Inverse modelling of groundwater flow using model reduction

P. T. M. VERMEULEN[1] & A. W. HEEMINK[2]

1 *Netherlands Institute of Applied Geoscience TNO – National Geological Survey (TNO-NITG), PO Box 80015, NL-3508 TA Utrecht, The Netherlands*
peter.vermeulen@tno.nl

2 *Delft University of Technology, Faculty of Electrical Engineering, Mathematics and Computer Science, Department of Applied Mathematical Analysis, PO Box 5031, NL-2600 GA Delft, The Netherlands*

Abstract This paper describes a reduced model for groundwater flow that reduces the computational burden necessary for inverse modelling. The formulation follows from a projection of the original groundwater flow equation upon a set of basis vectors (i.e. patterns). Those patterns are computed by a decomposition of the covariance matrix for an ensemble of model evaluations. These represent the model behaviour with respect to model forces and a set of estimation parameters. By projecting (mapping) the original equations upon those patterns, a reduced model is obtained that can be used to solve a minimization problem with negligible computational costs. For several synthetic cases this reduced model found the global minimum more efficiently than the original model using adjoints or finite differences.

Keywords adjoint; Galerkin projection; inverse modelling; reduced model

INTRODUCTION AND BACKGROUND INFORMATION

Inverse modelling (Tarantola, 1987) is known as computing a model that satisfies (invert) a given observation most optimally. It is widely accepted that a model therefore becomes more reliable. Extensive reviews of inverse models in geohydrology are given by Cooley (1985) and Yeh (1986), among others.

Nowadays, these models consist of large model networks that describe reality in more and more detail, and as a consequence, the computational demands are increased. The time efficiency can be roughly increased by: (a) using a more efficient solver (e.g. Mehl & Hill, 2001), (b) applying a coarse grid and/or a locally refined grid (e.g. Wen *et al.,* 2003), and/or (c) formulating a reduced model (e.g. Vermeulen *et al.*, 2004a,b; 2005a,b,c). In this paper we estimate parameters by means of an extended version of a reduced model via a Galerkin projection.

METHODOLOGY

A reduced model is based upon the assumption that the spatial distribution of the hydraulic head is very complex, contrary to its behaviour in time. In this situation, the hydraulic head can be *split* into a spatial and time-variant component, so:

$$h(x,t) = h'(x,t) + \sum_{j=1} p_j(x) r_j(t) \tag{1}$$

where h (m) is the hydraulic head, h' (m) is some kind of reference, r (m) is the reduced equivalent of h, and p_j is the jth pattern out of a set of patterns. Those patterns *span* a subspace of the original space, and describe the characteristic behaviour of the model, acquired by model evaluations (i.e. *snapshots*), e.g. Vermeulen *et al.* (2004a,b). Among others, these patterns are known as Empirical Orthogonal Functions (EOFs). The set of patterns are used to create a reduced model for which its application is limited to the variance described by the patterns, Fig. 1(a).

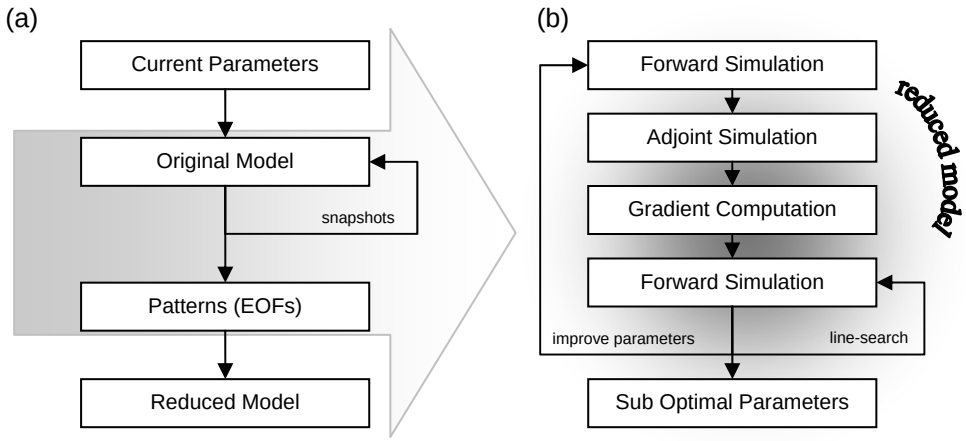

Fig. 1 Flow chart of the methodology: (a) Preparation phase that consists of simulating several snapshots by means of the original model that are described by patterns in order to form a reduced model. (b) The reduced model is then used to perform an optimization process that yields (sub) optimal parameters. The processes (a) and (b) reiterate until the innovation of the parameters is negligible.

The major advantage is that the optimization of the estimation parameters can be done with the reduced model only (Fig. 1(b)) and consumes a negligible amount of computation time. It yields an optimal set of parameters for the current reduced model that can be still sub optimal for the original model. Therefore, the sequence reiterates (Fig. 1(a) and (b)) until the maximal parameter update is less than a given abortion criterion.

REDUCED MODEL

Introduction

The term "reduced model" is here referred to as a dynamic model formulation that represents the original hydraulic head h by a reduced head r. This type of represent-tation has been applied in many different fields of science (Cazemier *et al.*, 1998; Park *et al.*, 1999) and in groundwater flow by Vermeulen *et al.* (2004a,b; 2005a,b). The

method can be concisely summarized by stating that the original partial differential equation (PDE) is substituted by equation (1) and the outcome is projected onto the set of patterns.

Patterns

A set of patterns **P** should reflect the model variance, but there is a trade-off as to what extent this variance should be represented. Since, the reduced model is used for inverse modelling, the patterns should describe the variance for the locations of observations with respect to the parameters, only. A common method is to simulate the original model for several samples of parameters (*Latin Hypercube Sampling* (Iman & Shortencarier, 1984)) and extract the hydraulic head *h* for the locations of observations. By storing them into a matrix **H**, the patterns are eventually equal to the eigenvectors of the covariance matrix $\mathbf{HH}^{\mathrm{T}}$. It should be noticed that the resulting reduced model will be only applicable to the simulation sequence, as used by the computation of **H**.

Reduced forward model

The Galerkin method substitutes the hydraulic head *h* in the original PDE by the linear equation (1). Since the *j*th estimation parameter α_j affects the transmissivity *T*, we add the first derivative of the hydraulic conductance *C* with respect to α_j, so:

$$\left\{\frac{\partial}{\partial x}\left[\mathbf{C}(\alpha')\frac{\partial \mathbf{P}}{\partial x}\right]+\sum_{j=1}\frac{\partial}{\partial x}\left[\frac{\partial \mathbf{C}(\alpha')}{\partial \alpha_j}\frac{\partial \mathbf{P}}{\partial x}\right]\Delta\alpha_j-\frac{1}{\Delta t}\mathbf{SP}-(\mathbf{C}_{\mathrm{d}})^{\varsigma}\right\}\mathbf{r}(t_i)= \\ =-\frac{1}{\Delta t}\mathbf{SPr}(t_{i-1})+\mathbf{q}(t_i)+(\mathbf{C}_{\mathrm{d}})^{\varsigma}\mathbf{z}(t_i) \tag{2}$$

where **C** is the harmonic hydraulic conductance (day^{-1}) and is a function of the current parameters α', **S** is a main diagonal matrix that contains the storage coefficient (–), vector **q** is the specified flux (m day^{-1}), $\mathbf{C}_{\mathrm{d}}$ represents the conductance (day^{-1}) of (non)-linear external forces (e.g. drainage conductance) and are therefore subject to an internal iteration cycle ζ, and **z** (m) is some reference level for external forces (e.g. drainage level). Equation (2) can be simplified by computing the second-order differentials of the pattern derivative of space in advance and multiplying each term with the patterns:

$$\left[\mathbf{N}^{\mathrm{r}}+\sum_{j=1}\mathbf{U}_{\mathrm{j}}^{\mathrm{r}}\Delta\alpha_j-\frac{1}{\Delta t}\mathbf{S}^{\mathrm{r}}-(\mathbf{C}_{\mathrm{d}}^{\mathrm{r}})^{\varsigma}\right]\mathbf{r}(t_i)=-\frac{1}{\Delta t}\mathbf{S}^{\mathrm{r}}\mathbf{r}(t_{i-1})+\mathbf{q}^{\mathrm{r}}(t_i)+(\mathbf{C}_{\mathrm{d}}^{\mathrm{r}})^{\varsigma}\mathbf{z}^{\mathrm{r}}(t_i) \tag{3}$$

where $(.)^{\mathrm{r}}=\mathbf{P}^{\mathrm{T}}(.)$ and denotes the reduced equivalent of the current matrix or vector $(.)$. The reduced model is initialized by $\mathbf{r}(t_0)=\mathbf{P}^{\mathrm{T}}\mathbf{h}(t_0)$ and $\Delta\alpha_j\neq 0.0$. During the simulation, the conductance $\mathbf{C}_{\mathrm{d}}$ needs to be recurrently evaluated for nonlinear elements that decrease the efficiency. However, high efficiencies are still achieved

(Vermeulen *et al.*, 2004a,b), since the dimensions of the reduced model (i.e. number of estimation parameters plus the number of patterns), is significantly less than the dimensions of the original model (i.e. number of nodes).

Reduced inverse model

There are many different ways to formulate the objective function and in this paper a general objective function J is defined that describes the weighted sum of squared residuals between a set of observations $\mathbf{y}^{\mathrm{o}}(t_1,t_n)$ and the computed observation $\mathbf{Pr}(t_1,t_n)$:

$$J[\alpha_j] = \sum_{i=1} \left[\mathbf{y}^{\mathrm{o}}(t_i) - \mathbf{Pr}(t_i)\right]^{\mathrm{T}} \mathbf{R}^{-1} \left[\mathbf{y}^{\mathrm{o}}(t_i) - \mathbf{Pr}(t_i)\right] \tag{4}$$

where $\mathbf{R}$ is the observational error covariance. Since the reduced model in equation (3) is quite simple, the tangent gradient can be obtained by Lagrange multipliers (Courant & Hilbert, 1953), yielding:

$$\frac{\Delta J}{\Delta \alpha_j} = \sum_{i=1} \left[\boldsymbol{\lambda}^{\mathrm{r}}(t_i)\right]^{\mathrm{T}} \mathbf{U}_j \mathbf{r}(t_i) \tag{5}$$

where λ^{r} is the reduced adjoint state computed by a reduced adjoint model that needs to be solved backwards in time (Vermeulen *et al.*, 2005b):

$$\left[\mathbf{N}^{\mathrm{r}} + \sum_{j=1} \mathbf{U}_j^{\mathrm{r}} \Delta\alpha_j - \frac{1}{\Delta t}\mathbf{S}^{\mathrm{r}} - (\mathbf{C}_{\mathrm{d}}^{\mathrm{r}})^{\varsigma}\right]^{\mathrm{T}} \boldsymbol{\lambda}(t_{i-1}) = -\frac{1}{\Delta t}\mathbf{S}^{\mathrm{r}}\boldsymbol{\lambda}(t_i) - \mathbf{P}^{\mathrm{T}} 2\mathbf{R}^{-1}\left[\mathbf{y}^{\mathrm{o}}(t_i) - \mathbf{Pr}(t_i)\right] \tag{6}$$

Once the gradient of J is known, equation (3) is used again to explore the evaluation of J along the current gradient (i.e. *line-search*). From a renewed location along that gradient, equation (5) is used again to update the current gradient; see Fig. 1(b). As both models are low-dimensional, the minimization requires a negligible amount of time.

SYNTHETIC CASE

Introduction

The efficiency and performance of the reduced model depends on whether the guessed estimation parameter α_j^{g} is in range of its optimal value α_j^{t}. To obtain insight in to the robustness of the method, different samples of α_j^{g} were generated (Iman & Shortencarier, 1984) from which J was minimized. The quality of these optimizations is determined by the error criterion:

$$\varepsilon(\alpha^{\mathrm{g}}) = 100\% \cdot 1.0 - \sqrt{\frac{1}{n}\sum_{i=1}^{n} \begin{cases} \left[\alpha_j^{\mathrm{g}}\right]^2 \text{ if } \alpha_j^{\mathrm{g}} \geq 1 \\ \left[\alpha_j^{\mathrm{g}}\right]^{-2} \text{ if } \alpha_j^{\mathrm{g}} \geq 1 \end{cases}} \tag{7}$$

that gives a relative error between the initial parameters and the estimated parameters.

For reasons of comparison, we have also minimized those problems with the original model using gradients according to adjoints or finite differences.

Problem description

The synthetic case corresponds to a network of seven columns by seven rows with $\Delta x = \Delta y = 10$ m, $\mathbf{T} = 100$ m^2 day^{-1}, and $\mathbf{S} = 0.27$ for the entire model domain. The left side of the model was determined by Dirichlet boundaries ($h = 0.0$ m). Each node within the other area was subject to $\alpha_j T;\ j \in \{1,...,42\}$ and yielded 42 synthetic observations $y_j^o(t_1,t_n)$, so eventually, 42 parameters were optimized.

Two pumping wells were located at (3,4) and (6,5) that varied randomly $-200 < q_i(t_1,t_n) < 200$ m^3 day^{-1}. A nonlinear drainage element was active along the entire column 4 with $\mathbf{z}(t_1,t_n) = 0.0$ m and $\mathbf{C}_d = 10$ m^2 day^{-1}. The system was simulated for 10 time steps with $\Delta t = 10$ days.

Results

For relatively simple disturbances of α_j^g, both the reduced model and the original models (gradient according to adjoints or finite differences) estimated the correct parameters (Fig. 2). However, the original model failed as the initial relative error was increased ($\varepsilon > 500$–1000%). Of course, this conclusion is rather specific for the application considered, but the reduced model always converged and was shown to be more robust.

The final efficiency of the reduced model is expressed by the number of original model evaluations. As expected, the efficiency for all models decreases whenever the

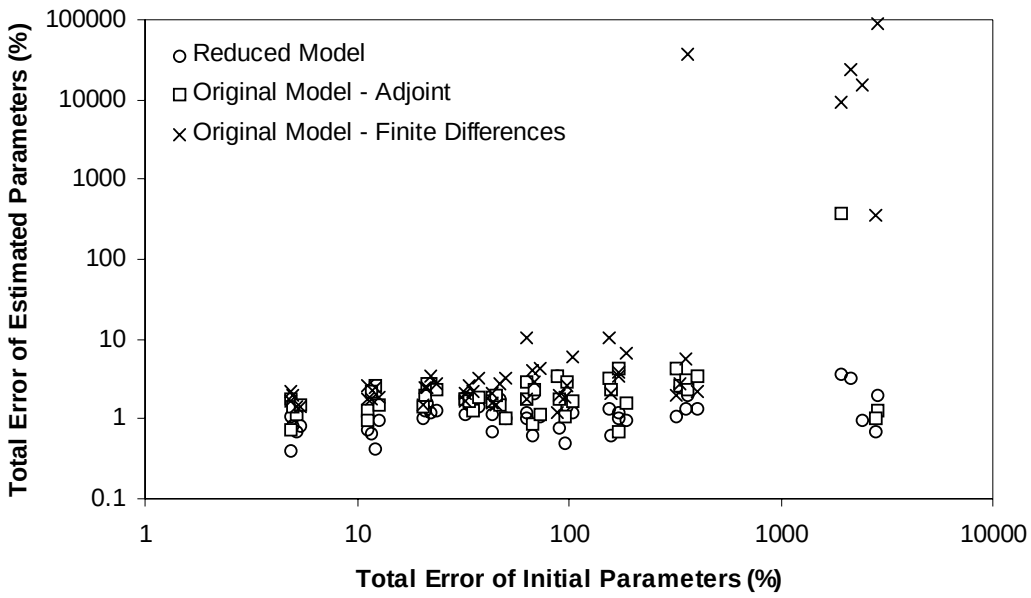

Fig. 2 Accuracy of the final estimation after optimizations by three different types of model for different initial parameters.

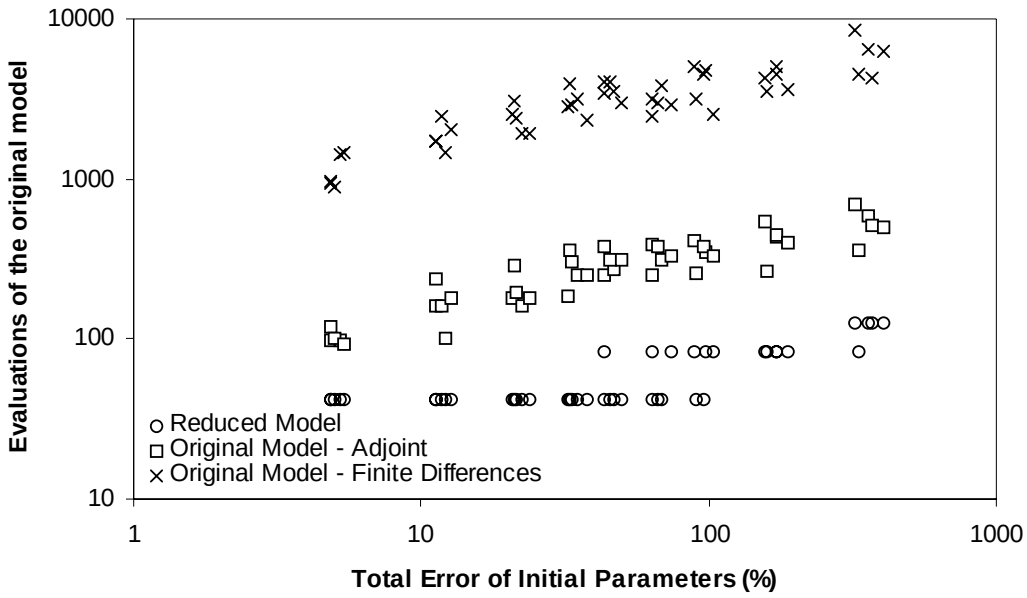

Fig. 3 Efficiency, expressed by the number of original evaluations, for three different types of models for different initial parameters.

relative initial error increases (Fig. 3). Moreover, the figure shows that the reduced model is more efficient than the adjoint method and even one order of magnitude more efficient than the finite difference method. Since this method is widely applied, it is a significant improvement.

CONCLUSIONS

This paper describes a parameter estimation procedure for numerical groundwater flow that was more efficient than classical optimization algorithms according to adjoints or finite differences. The procedure used a reduced model that was obtained by means of projection of the original PDE upon a set of patterns. These patterns were determined by model evaluations that described the relevant model variance with respect to the estimation parameters.

The algorithm was tested on a synthetic nonlinear case whereby the classical methods failed as the 42 initial parameters were significantly distorted. The reduced model converged more robustly and its final efficiency was higher than the efficiency of the original model using adjoints. Moreover, the reduced model was approximately one order of magnitude more efficient than the original model using the widely applied method of finite-differences. Even higher efficiencies are expected for more high dimensional problems and more complex models can be optimized using a more generic procedure (Vermeulen & Heemink, 2005c).

REFERENCES

Cazemier, W., Verstappen, R. W. C. P. & Veldman, A. E. P. (1998) Proper orthogonal decomposition and low-dimensional models for driven cavity flows. *Physics of Fluids* **10**, 1685–1699.

Cooley, R. L. (1985) A comparison of several methods of solving nonlinear regression groundwater flow problems. *Water Resour. Res.* **21**(10), 1525–1538.

Courant, R. & Hilbert, D. (1953) *Methods of Mathematical Physics.* Wiley Interscience, New York, USA.

Iman, R. L. & Shortencarier, M. J. (1984) A Fortran 77 program and user's guide for the generation of Latin Hypercube and random samples for use with computer models. *NUREG/CR-3624, Technical Report SAND83-2365.* Sandia National Laboratories, Albuquerque, New Mexico.

Mehl, S. W. & Hill, M. C. (2001) MODFLOW-2000, the U.S. Geological Survey modular ground-water model. User guide to the Link-AMG (LMG) package for solving matrix equations using an algebraic multigrid solver. *US Geol. Survey Open-File Report 01-177.*

Park, H. M., Chung, O. Y. & Lee, J. H. (1999) On the solution of inverse heat transfer problem using the Karhunen-Loève Galerkin method. *Int. J. Heat and Mass Transfer* **42**, 127–142.

Tarantola, A. (1987) *Inverse Problem Theory* (2nd edn). Elsevier, Amsterdam, The Netherlands.

Vermeulen, P. T. M., Heemink, A. W. & te Stroet, C. B. M. (2004a) Low-dimensional modeling of numerical groundwater flow. *Hydrol. Process.* **18**, 1487–1504.

Vermeulen, P. T. M., Heemink, A. W. & te Stroet, C. B. M. (2004b) Reduced models for linear groundwater flow models using empirical orthogonal functions. *Adv. Water Res.* **27**(1), 57–69.

Vermeulen, P. T. M., Heemink, A. W. & Valstar, J. R. (2005a) Inverse modeling of groundwater flow using model reduction. *Water Resour. Res.* **41**, W06003, doi:10.1029/2004WR003698.

Vermeulen, P. T. M. & Heemink, A. W. (2005b) Efficient inverse modelling of transient non-linear groundwater flow using model reduction. *Water Resour. Res.* (submitted).

Vermeulen, P. T. M. & Heemink, A. W. (2005c) Model-reduced variational data assimilation. *Month. Weath. Rev.* (accepted November 2005).

Wen, X. H., Durlofsky, L. J. & Edwards, M. G. (2003) Use of border regions for improved permeability upscaling. *Math. Geology* **35**, 521–547.

Yeh, W. W. G. (1986) Review of parameter identification procedures in groundwater hydrology: the inverse problem. *Water Resour. Res.* **22**(2), 95–108.