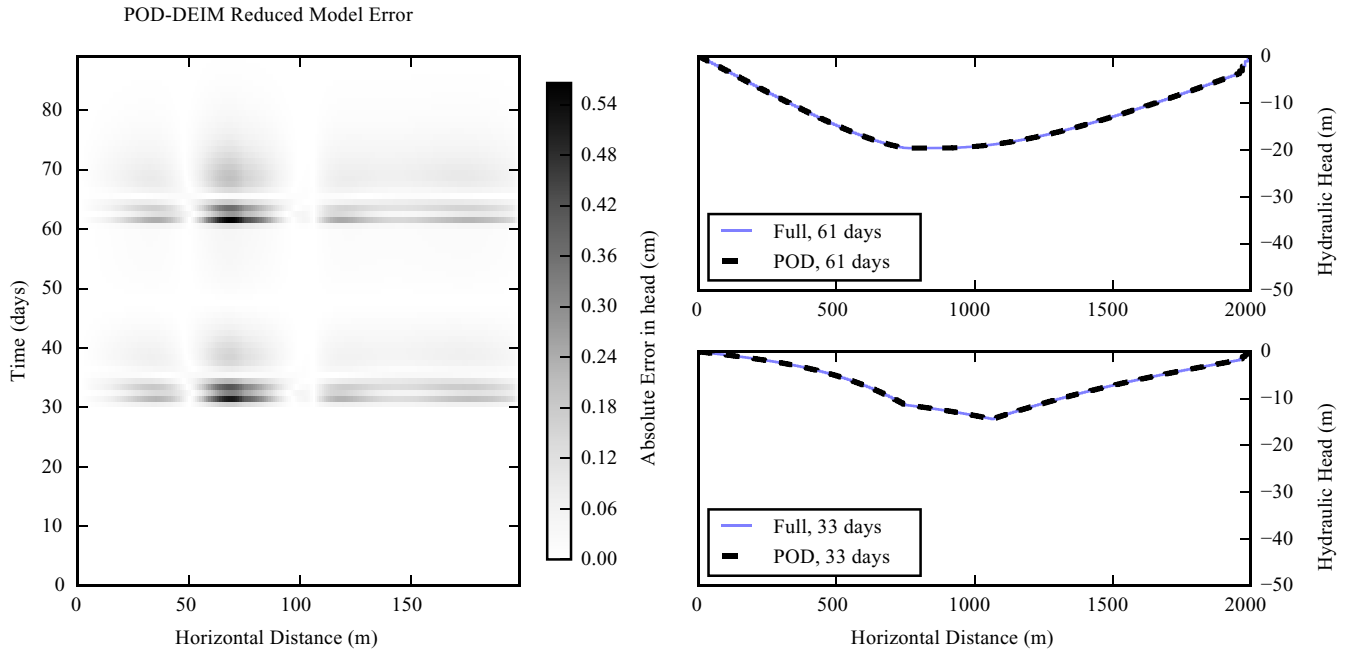

Fig. 5. (Left) The results from the 1-D test are shown as a residual between full model head and reduced model head, for each of the 90 time steps in the POD-DEIM reduced model. (Right) A cross-section of head also shows a match between the full and reduced models.

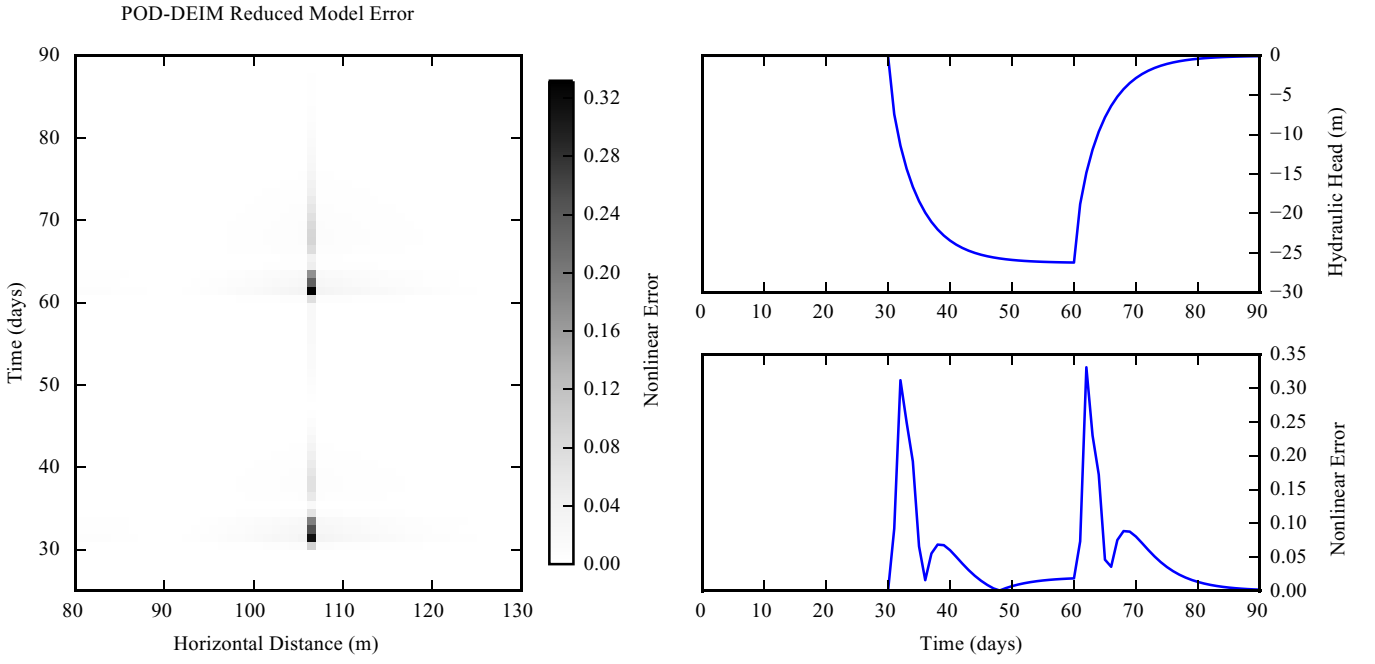

Fig. 6. (Left) The error in the nonlinear approximation is shown to be at a maximum where the well is located, cell 107, and at the beginning of each time step. (Right) A time series of head at the well is also shown above a time series of nonlinear error at the well.

needed to capture more of a response from a particular model feature.

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, Fig. 8 shows the water table plotted for two regions of the model that capture the drawdown from wells 2 and 3 (row 147 and 135, respectively). Even with significant drawdown inducing a steep head gradient toward the wells at the end of the third pumping quarter (day 305), the reduced model produces a head indistinguishable from the full model. After 395 days, the end of the simulation, the large temporal changes in head, as the water table recovers, does not produce discernable error either.

Further analysis of the error at specific time steps reveals different spatial distribution patterns. Examples of error in head and in the nonlinear term are illustrated in Fig. 9. On day 305, the largest errors in head are concentrated between Well 2 and the boundary, yet remain less than five millimeters. At 395 days, where pumping stops, the error pattern is similar, with maximum errors occurring in similar locations. A head error less than 5 mm is occurs throughout the domain. For the nonlinear error, similar spatial patterns emerge at the corresponding time steps. Even though interpolation

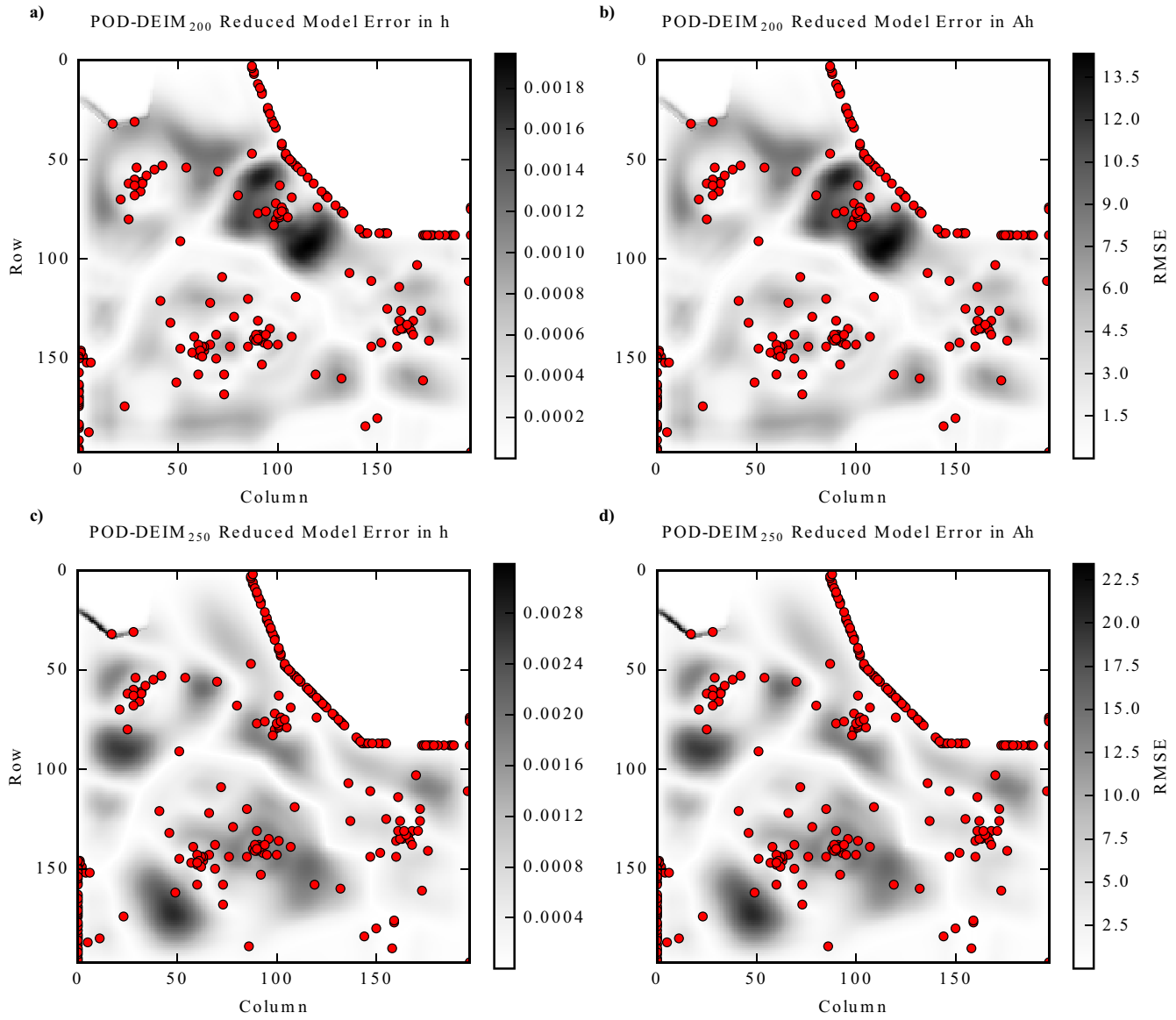

Fig. 7. The RMSE in head for each model cell is calculated over all time steps and shown over the model domain with DEIM interpolation points shown as dots. (a) Errors in **h** and (b) errors in **Ah** for the reduced model with $d=200$ are shown. Also shown are the (c) errors in **h** and (d) errors in **Ah** for the reduced model with $d=250$.

points are clustered around the wells, maximum nonlinear errors occur 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.

The error statistics propagate over time yet remain well below a reasonable threshold. The oscillatory error patterns are observed as small fluctuations in the time series of absolute errors shown in Fig. 10. This phenomena is commonly observed in POD-based model reduction and the propagation of error results in the largest errors occurring at the end of the simulation. The absolute error in head is shown for row 62 and column 90 (the location of the maximum error) at each time step for POD-DEIM$_{200}$ (the dual-reduced model using 200 interpolation points). The maximum absolute error spikes to 32 mm at the end of the simulation but is still within an acceptable error tolerance. For the POD-DEIM$_{250}$ model (250 interpolation points with the same dual-reduced approach), additional interpolation points level out the maximum errors at 56 mm, which now occurs at row 30 column 12. Oscillations in absolute error are attenuated as time progresses and are reduced with added interpolation points. The MAE (mean absolute error) tends to grow steadily over time but remains below 1 mm and is deemed negligible for the purposes of this study.

For a more robust test of the reduced model error, 50 model runs were performed using randomly generated pumping rates (as a fraction of the maximum). The RMSE is calculated for the head solution at each time step of each model run. Fig. 11 illustrates the exceedance curve of the RMSE for each of the reduced models. Here, another version of the reduced model, POD-DEIM$_{225}$, is also included. The POD-DEIM models are shown to reach a maximum RMSE below 1 cm a small percentage of the time and the majority of the RMSE (>60%) are below 1 mm. For the POD-DEIM$_{250}$ model, the RMSE is always less than or equal to the two other POD-DEIM versions. All RMSEs approach those of the traditional POD approach yet have superior speed.

The detailed statistical results of the 50 model runs are shown in Tables 4 and 5. Timing experiments show significant speed improvements, even for small test problems. Timing is calculated by summing the time for each call to the solver. Since both the POD

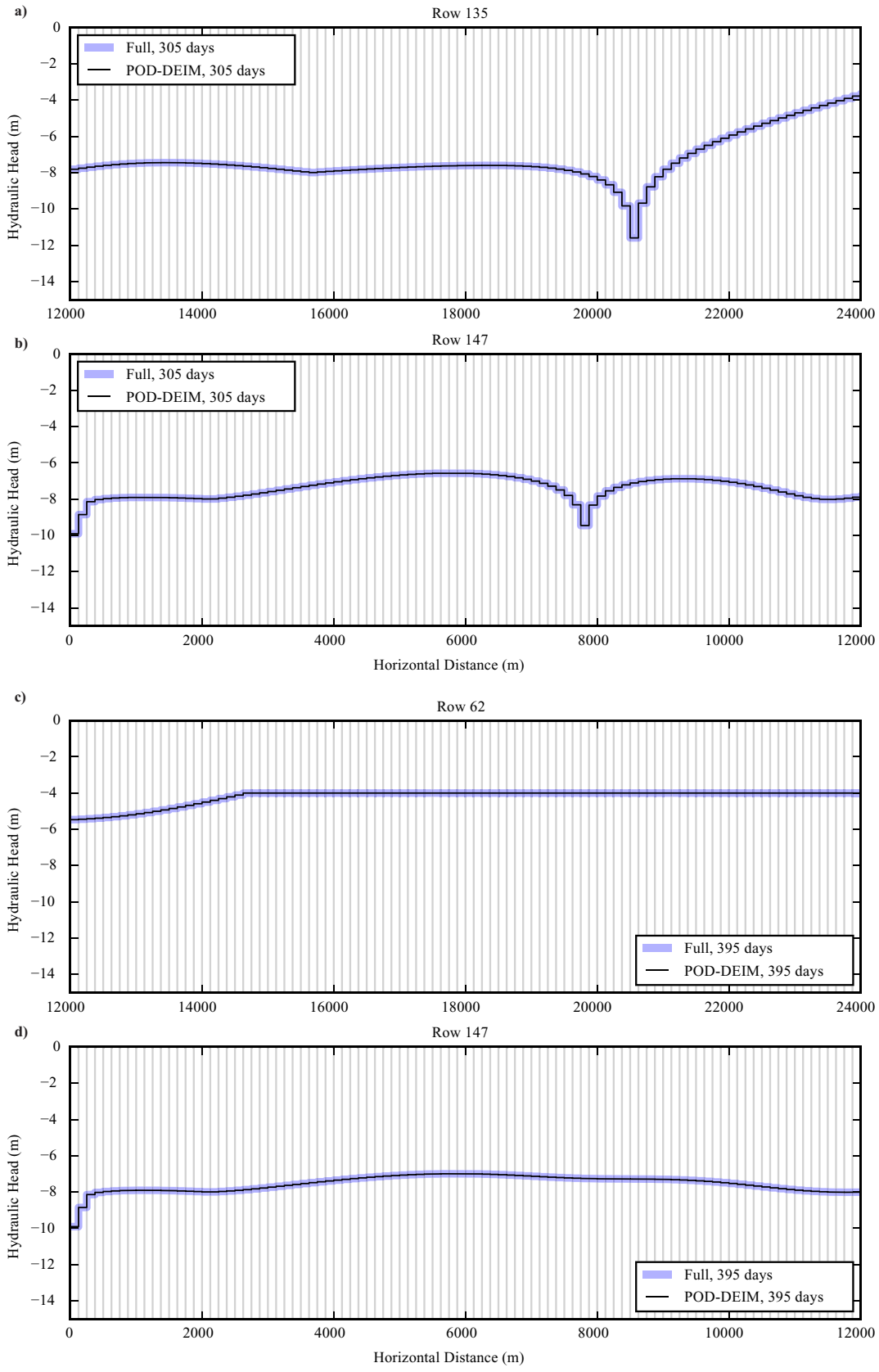

Fig. 8. The water table drawdown in selected regions is shown for the Full and POD-DEIM reduced model. Cross sections are shown for (a) 305 days at Well 3; (b) 305 days at Well 2; (c) 395 days at Well 1; and (d) 395 days at Well 2.

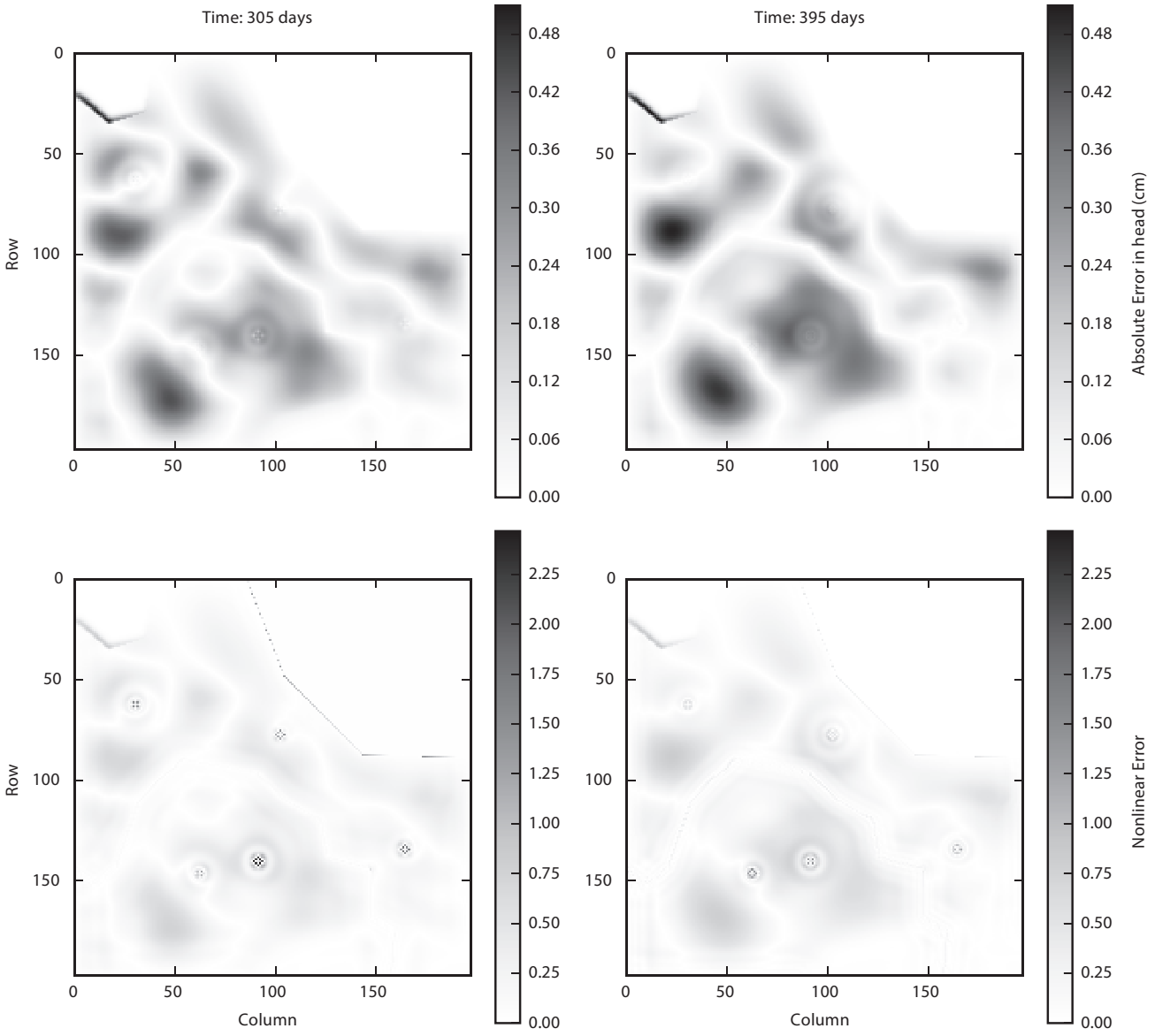

Fig. 9. (Top) The absolute error in head as well as (Bottom) the error in the nonlinear approximation are shown for (Left) day 305 and (Right) day 395, when large errors were observed.

and the POD-DEIM reduced models utilize the same solver, these times are comparable. The average speed up for the POD-DEIM$_{250}$ model over the POD model is 9.5 s while nearly equal error is obtained. Timing is not improved as additional interpolation points are added to the POD-DEIM reduced model. Though the full model utilizes a different solver, the mean total time for all calls to the PCG routine was about 3.6 s. This result indicates that for this model application, the POD reduced model causes an increase in CPU time compared to a near two-second decrease in CPU time for the POD-DEIM models. The maximum, minimum, and mean errors mostly decrease as more interpolation points are added. Similarly with the nonlinear errors, there is a unanimous decrease in errors between the POD-DEIM$_{200}$ model and the POD-DEIM$_{250}$ model.

4. 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 but a gain in computational speed (Table 4). With both the 1D and 2D numerical experiments, the slight increase in approximation error is negligible when adding DEIM with an appropriate number of interpolation points. Thus, performing the nonlinear operation in the reduced space is preferable. 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 importantly, the dimension of the POD-DEIM$_{250}$ model reduction is from 39,204 to 250. Consequently, with POD alone, matrix calculations with a dimension on the order of 10^4 still must be performed at each iteration of the solver. The highest dimension of any matrix calculation with the POD-DEIM approach is on the order of 10^2.

More than the magnitude of the errors, 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

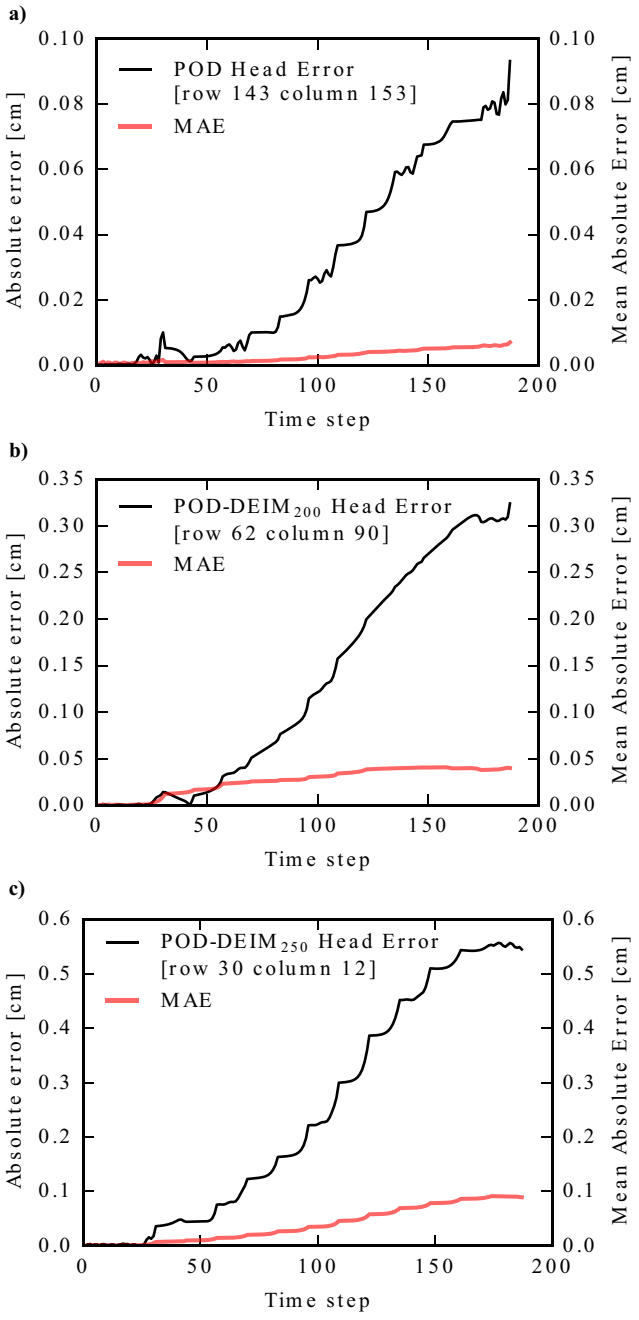

Fig. 10. Absolute error at the location found to have the largest error for the (a) POD, (b) POD-DEIM$_{200}$, and (c) POD-DEIM$_{250}$ reduced models. The MAE for the entire domain is computed for each time step and shown as a time series.

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 Figs. 7 and 9 and as oscillations in Fig. 10. This effect may be produced by the nature of the basis functions generated from POD and the Galerkin projection. If the errors approach levels that prohibit the application of the reduced model, alternative methods may be necessary. With careful error assessment, a new reduced model could be constructed to enhance accuracy at specific points in space and time. 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.

Table 4
Hydraulic head error statistics over 50 random samples for each of the 2D reduced models.

Model type	Minimum head			
	Min.	Max.	Mean	Std. dev.
POD	−18.6081	−11.3691	−14.9834	1.5532
POD-DEIM$_{200}$	−18.6083	−11.3680	−14.9838	1.5532
POD-DEIM$_{225}$	−18.6117	−11.3684	−14.9847	1.5536
POD-DEIM$_{250}$	−18.6080	−11.3693	−14.9847	1.5525
	Maximum absolute error			
	Min.	Max.	Mean	Std. dev.
POD	2.00E−03	7.48E−03	3.82E−03	1.45E−03
POD-DEIM$_{200}$	4.37E−03	1.76E−02	8.17E−03	3.25E−03
POD-DEIM$_{225}$	6.07E−03	1.76E−02	9.48E−03	2.41E−03
POD-DEIM$_{250}$	4.32E−03	1.22E−02	7.27E−03	1.82E−03
	Overall time-averaged RMSE			
	Min.	Max.	Mean	Std. dev.
POD	7.42E−05	1.89E−04	1.41E−04	2.41E−05
POD-DEIM$_{200}$	4.90E−04	1.93E−03	1.00E−03	2.89E−04
POD-DEIM$_{225}$	6.79E−04	1.46E−03	9.93E−04	1.68E−04
POD-DEIM$_{250}$	5.05E−04	1.14E−03	7.78E−04	1.33E−04
	NRMSE			
	Min.	Max.	Mean	Std. dev.
POD	2.23E−02	6.70E−02	4.02E−02	1.09E−02
POD-DEIM$_{200}$	3.86E−05	1.60E−04	8.50E−05	2.70E−05
POD-DEIM$_{225}$	5.15E−05	1.35E−04	8.45E−05	1.90E−05
POD-DEIM$_{250}$	3.50E−05	9.85E−05	6.64E−05	1.56E−05
	Time in solver			
	Min.	Max.	Mean	Std. dev.
POD	1.02E+01	1.23E+01	1.12E+01	3.83E−01
POD-DEIM$_{200}$	1.49E+00	2.49E+00	1.70E+00	1.28E−01
POD-DEIM$_{225}$	1.54E+00	1.76E+00	1.64E+00	5.43E−02
POD-DEIM$_{250}$	1.53E+00	1.82E+00	1.66E+00	5.91E−02

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 prior 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 any parameters when collecting snapshots can generate a parameter-independent reduced model (Boyce and Yeh, 2014).

For models with additional nonlinear processes, this POD-DEIM approach can be used. For more complex unconfined flow modeling in MODFLOW, these additional nonlinearities might come from more head-dependent boundary conditions (MNW, RCH, GHB, DRN, RIV) or more drawdown with additional wells. As more of these features are modeled, the simulation requires more nonlinear computations at each time step and reducing the dimension of the nonlinear term has increased benefit. If the model needs to be called 50,000 or 100,000 times—such is the case with many global heuristic search algorithms, such as a genetic algorithm—savings of

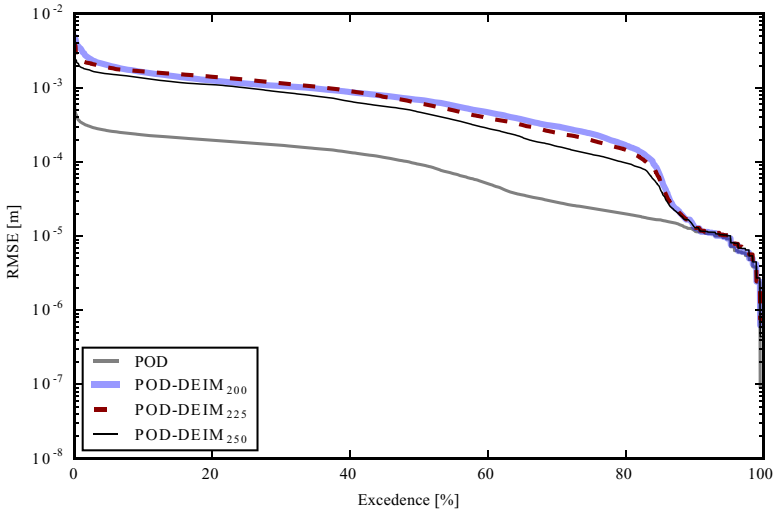

Fig. 11. A comparison of the RMSE for the residual error in head ($|\mathbf{h}-\mathbf{h_r}|$) for each of the reduced models, POD, POD-DEIM$_{200}$, POD-DEIM$_{225}$, and POD-DEIM$_{250}$ when measured against the full model.

Table 5
Nonlinear error statistics over 50 random samples for each of the 2D reduced models.

Model type	Maximum absolute error			
	Min.	Max.	Mean	Std. dev.
POD-DEIM$_{200}$	8.08E+01	3.07E+02	1.48E+02	5.32E+01
POD-DEIM$_{225}$	1.17E+02	2.92E+02	1.77E+02	4.14E+01
POD-DEIM$_{250}$	8.42E+01	2.37E+02	1.36E+02	3.16E+01
	Average RMSE			
	Min.	Max.	Mean	Std. dev.
POD-DEIM$_{200}$	3.59E+00	1.38E+01	7.26E+00	2.02E+00
POD-DEIM$_{225}$	4.92E+00	1.03E+01	7.23E+00	1.22E+00
POD-DEIM$_{250}$	3.74E+00	8.30E+00	5.70E+00	9.74E−01
	NRMSE			
	Min.	Max.	Mean	Std. dev.
POD-DEIM$_{200}$	1.11E−06	4.27E−06	2.24E−06	6.24E−07
POD-DEIM$_{225}$	1.52E−06	3.17E−06	2.23E−06	3.76E−07
POD-DEIM$_{250}$	1.15E−06	2.56E−06	1.76E−06	3.00E−07

just a few seconds in one model run can translate to a total time savings on the order of hours or days.

Hyper-reduction with POD-DEIM is an improvement over POD when it comes to versatility, efficiency and robustness. A wider range of groundwater flow processes can now be modelled in a reduced space with POD-DEIM. With only one POD projection, errors might grow too large when propagated over a long time horizon of a complex nonlinear model. To build a robust reduced model with traditional POD alone, a very large set of snapshots might be necessary to adequately approximate the nonlinear system. This often leads to only moderate dimension reduction and can increase computational cost as shown in Table 4. If reduced models are going to be used in practice, certainly online run times must be shorter than the full model. However, to decrease the overall online runtime of a full model, the online plus the offline computational cost of a reduced or hyper-reduced model ought to be considered. When snapshot selection and reduced basis construction become challenging for large-scale models, the overhead investment in the reduced model could be substantial. Therefore, POD-DEIM method allows the dimension of the online calculations to be as small as possible while still maintaining acceptable error.

Timing comparisons illustrate the practical result of dimension reduction within the context of this specific synthetic experiment. It is the reduction in dimension that has more implications for other applications. The POD-DEIM model takes about half as much time within the solver as the full model. The specific result may seem insignificant on this relatively small 2D application but when systems approach millions of nodes and decadal time horizons, any speed and dimensionality improvement will be highly beneficial. The time comparison between the full and reduced model is not entirely meaningful since different solvers are used. The PCG used for the full model is optimized for the type of problem it sees whereas the reduced model solver is more generic. Regardless of this fact, the time comparison between the reduced models is valid since they utilize the same solver. If a more efficient solver were to be used by the reduced models, the smaller dimensional complexity of the POD-DEIM model will almost always be faster to solve than the POD model.

Contrary to what one might expect, fewer interpolation points does not always lead to a faster reduced model, even though the computational complexity is less. An insufficient number of interpolation points induces more error and requires more internal solver iterations for convergence. In practice, keeping the volumetric water budget percent error below a nominal value set by MODFLOW has been a good indication of a robust interpolation.

While construction of the POD-DEIM reduced model can still be time intensive, the possible applications to a variety of problems can exploit the superior speed. For example, multi-objective optimizations or large computer experiments involving Monte Carlo can require thousands or millions of model calls. For many developed groundwater models, run times for a single simulation can approach days. Consequently, some types of optimizations or uncertainty analyses would be infeasible without model reduction. Any opportunity to reduce the dimensionality of model's calculations may permit what was previously infeasible. The methodology proposed could specifically be used within parameter estimation techniques used to estimate unknown pumping rates to match drawdown observations—a problem encountered during a calibration process when wells that lack pumping data are modeled within a groundwater basin. It may also be necessary to quantify the uncertainty of the pumping rate estimates to supplement any model predictions. If the proposed POD-DEIM methodology is used within an uncertainty analysis—such as the null-space Monte Carlo used by Siade et al. (2015)—further time savings could be

achieved. For more versatile applications to parameter estimation, the following steps can be taken: (1) take snapshots to capture a range of any parameter; (2) add them to the snapshot set; and (3) re-compute the basis.

There are some aspects that still need to be explored with POD-DEIM models as the size of the application scales up. For models that would not permit taking snapshots at every time step, methods for optimal snapshot selection ought to be considered. The choice of snapshots could be the difference between a reduced model that converges in the solver and one that does not. In cases where groundwater flow models are coupled to transport, rainfall-runoff, and/or atmospheric models, the reduced model's solution serves as input to these other models. The oscillatory nature of the reduced model's errors would have an unknown effect on models to which it may be coupled. Further research could also explore the trade-off in the size of dimensions r and d; the appropriateness of the percent energy criteria used to select the reduced dimensions; and applications of more efficient solvers for the reduced models.

5. Conclusion

A traditional model reduction 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). While more and more optimization algorithms are taking advantage of parallel computing power, 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.

Acknowledgments

This material is based on work supported by NSF under Award EAR-1314422. Partial support also was provided by an AECOM endowment. The in-depth reviews and constructive comments provided by two anonymous reviewers are greatly appreciated.

References

Abgrall, R., Amsallem, D., 2016. Robust model reduction by L1-norm minimization and approximation via dictionaries: application to linear and nonlinear hyperbolic problems. Adv. Model. and Simul. in Eng. Sci. 3 (1), 1–6. http://dx.doi.org/10.1186/s40323-015-0055-3.

Antoulas, A., Sorensen, D., Gugercin, S., 2001. A survey of model reduction methods for large-scale systems. Contemp. Math. 280, 193–220. http://dx.doi.org/10.1080/00207170410001713448.

Barrault, M., Maday, Y., Nguyen, N.C., Patera, A.T., 2004. An "empirical interpolation" method: application to efficient reduced-basis discretization of partial differential equations. C. R. Math. 339, 667–672. http://dx.doi.org/10.1016/j.crma.2004.08.006.

Boyce, S.E., Nishikawa, T., Yeh, W.W.-G., 2015. Reduced order modeling of the Newton formulation of MODFLOW to solve unconfined groundwater flow. Adv. Water Resour. 83, 250–262. http://dx.doi.org/10.1016/j.advwatres.2015.06.005.

Boyce, S.E., Yeh, W.W.-G., 2014. Parameter-independent model reduction of transient groundwater flow models: application to inverse problems. Adv. Water Resour. 69, 168–180. http://dx.doi.org/10.1016/j.advwatres.2014.04.009.

Cardoso, M.A., Durlofsky, L.J., Sarma, P., 2009. Development and application of reduced-order modeling procedures for subsurface flow simulation. Int .J. Numer. Methods Eng. 77, 1322–1350. http://dx.doi.org/10.1002/nme.2453.

Chaturantabut, S., Sorensen, D.C., 2010. Nonlinear model reduction via discrete empirical interpolation. SIAM J. Sci. Comput. 32, 2737–2764.

Drohmann, M., Haasdonk, B., Ohlberger, M., 2012. Reduced basis approximation for nonlinear parametrized evolution equations based on empirical operator interpolation. SIAM J. Sci. Comput. 34, A937–A969. http://dx.doi.org/10.1137/10081157x.

Forstall, V.H., 2015. Iterative solution methods for reduced-order models of parameterized partial differential equations, Doctoral dissertation. University of Maryland, College Park.

Graham, M., Kevrekidis, I., 1996. Alternative approaches to the Karhunen–Loeve decomposition for model reduction and data analysis. Comput. Chem. Eng. 20, 495–506. http://dx.doi.org/10.1016/0098-1354(95)00040-2.

Hanson R.T., Boyce S.E., Schmidt W., Hughes J.D., Mehl S.M., Leake S.A., et al., 2014. One-water hydrologic flow model (MODFLOW-OWHM), U.S. Geological Survey Techniques and Methods 6–A51, 120 p., http://dx.doi.org/10.3133/tm6A51.

Henneron, T., Clenet, S., 2014. Model order reduction of non-linear magnetostatic problems based on POD and DEI methods. IEEE Trans. Magn. 50, 33–36. http://dx.doi.org/10.1109/TMAG.2013.2283141.

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

Kasprzyk, J.R., Reed, P.M., Kirsch, B.R., Characklis, G.W., 2009. Managing population and drought risks using many-objective water portfolio planning under uncertainty. Water Resour Res 45 (12), W12401. http://dx.doi.org/10.1029/2009WR008121.

Keating, E., Zyvoloski, G., 2009. A stable and efficient numerical algorithm for unconfined aquifer analysis. Ground Water 47, 569–579. http://dx.doi.org/10.1111/j.1745-6584.2009.00555.x.

McPhee, J., Yeh, W.W.-G, 2008. Groundwater management using model reduction via empirical orthogonal functions. J. Water Resour. Plan. Manag. 134, 161–170. http://dx.doi.org/10.1061/(ASCE)0733-9496(2008)134:2(161).

Nigro, P.S.B., Anndif, M., Teixeira, Y., Pimenta, P.M., Wriggers, P., 2015. An adaptive model order reduction with quasi-Newton method for nonlinear dynamical problems. Int. J. Numer. Methods Eng. 106 (9), 740–759. http://dx.doi.org/10.1002/nme.5145.

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

Quarteroni, A., Rozza, G., 2007. Numerical solution of parametrized Navier–Stokes equations by reduced basis methods. Numer. Methods Partial Differ. Equ. 23, 923–948. http://dx.doi.org/10.1002/num.20249.

Radermacher, A., Reese, S., 2015. POD-based model reduction with empirical interpolation applied to nonlinear elasticity. Int. J. Numer. Methods Eng. 107 (6), 477–495. http://dx.doi.org/10.1002/nme.5177.

Reed, P.M., Hadka, D., Herman, J.D., Kasprzyk, J.R., Kollat, J.B., 2013. Evolutionary multiobjective optimization in water resources: the past, present, and future. Adv. Water Resour. 51, 438–456. http://dx.doi.org/10.1016/j.advwatres.2012.01.005.

Reyment, R.A., Joreskog, K.G., 1993. Applied Factor Analysis in the Natural Sciences. Cambridge University Press.

Siade, A., Nishikawa, T., Martin, P., 2015. Natural recharge estimation and uncertainty analysis of an adjudicated groundwater basin using a regional-scale flow and subsidence model (Antelope Valley, California, USA). Hydrogeol. J. 23, 1267–1291. http://dx.doi.org/10.1007/s10040-015-1281-y.

Siade, A.J., Putti, M., Yeh, W.W.-G., 2012. Reduced order parameter estimation using quasilinearization and quadratic programming. Water Resour. Res. 48 (6), W06502. http://dx.doi.org/10.1029/2011WR011471.

Sirovich, L., 1987. Turbulence and the dynamics of coherent structures. Part I: coherent structures. Q. Appl. Math. 45 (3), 561–571.

Ştefănescu, R., Navon, I.M., 2013. POD/DEIM nonlinear model order reduction of an ADI implicit shallow water equations model. J. Comput. Phys. 237, 95–114. http://dx.doi.org/10.1016/j.jcp.2012.11.035.

Ştefănescu, R., Sandu, A., Navon, I.M., 2015. POD/DEIM reduced-order strategies for efficient four dimensional variational data assimilation. J. Comput. Phys. 295, 569–595. http://dx.doi.org/10.1016/j.jcp.2015.04.030.

Tonkin, M., Doherty, J., 2009. Calibration-constrained Monte Carlo analysis of highly parameterized models using subspace techniques. Water Resour. Res. 45, 1–17. http://dx.doi.org/10.1029/2007WR006678.

Vermeulen, P.T.M., Heemink, A.W., Te Stroet, C.B.M., 2004. Reduced models for linear groundwater flow models using empirical orthogonal functions. Adv. Water Resour. 27, 57–69. http://dx.doi.org/10.1016/j.advwatres.2003.09.008.

Vermeulen, P.T.M., Heemink, A.W., Te Stroet, C.B.M., 2004. Low-dimensional modelling of numerical groundwater flow. Hydrol. Process. 18, 1487–1504. http://dx.doi.org/10.1002/hyp.1424.

von Storch, H., Hannoschöck, G., 1985. Statistical aspects of estimated principal vectors (EOFs) based on small sample sizes. J. Clim. Appl. Meteorol. 24, 716–724.

Willis, R., Yeh, W.W.-G., 1987. Groundwater Systems Planning and Management. Prentice Hall Inc., Old Tappan, NJ.