Low-dimensional modelling of numerical groundwater flow

P. T. M. Vermeulen,1* A. W. Heemink1 and C. B. M. te Stroet2

1 Delft University of Technology, Faculty of Information Technology and Systems, Department of Applied Mathematical Analysis, PO Box 5031, 2000 GA Delft, The Netherlands
2 Netherlands Institute of Applied Geoscience TNO, National Geological Survey, PO Box 80015, 3508 TA Utrecht, The Netherlands

Abstract:
Numerical models are often used for simulating groundwater flow. Written in state-space form, the dimension of these models is of the order of the number of grid cells used and can be very high (more than a million). As a result, these models are computationally very demanding, especially if many different scenarios have to be simulated. In this paper we introduce a model reduction approach to develop an approximate model with a significantly reduced dimension. The reduction method is based upon several simulations of the large-scale numerical model. By computing the covariance matrix of the model results, we obtain insight into the variability of the model behaviour. Moreover, by selecting the leading eigenvectors of this covariance matrix, we obtain the spatial patterns that represent the directions in state space where the model variability is dominant. These patterns are also called empirical orthogonal functions. We can project the original numerical model onto those dominant spatial patterns. The result is a low-dimensional model that is still able to reproduce the dominant model behaviour. Copyright © 2004 John Wiley & Sons, Ltd.

KEY WORDS groundwater; model reduction; Modflow; empirical orthogonal functions (EOFs); subspace identification; low-dimensional model

INTRODUCTION
The access and accuracy of digital data, nowadays, allows the use of numerical groundwater models with a huge amount of grid cells (often >1 million). Such large and often transient models have long computational times. Apparently, the advantage of digital data changes into a disadvantage when we need to evaluate the model for many different cases, e.g. to optimize a pumping strategy. Especially in such a case, our objective should not be to formulate a very accurate mathematical model, but to formulate a model that is useful for the intended study. The main idea of a reduction method is to seek a simpler model that is able to represent the numerical groundwater model for the intended study. From this perspective, Newman (1996a) stated: ‘We may wish to sacrifice some of the correctness of the model in order to make the equations easier to solve, or to allow a faster computation’. Such a simple model consists of a time-dependent coefficient and a distribution of spatial patterns. There are different approaches mentioned in the literature to specify the patterns, such as analytical-, system- and data-driven strategies:

• Analytical patterns/functions can be used, such as Fourier functions, wavelets or polynomials. The disadvantage of such functions is that they have to be defined a priori and that they are non-economical regarding the amount of analytical patterns required (Cazemier et al., 1998).
• More economical is the technique of computing patterns as an eigenvalue decomposition of the system matrix differential equations. Sahuquillo (1983) was one of the first to use eigenvalues and a partial part of...
The eigenvectors of the system matrix to compute the influence functions of pumpage on the piezometric head. To compute them within the entire domain the complete set of eigenvectors is necessary. When the number of cells is relatively high this can be computed efficiently with the Lanczos algorithm. This has been done for the groundwater flow equation (Dunbar and Woodbury, 1989), in simulating groundwater flow in a dual-porosity media (Zhang and Woodbury, 2000b), and in the solution of radionuclide decay chain transport in dual-porosity media (Zhang and Woodbury, 2000a). A significant reduction of the dimension of the problem can be obtained if the systems dynamics matrix contains relatively few dominant eigenmodes.

- The last technique uses specific model results (empirical data) to define the patterns. This data-driven technique produces a minimal set of patterns that span a specific data set optimally. In fact, the basis formed by the patterns is optimal in the sense that no other basis of the same dimension has a smaller mean square error. Empirical orthogonal function (EOF) analysis has been used for large data sets to find the denominators to depict the modes of variation (Reyment and Jöreskog, 1993) and, owing to the limited number of patterns, the behaviour of a complex system can be analyzed more easily. Ghanem (1998) used these patterns to model the heterogeneity in random porous media as a superposition of scales of heterogeneity that are statistically uncorrelated. In the different fields of science, this pattern identification technique is also called coherent structures (CS), principal component analysis (PCA), common factor analysis (CFA) or proper orthogonal decomposition (POD). Graham and Kevrekidis (1996) discuss some limitations of the pattern identification technique and propose certain modifications to it in cases where multiple stable steady states may exist and infrequent events occur, especially in turbulent flow.

In a recent study (Hooimeijer, 2001), the different pattern identification techniques (data driven and analytical) are described extensively. Hooimeijer (2001) selected the EOF technique as being most suitable to develop a reduced morphological model. The original model was not known explicitly and the best results were obtained using a linear auto-regressive model in which the EOFs were multiplied by coefficients. Hooimeijer (2001) stated: ‘It seems to be more elegant and natural to project the original equations directly onto the EOFs. In order for this method to work, the entire equation set of the original model must be written in matrix format.’ This type of projection is not new, and most of the concepts and methods presented in this paper have been mathematically proven by Newman (1996a). The technique has been applied in a wide range of scientific fields, such as turbulence and image processing (Sirovich, 1987), rapid thermal chemical vapour deposition (Adomaitis, 1995), in fluid dynamics for a lid-driven cavity with a rotating rod (Hoffman Jørgensen and Sørensen, 2000), in flow reactors governed by nonlinear equations (Park and Cho, 1996), and heat transfer problems (Park and Cho, 1999). They all projected the partial differential equation onto the patterns, which we call a Galerkin projection or Karhunen–Loève expansion. In groundwater hydrology, this projection technique has—as far as we know—never been applied. We present this type of model reduction to construct a low-dimensional groundwater model suitable for a predefined model application; see Figure 1. With the original full-dimensional model we generate specific model evaluations (snapshots) from which we extract patterns (EOFs). We use the data-driven technique because it is much more flexible and the most economical regarding the amount of patterns required. We use the patterns to create a low-dimensional model by performing a projection upon the finite differential equation for groundwater flow. A verification of the low-dimensional model can lead to a refinement of the snapshots and/or to the selection of more or different patterns. The performance and suitability of the low-dimensional model is illustrated by a real-life case.
MATHEMATICAL OUTLINE

Full-dimensional model

To describe groundwater flow in three dimensions we use the partial differential equation based upon Darcy’s law and the equation of continuity (McDonald and Harbaugh, 1988):

\[
\frac{\partial}{\partial x} \left( T_x \frac{\partial h}{\partial x} \right) + \frac{\partial}{\partial y} \left( T_y \frac{\partial h}{\partial y} \right) + \frac{\partial}{\partial z} \left( T_z \frac{\partial h}{\partial z} \right) - q = S \frac{\partial h}{\partial t}
\]

(1)

where \( h \) is the piezometric head, \( x, y \) and \( z \) are Cartesian coordinates, \( t \) is the time, \( T_x/y/z \) are the conductivities in \( x, y \) and \( z \) directions, \( S \) is the storage coefficient and \( q \) is the discharge or recharge term. With finite differences, such as used in Modflow (McDonald and Harbaugh, 1988), we discretize Equation (1) for a mesh of model cells (\( n_m \) is the total number of cells) and write the resulting model in state-space form as:

\[
Ah_k = b_k
\]

(2)

where vector \( h_k \) contains each \( h \) within the model for time step \( k \). The heptadiagonal matrix \( A \) with dimension \([n_m \times n_m]\) contains each \( T_x/y/z \) between adjacent model cells and their \( S \), which determine the behaviour of \( h \) over time. Vector \( b_k \) contains all the variables from the right-hand side of Equation (1), which we will unravel on purpose as

\[
b_k = Sh_{k-1} - q_k
\]

(3)

where \( S \) contains \((1/\Delta t)S\) and \( q_k \) contains the model impulses throughout the model for time step \( k \) (e.g. wells, precipitations, rivers). Equation (2) can be written as

\[
Ah_k = Sh_{k-1} + q_k
\]

(4)

so

\[
h_k = A^{-1}Sh_{k-1} + A^{-1}q_k
\]

(5)

Equation (5) is an auto-regressive formulation for groundwater flow. Simulating this model can be very time consuming. Within this paper we speak of a full-dimensional model that operates within a full-dimensional space \((h \in \mathbb{R}^{n_m})\). The dimension is equal to \( n_m \) and will, nowadays, often be \( > \) 1 million.

Low-dimensional model approximation

A low-dimensional model (approximate model) is based upon the assumption that Equation (5) can be replaced by a simpler model. This model structure divides the original model into a time-dependent coefficient and spatially distributed patterns:

\[
\hat{h}_k = Pr_k
\]

(6)

where \( \hat{h}_k \) is the approximated piezometric head for time step \( k \), \( P \) stores a collection of \( n_p \) spatial patterns, and \( r \) are \( n_p \) time-dependent coefficients. This low-dimensional model has been successfully applied within different sciences (Adomaitis, 1995; Newman, 1996b; Hoffman Jørgensen and Sørensen, 2000; Hooimeijer, 2001). The low-dimensional model restricts the model to vary within a space with limited dimensions \((r \in \mathbb{R}^{n_p})\). This low-dimensional space is described by the patterns \( P \) that span this space. To proceed to \( \hat{h}_{k+1} \) we only need to compute a new set of \( n_p \) coefficients \( r \). This saves an enormous amount of time compared with the original dimensional model, for which we need to compute \( n_m \) piezometric heads.

We add an extra term \( h_s \), such that the approximate solution \( \hat{h} \) exactly satifies the boundary conditions:

\[
\hat{h}_k = h_s + Pr_k
\]

(7)
This reference piezometric head \( h_\text{Ł} [\text{L}] \) represents a solution from the full-dimensional model, with only time-independent model impulses (e.g. wells, rivers, fixed-head conditions). We call them fixed impulses and store them within the vector \( q^f \). All other impulses, which are time dependent and represented by the patterns \( P \), are called pattern impulses \( \Delta q \bowtie \). Each low-dimensional model (Equation (7)) is restricted to a variation in \( q^P \). It is also restricted to \( h \)-independency for permeabilities and dewatering systems (linearity).

Another relationship exists between the patterns and the model parameterization described within \( A, S \) and \( q^f \). Whenever they are changed, we need to identify a new set of patterns.

**Pattern identification: snapshots method**

To find the patterns (EOFs) we create an empirical data set, in which \( n_e \) evaluations of the full-dimensional model are arranged in vectors \( h_1, h_2, \ldots, h_n \). These are called snapshot vectors (Sirovich, 1987). A special snapshot vector is the reference piezometric head \( h_\text{Ł} \). Before we estimate the covariance for this data set, we compute the difference between each snapshot vector and \( h_\text{Ł} \) and weight them upon their vector length:

\[
x_i = \frac{1}{\| h_i - h_\text{Ł} \|} h_i - h_\text{Ł} \quad i \in \{1, \ldots, n_e \}
\]

(8)

The necessity for weighting lies in the fact that, for example, a snapshot vector that contains the response for recharge can become dominant due to its larger vector length, compared with the response within another snapshot vector, e.g. one caused by a well.

We define the mean vector as

\[
\bar{x} = \frac{1}{n_e} \sum_{j=1}^{n_e} x_j
\]

(9)

which we subtract from each snapshot vector (zero-averaging/centring):

\[
d_i = x_i - \bar{x} \quad i \in \{1, \ldots, n_e \}
\]

(10)

This is mainly done for reasons of scale (Newman, 1996a) and because it reduces the number of dimensions by one beforehand (Hoffman Jørgensen and Sørensen, 2000). We collect the centred vectors as columns within the matrix \( D \) as

\[
D = [d_1, d_1, \ldots, d_n]
\]

(11)

The covariance matrix now becomes

\[
C_m = \frac{1}{n_e} DD^T
\]

(12)

with dimension \( n_m \). To shorten the calculation time necessary for solving the eigenvalue problem for the very high-dimensional empirical covariance matrix \( C_m \), we define a covariance \( C_e \) as a minor product (minor of a matrix) of \( D \):

\[
C_e = D^T D
\]

(13)

We solve the eigenvalue problem for the \( n_e \)-dimensional matrix \( C_e \) such that

\[
C_e g_i = g_i \lambda_i \quad i \in \{1, \ldots, n_e \}
\]

(14)

and we recover the eigenvectors \( G = [g_1, g_2, \ldots, g_{n_e}] \) with dimension \( n_e \) into eigenvectors \( V \) with dimension \( n_m \) by applying (Golub and van Loan, 1989)

\[
V = D G A^{-1/2}
\]

(15)
It is easy to compute $\Lambda^{-1/2}$ because $\Lambda$ is a main diagonal matrix with only positive eigenvalues $\lambda$. Finally, the patterns can be found by normalizing each eigenvector within $V$:

$$p_i = \frac{1}{\|v_i\|} v_i$$

such that all patterns have a unit length ($p^T p = 1.0$). The relative importance of each pattern/direction can be expressed by scaling the accompanying eigenvalue $\lambda$ by the total variance within the snapshot vectors ($\sqrt{\sigma}$):

$$\lambda_i(\%) = \frac{\Lambda_{i,i}}{\sqrt{\sigma}} \times 100 \quad i \in \{1, \ldots, n\}$$

with

$$\sigma_i^2 = \sum_{j=1}^{n} \Lambda_{j,j}$$

We collect $n_p$ patterns with eigenvalues $\lambda_1 > \lambda_2 > \ldots > \lambda_{n_p}$ and store their corresponding vectors as columns within $P = [p_1, p_2, \ldots, p_{n_p}]$. Matrix $P$, with dimension $[n_m \times n_p]$, is orthonormal, which means that all patterns are perpendicular to each other ($P^T P = I$). In other words, $P$ can be seen as a collection of vectors/directions that form the axes of the low-dimensional space.

**Time-dependent coefficients**

To find the time-dependent coefficients $r_k$ within Equation (7), we project the vectors and matrices within Equation (5) onto the patterns $P$. In this approach $P^T$ acts as reductor and $P$ as reconstructor. The reductor reduces an $n_m$-dimensional vector by multiplication into an $n_p$-dimensional vector; the reconstructor performs the reverse operation. To reduce a matrix we first multiply it by $P^T$ and thereafter with $P$. We project Equation (5) onto $P$ as

$$P^T \tilde{h}_k = P^T A^{-1} SP P^T \tilde{h}_{k-1} + P^T A^{-1} q_k^p$$

where $q_k^p$ represents the model impulses that can be used in the low-dimensional model. Unfortunately, we cannot solve Equation (19) as long as we need to compute $A^{-1}$. Owing to the symmetric form of $A$, we rewrite Equation (19) as

$$P^T \tilde{h}_k = (A^{-1}P)^T SPP^T \tilde{h}_{k-1} + (A^{-1}P)^T q_k^p$$

and, instead of computing $A^{-1}$ completely, we define $M$ with dimension $[n_m \times n_p]$ as

$$M = A^{-1}P$$

for which we solve

$$Am_i = p_i \quad i \in \{1, \ldots, n_p\}$$

We can now rewrite Equation (19) as

$$r_k = N r_{k-1} + M^T + M^T q_k^p$$

with

$$N = M^T S P$$

$$r = P^T \tilde{h}$$

Equation (23) is a formulation in state space for a low-dimensional groundwater model (Vermeulen et al., 2001). The computation of $r_k$ consists finally of two matrices, $M[n_m \times n_p]$ and $N[n_p \times n_p]$, and two vectors,
To compute $r_k$ we need a total of $n_p \times n_p + n_m \times n_p$ FLOPS (floating-point operations). To expand $r_k$ to $\hat{h}_k$ (Equation (7)) we need another $n_m$ FLOPS—although this is not necessary to proceed to the next time step.

The matrices $M$ and $N$ depend on $\Delta t$ and, therefore, we should compute them each time $\Delta t$ changes. This will increase the computational time dramatically. To avoid this, we solve $M$ for the smallest $\Delta t$ that is desired by the model, e.g. 1 day. Whenever $\Delta t$ changes during the simulation ($\Delta t = 3$ days), we compute the intermediate time steps ($t = 1, 2$ days) only in low-dimensional space. This will hardly increase the computational time, because we only need to recalculate a low-dimensional matrix multiplication ($N r_{k-1}$).

The product $M^T q_k$ remains identical. There is another motivation to support this strategy. Inherent to the method of finite differences, the evolution of $h$ in time ($\partial h/\partial t$) is discretized linearly between adjacent time steps as

$$\frac{h_{k+1} - h_k}{t_{k+1} - t_k} = \frac{\Delta h}{\Delta t}$$

and the smaller $\Delta t$ the more $\Delta h/\Delta t = \partial h/\partial t$.

**Initial conditions.** For transient simulations, the initial condition $h_0$ can be different from $h_*$. It is fairly easy to project $h_0$ upon the patterns as

$$r_0 = P^T (h_0 - h_*)$$

where $r_0$ is the initial condition for the low-dimensional model. It is only meaningful if we can indeed span $h_0$ by the $n_p$ selected patterns:

$$h_0 - h_* = c_1 p_1 + c_2 p_2 + \cdots + c_{n_p} p_{n_p}$$

where $c_1, \ldots, c_{n_p}$ are coefficients that should be found. Because $h_0$ is often a steady-state solution of the full-dimensional model, it is neccessary to include at least steady-state evaluations of the full-dimensional model in the snapshot vectors.

**Head-resistance relationship.** Features like surface water and other head–resistance relationships within a groundwater model have absolute levels (e.g. surface-water levels, drainage base). In some cases, these features are captured by $h_*$ and, therefore, part of $q^p$. But, if they must be varied in the low-dimensional model, then the absolute levels should be corrected by

$$h_{\text{feature}} = h_{\text{feature}} - h_*$$

before adding them to $q^p$ (Equation (23)). It is the same alternation that we applied to the snapshot vectors before extracting the patterns (Equation (8)).

**REAL-LIFE CASE**

**Introduction**

To evaluate the accuracy and suitability of a low-dimensional model for groundwater flow, we use a supra-regional model of the province of Noord-Brabant, the Netherlands. This province is characterized by regional areas with free-floating water tables (south, southeast) and regions with intense dewatering systems (north, northwest). The underground is a complex system of succeeding high ($>5000$ m$^2$ day$^{-1}$) and extremely low ($<1 \times 10^{-7}$ m$^2$ day$^{-1}$) permeabilities. The model has nine model layers which contain the highly permeable sediments (aquifers); see Figure 2. The low permeability in between the aquifers is processed as a vertical conductance; see Figure 3. Within each model layer (2, 3, ..., 9) several wells are active (770 in total). The
model network consists of 107 columns and 64 rows, each dimensioned $1 \times 1 \, \text{km}^2$. The entire model contains 61,632 model cells, in which 32,949 are active. All others are inactive (no-flow) due to an irregular boundary and/or several aquifers that thin out.

**Purpose of the low-dimensional model**

The main purpose of the low-dimensional model is to approximate the effects of different, predefined scenarios. They would consume too much time whenever computed—more accurately—with the full-dimensional model. We want to predict the long-term effects of global climate change (effect of precipitation), together with the impact of variable pumping rates for all extraction wells within the model layer 2, 3, ..., 9. This results in $n_s = 9$ groups which we can vary independently. We want to combine the groups randomly and simulate their combined effects for $n_t = 150$ time steps with $\Delta t = 10$. To be able to do so, we need to find patterns that are suitable to simulate at least those scenarios for the desired time scale.

**Determining patterns (EOFs)**

*Creating a data set.* Before we extract patterns, we need to fill a data set with $n_e$ snapshot vectors obtained from the full-dimensional model. The ensemble of snapshot vectors will influence the accuracy of a low-dimensional model (Newman, 1996b). Therefore, the snapshot vectors have to be chosen with care. On the other hand, it is a challenge to minimize the amount of snapshot vectors $n_e$ so that less computations have to be performed with the full-dimensional model. To create a minimal and ‘balanced’ data set, we take the linearity of the model into account. As a result, we are able to compute a combined effect for different groups, by combining their individual responses. Whenever we concern a steady state-simulation, we only need to fill
the data set with the steady-state response for each group; so \( n_e = n_v = 9 \). For our transient simulation, we need the patterns to capture the entire response behaviour over time; so \( n_e = \alpha n_v \). For each group, the factor \( \alpha \) depends on the response characteristics. Instead of collecting all time steps to describe the response, it is much more efficient to include only those time steps that differ significantly. In between these time steps we assume a linearity.

To determine which time steps are significant, we separately compute a ‘block’-impulse response for each group. Depending on their gradient in \( h \) between the last two time steps, we select model cells with different percentiles \( \omega \) for that gradient and plot their corresponding responses; see Figure 4. We observe that the response, caused by the collection of wells within model layer 2 (Figure 4a), reaches its equilibrium at \( t \approx 300 \) days. At that time we assume that all (\( \omega = 99\% \)) model cells have at least their steady-state solution. We can represent the behaviour of the response efficiently by computing the full-dimensional model at only four time steps: \( t = 10, 20, 70, 320 \) days. This time strategy does not satisfy the precipitation group whose response reaches its steady state for \( \omega = 99\% \) after \( t \approx 5000 \) days (Figure 4b). Its shape is different from Figure 4a and can be described best by nine time steps: \( t = 10, 30, 70, 150, 310, 630, 1270, 2550, 5210 \) days. The chosen time steps are not, by definition, the best; different combinations will improve or worsen the performance of the low-dimensional model. The most important issue herein is that the frequency should be high initially, and may be decreased as time elapses to obtain a set that fully characterizes the impulse-response optimally (Park and Cho, 1996). Thereafter, it should be continued until a steady-state solution is reached. This happens slowly for the recharge component, as the phreatic storage coefficient is much larger (\( S = 0.27 \)) than for those model layers where the wells are responding (\( S = 0.0007 \)). Finally, we fill the data set with \( n_e = 8 \times 4 + 9 = 41 \) snapshot vectors in total, which is in fact much less than one complete run with the full-dimensional model, which needs \( n_t = 150 \) evaluations.

**Reference piezometric head.** The reference piezometric head \( h_s \) is computed with only those model impulses that remain constant in time. Those impulses are stored within \( q^f \) and can, for example, be part or all precipitation, wells, dewatering systems, fixed-head boundaries. In our real-life case, all dewatering systems are part of \( q^f \) and they appear clearly in \( h_s \) (Figure 5a). The fixed-head conditions along most of the boundaries within each model layer are part of \( q^f \) too. They appear on the north and south boundaries (Figure 5b). The \( h_s \) is included as a snapshot vector in the data set; so \( n_e = 42 \).

**Results**

**Patterns.** With the \( n_e = 42 \) snapshot vectors from the two previous sections, we compute the patterns as described in the section pattern identification: snapshots method. The maximal explained variance for

---

Figure 4. Graphs showing the response caused by (a) wells within model layer 2, and (b) by precipitation. The responses are given for different model cells depending on their gradient in \( h \) between the last two time steps. Different time series are plotted for various percentiles (\( \omega \)) of that gradient.
a pattern ($\lambda_1 = 20.42\%$) declines rapidly ($\lambda_{>9} \ll 1\%$); see Figure 6. For different sciences, strategies are mentioned to obtain the number of patterns (Reyment and Jöreskog, 1993; Hooimeijer, 2001). For groundwater hydrology it is rather straightforward. If we include two patterns with $\lambda_1 = 20.42\%$ and $\lambda_2 = 16.81\%$, then we can reconstruct the full-dimensional model for $\lambda_1 + \lambda_2 = 37.23\%$. We call this the expected variance $\varphi^e(\%)$ corresponding to an approximation of order 2 ($n_p = 2$). It is remarkable that the behaviour of the full-dimensional model can be described with $\varphi^e = 99.780\%$ by only 10 patterns; see Figure 6. Apparently, the spatial distribution of $h$ is complex, but its behaviour in time is not complex at all (Sirovich, 1987; Park and Cho, 1996; Hoffmann Jørgensen and Sørensen, 2000). The number of patterns should be chosen such that $\varphi^e$ is more or less equal to the reconstructed variance $\varphi^r$:

$$
\varphi^r(\%) = \frac{\|s\|}{\|s\|} \times 100
$$

where $\hat{s}$ is the variance computed with the low-dimensional model and $s$ is the variance computed with the full-dimensional model. When $\varphi^e \neq \varphi^r$, the low-dimensional model lacks significant patterns (Figure 6, $n_p < 9$) and more patterns should be added.

**Alternative pattern selection:** Sometimes, we need a low-dimensional model that is accurate for at least $n_1$ specific model cells (e.g. near a well). In that case we compute $\xi$ for each pattern as the lumped pattern

![Figure 5. Spatial distribution of the reference piezometric head $h$. No-flow zones are crossed out](image)

![Figure 6. Graph of the order of approximation $n_p$ versus the relative eigenvalues $\lambda$, the expected $\varphi^e$ and reconstructed variance $\varphi^r$](image)
contribution for $n_1$ model cells, so:

$$\xi_j = \sum_{i=1}^{n_1} |P^{(i)}_{j}| \quad j \in \{1, \ldots, n_p\}$$  \hspace{1cm} (30)

Instead of $\lambda$, we can use $\xi$ to select appropriate patterns.

**Physical meaning of patterns:** The spatial distribution of a pattern with the largest eigenvalue ($p_{1}, \lambda_{1} = 20\% - 42\%$, Figure 7a and b) shows a strong similarity to $\iota$. We can observe the impact of precipitation on model layer 1, and the long-term effect of extraction wells on model layer 9. For a minor pattern ($p_{19}, \lambda_{19} = 0.009\%$, Figure 7c and d), such physical processes can be still recognized, though locally. If we look at patterns as if they represent such physical processes, we should expect most of the coefficients $r$ to be zero whenever we compute $q^p$, e.g. with only precipitation (Figure 8). Unfortunately, this is not the case, and almost all coefficients are not equal to zero; they range between $-17$ and $10$. We conclude that the meanings of patterns are not directly related to physical processes, but that they actually represent such basis functions that we can reconstruct each relevant solution by applying the appropriate coefficients. When we look at the distribution of the coefficients $r$ computed with a $q^p$ for the $n_v = 9$ groups individually, we observe that all major coefficients vary for each case (Figure 9). The absolute value of $r$ is mainly caused by the magnitude of the corresponding impulse; the wells in model layer 3 are significantly larger than those in the others. On the other hand, the coefficients $r_{i>11}$ are much less than $r_{i\leq11}$, which corresponds to their small $\lambda$ values.

![Figure 7. Spatial distribution of $p_1(\lambda_1 = 20\% - 42\%$) and $p_{19}(\lambda_{19} = 0.009\%)$](image-url)
Verification. We evaluate the capability of the low-dimensional model by computing \( q^p \) with different combinations of the \( n_v = 9 \) groups. We simulate \( n_t = 150 \) time steps and for each of the first 10 time steps we multiply the groups with different factors. Thereafter, we multiply them again after 10, 20, 30 and 40 time steps. In this manner, we evaluate the response for an impulse strength over a longer time period. It is important to notice that the factors are not exceeding the range of the snapshots on which the low-dimensional model is based.

The accuracy of the low-dimensional model can be quantified by examining the averaged absolute difference \( \zeta(m) \) between the full- \( (h) \) and the low-dimensional model \( \tilde{h} \) results over all time steps \( (n_t = 150) \) as:

\[
\zeta = \frac{1}{n_t} \sum_{k=1}^{n_v} \zeta_k
\]

with

\[
\zeta_k = \frac{1}{n_m} \| f_k \| \quad (31)
\]

and

\[
f_k = | h_k - \tilde{h}_k | \quad (32)
\]

Besides \( \zeta \) we also define the averaged absolute \( \beta \)-percentile error \( \zeta^\beta \) as

\[
\zeta^\beta = \frac{1}{n_t} \sum_{k=1}^{n_v} \zeta_k^\beta
\]

(33)

where $\xi^{\beta}$ is the $\beta$-percentile from $f_k$ as we order the values in $f_k$ from small to large. A disadvantage of using only the errors $\xi$ and $\xi^{\beta}$ to evaluate the performance of the low-dimensional model is that an acceptable error depends on the behaviour of the model. To improve this shortcoming, we define an averaged absolute percent error $\bar{\varepsilon}(\gamma)$ as

$$
\bar{\varepsilon}(\gamma) = \frac{1}{n_1} \sum_{k=1}^{n} \varepsilon_k(\gamma)
$$

with

$$
\varepsilon_k(\gamma) = \frac{1}{n_m} \| e_k(\gamma) \|
$$

and

$$
e^{(i)}_k(\gamma)(\%) = \begin{cases} 
0 & \text{if } f^{(i)}_k \leq \gamma \\
\frac{f^{(i)}_k}{|h^{(i)}_k - h^{(i)}_0|} \times 100 & \text{if } f^{(i)}_k > \gamma \quad i \in \{1, \ldots, n_m\}
\end{cases}
$$

For the computation of each element $i$ within $e^{(i)}_k(\gamma)$ we use a threshold with a minimal acceptable error $\gamma$. Here, we assign $\gamma = 0.01$ m. The error $e_k(\gamma)$ allows the low-dimensional model to decrease its performance whenever its solution is further away from the reference piezometric head $h_s$. The equivalents expressed as percentiles are defined as

$$
\bar{\varepsilon}^{\beta}(\gamma) = \frac{1}{n_1} \sum_{k=1}^{n} \xi^{\beta}_k(\gamma)
$$

with $\xi^{\beta}_k(\gamma)$ as the $\beta$-percentile within $e_k(\gamma)$.

We examine the behaviour of the errors described and we see that they are clearly related to the order of approximation; see Figure 10. The final order of approximation is more or less a matter of how much accuracy is desired. We observe for $n_p < 10$ that the $\bar{\varepsilon}^{0.01}(0.01)$ is unacceptably large ($>500\%$). With $n_p = 19$, all errors are reduced significantly (almost two orders of magnitude). We want the model to predict the behaviour of the model accurately for all time steps. If we plot $\varepsilon_k(0.01)$ for each time step for different orders of approximation (Figure 11), we see that $\varepsilon_k(0.01)$ for $n_p = 14$ fluctuates enormously with a maximum of $\sim50\%$. The difference between the error by $n_p = 19$ and $n_p = 26$ is negligible; the ‘extra’ seven patterns do not increase the performance of the entire low-dimensional model significantly. Those patterns describe a very local phenomenon and their coefficients are very tiny. It is possible that these patterns should be included in the low-dimensional model when special attention is needed on those local phenomenon; see Equation (30).

![Figure 10. Computed overall mean $\bar{\varepsilon}$ and percent $\bar{\varepsilon}(0.01)$ error and their percentiles $\bar{\varepsilon}^{0.01}$ and $\bar{\varepsilon}^{0.01}(0.01)$ versus the order of approximation $n_p$](image-url)
Figure 11. Computed norm error $\varepsilon_k^{99}(0.01)$ for different orders of approximation $n_p$ versus $n_t = 150$ time steps ($\Delta t = 10$)

![Figure 11](image)

We observe a similarity in the spatial distribution of $\varepsilon_k^{99}(0.01)$ for different orders of approximations; see Figure 12. The enlarged errors are mainly concentrated near regions that receive a significant model impulse (e.g., precipitation, wells). As expected, owing to its less dynamic behaviour due to its storage coefficient $S = 0.0007$, $\varepsilon_k^{99}(0.01)$ becomes smaller for model layer 9. We can see that indirectly as we compare Figure 4a and b. The behaviour of precipitation, which affects model layer 1, needs more snapshots, and, therefore,
indirectly more patterns than the description of the well responses within model layers 2, 3, \ldots, 9. If we compare \( \tau(0.01) \) for different orders of approximation (Figure 12a, c and d), then we see that the ‘extra’ patterns decrease the error mainly locally. With a maximum percentage error for model layer 1 of \( \leq 10\% \), the low-dimensional model is close to the full-dimensional model.

Finally, we plot two time series that both possess a maximum percentage error \( (\tau^{0.01}(0.01)) \). Although we can observe small acceptable differences, it is clear that the low-dimensional model simulates the transient behaviour of the model rather well; see Figure 13a and b. The low-dimensional model can simulate impulses that vary for each time step \( (t \leq 10) \); it can also handle block impulses \( (t > 10) \). Note that the error does not accumulate. We obtain an acceptable low-dimensional model with \( \tau^{0.01}(0.01) \ll 3.0\% \); \( \tau(0.01) = 0.043\% \); \( D = 0.019 \) m and \( D = 0.00036 \) m for \( n_p = 19 \). Including more patterns will improve the low-dimensional model slightly.

**Mass conservation:** The results of a groundwater model are often used to determine the velocity vectors that can be fed into a transport model. To evaluate whether the results obtained by a low-dimensional model are accurate enough, we compute an additional error that quantifies the difference in mass distribution; so:

\[
\Omega^j(\%) = \frac{1}{n_\theta} \sum_{k=1}^{n_\theta} \frac{\theta^{j(i)}_k}{6} \frac{\sum_{j=1}^{n} \left| Q^{(i)}_j - Q^{(i)}_j \right|}{Q^{(i)}_j} i \in \{1, \ldots, n_m\} \tag{39}
\]

with

\[
\theta^{j(i)}_k = \sum_{j=1}^{6} \left| Q^{(i)}_j - Q^{(i)}_j \right| i \in \{1, \ldots, n_m\} \tag{40}
\]

where \( Q^{(i)}_j [L^3T^{-1}] \) is the flow over cell face \( j \) for grid cell \( i \) and \( Q^{(i)}_j \) is the flow for the low-dimensional model. Within a finite difference scheme there are six cell faces. The error \( \theta^{(i)} \) has been computed as a time average for each grid cell \( i \) and plotted for model layer 1 and 9; see Figure 14a and b. For a single grid cell the maximum value for \( \theta \) is \( \sim 1560 \text{ m}^3 \text{ day}^{-1} \) and \( \sim 21 \text{ m}^3 \text{ day}^{-1} \) for model layers 1 and 9 respectively. Most of the grid cells possess a value for \( \theta \) that is considerable less. Whether this error in the mass balance is small enough, such that the coupling with a transport model and/or particle tracking program is sustained, the error \( \Omega \) scales \( \theta \) onto the total absolute face flows (Equation (39)) and quantifies the relevance of \( \theta \); see Figure 14c and d. The value for \( \Omega \) is acceptable within most of the model domain, with some isolated peaks that are still

![Figure 13](image-url)

Figure 13. Time series, selected for model layer 1 (a) and 9 (b), that satisfies \( \tau^{0.01}(0.01) \), showing the result for the full- (\( h \)) and low-dimensional model (\( \hat{h}, n_p = 19 \))
Table I. Total in- and out-flow for a full- and low-dimensional model with \( n_p = 19 \) patterns

<table>
<thead>
<tr>
<th>Layer</th>
<th>Full dimensions</th>
<th>Low dimensions</th>
<th>Difference (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>In (10^6 m³)</td>
<td>Out (10^6 m³)</td>
<td>ΔIn (10^6 m³)</td>
</tr>
<tr>
<td>1</td>
<td>770.9</td>
<td>−751.1</td>
<td>0.201</td>
</tr>
<tr>
<td>2</td>
<td>446.1</td>
<td>−705.4</td>
<td>0.039</td>
</tr>
<tr>
<td>3</td>
<td>182.8</td>
<td>−349.5</td>
<td>0.033</td>
</tr>
<tr>
<td>4</td>
<td>143.3</td>
<td>−225.7</td>
<td>0.007</td>
</tr>
<tr>
<td>5</td>
<td>121.9</td>
<td>−160.5</td>
<td>−0.006</td>
</tr>
<tr>
<td>6</td>
<td>103.0</td>
<td>−206.2</td>
<td>−0.001</td>
</tr>
<tr>
<td>7</td>
<td>85.3</td>
<td>−103.7</td>
<td>−0.030</td>
</tr>
<tr>
<td>8</td>
<td>117.9</td>
<td>−299.7</td>
<td>−0.033</td>
</tr>
<tr>
<td>9</td>
<td>29.7</td>
<td>−139.8</td>
<td>−0.099</td>
</tr>
<tr>
<td>Total</td>
<td>2000.9</td>
<td>−2941.6</td>
<td>0.084</td>
</tr>
</tbody>
</table>

less than \(~0.2\%). Table I presents the total water balance for the low- and full-dimensional models, subdivided for each model layer. The error in the water balance for each model layer is within acceptable ranges. Within the low-dimensional model there is \(~1.7 \times 10^6 \text{ m}^3\) more inflow and \(~2.3 \times 10^6 \text{ m}^3\) more outflow, which is only \(~0.16\%\) of the total water balance.
**CPU time reduction.** The main advantage of a low-dimensional model is its reduced computational time, compared with that of the original model. We express this by a time-reduction factor \( \tau = \tau_{\text{full}}/\tau_{\text{low}} \), which is the fraction between the computational time of the full-dimensional model \( (\tau_{\text{full}}) \) and the low-dimensional model \( (\tau_{\text{low}}) \). We compute \( \tau \) for four major processes within the computation of both models. In Figure 15 these are presented for the real-life case with 32,949 active cells and a low-dimensional model with \( n_p = 19 \) patterns.

The maximum time reduction with a low-dimensional model can be obtained by computing only the coefficients \( \tau_{1}^{\text{low}} \). This includes the reduction caused by processing modules that are already represented by \( h_x \) (dewatering systems) and the patterns (permeabilities). If we compute \( \tilde{h}_k \) (Equation (7)) for each time step \( k \), then we still achieve a considerable time reduction \( \tau_{2} \). If we save each time step \( k \) on disk, we achieve less time reduction \( \tau_{3} \). For the low-dimensional model, this process takes relatively much more time than the matrix computations \( \tau_{4} \). For the full-dimensional model, this process is negligible compared with the time taken by the solver \( \tau_{4} \). In practice, the final time reduction will be a combination of \( \tau_{1} \) and \( \tau_{2} \). The factor \( \tau \) relates to \( n_m \) and the numerical complexity of the full-dimensional model. As the CPU time for Modflow tends to increase with a factor of \( n_m^2 \), the CPU time for a low-dimensional model increases with \( n_m^2 \). Therefore, the reduction approach increases in efficiency with \( n_m \).

**COMPARISON OF THE EOF METHOD WITH THE LANCZOS–ARNOLDI METHOD**

The Lanczos method (Dunbar and Woodbury, 1989; Zhang and Woodbury, 2000b) and the Arnoldi method (Zhang and Woodbury, 2000a) can also be used for model reduction. The idea behind the Lanczos or Arnoldi vectors is that groundwater systems, which are diffuse in nature, can have eigenvalues that are relatively small. The reduction is achieved by selecting only the leading Lanczos vectors of the system dynamics matrix. The Lanczos vectors that are left out represent a dynamic behaviour within the model that dissipates more quickly than with other vectors. Consequently, the maximum error always occurs in the first several time steps, or the moment the pumping rate is changed at a particular well. The main difficulty arises in the fact that the eigenvalue decomposition is difficult to implement, as the matrices are large. This is solved by the Lanczos process, which is recursive in nature and gives an additional largest Lanczos vector (pattern) after each step in the recursion. To determine when to terminate the recursion process a ‘participation factor’ is derived, which computes the contribution of an ‘extra’ Lanczos vector. The Lanczos approach is very general, since the resulting reduced model is independent of source terms such as wells and Neumann (flow) boundary conditions. On the other hand, in most realistic applications, the system matrix does not contain a relatively few dominant eigenvalues and the possibilities for reduction are limited.

![Figure 15](image-url)
In our approach, the reduction is based upon the snapshots available. As a result, the reduced model cannot be used for scenarios where specific source terms are active and which were not included in the computation of the snapshots. The reduced model is, therefore, less generally applicable than in the case of the Lanczos method, but the reduction in computational time is much larger. Even in the case where the system dynamics matrix does not have eigenmodes that can be neglected (and the Lanczos method cannot be used), a significant reduction can be achieved with our method.

A comparison between our approach and the Lanczos method has been made for a simple Theis problem, as explained in the section Examples in Dunbar and Woodbury (1989). The Lanczos method needed, for an identical accuracy, 10 Lanczos vectors instead of three EOFs in our approach. The maximum reduced computational time for this specific test problem was two orders of magnitude for the Lanczos method and four orders of magnitude for our method (both methods computing only the time-dependent coefficients $r_k$).

CONCLUSIONS

This paper describes a methodology to approximate a linear numerical groundwater model by constructing a low-dimensional model. This approximate model is derived by projecting the finite differential equation upon an orthogonal basis of patterns. Those patterns (EOFs) are the most important eigenvectors of a covariance matrix that is computed from several simulations with the full-dimensional model. The performance of the low-dimensional model is directly related to the amount of patterns. In the real-life case we obtained an acceptable low-dimensional model that was able to reconstruct a wide variety of scenarios, with $n_p = 19$ pattern ($\phi = 99.99\%$). Instead of solving millions of equations iteratively, the low-dimensional model computes the piezometric heads by matrix multiplication only. Without affecting the accuracy of the model significantly, the computational time can be reduced enormously: approximate one order of magnitude to reconstruct the entire full-dimensional model for each time step, and approximately two orders of magnitude whenever the time-dependent coefficients are computed only.

In future research we intend to focus on non-linear problems, such as unconfined aquifers (Vermeulen, unpublished) and the non-linearity that exists within the relationship between groundwater and dewatering systems, such as drains. Experiences from other sciences showed that the low-dimensional model is truly valuable in parameter estimation (Park and Cho, 1996) and in the computation of the reduced-order Kalman filters (Heemink et al., 2001). These topics also have our interest.

REFERENCES


