Model-Reduced Variational Data Assimilation

P. T. M. VERMEULEN

Netherlands Institute of Applied Geoscience TNO–National Geological Survey, Utrecht, Netherlands

A. W. HEEMINK

Delft University of Technology, Delft, Netherlands

(Manuscript received 3 August 2005, in final form 29 November 2005)

ABSTRACT

This paper describes a new approach to variational data assimilation that with a comparable computational efficiency does not require implementation of the adjoint of the tangent linear approximation of the original model. In classical variational data assimilation, the adjoint implementation is used to efficiently compute the gradient of the criterion to be minimized. Our approach is based on model reduction. Using an ensemble of forward model simulations, the leading EOFs are determined to define a subspace. The reduced model is created by projecting the original model onto this subspace. Once this reduced model is available, its adjoint can be implemented very easily and can be used to approximate the gradient of the criterion. The minimization process can now be solved completely in reduced space with negligible computational costs. If necessary, the procedure can be repeated a few times by generating new ensembles closer to the most recent estimate of the parameters. The reduced-model-based method has been tested on several nonlinear synthetic cases for which a diffusion coefficient was estimated.

1. Introduction

Data assimilation is a methodology to combine the results of a large-scale numerical model with the measurements available in order to obtain an optimal reconstruction of the state of the system. It can be applied both for state estimation and predicted problems, and for model calibration. In this paper we concentrate our attention on calibration problems: the estimate of a large number of constant model parameters from data.

Variational data assimilation or "the adjoint method" has been used very often for model calibration (e.g., Carrera and Neuman 1986; Lardner and Song 1995; Ullman and Wilson 1998; Heemink et al. 2002). The method is based on optimal control theory. Here, a number of unknown "control" parameters are introduced into the numerical model. Using the available data, these control parameters are identified by minimizing a certain cost function that measures the difference between the model results and the data. To obtain a computationally efficient procedure, the function is minimized with a gradient-based algorithm that determines the gradient by solving the adjoint problem.

Variational data assimilation requires implementation of the adjoint model. Even with the use of the adjoint compilers that have become recently available (Kaminski et al. 2003) this is a tremendous programming effort for existing model systems. This is a serious practical disadvantage that hampers new applications of the method.

A number of existing data assimilation schemes are based on Kalman filtering. The ensemble Kalman filter (Evensen 1994) is based on a representation of the probability density of the state estimate by a finite number of randomly generated states. The scheme is very easy to implement and does not require the adjoint implementation of the model. The Kalman filter can also be used for parameter estimation problems by augmenting the state vector with the uncertain parameters (Jazwinski 1970). Generally this approach performs well in case of noise parameters that slowly vary in time (and, therefore, act like states). For constant parameters not affected by noise, however, the filter is biased (Ljung 1979). It is known (Jazwinski 1970) that even for simple linear systems the Kalman filter tends to diverge

Corresponding author address: P. T. M. Vermeulen, Netherlands Institute of Applied Geoscience TNO–National Geological Survey, P.O. Box 80015, 3508 TA Utrecht, Netherlands.
E-mail: peter.vermeulen@tno.nl

from the optimal results when the system is not controllable (i.e., when not all states are affected by noise, which is the case for constant parameters).

In this paper we introduce a new approach to variational data assimilation having a comparable computational efficiency that does not require implementation of the adjoint of the original model. Our approach is based on model reduction. An ensemble of forward model simulations is used to determine an approximation of the covariance matrix of the model variability. A limited number of leading eigenvectors (empirical orthogonal functions) of this matrix are selected to define a model subspace. By projecting the original model onto this subspace, an approximate linear model is obtained. Once this reduced model is available, its adjoint can be implemented very easily and the minimization process can be solved completely in reduced space with negligible computational costs. If necessary, the procedure can be repeated a few times by generating new ensembles closer to the most recent estimate of the parameters.

The computational costs of the reduced model approach are dominated by the generation of the ensemble of forward model simulations. To obtain an accurate reduced model, however, often a significantly smaller simulation period can be chosen for the ensemble of model simulations than that used in solving the optimization problem. For the original model, the period chosen for the simulations is related to the time scale of this model, while that chosen for the reduced model in the optimization problem is the period over which data is available.

Model-reduced variational data assimilation is computationally very efficient if the number of uncertain control parameters is much smaller than the dimension of the state. Then the controllable subspace is very likely to be of low rank and an accurate reduced model can indeed be obtained. An efficient implementation can also be obtained when the number of data locations is much smaller than the dimension of the state. Then the observable subspace (i.e., the one that can be observed from the data) is very likely to be of low rank and as a result an accurate reduced model can be obtained that is able to reproduce the model results at least in the data locations.

Model-reduced variational data assimilation is more robust than the classical approach. While the adjoint of the tangent linear approximation of the original model produces the exact local gradient, the model reduction approach is based on a statistically linearized model and tends to produce a spatially averaged gradient. As a result, the model reduced approach is less sensitive to local minima.

Our model reduced approach was inspired by the work done in reduced order control, (e.g., Ravindran 2002; Antoulas et al. 2004; Atwell and King 2004; van Doren et al. 2006). In control problems, an additive low-dimensional control input is introduced into the model and identified by minimizing a given performance index. We have modified this approach for large-scale parameter estimation problems, which generally result in highly nonlinear optimization problems.

In this paper we first summarize briefly in section 2 the classical inverse modeling methods (finite-difference method and adjoint method). In section 3 we describe in detail our model reduced variational data assimilation approach. A number of case studies are discussed in section 4. Here, too, the spatially varying diffusion coefficient in a diffusion model is identified. We finally present a number of conclusions and some ideas for further improvement of the method in section 5. Throughout the manuscript we used the unified notations as proposed for atmospheric and oceanic modeling (Ide et al. 1997).

2. Inverse modeling

A discrete model for the evolution of an atmospheric, oceanic, or coupled system from time t_i to time t_{i+1} is governed by the equation

$$\mathbf{X}^f(t_{i+1}) = \mathbf{M}_{i+1}[\mathbf{X}^f(t_i)], \tag{1}$$

$$\mathbf{Y}(t_i) = \mathbf{H}\mathbf{X}^f(t_i), \tag{2}$$

where $\mathbf{X}^f$ is the forecast state vector, $\mathbf{M}$ is the dynamics operator, $\mathbf{H}$ is an operator that maps the model field to observation space, and $\mathbf{Y}$ is the estimated observation. In many situations, the forecast becomes more reliable when a set of model parameters within $\mathbf{M}$ can be found that reduces the misfit between a set of observations $\mathbf{Y}^o$ and its estimated observation value $\mathbf{Y}$. A quantity that determines this difference is, for example, the weighted sum of squared residuals:

$$J[\alpha_j] = \sum_{i=1} [\mathbf{Y}^o - \mathbf{Y}]^{\mathrm{T}}\mathbf{W}[\mathbf{Y}^o - \mathbf{Y}](t_i), \tag{3}$$

where $\mathbf{W}$ is the observational weighting and J is the cost function value that is minimized with respect to a variable α_j. The presented discrete model (1) can be rewritten as

$$\mathbf{X}^f(t_{i+1}) = \mathbf{M}_{i+1}[\mathbf{X}^f(t_i), \alpha_j]. \tag{4}$$

There are several ways to find an optimal value for α_j such that J is minimal (Press et al. 1992; Tarantola 1987). Several techniques compute the gradient of J

(denoted as ∇J) as a finite-difference approximation of the sensitivity of J subject to a perturbation of an estimation variable α_j, so

$$\nabla J_j \equiv \frac{\Delta J}{\Delta \alpha_j} \approx \sum_{i=1} \frac{J\{\mathbf{M}_{i+1}[\mathbf{X}^f(t_i), \Delta\alpha_j]\} - J\{\mathbf{M}_{i+1}[\mathbf{X}^f(t_i)]\}}{\Delta \alpha_j}. \tag{5}$$

A disadvantage of this method is that for each perturbed variable it requires a simulation of the model, and an additional simulation for a reference simulation. From a computational point of view, the method is therefore restricted to a limited number of variables.

The adjoint or variational method (Courant and Hilbert 1953) is more efficient as it simultaneously finds the gradient of the cost function with respect to all variables:

$$\nabla J_j \equiv \frac{\Delta J}{\Delta \alpha_j} = \sum_{i=1} - [\boldsymbol{\lambda}(t_{i+1})]^{\mathrm{T}} \left\{ \frac{\partial \mathbf{M}_{i+1}[\mathbf{X}^f(t_{i+1})]}{\partial \alpha_j} \right\}, \tag{6}$$

where $\boldsymbol{\lambda}(t_n, t_1)$ is the solution of the adjoint model. It needs one forward simulation and a second additional simulation backward in time with the adjoint model. It is independent of the number of variables to be estimated. However, implementation of the adjoint method is very complicated. Besides that, the minimization can fail as the cost function is not strictly convex and the exact gradient is sensitive to local minima.

3. Reduced model methodology

a. Introduction

The model-reduced method used in this paper falls into the category of spectral methods. These methods describe the original state as a truncated series of known basis functions and independent coefficients (i.e., the Galerkin method). Briefly, this type of model is based upon the simulation of specific model configurations [i.e., the method of *snapshots* Sirovich 1987)]. By applying functions that describe the variance in these snapshots, the superfluous state space dimensions can be eliminated. These functions are widely known as proper orthogonal decompositions (PODs) and empirical orthogonal functions (EOFs). Eventually, the model operates within the remaining dimensions $\mathbb{R}^n$ where $n \ll O$, and therefore consumes significantly less time. This type of model has been extensively mathematically described (Newman 1996) and successfully applied within different fields of science, among others, the fields of turbulence and image processing (Sirovich 1987), rapid thermal chemical vapor deposition (Adomaitis 1995), fluid dynamics (Hoffmann Jørgensen and Sørensen 2000), and groundwater flow (Vermeulen et al. 2004). This concept was extended for a groundwater application to handle an inverse modeling problem (Vermeulen et al. 2005). However, the method was restricted to a single model system matrix that needs to be explicitly available. In this paper a quite generic approach is conducted that does not need this constraint.

b. Methodology

A disadvantage of the classical inverse methods is that they need to evaluate the original model for each gradient computation; see Fig. 1a. As a result, a lot of effort is put into single gradient estimation. In contrast, we propose to exploit more of the information contained in the model simulation by creating an approximate linear reduced model. The flowchart (Fig. 1b) shows that the original model is used to generate several snapshots. These snapshots form the basis of a reduced model for which it is easy to implement an adjoint algorithm. Eventually, this model is used to improve the parameters by an inner loop η until a minimal cost function value is found. This is optimal for the linearized reduced model but can be suboptimal for the original model. The outer loop ζ is entered again and the reduced model is adjusted to reflect the updated parameters. This sequence continues and converges when the maximal innovation of the parameters between two adjacent ζ cycles is less than a given terminal criterion.

c. Reduced model structure

We start by linearizing the nonlinear high-order model (1) with respect to the variable α_j:

$$\overline{\mathbf{X}}^f(t_{i+1}) = \mathbf{M}_{i+1}[\mathbf{X}^b(t_i)]\overline{\mathbf{X}}^f(t_i) + \sum_{j=1} \frac{\partial \mathbf{M}_{i+1}[\mathbf{X}^b(t_i)]}{\partial \alpha_j} \Delta\alpha_j, \tag{7}$$

where $\overline{\mathbf{X}}^f$ is the linearized forecast state, and $\mathbf{X}^b$ is the background state for which a specific estimation vector $\boldsymbol{\alpha}^b$ is linearized. Assuming that the function $\partial\mathbf{M}/\partial\alpha_j$ can be computed analytically, a comparable model is constructed in reduced resolution for the incremental four-dimensional variational data assimilation (4DVAR) (Lawless et al. 2005). However, for the majority of simulation models, the partial differential is not known explicitly, and the function can be approximated by

$$\frac{\partial \mathbf{M}_{i+1}}{\partial \alpha_j} \approx \frac{\mathbf{M}_{i+1}[\mathbf{X}^b(t_i), \Delta\alpha_j] - \mathbf{M}_{i+1}[\mathbf{X}^b(t_i)]}{\Delta \alpha_j}. \tag{8}$$

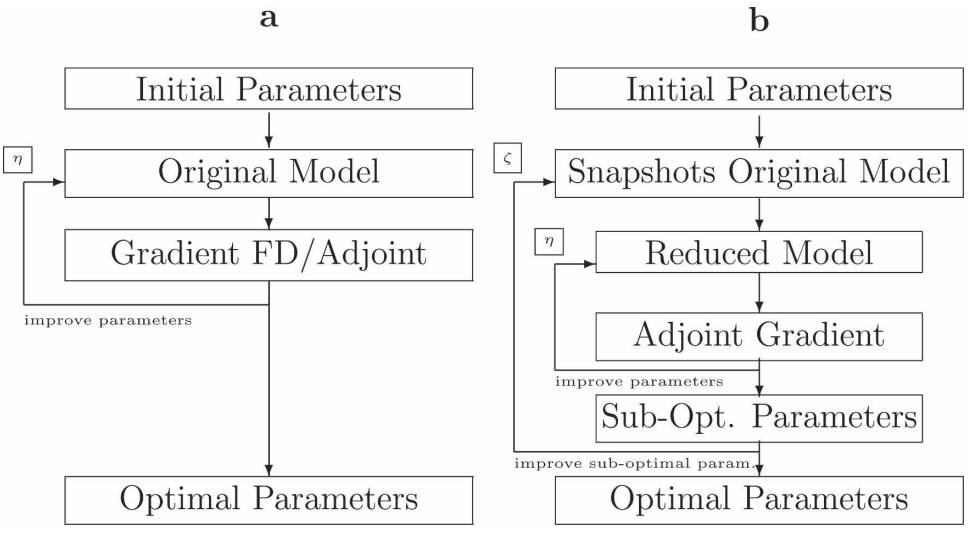

FIG. 1. Flowcharts of the methodology of inverse modeling using (a) the classic methods and (b) a reduced model method. The number of gradient iterations is denoted by η and outer iteration cycles by ζ.

A model can be reduced whenever the state $\overline{\mathbf{X}}^{\mathrm{f}}$ can be reconstructed as the linear combination

$$\hat{\overline{\mathbf{X}}}^{f}(t_i) = \mathbf{X}^{b}(t_i) + \mathbf{P}\mathbf{R}(t_i), \tag{9}$$

where $\hat{\overline{\mathbf{X}}}$ is the approximate linearized state, $\mathbf{P}$ is a set of basis vectors that span a subspace $\mathcal{S}$ of the state space $\mathcal{X}$, and $\mathbf{R}$ is a reduced time-varying state that is computed by

$$\mathbf{R}(t_{i+1}) = \mathbf{N}_{i+1}[\mathbf{X}^{b}(t_i)]\mathbf{R}(t_i) + \sum_{j=1} \frac{\partial \mathbf{N}_{i+1}}{\partial \alpha_j} \Delta\alpha_j, \tag{10}$$

where $\mathbf{N}$ is a reduced dynamics operator. This model requires less computational time, as it simulates a reduced state within the dimensions of a subspace $\mathcal{S} \in \mathbb{R}^n$ of the original state space $\mathcal{X} \in \mathbb{R}^O$ with $n \ll O$. Note that $\mathbf{R}$ represents a perturbation of α_j, and (10) is therefore referred to as a reduced perturbation model since $\mathbf{R}$ remains zero as $\Delta\alpha_j = 0.0$. In the following subsections the sets of basis vectors $\mathbf{P}$ (i.e., *patterns*) and $\mathbf{N}$ are elaborated.

d. Pattern identification

Pattern identification consists of computing a subspace $\mathcal{S}$ for an original model space $\mathcal{X}$. Among others, this technique is known as POD or EOFs. These basis vectors (functions) are the eigenvectors of a covariance matrix constructed from several empirical snapshots that are acquired by means of original model simulations. The basis is optimal as it has a smaller mean square error than any other basis of the same dimensions (Holmes et al. 1996). Eventually, the reduced model is limited to the subspace spanned by the basis.

The application considered in this paper prescribes a reduced model that includes a dependency of α_j. Hence, the snapshots should provide insight in the sensitivity of $\mathbf{X}^f \in \mathcal{X}$ with respect to $\alpha_j \in \mathcal{A}$. The projection matrix $\mathbf{A} = \{\mathbf{A}_1, \ldots, \mathbf{A}_n\}$ spans the parameter space $\mathcal{A}$ and maps each estimation variable to the model field. As in the classic finite-difference method (5), each sensitivity subject to α_j is computed along the corresponding definition $\mathbf{A}_j$. However, this strategy is disadvantageous for the quality of the basis vector as the algorithm for this is sensitive to any correlation in the snapshots (Cazemier et al. 1998). This means that whenever a perturbation of a specific parameter occurs, it influences a limited subdomain of the entire model domain. To avoid this, the snapshots should reflect the influence of all parameters simultaneously (Vermeulen et al. 2005). A practical method for this is to apply a coordinate transformation by an orthonormal projection matrix $\mathbf{B} = \{\mathbf{B}_1, \ldots, \mathbf{B}_n\}$ that yields

$$\Delta\beta = \mathbf{B}^{\mathrm{T}}\mathbf{A}\Delta\alpha. \tag{11}$$

Instead of perturbing along the original definition of each variable $\mathbf{A}_j\Delta\alpha_j$ (8), we perturb along the vectors $\mathbf{B}_j$ by $\Delta\beta_j$. So a single snapshot vector $\mathbf{S}$ becomes

$$\mathbf{S}_j(t_{i+1}) = \frac{\partial \mathbf{M}_{i+1}}{\partial \beta_j} \approx \frac{\mathbf{M}_{i+1}[\mathbf{X}^{b}(t_i), \Delta\beta_j\mathbf{B}_j] - \mathbf{M}_{i+1}[\mathbf{X}^{b}(t_i)]}{\Delta\beta_j}. \tag{12}$$

The results of (12) are stored within $\mathbf{S}_j(t_1, t_n)$ and each vector is scaled such that $\|\mathbf{S}_j(t_1)\| = 1$ (i.e., normalizing) since the amplitude within each vector is less important than its variation (Newman 1996). The collection of snapshot vectors are then collected in a matrix $\mathbf{S} = \{\mathbf{S}_1(t_1), \ldots, \mathbf{S}_n(t_n)\}$. Instead of an eigenvalue problem (Press et al. 1992) for the high-dimensional covariance matrix $\mathbf{SS}^{\mathrm{T}}$, we solve the following low-dimensional eigenvector problem (Krysl et al. 2001):

$$\mathbf{S}^{\mathrm{T}}\mathbf{S}\mathbf{V}_i = \lambda_i \mathbf{V}_i, \tag{13}$$

where $\mathbf{V}_i$ and λ_i are the ith eigenvector and eigenvalue, respectively. Each eigenvalue can be scaled according to

$$\varphi_i = \lambda_i/\mathrm{sum}(\lambda) \times 100\% \tag{14}$$

to express the relative amount of variance φ_i that the corresponding eigenvector represents. For each selected eigenvalue, the corresponding basis vector $\mathbf{P}_i$ can be obtained by applying (Golub and van Loan 1989)

$$\mathbf{P}_i = \mathbf{S}\mathbf{V}_i\lambda_i^{-1/2}, \tag{15}$$

and the selected basis becomes $\mathbf{P} = \{\mathbf{P}_1, \ldots, \mathbf{P}_n\}$ subject to orthonormality. The dimension of $\mathbf{P}$ depends on the desired accuracy of the reduced model and is expressed by the expected variance $\varphi^{\mathrm{e}} = \Sigma_{i=1}\, \varphi_i$. Since the highest accuracy of a reduced model is achieved by $\varphi^e = 100.0\%$, there are realistic situations where the time efficiency increases by $\varphi^e = 99.9\%$ without any decrease in accuracy. This particular aspect depends on the distribution of φ_i (e.g., many patterns that describe a very small amount of variance can be eliminated without affecting the quality of the reduced model) and is therefore subject to the application considered. Moreover, small eigenvalues are contaminated with roundoff errors and "describe" a numerical space that is unreliable (e.g., $\varphi_i \ll 0.001\%$).

e. Generic reduced model formulation

A reduced model operates within the subspace $\mathcal{S}$ that is described by $\mathbf{P}$. The reduced model formulation is initially presented in (10) and can be written as a partitioned matrix as

$$\begin{bmatrix} \mathbf{R} \\ \Delta\beta \end{bmatrix}(t_{i+1}) = \begin{bmatrix} \mathbf{N} & \mathbf{N}^{\beta} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}_{i+1} \begin{bmatrix} \mathbf{R} \\ \Delta\beta \end{bmatrix}(t_i), \tag{16}$$

where $\Delta\beta$ is the projected prior estimate (11); $\mathbf{N}$ and $\mathbf{N}^{\beta}$ are reduced dynamics operators that can be computed by a linear transformation (i.e., a change of coordinates) of $\mathbf{M}$ and $\partial\mathbf{M}/\partial\beta_j$ (12) upon a set of basis vectors $\mathbf{P}$, so

$$\mathbf{N}_{i+1} = \mathbf{P}^{\mathrm{T}}\mathbf{M}_{i+1}\mathbf{P}, \tag{17}$$

$$\mathbf{N}^{\beta}_{i+1} = \mathbf{P}^{\mathrm{T}}\left\{\frac{\partial\mathbf{M}_{i+1}}{\partial\beta_1}, \ldots, \frac{\partial\mathbf{M}_{i+1}}{\partial\beta_n}\right\}. \tag{18}$$

For these applications, where the operator $\mathbf{M}_{i+1}$ is not explicitly available, the reduced operator $\mathbf{N}_{i+1}$ (17) can be written as

$$\mathbf{N}_{i+1} = \mathbf{P}^{\mathrm{T}}\left\{\frac{\partial\mathbf{M}_{i+1}}{\partial\mathbf{X}(t_{i+1})}\mathbf{P}_1, \ldots, \frac{\partial\mathbf{M}_{i+1}}{\partial\mathbf{X}(t_{i+1})}\mathbf{P}_n\right\}, \tag{19}$$

for which each column j can be solved by a finite difference approximation

$$\frac{\partial\mathbf{M}_{i+1}}{\partial\mathbf{X}(t_{i+1})}\mathbf{P}_j \approx \frac{\mathbf{M}_{i+1}[\epsilon\mathbf{P}_j, \mathbf{X}^b(t_{i+1})] - \mathbf{M}_{i+1}[\mathbf{X}^b(t_{i+1})]}{\epsilon}, \tag{20}$$

where ϵ is the interval for which the partial differential equation is linearized. It approximates the product $\mathbf{M}_{i+1}\mathbf{P}$ by slightly perturbing the function $\mathbf{M}_{i+1}$ in the direction of $\epsilon\mathbf{P}$. Experimentation showed that large values for ϵ could improve the reduced model; however, it could easily worsen the model too. Most robust results were obtained for $\epsilon = 0.01$. The dimension of the reduced model depends eventually on the number n of estimation variables and the number m of basis vectors in $\mathbf{P}$. Since the reduced model is autoregressive and operates in $\mathbb{R}^{m+n}$, it needs significantly less computational time compared to the original model, which operates in $\mathbb{R}^{O}$.

f. Adjoint of the reduced model

The results of the simulations are used to minimize an approximate cost function for the reduced model, defined as

$$\hat{J}[\Delta\alpha_j] = \sum_{i=1} [\mathbf{P}^{\mathrm{T}}(\mathbf{Y}^o - \mathbf{H}\mathbf{X}^b) - \mathbf{R}]^{\mathrm{T}}\mathbf{P}^{\mathrm{T}}\mathbf{W}\mathbf{P}[\mathbf{P}^{\mathrm{T}}(\mathbf{Y}^o - \mathbf{H}\mathbf{X}^b) - \mathbf{R}](t_i), \tag{21}$$

where the observation value $\mathbf{Y}^o$ is projected on the basis $\mathbf{P}$ and needs to be corrected for the background state $\mathbf{X}^b$ that is projected on the observation space $\mathbf{H}$ since $\mathbf{R}$ yields a perturbation upon that same state (9). It seems illogical to optimize $\Delta\alpha_j$ instead of $\Delta\beta_j$. However, the main reasons for optimizing $\Delta\alpha_j$ are that it hardly affects the efficiency, distinguishes the method from oth-

ers more clearly, and allows direct comparison of the gradient to other algorithms.

The cost function (21) is minimized with respect to $\Delta\alpha_j$ by means of a reduced adjoint-based gradient algorithm

$$\nabla\hat{J}_j \equiv \frac{\Delta\hat{J}}{\Delta\alpha_j} = \sum_{i=1} [\boldsymbol{\lambda}^r(t_{i+1})]^{\mathrm{T}} - \left\{ \begin{bmatrix} \mathbf{N}^{\beta} \\ \mathbf{0} \end{bmatrix}_{i+1} \mathbf{B}^{\mathrm{T}} d(\mathbf{A}_j \alpha_j) \right\}, \tag{22}$$

where $\mathbf{A}_j$ is the jth vector in $\mathcal{A}$ that maps the jth estimation variable on the model field, $\mathbf{B}$ maps the estimation variable on the coordinate system that is affected by $\Delta\beta_j$, and $\boldsymbol{\lambda}^r(t_n, t_1)$ is the reduced adjoint state model. The reduced adjoint state is easy to implement owing to the linear character of the reduced model (elaborated in the appendix). Once the gradient is obtained, the reduced model (16) is used again to explore the evaluation of $\hat{J}$ along the direction of the gradient (i.e., a line search is carried out). From a new location along that direction, a different gradient is determined (22). As both models are low dimensional, the minimization acquires a negligible amount of simulation time.

g. Efficiency

The efficiency of the optimization process is mainly influenced by the following.

- *Snapshots*: The maximal number of snapshot simulations n is equal to the number of estimation variables times the number of time steps. However, it is difficult to give a minimal number since there is a tradeoff between the effort put in the snapshot simulations and the "$\hat{J}$-reduction capacity" of the reduced model evolved. For example, it is possible to start with half the maximal number of simulations such that the optimization procedure will first focus on the major directions and add more snapshots as the process proceeds.
- *Outer iteration*: The number of outer iterations ζ is influenced by the chosen abortion criterion γ. It should be chosen not too small ($\gamma \geq 0.1$) as this causes jumping of suboptimal posterior estimates around the optimal global solution. Since the reduced model probably overestimates α_j^t due to the applied linearizations, it probably yields a redundancy of ζ whenever a threshold ψ_j is applied. This threshold prevents the model from updating the variables to extraordinary values, far beyond the background estimates α_j^b that were used for the linearization.

The next section will illustrate that these "extra" computations will be regained eventually.

4. Synthetic cases

a. Introduction

Commonly, large-scale numerical models can be used to describe a particular transport through a certain medium. The generic time-dependent diffusion equation can be written as a partial differential equation for a discretized numerical scheme, so

$$\boldsymbol{\Phi}\frac{\partial \mathbf{X}}{\partial t} = \nabla^2 \boldsymbol{\kappa}\mathbf{X} + \mathbf{Q}(\mathbf{X}, t), \tag{23}$$

where $\boldsymbol{\kappa}$ [T^{-1}] contains the diffusion coefficient, $t \in [0, \infty]$ [T] is the time, $\boldsymbol{\Phi}$ [$-$] is a storage coefficient, and $\mathbf{Q}$ [$\mathrm{L\,T}^{-1}$] is a linear and/or nonlinear Neumann boundary condition that determines imposed fluxes. The operator $\nabla^2 = \Sigma_{i=1}^n \partial^2/\partial x_i^2$ and is the Laplacian for the current model domain $\Omega \in \mathbb{R}^n$. A solution to the partial differential equation can be obtained by imposing certain boundary conditions and applying a finite difference approximation for the space and time derivative of $\mathbf{X}$.

In the next subsections, several synthetic cases are described for the diffusion equation presented. The main purpose of this is to illustrate the methodology and to demonstrate its efficiency and robustness compared to classic optimization algorithms, that is, the method of finite differences (5) and the adjoint-based method (6). The parameter optimization for each case is computed by the variable metric method (quasi-Newton) (Press et al. 1992), which adjusts the acquired gradient by an approximated Hessian. The iteration cycle is eventually aborted whenever the maximal innovation of the parameters is

$$\gamma = \sum_j = \|[\alpha_j^b]^{\eta} - [\alpha_j^b]^{\eta-1}\| \leq 0.1. \tag{24}$$

Unless stated differently, the number of evaluations of the original model τ is used to compare the different methods. This includes the number of evaluations that were eventually needed for the line searches. Another quality criterion is the accuracy of the posterior estimate α_j^{a}; this is a relative error ε determined by

$$\varepsilon = 100\% \cdot \|\boldsymbol{\alpha}^t\| - \sqrt{\frac{1}{n}\sum_{j=1}^{n} \begin{cases} [\alpha_j^a]^2 & \text{if} \quad \alpha_j^a \geq 1 \\ [\alpha_j^a]^{-2} & \text{if} \quad 0 < \alpha_j^a < 1, \end{cases}} \tag{25}$$

where $\boldsymbol{\alpha}^t$ is the "truth" value for the estimation variable, which is equal to 1.0 for all cases.

b. Problem description

The synthetic cases in the following subsections increase in complexity with respect to the number n of

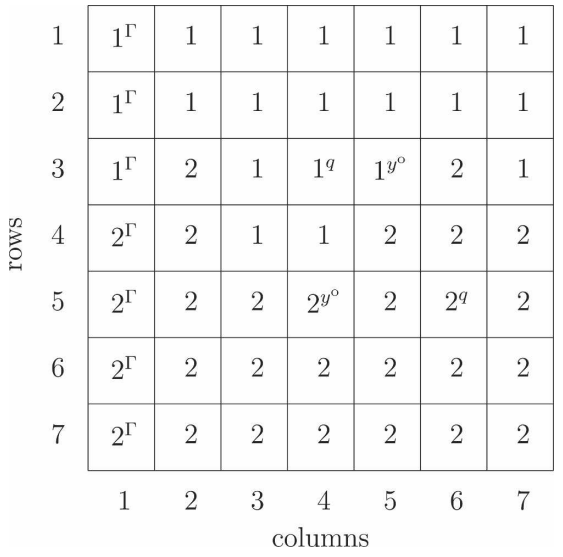

FIG. 2. Model network for Case I ($\mathcal{A} \in \mathbb{R}^2$); the location of linear Neumann boundary conditions $q \in \mathbf{Q}$, Dirichlet boundary conditions $\Gamma \in \mathbf{\Gamma}$, and the observations $y^o \in \mathbf{Y}^o$ are given. The definitions of the zones are given in the grid centers.

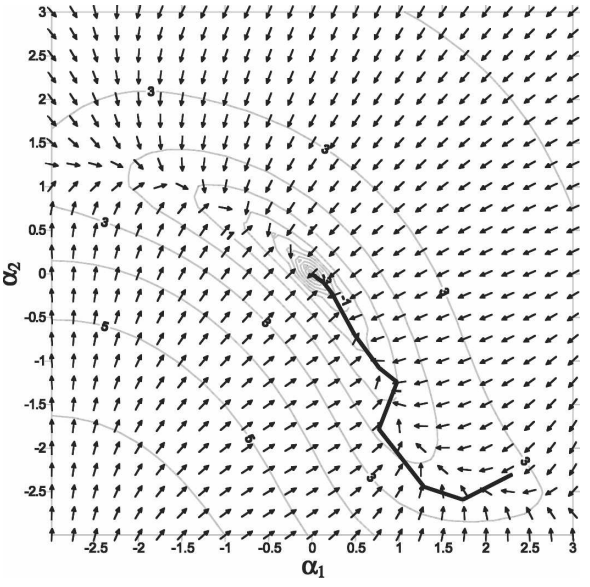

FIG. 3. Surface of J for $\log(\alpha_1)$ with respect to $\log(\alpha_2)$; the arrows represent the adjoint gradient ∇J. Surface of $\hat{J}$ for $\log(\alpha_1)$ with respect to $\log(\alpha_2)$; the arrows represent the gradient $\nabla \hat{J}$. The reduced model is linearized for $\alpha_1^b = 2.5$ and $\alpha_2^b = 0.3$ and explains $\varphi^e = 99.99\%$ model variance.

estimate variables $\mathcal{A} \in \mathbb{R}^n$, among others. The corresponding network consisted of seven columns and seven rows dimensioned by $\Delta x \equiv \Delta y = 10.0$ L, and $\boldsymbol{\kappa}$ and $\mathbf{\Phi}$ were kept at 1.0 T^{-1} and 0.27 for the entire model domain Ω, respectively. In each case, the estimate α_j affected the parameter $\boldsymbol{\kappa}$. Two linear Neumann boundaries were defined near the center of the model that varied randomly between $-2 \leq \mathbf{Q}(t_1, t_n) \leq 2$ L T^{-1}. The time-invariant states for the Dirichlet boundaries were $\mathbf{\Gamma}(t_1, t_n) = 0.0$ L and located at the entire left side

TABLE 1. (a)–(c) Results for case I for an optimization of two estimation variables α_1 and α_2 by using (a) the classic methods and (b), (c) a reduced model method. This is expressed by the number of gradient iterations η, number of outer iteration cycles ζ, number of simulations with the original model τ, relative estimation error ε, and the innovation γ. For each ζ cycle, $\hat{J}^b$ and $\hat{J}^a$ express the prior and posterior cost function value, respectively.

(a) Original model

η	τ	α_1	α_2	J^η	ε	γ	$\mathbb{R}^n$
0		2.50	0.30	0.4557	115.55		42
6*	26	1.011 22	0.992 25	0.3684×10^{-4}	0.1597	0.496×10^{-3}	
11**	32	0.999 88	1.000 22	0.2274×10^{-7}	0.4804×10^{-2}	0.6366×10^{-3}	

* Finite difference gradient [Eq. (5)]
** Adjoint gradient [Eq. (6)]

(b) Reduced model—$\varphi^e = 99.99\%$

ζ	τ	α_1	α_2	$\hat{J}^b$	$\hat{J}^a$	ε	γ	$\mathbb{R}^n$
1	7	2.50	0.30	0.3603	0.096 57	194.63	1.50	8
2	14	1.5259	0.8353	0.1215	0.5471×10^{-2}	37.14	0.974	8
3	21	0.9733	1.0373	0.4179×10^{-3}	0.7051×10^{-4}	3.24	0.5526	8
4	28	0.9996	0.9997	0.6218×10^{-6}	0.5704×10^{-7}	0.031	0.0375	8

(c) Reduced model—$90.39 \leq \varphi^e \leq 95.62\%$

ζ	τ	α_1	α_2	$\hat{J}^b$	$\hat{J}^a$	ε	γ	$\mathbb{R}^n$
1	4	2.50	0.30	0.3603	0.107 36	194.63	1.50	4
2	8	1.4689	0.9198	0.1523	0.011 25	29.27	1.031	4
3	12	1.0006	1.0028	0.1913×10^{-4}	$0.234\ 37 \times 10^{-5}$	0.1708	0.468	4
4	16	0.9992	1.0009	0.1996×10^{-6}	$0.187\ 53 \times 10^{-6}$	0.0873	0.001 91	4

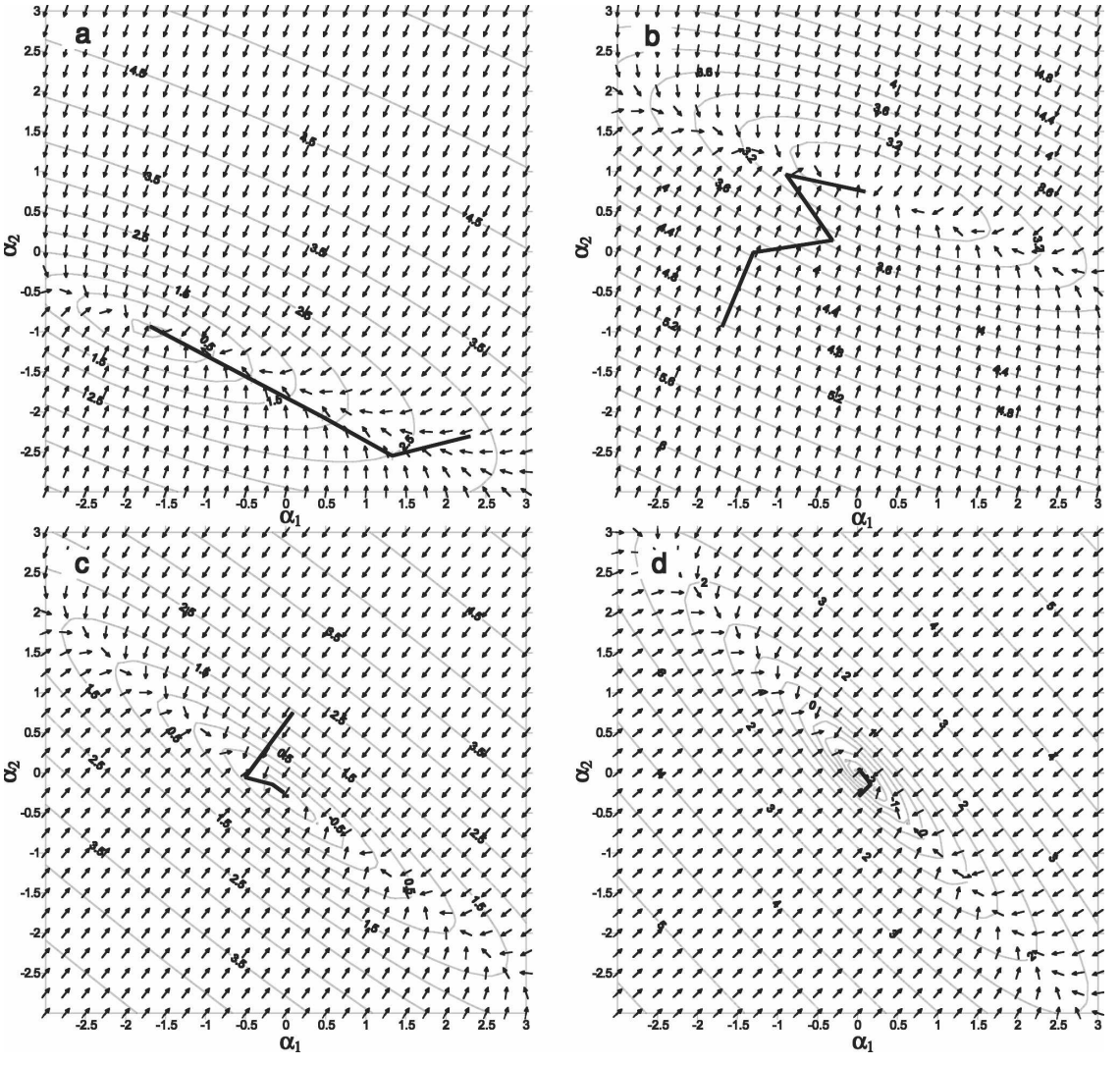

FIG. 4. Model network for case II ($\mathcal{A} \in \mathbb{R}^{42}$); the location of linear Neumann boundary conditions $q \in \mathbf{Q}$, nonlinear Neumann boundary conditions $\tilde{q} \in \tilde{\mathbf{Q}}$, Dirichlet boundary conditions $\Gamma \in \mathbf{\Gamma}$, and observations $y^o \in \mathbf{Y}^o$ are given. The definitions of the zones are given in the grid cell centers.

of the model domain. Each case was simulated for 10 time steps with $\Delta t_i = 10$ T. Finally, a set of synthetic measurements $\mathbf{Y}^o(t_1, t_n)$ for different locations was generated.

1) CASE I: $\mathcal{A} \in \mathbb{R}^2$

In this case, the diffusion Eq. (23) is used to estimate two variables that affect the parameter $\boldsymbol{\kappa}$ for two zones. Within each of these zones, a synthetic observation was created; see Fig. 2.

The cost function J is depicted in Fig. 3 for the range $0.1 \leq \log(\alpha_j) \leq 3.1$, where the optimal value is located at $\boldsymbol{\alpha}^t = 1.0$. In the situation where the prior estimates were assumed to be $\alpha_1 = 2.5$ and $\alpha_2 = 0.3$, this minimal value was eventually found after $\eta = 11$ iterations with the original adjoint method (6); see Table 1a.

To optimize the prior estimates with the reduced model, we constructed a model for $\alpha_1^b = 2.5$ and $\alpha_2^b = 0.3$. It yielded a reduced model with six patterns that describe $\varphi^e = 99.99\%$ of the model variance. Combined with two estimation variables, the reduced model operated eventually in $\mathbb{R}^{6+2}$ instead of $\mathbb{R}^{42\times 2}$. This reduced model decreased $\hat{J}^\eta$ significantly and found a new suboptimum at $[\alpha_1^a]^\eta = 1.53$ and $[\alpha_2^a]^\eta = 0.84$; see Fig. 4a. The process reentered the outer loop ζ (Fig. 1), and hence the reduced model and the surface of $\hat{J}$ changed. The suboptimum from the previous model, $\hat{J}^\eta$, did not appear to be an optimum for the updated model and after another minimization, a different suboptimum

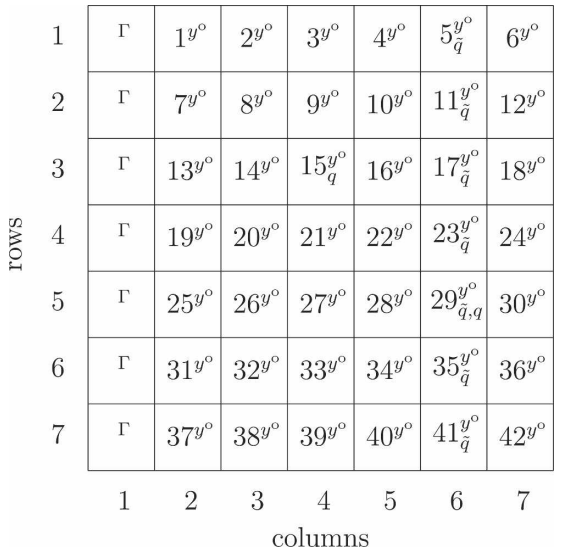

FIG. 5. Graph of the relative error ε of the prior estimate vs the posterior estimate, for the reduced model ($\varphi^e = 99.99\%$) as well as for two classic methods.

was found again at $[\alpha_1^a]^{\eta+1} = 0.97$ and $[\alpha_2^a]^{\eta+1} = 1.037$. This sequence reiterated, and after $\eta = 4$ cycles the optimal values for α_j were found; see Table 1b. Compared to the classic methods, almost an identical number of simulations with the reduced model were needed. For a different situation, where less variance was included ($90.39\% \leq \varphi^e \leq 95.62\%$), the reduced model still converged; see Table 1c. An explanation for this is that, for the considered case, the surface of $\hat{J}$ did not depend on the patterns that were eliminated. These patterns describe some model variance, irrelevant to the observations. This implies that further reduction is possible as patterns are computed that only capture the variance of observations. This is more or less related to the frequently applied "method of balancing" in optimal control theory (Lee et al. 2000).

2) CASE II: $\mathcal{A} \in \mathbb{R}^{42}$

The efficiency and performance of the optimization problem depends for the most part on the dimension and complexity of the parameter space $\mathcal{A}$. Furthermore, it relates to the identifiability of estimates and whether the prior estimate α_j is already near its optimal value α_j^t. To illustrate this, we applied a parameter optimization to a 42-dimensional parameter problem, so $\mathcal{A} \in \mathbb{R}^{42}$. For this case, each active cell (i.e., each grid cell without a Dirichlet boundary condition) possesses an observation $\mathbf{Y}^o$ and is part of an individual zone $\mathbf{A}_j$; see Fig. 5. Furthermore, several nonlinear boundary conditions $\tilde{\mathbf{Q}}$ are added that describe a flux that behaves nonlinearly to the state, so

$$\tilde{\mathbf{Q}}(\mathbf{X}, t) = \begin{cases} \mathbf{X} \cdot 0.1, & \text{if} \quad \mathbf{X} \leq 0.0 \\ 0.0, & \text{if} \quad \mathbf{X} > 0.0. \end{cases} \tag{26}$$

To obtain insight in the robustness of the reduced model, we generated different samples of prior estimates from which the cost function was minimized. A constrained Monte Carlo sampling scheme was used known as the Latin Hypercube Sampling (LHS) (McKay et al. 1979; Iman and Shortencarier 1984). This method divides the range of each α_j into n nonoverlapping intervals on the basis of equal probability. One value is randomly sampled from each interval with respect to the probability density in the interval. The values are then paired in a random manner with the other valued α. We sampled five times from 10 different lognormal distributions $\mathcal{N}^{\log}(\mu, \sigma)$, where $\mu = 1.0$ and $\sigma^2 \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.8, 0.9, 1.0, 10.0\}$. For those samples, the optimization results are depicted in Fig. 6.

For relatively simple disturbances of the prior estimate ($\varepsilon \leq 60\%$), each of the methods described in this paper succeeded in computing the correct posterior estimate value. Since the adjoint method yields an exact gradient, it is sensitive to local disturbances in the surface of J and this is the main reason of failure when $\varepsilon > 60\%$. Although the finite difference method is sensitive to this phenomenon too, its robustness can be artificially increased by decreasing the step size of the finite difference approximation while proceeding. Of course, these results are specific for the current application and can be rather different for other applications. Nevertheless, the reduced model converged fairly constantly and was not sensitive to the relative error ε of the prior estimate. The explanation for this result is twofold: 1) the applied linearization "flattens" the cost function and 2) the reduced model describes a part of the total model variance, and local disturbances in the cost function are mainly caused by minor and detailed variances in the model.

The efficiency of the reduced model is expressed by τ and represents the number of model simulations of the original model. These are all simulations that are required in order to find the mathematical formulation of the reduced model. As expected, τ increases with ε; see Fig. 7. It shows that for the reduced model τ is almost one order of magnitude less than for the finite difference method. Moreover, the reduced model offers an efficiency comparable to the adjoint method, without the total burden of implementation.

5. Conclusions

This paper describes a new approach to the calibration of complex numerical simulation models. The ma-

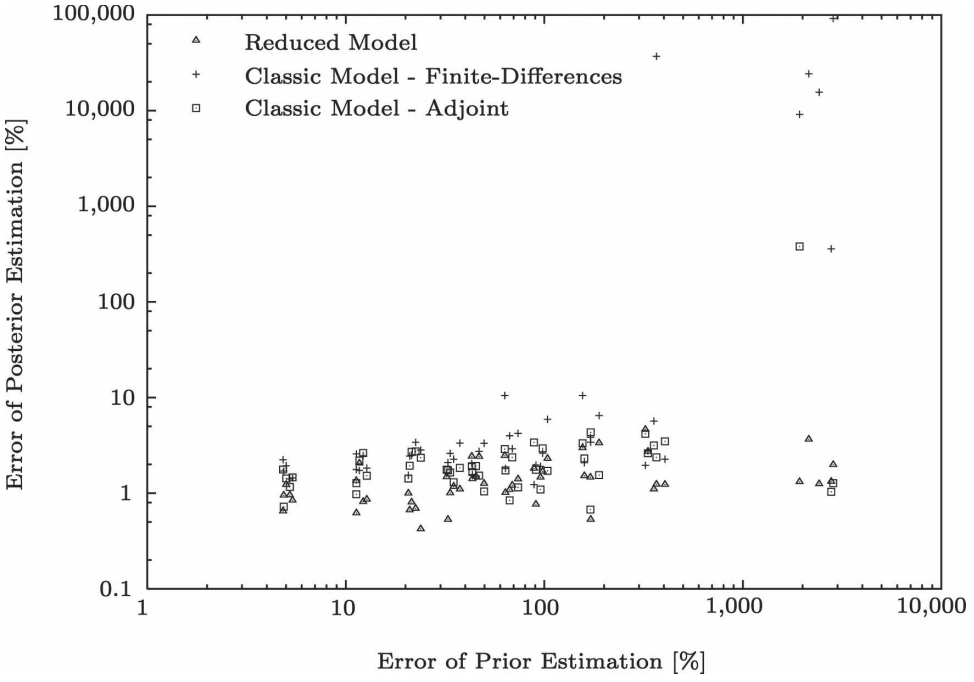

FIG. 6. Graph of the relative error ε of the prior estimate vs the number of original simulations (τ), for the reduced model ($\varphi^e = 99.99\%$) as well as for two classic methods.

jor advantages of the proposed concept are that it is generic in kind, easy to implement, applicable to complex models that consist of multiple systems, and very time efficient compared to current classic optimization sequences.

The concept is based on the computation of a reduced model using a number of simulations (i.e., snapshots) of the original model. Those snapshots describe the variance of the original model with respect to the dynamics and parameters. By selecting the main eigen-

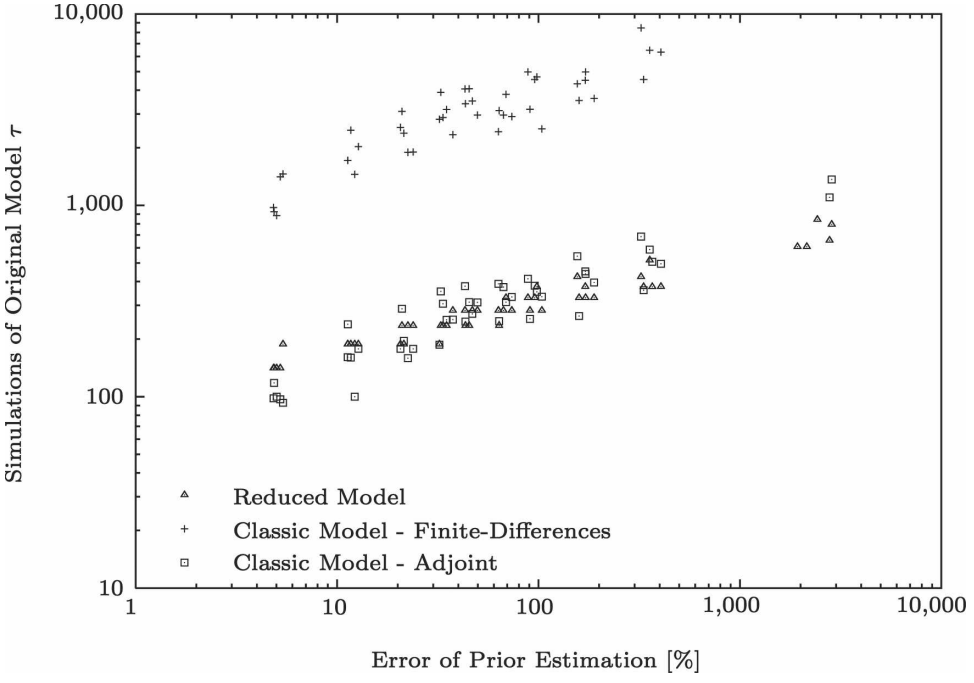

FIG. 7. Graph of the simulations of the original model vs error of the prior estimation for the reduced model and the classic model using finite differences and the adjoint.

vectors of the covariance of the snapshots, a reduced model is created that operates within a reduced space. It models the effect of perturbations of the parameters with a negligible amount of simulation time. As the adjoint algorithm for this reduced model is easy to implement, the reduced model approach is capable of finding the suboptimal set of parameters that yields a minimal cost function value for the reduced model. With these parameters, the sequence reiterates until the parameter values converge.

The algorithm was tested on several nonlinear synthetic cases where the diffusion coefficient has been estimated in a diffuse-type model. The results were compared with two widely used optimization algorithms: the method of finite differences and the adjoint method. From several optimization problems that increased in complexity, the reduced model always converged in contrast to the classic methods. These methods failed if the prior estimate was too distorted compared to its "truth" value. Moreover, the reduced model approach required more or less the same number of simulations of the original model as the adjoint method.

Since the results from the synthetical cases are promising, some issues can be addressed to improve the method's efficiency even more:

- Because the approximation of the Hessian is used to update the parameters within the inner iteration cycle η, a comparable algorithm can be applied to the outer iteration loop ζ.
- The accuracy of the reduced model is directly related to the step size and direction of the linearization (8). These could be exploited during the outer iteration by adding some intelligence.
- The robustness of the method is related to the "quality" of the snapshots that assume a linearity between state vector and estimated parameters. To ensure convergence of the algorithm we should prevent the model from updating the parameters to extraordinary values, far beyond the background estimations. However, some "violation" is part of a trade-off between efficiency and robustness.
- The patterns could be varied over time since this can reduce the overall number of patterns. On the other hand, time propagating phenomena (e.g., waves) are inefficiently described by POD (Hooimeijer 2001). For these cases, the POD should be time varying and during the reduced model simulation, the time-dependent coefficients should be transformed from the subspace described by a POD for time step i into the subspace of the adjacent POD for time step $i + 1$.
- To minimize the cost function, the reduced model needs to forecast the state accurately for only the observations. Hence, it is possible to construct a reduced model for only the observations [i.e., this will be more or less comparable to the "balancing" principle applied in optimal control applications (Lee et al. 2000)]. A major advantage in this is that the effort of the algebraic computations (i.e., the eigenvalue decomposition and linear transformations) is reduced, and eventually fewer patterns will be needed.

APPENDIX

Derivation of the Reduced Adjoint State Variable

In (22), the reduced adjoint state variable $\boldsymbol{\lambda}^{\mathrm{r}}_{i+1}$ is introduced in order to compute the gradient of the approximate cost function $\Delta\hat{\mathcal{J}}$ with respect to the estimate $\Delta\boldsymbol{\alpha}_j$, so

$$\hat{J}[\Delta\alpha_j] = \hat{J}[\Delta\alpha_j] + \sum_{i=1} [\boldsymbol{\lambda}^r(t_i)]^{\mathrm{T}}\{\mathbf{R}^p(t_{i+1}) - \mathbf{N}^p_{i+1}[\mathbf{R}^p(t_i)]\}, \tag{A1}$$

with the partioned matrix $\mathbf{N}^{\mathrm{p}}$ and vector $\mathbf{R}^{\mathrm{p}}$ defined as

$$\mathbf{N}^{\mathrm{p}}_{i+1} = \begin{bmatrix} \mathbf{N} & \mathbf{N}^{\beta} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}_{i+1}, \tag{A2}$$

$$\mathbf{R}^{\mathrm{p}}(t_i) = \begin{bmatrix} \mathbf{R}^f \\ \Delta\beta \end{bmatrix}(t_i), \tag{A3}$$

where $\Delta\beta$ represents a projected perturbation of the original perturbed estimation variables (11). After a Taylor series expansion for the variables, $\boldsymbol{\lambda}^{\mathrm{r}}(t_{i+1})$, $\mathbf{R}^{\mathrm{p}}(t_i)$, $\mathbf{R}^{\mathrm{p}}(t_{i+1})$, and $\Delta\alpha_j$, the expression yields

$$\Delta\hat{J} = \sum_{i=1} [\boldsymbol{\lambda}^r(t_{i+1})]^{\mathrm{T}}\left[-\mathbf{N}^p_{i+1}\Delta\mathbf{R}^p(t_i) + \Delta\mathbf{R}^p(t_{i+1}) + \frac{\partial\mathbf{R}(t_{i+1})}{\partial\Delta\alpha_j}\Delta\alpha_j \right] \tag{A4}$$

$$+ \left[\frac{\partial J}{\partial\mathbf{R}^p(t_{i+1})}\right]^{\mathrm{T}}\Delta\mathbf{R}^p(t_{i+1}). \tag{A5}$$

The reduced adjoint state can be solved for by reexpressing $\mathbf{R}^p(t_i)$ to be $\mathbf{R}^p(t_{i+1})$:

$$-[\boldsymbol{\lambda}^r(t_{i+1})]^{\mathrm{T}}\mathbf{N}_{i+1} + [\boldsymbol{\lambda}^r(t_i)]^{\mathrm{T}} = \left[\frac{\partial J}{\partial\mathbf{R}^p(t_{i+1})}\right]^{\mathrm{T}}, \tag{A6}$$

which yields the reduced adjoint model $\boldsymbol{\lambda}^{\mathrm{r}}(t_n, t_1)$:

$$\boldsymbol{\lambda}^r(t_i) = \mathbf{N}^{\mathrm{T}}_{i+1}\boldsymbol{\lambda}^r(t_{i+1}) + \mathbf{E}(t_i), \tag{A7}$$

where $\boldsymbol{\lambda}^{\mathrm{r}}(t_n) = 0.0$ and

$$\mathbf{E}(t_i) = \begin{bmatrix} \partial J/\partial\mathbf{R}^f(t_i) \\ \partial J/\partial\Delta\beta \end{bmatrix}, \tag{A8}$$

with

$$\frac{\partial J}{\partial \mathbf{R}^f(t_i)} = -2\mathbf{P}^{\mathrm{T}}\mathbf{W}\mathbf{P}[\mathbf{R} - \mathbf{P}^{\mathrm{T}}(\mathbf{Y}^o - \mathbf{X}^b)](t_i), \tag{A9}$$

$$\frac{\partial J}{\partial \Delta\beta} = 0. \tag{A10}$$

The gradient of $\partial J/\partial\alpha_j$ eventually becomes

$$\frac{\Delta J}{\Delta\alpha_j} = \sum_{i=1} [\boldsymbol{\lambda}^r(t_{i+1})]^{\mathrm{T}} \frac{\partial \mathbf{R}^p(t_{i+1})}{\partial \Delta\alpha_j}, \tag{A11}$$

where

$$\frac{\mathbf{R}^p(t_{i+1})}{\partial \Delta\alpha_j} = \frac{\partial \mathbf{R}^p(t_{i+1})}{\partial \beta_j} \cdot \frac{\partial \beta_j}{\partial \alpha_j}, \tag{A12}$$

$$\frac{\mathbf{R}^p(t_{i+1})}{\partial \Delta\beta_j} = -\begin{bmatrix} \mathbf{N}^\beta \\ \mathbf{0} \end{bmatrix}_{i+1}, \tag{A13}$$

$$\frac{\partial \beta_j}{\partial \alpha_j} = \mathbf{B}^{\mathrm{T}} d(\mathbf{A}_j \alpha_j), \tag{A14}$$

where $\mathbf{B}_j$ is a transformation matrix that maps the first derivative of the mapping for the jth estimation variable on the model field.

REFERENCES

Adomaitis, R. A., 1995: RTCVD model reduction: A collocation on empirical eigen-functions approach. Tech. Rep. 95-64, 18 pp.

Antoulas, A., S. Sørensen, and K. A. Gallivan, 2004: Model reduction of large-scale dynamical systems. *Lecture Notes Comput. Sci.,* **3038,** 740–747.

Atwell, J. A., and B. B. King, 2004: Reduced order controllers for spatially distributed systems via proper orthogonal decomposition. *SIAM J. Sci. Comput.,* **26,** 128–151.

Carrera, J., and S. P. Neuman, 1986: Estimation of aquifer parameters under transient and steady-state conditions. 1. Maximum-likelihood method incorporating prior information. *Water Resour. Res.,* **22,** 199–210.

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

Courant, R., and D. Hilbert, 1953: *Methods of Mathematical Physics.* Wiley Interscience, 578 pp.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics. *J. Geophys. Res.,* **99** (C5), 10 143–10 162.

Golub, G., and A. van Loan, 1989: *Matrix Computations.* 2d ed. The Johns Hopkins University Press, 732 pp.

Heemink, A. W., E. E. A. Mouthaan, and M. R. T. Roest, 2002: Inverse 3D shallow water flow modelling of the continental shelf. *Cont. Shelf Res.,* **22,** 465–484.

Hoffmann Jørgensen, B., and J. N. Sørensen, 2000: Proper orthogonal decomposition and low-dimensional modelling. *ERCOFTAC Bull.,* **46,** 44–51.

Holmes, P., J. L. Lumley, and G. Berkooz, 1996: *Turbulence, Coherent Structures, Dynamical Systems and Symmetry.* Cambridge University Press, 440 pp.

Hooimeijer, M. A., 2001: *Reduction of Complex Computational Models.* Siecra Repro, 123 pp.

Ide, K., P. Courtier, M. Ghill, and A. C. Lorenc, 1997: Unified notation for data assimilation: Operational, sequential and variational. *J. Meteor. Soc. Japan,* **75B,** 181–189.

Iman, R. L., and M. J. Shortencarier, 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, Tech. Rep. SAND83-2365, Sandia National Laboratories, Albuquerque, NM, 148 pp.

Jazwinski, A. H., 1970: *Stochastic Processes and Filtering Theory.* Academic Press, 371 pp.

Kaminski, T., R. Giering, and M. Scholze, 2003: An example of an automatic differentiation-based modelling system. *Lecture Notes Comput. Sci.,* **2668,** 95–104.

Krysl, P., S. Lall, and J. E. Marsden, 2001: Dimensional model reduction in non-linear finite element dynamics of solids and structures. *Int. J. Numer. Methods Eng.,* **51,** 479–504.

Lardner, R. W., and Y. Song, 1995: Optimal estimation of eddy viscosity and friction coefficients for a quasi-3-dimensional numerical tidal model. *Atmos.–Ocean,* **33,** 581–611.

Lawless, A. S., S. Gratton, and N. K. Nichols, 2005: Approximate iterative methods for variational data assimilation. *Int. J. Numer. Methods Fluids,* **47,** 1129–1135.

Lee, K. S., E. Yongtae, J. W. Chung, J. Choi, and D. Yang, 2000: A control-relevant model reduction technique for nonlinear systems. *Comput. Chem. Eng.,* **24,** 309–315.

Ljung, L., 1979: Asymptotic-behavior of the extended Kalman filter as a parameter estimator for linear systems. *IEEE Trans. Automatic Control,* **24,** 36–50.

McKay, M. D., W. J. Conover, and R. J. Beckman, 1979: A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics,* **221,** 239–245.

Newman, A. J., 1996: Model reduction via the Karhunen–Loève expansion. Part I: An exposition. Institute for Systems Research Tech. Rep. 96-32, 15 pp.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 1992: *Numerical Recipes in Fortran: The Art of Scientific Computing.* Cambridge University Press, 966 pp.

Ravindran S. S., 2002: Control of flow separation over a forward-facing step by model reduction. *Comput. Methods Appl. Mech. Eng,* **191,** 4599–4617.

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

Tarantola, A., 1987: *Inverse Problem Theory.* 2d ed. Elsevier, 356 pp.

Ullman, D. S., and R. E. Wilson, 1998: Model parameter estimation from data assimilation modeling: Temporal and spatial variability of the bottom drag coefficient. *J. Geophys. Res.,* **103** (C3), 5531–5549.

van Doren, J., R. Markovinović, and J. D. Jansen, 2006: Reduced-order optimal control of water flooding using proper orthogonal decomposition. *Comput. Geosci.,* **10,** doi:10.1007/s10596-005-9014-2.

Vermeulen, P. T. M., A. W. Heemink, and C. B. M. Te Stroet, 2004: Reduced models for linear groundwater flow models using empirical orthogonal functions. *Adv. Water Res.,* **27,** 57–69.

——, ——, and J. R. Valstar, 2005: Inverse modeling using model reduction. *Water Resour. Res.,* **41,** W06003, doi:10.1029/2004WR003698.