Snapshot selection for groundwater model reduction using proper orthogonal decomposition

Adam J. Siade,[1] Mario Putti,[2] and William W.-G. Yeh[1]

Received 19 October 2009; revised 15 February 2010; accepted 5 April 2010; published 24 August 2010.

[1] Water resources systems management often requires complex mathematical models whose use may be computationally infeasible for many advanced analyses, e.g., optimization, data assimilation, model uncertainty, etc. The computational demand of these analyses can be reduced by approximating the model with a simpler reduced model. Proper Orthogonal Decomposition (POD) is an efficient model reduction technique based on the projection of the original model onto a subspace generated by full-model snapshots. In order to implement this method, an appropriate number of snapshots of the full model must be taken at the appropriate times such that the resulting reduced model is as accurate as possible. Since confined aquifers reach steady state in an exponential manner, we present a simple exponential function that can be used to select snapshots for these types of models. This selection method is then employed to determine the optimal snapshot set for a unit, dimensionless model. The optimal snapshot set is found by maximizing the minimum eigenvalue of the snapshot covariance matrix, a criterion similar to those used in experimental design. The resulting snapshot set can then be translated to any complex, real world model based on a simple, approximate relationship between dimensionless and real-world times. This translation is illustrated using a basin scale model of Central Veneto, Italy. We show that this method is very easy to implement and produces an accurate reduced model that, in the case of Central Veneto, Italy, runs approximately 1,000 times faster than the full model.

Citation: Siade, A. J., M. Putti, and W. W.-G. Yeh (2010), Snapshot selection for groundwater model reduction using proper orthogonal decomposition, *Water Resour. Res.*, *46*, W08539, doi:10.1029/2009WR008792.

1. Introduction

[2] Groundwater management requires the development and implementation of mathematical models that, through simulation, evaluate the anthropogenic impacts on an aquifer system. These models must exhibit a significant degree of accuracy in order to provide reliable results that can be used for prediction and management purposes. The accuracy and reliability of a model is predicated on many factors including the quality and quantity of observed data, the complexity of the model parameterization, and correct identification of model parameters and initial and boundary conditions. There are many well developed methods for dealing with these factors; however, computing power limits some of them. For example, a groundwater model with a coarse spatial discretization will yield inaccurate results, yet a model with too fine of a discretization will require so much computer time that its use may be impractical.

[3] Highly heterogeneous groundwater models require a degree of model complexity that corresponds to that of the subsurface lithology. In other words, aquifers whose properties change significantly within a certain distance are best represented using a model with a discretization on that same order of magnitude. However, matching the model discretization with the complexity of the aquifer can be impractical due to computational demand. Therefore, there exists a trade-off between accuracy and model computation time. Even though current computing technology affords the use of complex models, there are many cases where forward simulation is still inefficient. Even though a single forward simulation of a groundwater model may seem effective, the computational expense resides in the need to call the model numerous times depending on the analysis under consideration e.g., parameter identification, management/optimization, data assimilation, and model uncertainty analysis [*Hill and Tiedeman*, 2007].

[4] There exist a multitude of techniques designed to alleviate the computational demand required of a numerical model. These techniques are known as model reduction techniques where the objective is to develop a reduced model with a much smaller dimension for fast execution. The reduced model is of course only an approximation of the full model and the introduced model error must be closely examined. Model reduction can be classified into three major categories: data driven, model driven and combined methods. Data driven methods assume a black box model where the objective is only to match inputs with their respective outputs. Model driven methods make use of the mathematical structure of the model itself when developing the reduced model. Combined methods make use of input/output data relationships as well as incorporate the physics of the model.

[1]Department of Civil and Environmental Engineering, University of California, Los Angeles, California, USA.
[2]Department of Mathematical Methods and Models for Scientific Applications, University of Padua, Padua, Italy.

0043-1397/10/2009WR008792

[5] Artificial neural networks (ANNs) are an example of data driven methods. In this approach, the response of a full model is simulated explicitly by an ANN model based on biological neural networks, which is much easier and faster to execute. A number of full model simulations are conducted to generate an input and output data set, which is then used to "train" a black box model. *Rogers and Dowla* [1994] used an ANN model to replace a groundwater flow and contaminant transport model in a management scenario in which the objective was to prevent migration of a contaminant plume. However, in order to train the ANN model, numerous calls to the full model must be evaluated to provide enough data for an accurate surrogate model. *Yan and Minsker* [2006] developed a methodology for reducing the number of full-model calls using a dynamic learning technique. A technique similar to ANN modeling is a method proposed by *Baú and Mayer* [2006] in which the objective functions are approximated using kriging interpolation, resulting in surrogate functions which reduce the computational burden of the management problem presented in their work.

[6] Inverse eigenvalue methods are an example of model driven methods designed to exploit the mathematical structure of the full model. The full numerical model often involves large, general matrices. The inverse eigenvalue approach projects the model onto a subspace by finding smaller matrices that share the same nonzero/significant eigenvalue spectrum as their corresponding full-model matrices. We refer to *Chu* [1998] for a complete introduction to inverse eigenvalue problems. *Orbak et al*. [2004] evaluated the sensitivities of the system parameters with respect to eigenvalues for some simple example problems. "Effect" matrices were then developed allowing for physical model reduction via inverse eigenvalue methods.

[7] The third category of model reduction combines concepts from both the data and model driven methods. The distribution of the state variables is sampled from the full model for various values of the decision variables and time. These samples are called snapshots of the full model and are used in an interpolation scheme to approximate the full model. The act of interpolating reduces the dimensionality of the problem allowing for model reduction. This is equivalent to considering only a subspace of the vector space where the full-model solution resides. This subspace is evaluated so that it captures the majority of the variation in the model solution. The entire model is then projected onto this subspace and solved, resulting in a fast approximate solution to the original model. Depending on the application, the method in which the spatial patterns (basis functions that define the reduced model space) are developed has been called Proper Orthogonal Decomposition (POD) [*Cazemier et al*., 1998; *Willcox and Peraire*, 2002], Principal Component Analysis (PCA) or the discrete Karhunen Loève Transform [*Loève*, 1978; *Zhang and Michaelis*, 2003]. The spatial basis functions have been called Principal Vectors, Empirical Orthogonal Functions (EOFs) [*Vermeulen et al*., 2004; *McPhee and Yeh*, 2008], or Coherent Structures (CS) [*Mijalkovic*, 2004].

[8] The focus of this study is on the efficient development of combined model reduction methods for groundwater flow models. *Vermeulen et al*. [2004] and *McPhee and Yeh* [2008] developed reduced groundwater models by sampling hydraulic head distributions for some constant, reference pumping rate at specified time intervals. These snapshots were then used to determine the EOFs of the system. The evaluation of these snapshots, a phase shared by all combined model reduction methods, requires one full-model run per stress location (pumping/injection well) from early times to steady state. However, there is no established method for determining a best snapshot selection scheme, i.e., at which time steps the full-model solution is to be recorded and used to develop the reduced model. The optimal selection strategy would yield the smallest set of full-model state variable snapshots such that the resulting reduced-model solution achieves a predefined accuracy. However, in order to optimally select snapshots, the dynamics of the system must be well understood [*Vermeulen et al.*, 2004; *McPhee and Yeh*, 2008]. Once a good snapshot set is obtained, the model can be reduced via POD and time-varying extraction rates can be applied to the reduced model, i.e., the superposition principle applies due to the linearity of the governing equations.

[9] Snapshot selection for POD model reduction has received little attention and requires further research. *Kowalski and Jin* [2003] performed POD model reduction for Maxwell's equations, used to model the performance of a medical device in the human body. They considered three snapshot selection techniques: uniform in time, logarithmic with a focus on early time steps and logarithmic with a focus on later time steps. The early-time logarithmic selection scheme produced the best results. They concluded that the optimal snapshot selection scheme depends on the mathematical model under investigation and the parametric structure of that model. The primary objective of our study is to conduct POD model reduction for confined groundwater aquifers and to develop a strategy for snapshot selection based on both the governing equation and the parametric structure of the model.

2. Confined Aquifer Groundwater Flow Model

[10] The following partial differential equation (PDE) describes three-dimensional groundwater flow for a confined, anisotropic aquifer with pumping [*Bear*, 1979],

$$\frac{\partial}{\partial x}\left(K_x\frac{\partial h}{\partial x}\right)+\frac{\partial}{\partial y}\left(K_y\frac{\partial h}{\partial y}\right)+\frac{\partial}{\partial z}\left(K_z\frac{\partial h}{\partial z}\right)-q-S_s\frac{\partial h}{\partial t}=0 \quad (1)$$

with initial and boundary conditions,

$$h(x,y,z,0)=f_1(x,y,z)$$

$$h(x,y,z,t)=f_2(x,y,z,t),\quad (x,y,z,t)\in(\Gamma_1)$$

$$q_n(x,y,z,t)=f_3(x,y,z,t),\quad (x,y,z,t)\in(\Gamma_2)$$

where h is the hydraulic head (L), K_x, K_y, K_z are hydraulic conductivities in the x-, y-, and z-directions, respectively (L/T), S_s is the specific storage (L^{-1}), q is the specific volumetric pumping rate (T^{-1}), q_n is the specific discharge normal to the flux boundary Γ_2 (T^{-1}), Γ_1 is the fixed head boundary, f_1, f_2 and f_3 are known functions.

[11] Application of the superposition principle followed by spatial discretization of the resulting PDE (e.g., by finite differences, finite elements, etc.) yields a system of linear ordinary differential equations (ODEs) for the drawdown:

$$\mathbf{B}\frac{d\mathbf{s}}{dt}+\mathbf{A}\mathbf{s}=\mathbf{q} \quad (2)$$

where $\mathbf{A}$ is the $n \times n$ stiffness matrix, $\mathbf{B}$ is the $n \times n$ mass matrix, $\mathbf{s}$ and $\mathbf{q}$ are the n-dimensional vectors of nodal (cell) drawdowns and sinks, respectively, and n is the number of spatial computational nodes (cells), which is generally very large. Drawdown is the difference between the initial head and the head after pumping, i.e., $s = H - h$, where H is the initial head (e.g. steady state or the natural system dynamics). In the majority of cases of practical interest, matrices $\mathbf{A}$ and $\mathbf{B}$ are large, sparse, symmetric and positive definite.

3. Model Reduction via Proper Orthogonal Decomposition

[12] In this section we provide a summary of the method of POD model reduction. This summary is a culmination of many previous works, including those by *Cazemier et al.* [1998], *Willcox and Peraire* [2002], *Kowalski and Jin* [2003], *Vermeulen et al.* [2004], and *McPhee and Yeh* [2008].

3.1. Subspace Projection

[13] Model reduction via POD is achieved by projecting the state vector into an n_p-dimensional subspace ($n_p \ll n$) such that $\hat{\mathbf{s}} = \mathbf{P}\mathbf{s}_r$, where $\hat{\mathbf{s}} \in \mathbb{R}^n$ is an approximation of the state vector after subspace projection, $\mathbf{P}$ is a projection operator represented by a matrix whose columns form an orthonormal basis spanning the subspace $\mathrm{V} \subset \mathbb{R}^n$, and $\mathbf{s}_r \in \mathbb{R}^{n_p}$ is the state vector of the reduced model. This can also be written as follows [*Vermeulen et al.*, 2004]:

$$\hat{\mathbf{s}}(t) = \mathbf{s}_{nat}(t) + \sum_{k=1}^{n_p} \mathbf{p}_k s_{r,k}(t) \tag{3}$$

where, $\mathbf{s}_{nat}(t)$ is the natural system dynamics without forcing (e.g. the steady state solution) and $s_{r,k}$ indicates the k-th component of $\mathbf{s}_r$. Note here that generally, $\mathbf{s}_{nat}(t) = \mathbf{0}$.

[14] The definition of the n_p basis vectors that span the subspace V proceeds by selecting from a full-model model run n_s numerical solution vectors (snapshots) at predefined times. This must be done for each well individually with constant pumping rates. These snapshots (here identified by the n-dimensional vectors $\mathbf{s}_i$, $i = 1,\ldots,n_s$) must be significantly different and cover the overall range or variability of the full model. Note that $n_s \geq n_p$ where n_s is often much greater than n_p. The basis vectors for V are then obtained via PCA as follows. Once the snapshots have been taken, the so called covariance matrix [*Sirovich*, 1987] can be calculated as $\mathbf{C} = \mathbf{X}\mathbf{X}^T$, where $\mathbf{X} = \{\bar{\mathbf{s}}_1, \ldots, \bar{\mathbf{s}}_{n_s}\}$ and $\bar{\mathbf{s}}_i$ is the i-th normalized snapshot [*McPhee and Yeh*, 2008; *Vermeulen et al.*, 2004]. From PCA, the eigenvectors (or principal vectors) $\mathbf{p}_k$ of $\mathbf{C}$ are linearly independent and mutually orthogonal. These principal vectors can be considered as candidates for the spatial basis functions that are used to express the problem solution in the reduced space.

[15] In practice, the length of each snapshot vector is large, as it is equal to the number of nodes in the model; this will cause the eigenvalue decomposition of the resulting $\mathbf{C}$ matrix to be computationally infeasible. This is remedied by calculating the eigenpairs of $\mathbf{C}_s = \mathbf{X}^T\mathbf{X}$, that are related to the eigenpairs of $\mathbf{C}$. It is easy to prove, in fact, that if $(\lambda_k, \mathbf{g}_k)$ is the k-th eigenpair of $\mathbf{C}_s$, then $\lambda_k, \mathbf{u}_k = \mathbf{X}\mathbf{g}_k$ characterizes the eigenpair of $\mathbf{C}$ corresponding to the k-th nonzero eigenvalue.

[16] Once the eigenvalues of $\mathbf{C}$ are calculated, insignificant principal vectors can be removed from the basis for further subspace dimension reduction. This is accomplished by normalizing the eigenvalues to represent relative magnitudes as follows:

$$\phi_j = \frac{\lambda_j}{\sum_{i=1}^{n_s} \lambda_i} \tag{4}$$

The largest n_p normalized eigenvalues and their corresponding eigenvectors are retained for use in the reduced model and the rest are discarded. The parameter n_p is chosen such that $\sum_{i=1}^{n_p} \phi_i \geq \Phi$ where Φ is user-specified [*McPhee and Yeh*, 2008; *Vermeulen et al.*, 2004]. Finally, imposing $\mathbf{P}^T\mathbf{P} = \mathbf{I}$, the matrix of the basis vectors for V is calculated as

$$\mathbf{P} = \mathbf{X}\mathbf{G}\Lambda^{-1/2} \tag{5}$$

where G and Λ are the rectangular and diagonal matrices that consist of the n_p retained eigenvectors, $\mathbf{g}_k$, and the corresponding eigenvalues, λ_k, respectively.

[17] In other words, the elements of the reduced state vector reflect the weight or importance of each spatial basis function or principal vector at each time step. Some of these principal vectors are insignificant for all times and can be excluded from the problem entirely, achieving further model reduction. The "importance" of each principal vector depends on the relative amount of variability the corresponding principal component captures within the snapshot dataset. From PCA, the relative variance of each principal component is equivalent to the eigenvalue associated with its principal vector (eigenvector). Therefore, one can control the degree of accuracy by keeping only those eigenvalues whose sum encompasses a desired amount of the variability within the snapshot dataset.

[18] In summary, the determination of the matrix $\mathbf{P}$ and hence the subspace onto which the model is to be projected is as follows.

[19] 1. Perform a full model run for each well with a constant unit flow rate. At the appropriate time steps record the solution (i.e., drawdown distribution). These solutions are the snapshots. Note that each run should only have one active well.

[20] 2. Normalize and group *all* the snapshots into the matrix, $\mathbf{X}$. Conduct a spectral decomposition of $\mathbf{C}_s = \mathbf{X}^T\mathbf{X}$ and use equation (5) to obtain the matrix of principal vectors.

[21] 3. Omit any insignificant principal vectors to from the basis, $\mathbf{P}$.

3.2. Galerkin Projection

[22] Once the projection matrix, $\mathbf{P}$, has been determined, the full model must be projected onto the subspace V. Let $L(\mathbf{s}) \in \mathbb{R}^n$ be the residual from equation (2):

$$L(\mathbf{s}) = \mathbf{B}\frac{d\mathbf{s}}{dt} + \mathbf{A}\mathbf{s} - \mathbf{q}$$

The problem of finding the solution of the full model (i.e., solving $L(\mathbf{s}) = \mathbf{0}$) can be equivalently written as finding the vector $\mathbf{s} \in \mathbb{R}^n$ such that $L(\mathbf{s})^T\mathbf{z} = 0$, $\forall\ \mathbf{z} \in \mathbb{R}^n$. The Galerkin projection step of POD is now easily defined as

find the vector $\hat{\mathbf{s}}$ *such that*:

$$L(\hat{\mathbf{s}})^T\mathbf{v} = 0, \forall\ \mathbf{v} \in V$$

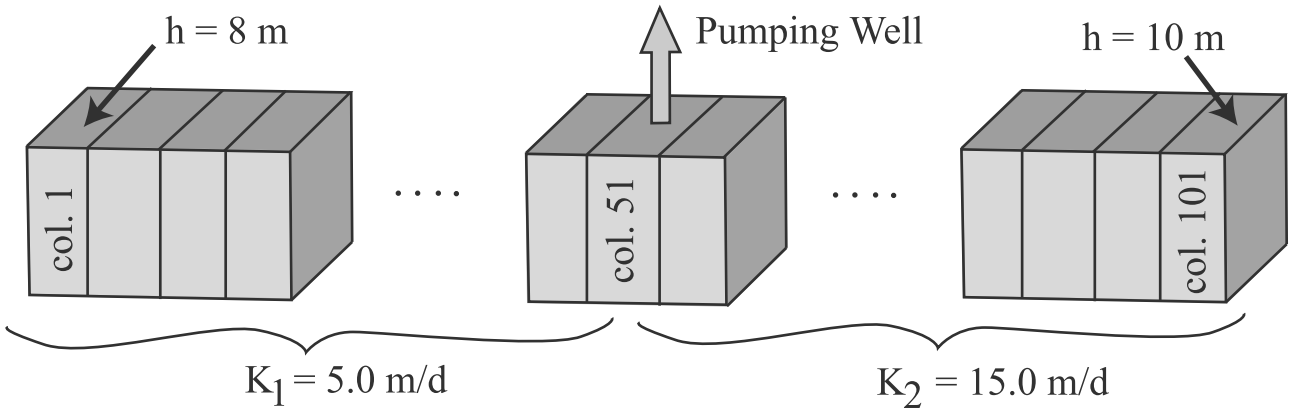

Figure 1. One-dimensional groundwater flow model.

[23] Since any vector, $\mathbf{v} \in V$, can be written as some linear combination of the vectors in the basis, $\mathbf{P}$, the Galerkin projection is simply: $\mathbf{P}^T L(\hat{\mathbf{s}}) = \mathbf{0}$. Applying this projection to equation (2) results in the following,

$$\mathbf{P}^T\mathbf{B}\frac{d\hat{\mathbf{s}}}{dt} + \mathbf{P}^T\mathbf{A}\hat{\mathbf{s}} = \mathbf{P}^T\mathbf{q} \tag{6}$$

Substituting the interpolation shown in equation (3), equation (6) becomes a system of n_p, linearly independent, ODEs and n_p unknowns (assuming the natural system dynamics have been removed):

$$\mathbf{P}^T\mathbf{B}\mathbf{P}\frac{d\mathbf{s}_r}{dt} + \mathbf{P}^T\mathbf{A}\mathbf{P}\mathbf{s}_r = \mathbf{P}^T\mathbf{q} \tag{7}$$

Letting, $\tilde{\mathbf{B}} = \mathbf{P}^T\mathbf{B}\mathbf{P}$, $\tilde{\mathbf{A}} = \mathbf{P}^T\mathbf{A}\mathbf{P}$ and $\tilde{\mathbf{q}} = \mathbf{P}^T\mathbf{q}$, we obtain

$$\tilde{\mathbf{B}}\frac{d\mathbf{s}_r}{dt} + \tilde{\mathbf{A}}\mathbf{s}_r = \tilde{\mathbf{q}} \tag{8}$$

This reduced system of ODEs can be solved by any stable time stepping technique, such as Implicit Euler or Crank-Nicolson methods. However, because of its drastically reduced size ($n_p \ll n$), often by several orders of magnitude, the system can also be solved very efficiently by analytical methods via matrix exponential.

3.3. One-Dimensional Test Case

[24] A synthetic experimental setup, illustrated in Figure 1, from *McPhee and Yeh* [2008] was used to exemplify the use of POD. The specific storage is 1.0 m^{-1} and the width of each column is 1.0 m. There are Dirichlet boundary conditions at columns 1 and 101. This test case and all other experiments in this study were modeled by separating the natural system dynamics from the drawdown due to pumping using the superposition principle [*Bear*, 1979]. Model reduction is applied to the groundwater model of drawdown. Therefore, the initial conditions are 0 m everywhere and the Dirichlet boundary conditions at columns 1 and 101 are 0 m. The extraction rate applied to column 51 is 1 m^3 / d. The reduced model was developed using 10 snapshots taken at 0.6, 1.5, 2.9, 5.1, 8.5, 13.9, 22.4, 35.6, 56.4 and 89.1 days ($n_s = 10$).

[25] Figure 2a displays the drawdown results of both the full and reduced models of the one-dimensional test case. The results of the reduced model are indistinguishable with

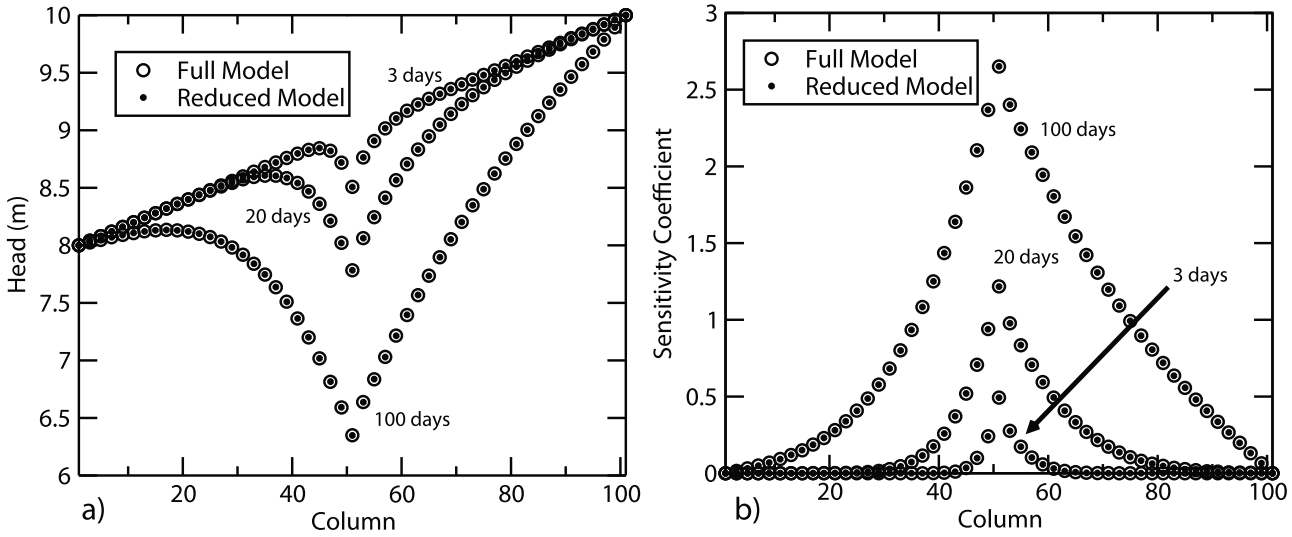

Figure 2. (a) Drawdown results of both the full and reduced models after 3, 20 and 100 days for the one-dimensional test case. (b) Sensitivity coefficient results of both the full and reduced models after 3, 20 and 100 days for the one-dimensional test case.

that of the full model. Note that the Dirichlet boundary conditions are met naturally when drawdown is used as the dependent variable. *McPhee and Yeh* [2008] applied POD model reduction to head and reported difficulty in satisfying the boundary conditions.

[26] The sensitivities of the heads with respect to pumping (i.e., the impact changes in pumping will have on drawdown) were also calculated using a simple finite difference approach. Figure 2b illustrates these sensitivities for both the full and reduced models. Again, the results of the reduced model are indistinguishable with those of the full model. Note that the time steps for which the results have been displayed do not coincide with those of the snapshots used to develop the reduced model. In this example, the full model consists of 101 equations and the reduced model consists of 10 equations, creating a reduction in model dimension of one order of magnitude.

4. Snapshot Selection

4.1. Overall Strategy

[27] The ability of a reduced model, obtained from POD, to accurately represent and, in practice, replace the full model is based solely on the manner in which the full model snapshots are obtained. Both the number of snapshots, as well as the time steps in which they are taken, affect the accuracy of the reduced solution. However, a large number of snapshots will not necessarily result in a high level of accuracy. Therefore, the goal is to find the best sampling strategy for a given snapshot size such that the accuracy of the reduced model is maximized.

[28] There are many ways to approach this problem. In this paper we define the "best" sampling times for snapshots by requiring that all snapshots be significant in the amount of information that they possess. In other words, after conducting PCA, the number of normalized eigenvalues significantly different from zero should be maximized in order to maintain the desired level of accuracy within the reduced model (provided the snapshot size, n_s, is large enough). The number of snapshots is optimal when the addition of another snapshot does not add a significant amount of information to the reduced model. For example, the addition of one snapshot which, after PCA, results in the same number of significant principal vectors as before the snapshot was added is indicative that increasing the snapshot size (i.e., the number of snapshots) was unnecessary. This can be thought of as adding a snapshot which is approximately a linear combination of the other snapshots; thus, providing little additional information. Of course, when the snapshot size is increased, the optimal times in which all of the snapshots are taken must be reevaluated.

4.2. Exponentially Distributed Snapshots

[29] In order to determine the best snapshot times, for any POD model reduction, the dynamic properties of the governing equation must be taken into consideration. In the case of saturated groundwater flow, the governing equation is a parabolic PDE, indicating the system is diffusive in nature. These types of transient models tend to exhibit rapid changes in the state variable for early times and small changes for later times. In other words, these types of systems reach steady state at an exponential rate [*John*, 1982]. Therefore, to fully capture the dynamics of the system, one must acquire many snapshots early in time and fewer snapshots later in time. For example, if many snapshots were taken near steady state, they would be approximately equal and the corresponding covariances would be small, resulting in small eigenvalues and insignificant principal vectors. If few snapshots were taken at early time steps where the shape of the state variable distribution is changing rapidly, interpolations between these snapshots (i.e., the reduced model) could be inaccurate.

[30] To study these properties we look at the analytical solution of equation (2), which can be written as

$$\mathbf{s}(t) = e^{-\mathbf{B}^{-1}\mathbf{A}t}\left[\mathbf{s}_0 - \mathbf{A}^{-1}\mathbf{q}\right] + \mathbf{A}^{-1}\mathbf{q} \tag{9}$$

The above matrix exponential can be evaluated by means of the solution to the generalized eigenproblem $\mathbf{Au} = \lambda\mathbf{Bu}$, where matrices $\mathbf{A}$ and $\mathbf{B}$ are symmetric and positive definite. The drawdown vector is then given as

$$\mathbf{s}(t) = \mathbf{U}e^{-Mt}\mathbf{U}^T\left[\mathbf{s}_0 - \mathbf{A}^{-1}\mathbf{q}\right] + \mathbf{A}^{-1}\mathbf{q} \tag{10}$$

where matrix $\mathbf{U}$ is the unitary matrix of the generalized eigenvectors and M is a diagonal matrix containing the generalized eigenvalues $\mu_i > 0$. Thus it is clear that only the smaller eigenvalues and corresponding eigenvectors are necessary to capture the transient behavior of the solution. The question is then how can we select the times of the snapshots so that the information contained in this set of vectors is maximized. We illustrate our approach by working on a simple scalar ODE, whose solution can be directly written as

$$y(t) = ce^{-\alpha t} + b \tag{11}$$

[31] Introducing the scaled time $t^* = e^{-\alpha t}$ the behavior $y(t^*)$ becomes linear. Thus, we can subdivide the interval of the scaled time uniformly without loss of accuracy. Hence we define the times of snapshots with the following function:

$$t_s = \beta e^{-\alpha t_u} + \gamma \tag{12}$$

where α, β and γ are parameters to be determined on the basis of the equation properties (i.e. hydraulic conductivity, elastic storage, etc.), while t_u varies uniformly within an appropriate interval of length T such that $t_u = uh$, $u = 0,\ldots,n_s - 1$ where $h = T/(n_s - 1)$. For example, we can choose $T = 10$ and $n_s = 5$ such that $t_u = 0, 2.5, 5, 7.5, 10$. We observe that the snapshot time coincides with $\beta + \gamma$ for $u = 0$, and with γ for large t_u, i.e., "pseudo" steady state conditions. Thus, given a fixed number of snapshots and values for α, β and γ, we can determine a snapshot set with an exponential distribution in time.

[32] We define the solution error of the reduced model at a particular time step t_k as the Euclidian norm of the error between the reduced-model and the full-model solutions:

$$\tau_r(t_k) = \sqrt{\sum_{i=1}^{n}\left(\mathbf{s}_{r,i}(t_k) - \mathbf{s}_{f,i}(t_k)\right)^2} \tag{13}$$

Using our snapshot selection strategy, the solution error will depend only on the number n_s of snapshots and the values

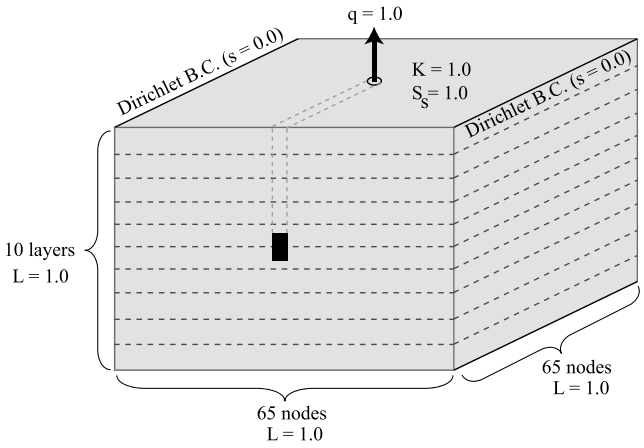

Figure 3. Experimental setup for the three-dimensional dimensionless model.

of the parameters α, β and γ. Thus the "optimal" reduced model can be identified by the following minimization problem:

$$\min_{n_s,\alpha,\beta,\gamma}\left[\text{ARMSE} = \frac{1}{T}\int_0^T \tau_r(t)dt\right] \tag{14}$$

where ARMSE is the Average Root Mean Square Error and T is the ending time of the simulation. We evaluate the time integral in the formula for ARMSE by means of a 3-point Gaussian quadrature rule. Given that τ_r (t_k) varies smoothly with time, this Gaussian quadrature rule is sufficiently accurate to ensure that its error is always smaller than the error we are evaluating. It is obvious that the optimal n_s, α, β and γ depend upon the hydrological parameters, among other factors, and must be determined case by case. Extending these results to the general case, we need to relax absolute optimality and look for a "sub-optimal" but still accurate reduced-model solution. To accomplish this, we apply our selection procedure to a dimensionless model and translate this result to any real model using an approximation or averaging technique to account for heterogeneities and complex geometries within the real model. This technique is described in subsequent sections.

4.3. Optimal Snapshot Set

4.3.1. Dimensionless Model

[33] Equation (12) was applied to the one-dimensional test case in Figure 1. The preliminary results of this experiment indicated that the best snapshot sets would include the very first time step of the numerical model, i.e., a snapshot at a very early time. This makes sense since having a snapshot as close to the initial condition as possible will accurately capture the behavior of the system at early times. One may reasonably assume that the optimal snapshot set would always include the drawdown distribution at early times and the drawdown distribution representative of steady state conditions. Therefore, an effective exponential snapshot selection method will be one in which the initial and final snapshot times are fixed accordingly. For a given value of γ, there is one value for α and one value for β such that the initial and final snapshot times produced by equation (12) agree with the corresponding desired values. Therefore, the optimal exponential snapshot set is found by determining the optimal value for the parameter γ.

[34] Even though this method will result in an accurate reduced model, determining the optimal γ value for a large-scale model can be computationally demanding. One would have to use a fine temporal discretization and record the solution for the entire model domain at every time step, then construct the reduced model for every value of γ and compute an error criterion that would be used to determine the optimal value for γ. This process would have to be repeated for every well. However, if one were to use this process for the dimensionless model and could then translate the result to any real-world model, the computational requirement would be greatly reduced. Note that this translation can only be approximated at best. Since the optimal snapshot set varies with the parameters of the system (hydraulic conductivity, specific storage, etc.) and the physical geometry of the system, in principle we cannot exactly translate the optimal snapshot set developed from a dimensionless model to that of a heterogeneous model with a non-regular domain.

[35] We address this problem as follows. Define the dimensionless time as [*Bear*, 1979]

$$\bar{t}_s = \frac{K}{S_s L^2} t_s = \text{C} t_s \tag{15}$$

where $\bar{t}_s$ is the snapshot time of the dimensionless model and L, K, and S_s are the characteristic length, hydraulic conductivity, and specific storage of the real-world system, respectively. This equation is exact only if the real-world model is homogeneous and has the same shape as the dimensionless model. Using representative values for the parameters and system size (e.g., average hydraulic conductivity, etc.), the optimal snapshot selection determined for the dimensionless model can be translated, approximately, to any real-world model using equation (15). However, instead of looking for representative values of the parameters, we can proceed by noting that the ratio between the dimensionless times and real-world time is a constant, $\bar{t}_s$ / t_s = C, encompassing the representative values of the parameters and the geometry. Hence, we can evaluate this constant by looking at the pseudo-steady state times for the dimensionless and the real-world model, and then all the snapshot times can be translated using this constant. Of course, the resulting real-world snapshot set would only be approximately optimal; however, we expect that slight deviations in snapshot times from the optimal set would yield insignificant changes in the reduced-model accuracy.

[36] This methodology was applied to the dimensionless, confined, homogeneous, isotropic, aquifer in three dimensions (Figure 3). SAT3D, a linear three-dimensional finite element, saturated groundwater flow model was used to calculate drawdown given forcing [*Gambolati et al.*, 1999]. All parameters and spatial dimensions were set equal to 1.0 in the dimensionless model, which has been uniformly discretized with a nodal arrangement of 65 × 65 × 11 resulting in 46,475 nodes and 245,760 tetrahedral elements. Zero Dirichlet boundary conditions are applied across two faces of the cube opposite one another. The initial conditions are $s = 0$ everywhere. To determine the time of pseudo-steady state, the observation network depicted in Figure 4 was used. The

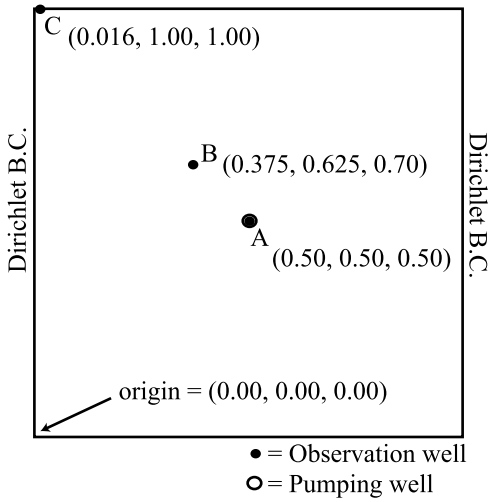

Figure 4. Locations of observation and pumping wells for the dimensionless model.

results of a single forward simulation are shown in Figure 5. Accordingly, the system reaches steady state quite rapidly near the pumping well, so rapidly that the initial snapshot must be taken at $\bar{t}_1 = 1.0 \times 10^{-7}$. At this initial time step the drawdown is on the order of 5.0×10^{-3} at the location of the well, whereas at steady state, the drawdown is on the order of 5.0. Based on Figure 5, the system is clearly at approximate steady state conditions when $\bar{t} = 0.90$. Therefore, the initial and final snapshots will be taken at these times.

[37] Equation (12) can be used to generate snapshot sets for various values of γ. Figure 6 displays the behavior of the snapshot times for different values of γ. We see that the smaller the value of γ, the heavier the emphasis on early time steps. When $\gamma \approx 1.80$, the snapshot selection is approximately linear between the initial and final snapshot times.

[38] Obtaining the optimal snapshot set when using a numerical model presents some practical difficulties. For example, the numerical model solves the system at discrete time steps, whereas the exponential snapshot distribution, described by equation (12), is continuous. Therefore, the results of the continuous snapshot selection must be rounded to the corresponding nearest model time steps. In addition, capturing snapshots at very early time steps requires variable time stepping in the numerical model. For this exercise, the time stepping is variable and consists of 2,075 time steps to ensure accuracy in the calculations. In practice, coarser time steps can actually be used. We chose to work with $n_s = 10$ snapshots. The initial-snapshot is fixed at $\bar{t}_1 = 1.0 \times 10^{-7}$ and the final-snapshot is always $\bar{t}_{10} = 0.90$. However, in between, the snapshot times are rounded to the nearest model time step.

4.3.2. Optimality Criterion

[39] The criterion needed to define optimality for the values of n_s and γ must now be determined. As mentioned before, in principle, we would like to find the snapshot set that minimizes the solution error, i.e., solves problem (14). However, the relationship between γ and ARMSE is complicated, being strongly affected in particular by the PCA step of the POD algorithm. Therefore, instead of using ARMSE as a criterion, one could employ experimental design criteria in which the goal is to maximize the information contained in the snapshot covariance matrix. The approach would be similar to using the E-optimality criterion where, in the context of experimental design, the objective is to minimize the maximum eigenvalue of the parametric covariance matrix [*Yeh*, 1992]. We may think of the best snapshot set as being the one that yields PCA principal components that are all significant in terms of variance accounted for, i.e., for which all the values of ϕ_j of equation (4) are significant. Hence, our objective is to maximize the minimum eigenvalue $\lambda_{\min}$ of the state covariance matrix **C** generated from the snapshots. The latter is the criterion of choice.

4.3.3. Numerical Results

[40] We apply this criterion to the dimensionless model described above. The best value for γ is numerically obtained via exhaustive search. Snapshot sets of 10 snapshots each are generated for 40 values of γ and the corresponding minimum eigenvalues are calculated. The results are displayed in Figure 7, where the behavior of $\lambda_{\min}$ versus γ is shown. The curve displays a distinct maximum when $\gamma = 3.87 \times 10^{-6}$, which is thus our optimal value. Some oscillations and negative eigenvalues begin to appear when γ is large, due to numerical errors for very small eigenvalues, a sign that the covariance matrix becomes ill-conditioned.

[41] In addition to determining optimal snapshot times, one must also address the size of the snapshot set, i.e., how large must the snapshot set be such that the reduced model is sufficiently accurate. To explore this problem, we develop optimal snapshot sets, using the aforementioned method and dimensionless model, for sizes of 4, 7 and 10 snapshots. To compare each snapshot size, we need to ascertain the solution accuracy (ARMSE defined in equation (14)) with respect to the full model. Although, the ARMSE is not a good criterion for snapshot determination, it can still be used to gauge model agreement. The ARMSE values were 3.09×10^{-3}, 2.74×10^{-5}, and 3.36×10^{-6} for snapshot sizes of 4, 7 and 10,

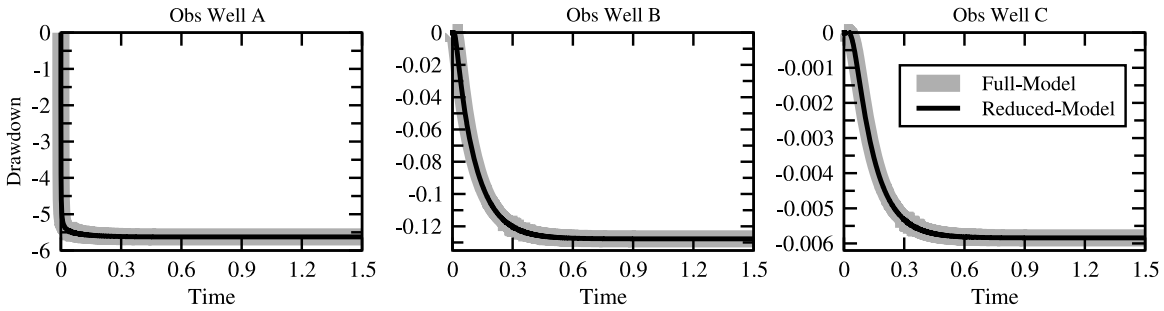

Figure 5. Time behavior of the drawdown of the dimensionless model at well locations shown in Figure 4.

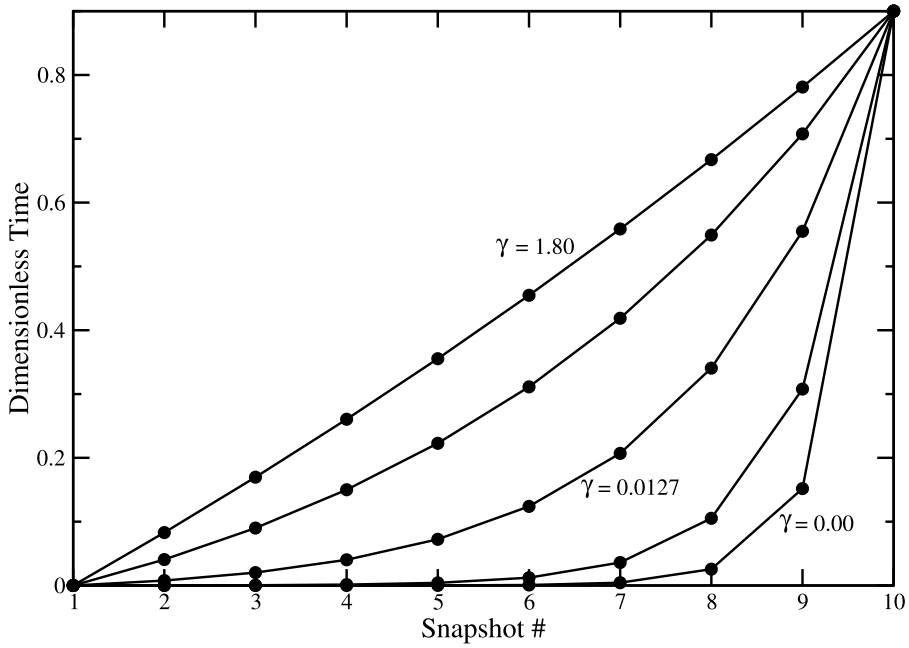

Figure 6. Behavior of snapshot times for different values of the parameter γ in equation (12).

respectively. Clearly, a snapshot size of 10 will produce a sufficiently accurate reduced model. Figure 5 shows both the full and reduced-model results for 3 nodes at all time steps using 10 snapshots and Table 1 lists the optimal snapshot times. Based on these results, the 10-snapshot reduced model is in excellent agreement with the full model. Note that the full-model system of 46,475 ODEs (n) has been reduced to a system of 10 ODEs (n_s) for a single well, creating a reduction in model dimension of more than three orders of magnitude. This can be reduced even further by omitting insignificant principal vectors from the basis, **P**; n_p is usually equal to 3 for a single well.

4.3.4. Dimensionless Versus Heterogeneous Models

[42] In this section, we verify the proposed method of translating the snapshot times of the dimensionless model to a real-world model. The real-world model is represented by a "realistic" synthetic model with dimensions 5,000m × 3,000m × 100m. The discretization for this model is the same as that of the dimensionless model: 46,475 nodes and 245,760 tetrahedral elements. The model contains 10, uniformly spaced, layers and each layer has 4 zones. Figure 8 shows the zonation of each layer. A well is placed in the center of the model domain with a constant extraction rate. The initial conditions are zero drawdown throughout the model domain.

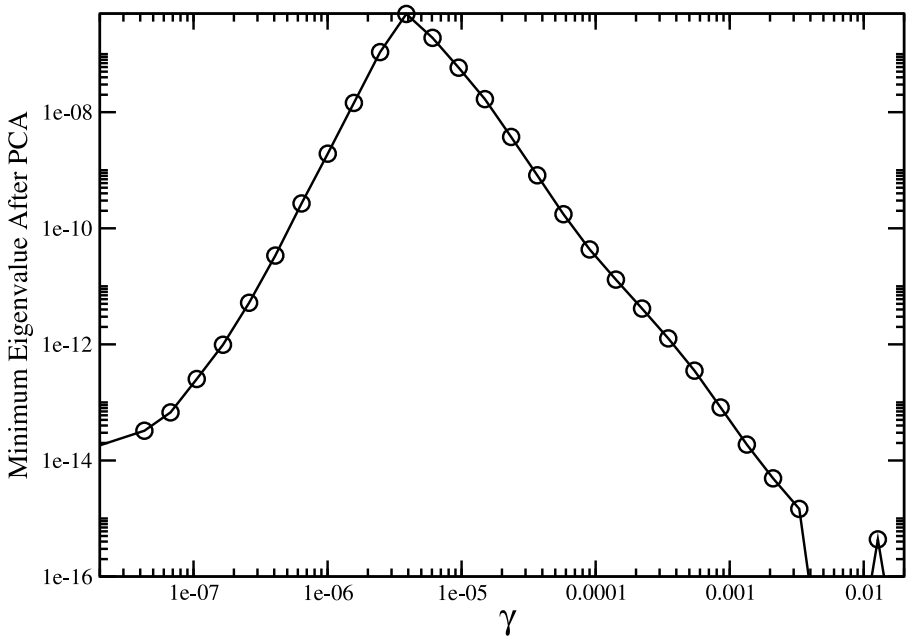

Figure 7. Relationship between γ and the minimum eigenvalue of the covariance matrix after PCA.

Table 1. Optimal Snapshot Sets for the Dimensionless and Realistic Models Along With Approximate Optimal Snapshots for the Realistic Model Determined From Translating Dimensionless Times Into Realistic Times

Snapshot Number	Optimal Snapshot Set		Approximate Optimal Snapshot Set for Realistic Model (d)
	Dimensionless	Realistic (d)	
1	1.00E−07	1.42E−04	1.42E−04
2	1.18E−05	8.06E−03	1.67E−02
3	5.76E−05	4.24E−02	8.18E−02
4	2.38E−04	1.91E−01	3.38E−01
5	9.49E−04	8.35E−01	1.35E+00
6	3.75E−03	3.62E+00	5.32E+00
7	1.48E−02	1.57E+01	2.10E+01
8	5.81E−02	6.81E+01	8.25E+01
9	2.29E−01	2.95E+02	3.25E+02
10	9.00E−01	1.28E+03	1.28E+03

Upon running the simulation and observing the drawdowns at various locations within the model domain, we conclude that the system reaches pseudo-steady state conditions after approximately 3.5 years (1,278 days). The optimal snapshot set (n_s = 10) was rigorously determined for this realistic model by maximizing λ_{min}; the optimal snapshot times are listed in Table 1. Using the pseudo-steady state times of both the realistic and dimensionless models, the constant, C (equation (15)), can be determined, and the optimal snapshots of the dimensionless model can be translated to approximate those of the realistic model. These approximated snapshots are listed in Table 1. The snapshot set from the approximation is very similar to that determined by numerically maximizing λ_{min}. The ARMSE for the approximation and the formally optimized snapshots sets, with respect to the full model, are 3.68×10^{-4} m and 5.52×10^{-4} m, respectively. Note that the ARMSE for the approximation is slightly smaller than that determined from maximizing λ_{min}. However, the ARMSE for both cases is very small suggesting that the proposed translation of snapshot times from the dimensionless model to any real-world model will likely yield a reduced model just as accurate as if one were to conduct a formal optimization of snapshot times for the real-world model itself. For this exercise, determining the optimal snapshot set for the realistic model by maximizing λ_{min} required 14 hours of CPU time using an Intel Pentium 4, 3.0 GHz, processor with a 512 KB L2 cache; whereas, the proposed translation method requires less than a second because the optimal snapshot times of the dimensionless model are predetermined.

5. Application: Central Veneto, Italy

[43] The applicability of the proposed methodology is demonstrated on a large aquifer model developed for management purposes. The aquifer is located in Northeast Italy and serves a population of a few million people in addition to agricultural and industrial demands.

5.1. Site Description

[44] The Upper and Central Veneto Plain in Northern Italy (Figure 9, top) is characterized by abundant subsurface water resources that are heavily exploited for industrial, agricultural, and civil uses, serving a population of about 3 million people. The area of interest is situated in the piedmont plain and is delimited on the west from the Lessini mountains and the Berici hills, on the north from Asiago plateau, on the east from the Brenta river and on the south from the Adriatic Sea. The aquifer system is characterized to the north-west by an undifferentiated unconfined formation recharged directly by copious precipitation collected by the Asiago Plateau. The area is characterized by the presence of spatially distributed groundwater springs localized in an area overlying the transition zone from an unconfined aquifer (in the north of the area) to confined aquifers (in the south) (Figure 9, top). The multiaquifer system originating in this area is formed by five highly productive aquifers that are the main source for potable water in the surrounding region. These aquifers extend aerially towards the south-east until they are pinched out by the Santerno Formation, a thick clay layer that confines from below the Quaternary deposits and surfaces on the eastern boundary of the Adriatic Sea. The well-developed basement of the Central Veneto aquifer system surfaces on the northern boundary and reaches about 500 m in depth on the southern boundary [*Passadore et al.*, 2007].

[45] In recent years, water extraction has exceeded recharge and as a result, most of the observation wells in the area show a trend of decreasing head levels. For this reason, public authorities have commissioned the development of a fully 3-dimensional groundwater flow model to help in the determination of management strategies. The model, described in detail by *Passadore et al.* [2007], uses a finite element method based on linear tetrahedral elements defined on a mesh of

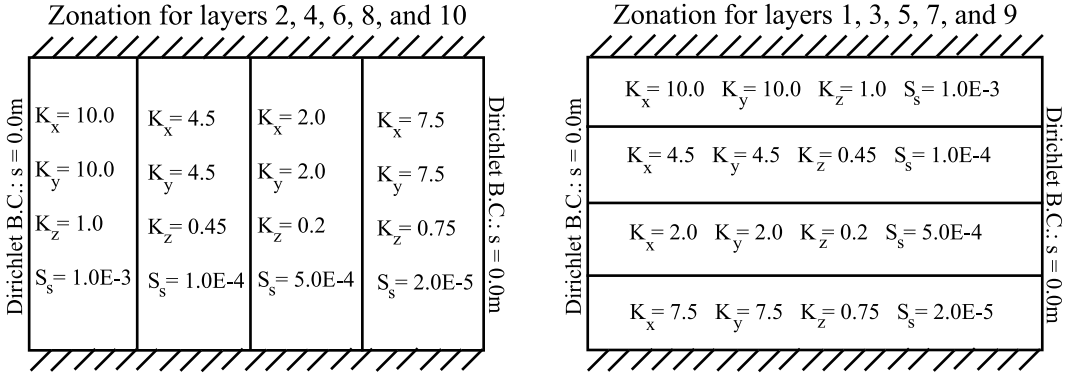

Figure 8. This diagram displays, in plan view, the zonation pattern and parameter values used in the "realistic" model.

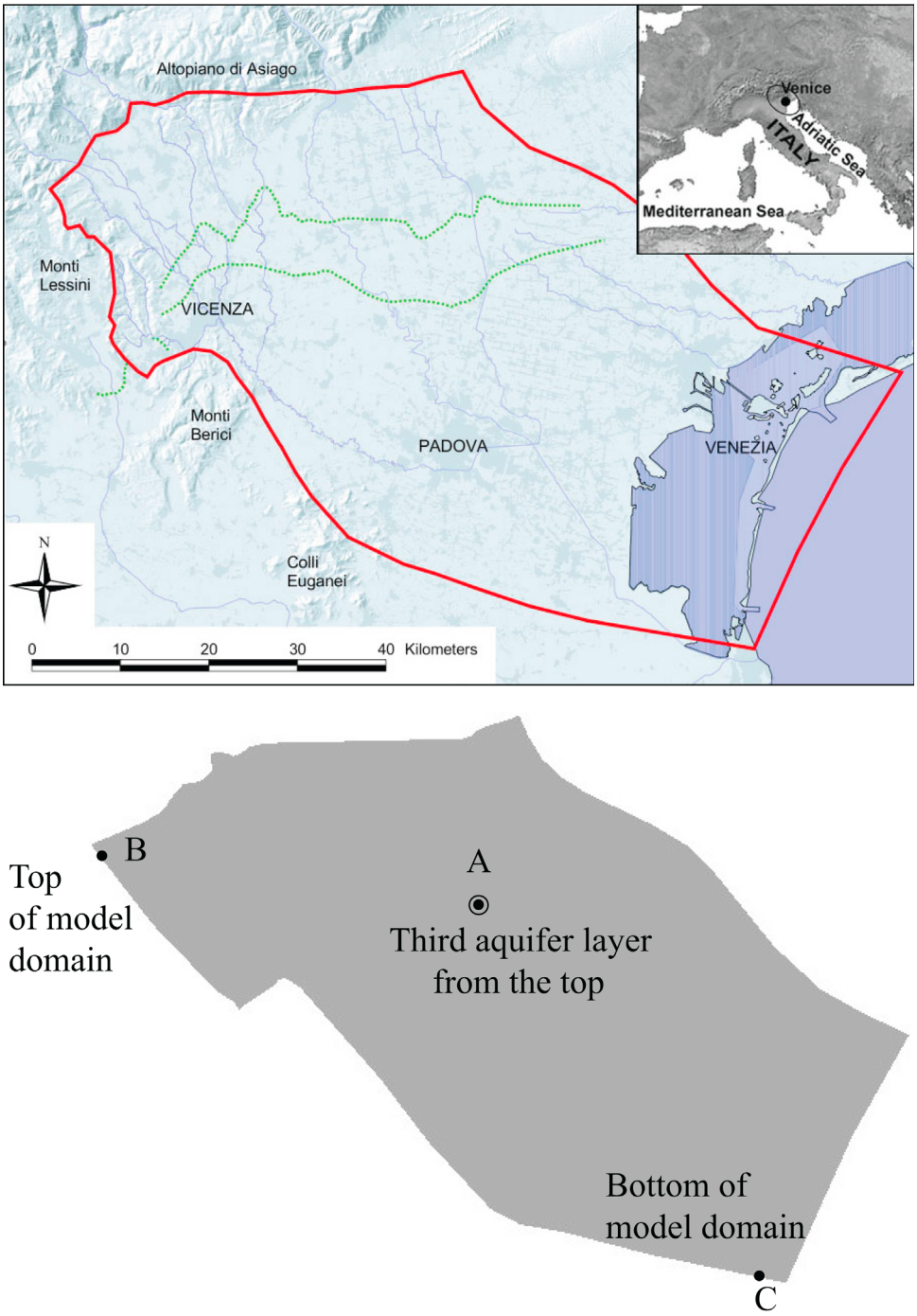

Figure 9. (top) Planar view of the Central Veneto model domain. (bottom) Locations of the pumping and observation wells for model reduction of the Central Veneto model.

143,496 nodes and 831,790 tetrahedral elements. The version of the model used in this study was refined to 222,687 nodes and 1,289,570 tetrahedral elements. The highly unstructured mesh and the large number of elements were dictated by the need to accurately capture the behavior of the unconfined-confined transition zone. This requires an accurate geometrical description of the pinch-outs in the low permeability lenses. Calibration was performed by minimizing the squared difference between simulation results and observation data collected at more than 100 wells in the area. The unconfined aquifer was subdivided into 50 material zones to capture the highly heterogeneous structure of a geological formation complicated by the presence of several paleoriverbeds.

5.2. Application of POD

[46] To demonstrate the application of the aforementioned snapshot selection method, we consider a single well located near the center of the model domain (Figure 9). The timing of approximate steady state conditions is calculated using several observation wells located throughout the model

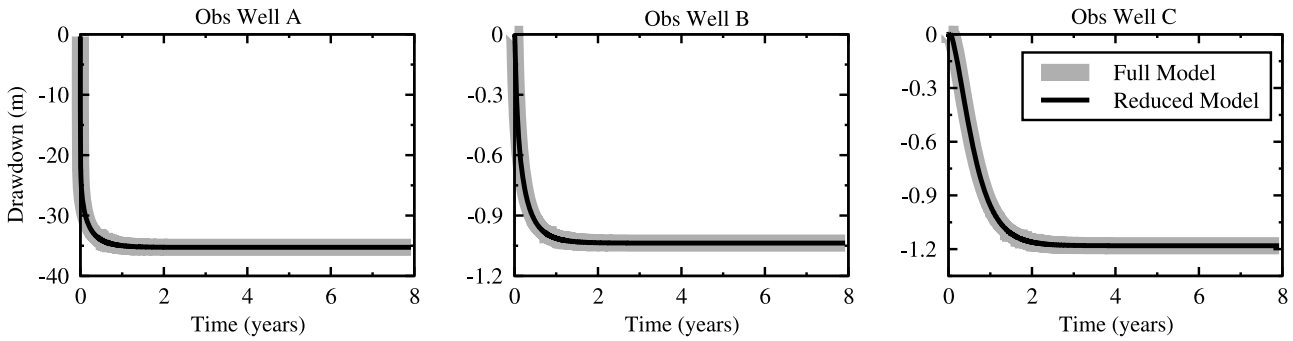

Figure 10. Time behavior of the calculated drawdowns at the well locations shown in Figure 9 for the Central Veneto Models.

domain. After some experimentation, location C in Figure 9 was determined to be an appropriate indicator as to when approximate steady state conditions are reached for the entire model domain (i.e., drawdown at this location will take the longest time to reach pseudo-steady state). After visual inspection of the drawdown curves in Figure 10, approximate steady conditions, similar to those in the dimensionless case, occur four years after constant pumping has begun. Table 2 lists the optimal snapshot times of the dimensionless model and those corresponding to the Central Veneto model.

[47] The first time step of the Central Veneto numerical model was set to 4.4×10^{-7} years (13.9 sec) and the length of each time step is multiplied by a factor of 1.03 until this length reaches 1 day, after which the time step is fixed at 1 day. The snapshots were then acquired as closely to the times listed in Table 1 as possible. The CPU time required to solve the system of equations at all time steps using a workstation with an Intel Pentium 4, 3.0 GHz, processor with a 512 KB L2 cache was 15.3 hours. Using the snapshots acquired from this simulation, the reduced model required only 56 seconds to run the same simulation on the same machine. This is a reduction in CPU time of 1000 times. Note that the full-model system of 222,687 ODEs (n) has been reduced to 10 ODEs (n_s) with a single well, which can be reduced even further to n_p ODEs by omitting insignificant principal vectors from the basis, **P**. The results of both simulations are displayed in Figure 10 for the locations depicted in Figure 9. The RMSE was also calculated for the entire domain at 10 time steps and listed in Table 3.

Table 2. Optimal Snapshot Set for the Dimensionless Model and Approximate Optimal Snapshot Set for the Central Veneto Model

Snapshot Number	Optimal Snapshot Set for Dimensionless Model	Approximate Optimal Snapshot Set for Central Veneto Model (yrs)
1	1.00E−07	4.40E−07
2	1.18E−05	5.18E−05
3	5.76E−05	2.54E−04
4	2.38E−04	1.05E−03
5	9.49E−04	4.18E−03
6	3.75E−03	1.65E−02
7	1.48E−02	6.50E−02
8	5.81E−02	2.56E−01
9	2.29E−01	1.01E+00
10	9.00E−01	3.96E+00

[48] The slightly larger ARMSE values in the early time steps are a result of some oscillations, on the order of a few millimeters, at locations far from the pumping well where the drawdown is essentially 0 m. This phenomenon is also apparent in the dimensionless case as well. However, these oscillations are so small as to be insignificant in practice. The reduced model is an extremely accurate surrogate model that runs 1000 times faster than the full model. Therefore, the reduced model now allows one to analyze scenarios that require several hundreds or even thousands of model runs such as, optimization, data assimilation, Monte Carlo uncertainty analysis, etc.

6. Conclusion

[49] The accuracy of a reduced model, developed via POD, depends strongly on the manner in which full-model snapshots are taken in time. We have developed a snapshot selection technique based on the fact that confined aquifers, subjected to constant pumping, reach approximate steady state conditions in an exponential manner. This exponential selection of snapshots assumes fixed initial and final snapshot times. Given a pre-selected fixed number of snapshots to be used in the POD procedure, we determine the optimal snapshot times by maximizing the smallest eigenvalue of the covariance matrix. This optimality criterion, similar to the E-optimality criterion used in experimental design, ensures that all the snapshots in the computed set are as significant as possible after the PCA phase of POD. The initial snapshot is set small enough such that there is only a very slight change in drawdown from the initial conditions. The final snapshot is fixed at time representative of approximate steady state conditions. A simple exponential function then defines

Table 3. RMSE Values, Over the Entire Model Domain, for Comparison Between the Full and Reduced Central Veneto Models

Time (yrs)	RMSE (m)
3.20E−04	2.15E−03
3.20E−02	2.02E−02
1.60E−01	1.82E−02
3.20E−01	8.31E−03
4.80E−01	1.18E−02
6.30E−01	7.80E−02
9.50E−01	1.08E−03
1.30E+00	1.47E−03
1.90E+00	8.41E−04
3.20E+00	5.56E−05

the manner in which snapshots are selected in between these time steps. There are three parameters in this function; however, since the first and last snapshot is fixed, only one of the parameters needs to be optimized, i.e., for a particular value of one parameter, the other two can be calculated.

[50] Solving this problem of optimal snapshot selection for a large scale, real world, problem can be computationally demanding. This would require a very fine temporal discretization, in which the details are not straight forward, and the solution must be recorded at each time step. Then the exponential selection method developed in this study can be applied. However, this method requires a search algorithm to determine the best parameter values in the exponential snapshot selection function. This process would have to be repeated for each well. In addition, this analysis would have to be conducted for each real world model. However, the optimal parameters of the exponential formula, for any real-world model, are primarily a function of both the model parameters and model geometry. To approximate the optimal snapshot set for any domain shape and heterogeneous porous media, we have developed a simple protocol for snapshot selection based on the optimal snapshot set of a unit cubic domain and a dimensionless equation. Since we are looking at selecting snapshot times, we need to translate the dimensionless times into real-world times. The transformation from dimensionless time to real-world time is approximately linear and is based on the properties of the real-world model. The determination of the proportionality constant in this relationship is obtained as the ratio between the lengths of time in which approximate steady state conditions are reached for both the dimensionless and real-world models.

[51] We have determined the optimal snapshot set for the dimensionless model in three dimensions by determining the appropriate parameter values in the exponential snapshot selection function via exhaustive search. The accuracy of the reduced model was assessed by evaluating the norm of the error between full-model and reduced-model solutions. The results have shown that this technique yields an accurate reduced model. The optimal snapshot set was then used to evaluate the accuracy of a simple heterogeneous test problem defined on a regular domain, mimicking a real-world model. Pseudo-steady state times allowed the determination of the transformation between the dimensionless and heterogeneous models. The results showed that the sub-optimality of the resulting snapshot set does not significantly influence the accuracy of the reduced model.

[52] The applicability of the proposed approach was tested on a real aquifer model. The resulting optimal snapshot times for the dimensionless model were mapped to their equivalent times in a basin scale model of Central Veneto, Italy. This model consists of 222,687 nodes and 1,289,570 elements with more than 50 zones each containing different values for hydrogeologic properties. The reduced-model solution contains very small errors when compared to the full model, and achieved a reduction in CPU time of approximately 1000 times. The full-model system, with a single well, was reduced from 222,687 ODEs to 10 ODEs which can be reduced even further by omitting insignificant principal vectors from the basis spanning the reduced-model solution space. This result shows that the proposed methodology can be used very effectively for groundwater simulation models, in particular when repeated runs with different forcings or different parameter values are required, such as in the case of scenario evaluation, groundwater management models, Monte Carlo simulations, etc.

[53] In conclusion, POD is a very effective tool for reducing the computational burden of groundwater simulations for linear problems, although its applicability and robustness in nonlinear cases is still debated and needs to be studied in more depth. Efficiency can be further enhanced using matrix exponentiation algorithms as opposed to time-stepping strategies. Matrix exponentiation schemes using, for example, Leja polynomials [*Bergamaschi and Vianello*, 2000; *Caliari et al.*, 2004] allow for the evaluation of the solution at selected times without the need to go through more commonly used time-stepping schemes. This will significantly aid in snapshot selection as it can provide the model solution at any specific time, thus alleviating problems associated with matching snapshot times with the time steps of a time-stepping strategy.

[54] **Acknowledgments.** This material is based on work supported by NSF under award EAR-0910507 and ARO under award W911NF-10-1-0124. We would like to thank two anonymous reviewers for their in-depth and constructive reviews.

References

Baú, D. A., and A. S. Mayer (2006), Stochastic management of pump-and-treat strategies using surrogate functions, *Adv. Water Resour.*, *29*, 1901–1917, doi:10.1016/j.advwatres.2006.01.008.

Bear, J. (1979), *Hydraulics of Groundwater*, McGraw-Hill, London.

Bergamaschi, L., and M. Vianello (2000), Efficient computation of the exponential operator for large, sparse, symmetric matrices, *Numer. Linear Algebra Appl.*, *7*(1), 27–45, doi:10.1002/(SICI)1099-1506(200001/02)7:1<27::AID-NLA185>3.0.CO;2-4.

Caliari, M., M. Vianello, and L. Bergamaschi (2004), Interpolating discrete advection-diffusion propagators at spectral Leja sequences, *J. Comput. Appl. Math.*, *172*(1), 79–99.

Cazemier, W., R. Verstappen, and A. Veldman (1998), Proper orthogonal decomposition and low-dimensional models for driven cavity flows, *Phys. Fluids*, *10*(7), 1685–1699, doi:10.1063/1.869686.

Chu, M. T. (1998), Inverse eigenvalue problems, *SIAM Rev.*, *40*(1), 1–39, doi:10.1137/S0036144596303984.

Gambolati, G., M. Putti, and C. Paniconi (1999), Three-dimensional model of coupled density-dependent flow and miscible salt transport in groundwater, in *Seawater Intrusion in Coastal Aquifers: Concepts, Methods, and Practices*, edited by J. Bear et al., pp. 315–362, Kluwer Acad., Dordrecht, Netherlands.

Hill, M. C., and C. R. Tiedeman (2007), *Effective Groundwater Model Calibration: With Analysis of Data, Sensitivities, Predictions, and Uncertainty*, Wiley-Interscience, Hoboken, N. J.

John, F. (1982), *Partial Differential Equations*, Springer, New York.

Kowalski, M. E., and J.-M. Jin (2003), Model-order reduction of nonlinear models of electromagnetic phased-array hyperthermia, *IEEE Trans. Biomed. Eng.*, *50*(11), 1243–1254, doi:10.1109/TBME.2003.818468.

Loève, M. (1978), *Probability Theory*, vol. 2, Springer, New York.

McPhee, J., and W. W.-G. Yeh (2008), Groundwater management using model reduction via empirical orthogonal functions, *J. Water Resour. Plann. Manage.*, *134*(2), 161–170, doi:10.1061/(ASCE)0733-9496(2008)134:2(161).

Mijalkovic, S. (2004), Using frequency response coherent structures for model-order reduction in microwave applications, *IEEE Trans. Microwave Theory Tech.*, *52*(9), 2292–2297, doi:10.1109/TMTT.2004.834567.

Orbak, Â. Y., E. Eskinat, and O. S. Turkay (2004), Physical parameter sensitivity of system eigenvalues and physical model reduction, *J. Franklin Inst.*, *341*, 631–655, doi:10.1016/j.jfranklin.2004.07.001.

Passadore, G., M. Monego, M. Sartori, M. Putti, L. Altissimo, A. Sottani, and A. Rinaldo (2007), 3D flow model of aquifer systems of central Veneto (Italy), in *2007 IAHR Congress Proceedings: Venice, Italy* [CD-ROM], edited by G. Di Silvio and S. Lanzoni, CORILA, Venice, Italy.

Rogers, L. L., and F. U. Dowla (1994), Optimization of groundwater remediation using artificial neural networks with parallel solute transport modeling, *Water Resour. Res.*, *30*(2), 457–481, doi:10.1029/93WR01494.

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

Vermeulen, P., A. Heemink, and C. T. Stroet (2004), Reduced models for linear groundwater flow models using empirical orthogonal functions, *Adv. Water Resour.*, *27*, 57–69, doi:10.1016/j.advwatres.2003.09.008.

Willcox, K., and J. Peraire (2002), Balanced model reduction via the proper orthogonal decomposition, *AIAA J.*, *40*(11), 2323–2330, doi:10.2514/2.1570.

Yan, S., and B. Minsker (2006), Optimal groundwater remediation design using an adaptive neural network genetic algorithm, *Water Resour. Res.*, *42*, W05407, doi:10.1029/2005WR004303.

Yeh, W. W.-G. (1992), Systems analysis in ground-water planning and management, *J. Water Resour. Plann. Manage.*, *118*(3), 224–237, doi:10.1061/(ASCE)0733-9496(1992)118:3(224).

Zhang, W., and B. Michaelis (2003), Shape control with Karhunen-Loève-decomposition: Theory and experimental results, *J. Intell. Mater. Syst. Struct.*, *14*(7), 415–422, doi:10.1177/1045389X03034059.

M. Putti, Department of Mathematical Methods and Models for Scientific Applications, University of Padua, I-35121 Padua, Italy.

A. J. Siade and W. W.-G. Yeh, Department of Civil and Environmental Engineering, University of California, 5732 Boelter Hall, Los Angeles, CA 90024, USA. (williamy@seas.ucla.edu)