SANGOMA: Stochastic Assimilation for the Next Generation Ocean Model Applications EU FP7 SPACE-2011-1 project 283580

Deliverable 3.2: Living Document

Due date: 01/05/2012
Delivery date: 01/05/2012

Delivery type: Report , public

Arnold Heemink Nils van Velzen Martin Verlaan
Umer Altaf
Delft University of Technology, NETHERLANDS

Jean-Marie Beckers Alexander Barth
University of Liège, BELGIUM

Peter Jan Van Leeuwen
University of Reading, UK

Lars Nerger
Alfred-Wegener-Institut, GERMANY

Pierre Brasseur Jean-Michel Brankart
CNRS-LEGI, FRANCE

Pierre de Mey
CNRS-LEGOS, FRANCE

Laurent Bertino
NERSC, NORWAY

March 30, 2016

Contents

0.1 Introduction

This deliverable 'Live document' now is closed and the final version of it can be accessed in deliverable DL3.5, the final document.

MyOcean is the E.U. project dedicated to the implementation of the GMES Marine Core Service for ocean monitoring and forecasting. MyOcean aims at accurately delivering regular and systematic information on the state of global oceans and European regional seas at the required resolution. This information includes hindcasts, nowcasts and forecasts describing the physical state of the ocean and its primary ecosystem. The project also contributes to climate research by providing timeseries of analysed parameters. A new FP7 project (R&D to enhance future GMES applications in the Marine and Atmosphere areas) by the MyOcean consortium aims at expansion to the MyOcean project is recently initiated. As part of this FP7 project a proposal "Stochastic assimilation for the next generation ocean model application (SANGOMA)" to prepare an assimilation component of the next generation operational system of the GMES marine core service is funded.

Data assimilation (DA) is a group of methods in which the observations of the state of a system are combined with the results from numerical model to produce accurate estimates of all the current (and future) state variables of the system. A data assimilation system consists of three components: a set of observations, a dynamical model, and a data assimilation scheme.

The central concept of the data assimilation is the concept of errors, error estimation and error modelling. The observations have errors arising from various sources: e.g. instrumental noise and the representativeness errors. All dynamical models are imperfect with errors arising from: the approximate physics (or biology or chemistry), which parametrises the interaction of the state variables and the discretisation of continuum dynamics into a numerical model. An aspect common for all the data assimilation schemes is that the quantitative basis of the assimilation is formed by the relative uncertainties of the dynamics and observations. Thus, the new estimate does not degrade the reliable information of the observational data but rather enhances that information content.

The most well-known application of DA is in weather forecasting problems in which it was applied in real life for the first time in $1950's$ and $1960's$ to improve the weather forecasts. A good description of the development of DA in meteorology can be found in Delay (1991). The DA has already proved to be useful in other fields of application like tidal models Heemink and Kloosterhuis (1990), oceanography Evensen (1994b), nonlinear shallow-water storm surge models Verlaan and Heemink (1996) and atmospheric chemistry and transport modelling (e.g. Elbern *et al.* (1997), Segers *et al.* (2000)). Among all the DA methods, four dimensional variational data assimilation (4DVAR) called as adjoint method is the one of the most effective and powerful approaches. The method has an advantage of directly assimilating all the available observations distributed in time and space into the numerical model while maintaining dynamical and physical consistency with the model Talagrand (1997). On the other hand since the adjoint of the numerical model needs to be determined, which is usually complicated and time consuming effort for a nonlinear model, the use of 4DVAR is still limited in various fields.

The Kalman filter (KF) Kalman (1960) is a sequential data assimilation al-

gorithm. For linear stochastic systems, it can be shown that the KF is an optimal linear estimator that minimises the variance of the estimation error Simon (2006). Because of its relative simplicity in implementation, the KF is suitable for many data assimilation problems. However, for high dimensional systems such as weather forecasting models, direct application of the KF is prohibitively expensive as it involves manipulating covariance matrices of the system states. For this reason, different modifications of the KF were proposed to reduce the computational cost. These include various ensemble Kalman filters (EnKF) Anderson (2001); Bishop *et al.* (2001); Burgers *et al.* (1998); Evensen (1994a); Evensen and van Leeuwen (1996); Houtekamer and Mitchell (1998); Whitaker and Hamill (2002), the error subspace-based filters Cohn and Todling (1996); Hoteit *et al.* (2001, 2002); Pham *et al.* (1998); Verlaan and Heemink (1997), to name but a few. For a detailed description of the above filters, readers are referred to Evensen (2003); Nerger *et al.* (2005) for reviews of some of the aforementioned filters. Roughly speaking, these modifications exploit the information of a subset in the state space of a dynamical system, while the information of the complement set is considered less influential, and thus ignored. Consequently, the computations of these modified filters are normally conducted on the chosen subsets, instead of the whole state space, so that their computational costs are reduced. For simplicity, we may sometimes abuse the terminology by referring to all the aforementioned filters as the EnKF-based methods (EnKF methods for short).

0.2 Purpose of this Document

As describe in the proposal one of the main aim of the SANGOMA project is to accelerate the implementation of flexible DA toolboxes to strengthen the connection between academics and oceanographic community. This will also allow fast implementation and evaluation of the new DA techniques. Present day high resolution ocean models are very nonlinear and require a strong need for data assimilation methods that can handle these non-linearities. Apart from ensemble based methods, completely nonlinear DA techniques have also been available such as Particle Filters (PF) that directly utilises Bayes Theorem but suffering from high dimensionality problems (Bengtsson *et al.* (2008)). Recently methods are developed with the combinations of EnKF and PF (e.g., Van Leeuwen (2009)). Similarly, efforts are made to develop robust filters that emphasise on the robustness of their error estimates, so that they may have better tolerances to possible uncertainties in assimilation. As an example, the H_∞ filter (HF) Francis (1987); Simon (2006). An Ensemble time-local H_∞ filter (EnTLHF) is proposed recently in Luo and Hoteit (2011) as an analogy to the EnKF for high dimensional data assimilation problems.

This document gives a detailed description of the DA methods that include uncertainty estimation and that can be implemented for large dimensional ocean models. The methods presently available in the present toolboxes (PDAF, OpenDA, Beluga/Sequoia, SESAM, NERSC EnKF, OAK etc) are the most likely candidates to be adapted during this project. Apart from these methods the above mentioned new developments which include new extension to EnKF, ensemble based parameter identification methods and promising PF developments are also most

likely candidate and examples of the methods that can be implemented. The next section gives a detail description of each individual method. This document will be updated throughout the project duration whenever a method is selected for a common DA toolbox.

0.3 Shared Software Modules

This section gives a detailed description of the shared modules. We start by giving common symbol notations. These main symbol notations will be used for all the modules throughout this document.

M	non-linear model operator
$\mathbf{M}$	Tangent linear model
X	model state vector
X_t	true value of the model state vector
X_b	background model state
X_a	analyzed model state
Y	observation vector
H	observation operator
$\mathbf{H}$	linearized observation operator
$\mathbf{B}$	background error covariances
$\mathbf{A}$	analysis error covariances
$\mathbf{R}$	observation error covariances
$\mathbf{K}$	analysis gain Matrix
$\mathbf{I}$	identity Matrix
J	cost function
J_b	background term of the cost function
J_o	observation term of the cost function
J_p	penalty term of the cost function
γ	parameter vector
n	size of state vector
n^q	size of observation vector

0.3.1 POD Calibration Method

The adjoint method is a well-known approach to inverse modelling. The method aims at adjusting a number of unknown control parameters on the basis of given data. The control parameters might be model initial conditions or model parameters Le Dimet and Talagrand (1986), Thacker and Long (1988). An objective function is defined which measures the misfit between the solution and the available data for any model solution over the assimilation interval. To obtain

a computationally efficient procedure this objective function is minimized with a gradient-based algorithm where the gradient is determined by solving the adjoint problem. The adjoint approach is computationally very efficient because one gradient calculation requires just a single simulation of the forward model and a single simulation of the adjoint model backward in time, irrespective of the number of parameters. The adjoint method has been used and applied successfully to many types of inverse problems in ground water flow studies (e.g. Carrera and Neuman (1986)), in meteorology (e.g. Courtier and Talagrand (1990)), in oceanography (e.g. Tziperman *et al.* (1992)) and in shallow water flow models (e.g. Ten-Brummelhuis *et al.* (1993), Lardner *et al.* (1993), Ulman and Wilson (1998), Heemink *et al.* (2002)). One of the drawbacks of the adjoint method is the programming effort required for the implementation of the adjoint model. Research has recently been carried out on automatic generation of computer codes for the adjoint, and adjoint compilers have now become available (see Kaminski *et al.* (2003)). Even with the use of these adjoint compilers developing an adjoint model is often a significant programming effort that hampers new applications of the method.

Proper orthogonal decomposition (POD) is a model reduction method considered as an application of the singular value decomposition (SVD) to the approximation of general dynamical systems Antoulas (2005). It is a data driven projection based method originally developed by Karl Pearson Pearson (1901). Karhunen Karhunen (1946) and Loeve Loeve (1946) had used it as statistical tool to analyze random process data. Lumley Lumley (1967) gave the name POD, and used the method to study turbulent flow. The POD method has application in many fields like image processing, signal processing, data compression, oceanography, chemical engineering and fluid mechanics (see Gunzburger (2004)). In the POD method the projection subspace is determined by processing data obtained from numerical simulations of high dimensional model which is expected to provide information about the dynamical behaviour of the system. The high dimensional equations are projected onto the low dimensional subspace resulting in a low dimensional model and thus reduces the CPU time of model simulation.

Vermeulen and Heemink Vermeulen and Heemink (2006) proposed a method based on POD which shifts the minimisation into lower dimensional space and avoids the implementation of the adjoint of the tangent linear approximation of the original nonlinear model. In their approach, an ensemble of snapshot vectors of forward model simulations is used to determine an approximation of the covariance matrix of the model variability and a small number of leading eigenvectors of this matrix is used to define a model subspace. By projecting the original model onto this subspace an approximate linear reduced model is obtained. Due to the linear character of the reduced model its adjoint can be implemented easily and the minimisation problem is solved completely in reduced space with very low computational cost. The method has recently been applied successfully to the Dutch continental shelf model to estimate water depth Altaf *et al.* (2012).

Background

The discrete model for the evaluation of shallow water system from time t_i to time t_{i+1} can be described by an equation of the form

$$X(t_{i+1}) = M_i[X(t_i), \gamma], \tag{1}$$

where state vector $X(t_{i+1}) \in \Re^n$ denotes the state vector at time t_{i+1} and γ is the vector of the uncertain parameters which needs to be determined. M_i is nonlinear and deterministic dynamics operator that includes inputs. Suppose now that we have imperfect observations $Y(t_i) \in \Re^{n^q}$ of the dynamical system (1) that are related to the model state at time t_i through

$$Y(t_i) = \mathbf{H}X(t_i) + \eta(t_i), \tag{2}$$

where $\mathbf{H} : \Re^n \rightarrow \Re^{n^q}$ is a linear observation operator that maps the model fields on observation space and $\eta(t_i)$ is an unbiased, random Gaussian error vector with covariance matrix $\mathbf{R}_i$. The aim of data assimilation is to find the best estimate of the system initial parameters γ based on given observations. We assume that the difference between data and simulation results is only due to measurement errors and incorrectly prescribed model parameters. The problem of the estimation is then solved by directly minimising the objective function J

$$J(\gamma) = \sum_i [Y(t_i) - \mathbf{H}(X(t_i))]^T \mathbf{R}_i^{-1} [Y(t_i) - \mathbf{H}(X(t_i))] \tag{3}$$

with respect to the parameters γ satisfying the discrete nonlinear forecast model (1). The efficient minimisation of the objective function requires the computation of the gradient of the objective function (3). The adjoint method computes the exact gradient efficiently. Regardless of the number of parameters, the time required to compute the gradient using adjoint technique is more or less identical and is comparable with the computational time needed for a single simulation run of the nonlinear model (1). It requires one forward simulation with the original nonlinear model (1) and a second additional simulation backward in time with the adjoint model

$$\nu(t_i) = \mathbf{M}_{i+1}^T \nu(t_{i+1}) - 2\mathbf{H}\mathbf{R}_i^{-1}[Y(t_i) - \mathbf{H}(X(t_i))], \tag{4}$$

where $\nu(t_i)$ represents the solution of the adjoint model. The gradient ∇J of the objective function J with respect to each component γ_k of the uncertain parameters vector γ is given by

$$\nabla J_k = \sum_i - [\nu(t_{i+1})]^T [\frac{\partial M_i[X(t_i), \gamma]}{\partial \gamma_k}], k = \{1, \cdots, n^p\}. \tag{5}$$

The main hurdle in the use of adjoint method is its implementation, especially when the forward model contains non-linearities.

Linearisation and reduced model formulation

The classical adjoint problem can be simplified with the hypothesis that the objective function J can be made quadratic by assuming that the nonlinear dynamics

operator M_i can be linearised. The linearisation of nonlinear high-order model (1) using the first order Taylor's formula around the background parameter γ_k^b gives

$$\triangle\overline{X}(t_{i+1}) = \frac{\partial M_i[X_b(t_i),\gamma^b]}{\partial X_b(t_i)}\triangle\overline{X}(t_i) + \sum_k \frac{\partial M_i[X_b(t_i),\gamma^b]}{\partial\gamma_k}\Delta\gamma_k \tag{6}$$

where $\overline{X}$ is linearised state vector, X_b is the background state vector with the prior estimated parameters vector γ^b and $\triangle\overline{X}$ is a deviation of the model from background trajectory.

A model can be reduced if the incremental state $\triangle\overline{X}(t_{i+1})$ can be written as linear combination

$$\triangle\overline{X}(t_i) = P\xi(t_{i+1}), \tag{7}$$

where $P = [p_1, p_2, \cdots, p_r]$ is a projection matrix such that $P^TP = I_r$ and ξ is a reduced state vector given by

$$\xi(t_{i+1}) = \widetilde{\mathbf{M}}_i\xi(t_i) + \sum_k \frac{\partial\widetilde{M}_i}{\partial\gamma_k}\Delta\gamma_k \tag{8}$$

or in matrix form

$$\begin{pmatrix} \xi(t_{i+1}) \\ \Delta\gamma \end{pmatrix} = \begin{pmatrix} \widetilde{\mathbf{M}}_i & \widetilde{\mathbf{M}}_i^\gamma \\ 0 & \mathbf{I} \end{pmatrix} \begin{pmatrix} \xi(t_i) \\ \Delta\gamma \end{pmatrix} \tag{9}$$

Here $\Delta\gamma$ is the control parameter vector, $\widetilde{\mathbf{M}}_i$ and $\widetilde{\mathbf{M}}_i^\gamma$ are simplified dynamics operators which approximate the full Jacobians $\mathbf{M}_i$ and $\frac{\partial M_i}{\partial\gamma_k}$ respectively:

$$\widetilde{\mathbf{M}}_i = P^T\mathbf{M}_iP, \tag{10}$$

$$\widetilde{\mathbf{M}}_i^\gamma = P^T(\frac{\partial M_i}{\partial\gamma_1}, \cdots, \frac{\partial M_i}{\partial\gamma_{n^p}}), \tag{11}$$

The Jacobian $\mathbf{M}_i$, is obtained by approximating the nonlinear dynamics operator M_i by linearising it with respect to background state X^b. Instead of computing this huge Jacobian by approximating the partial differential with finite difference by perturbing the nonlinear operator M_i in the direction of each node, we perturb along the direction of p_h only:

$$\mathbf{M}_ip_h = \frac{M_i[X_b(t_i) + \varepsilon p_h, \gamma^b] - M_i[X_b(t_i), \gamma^b]}{\varepsilon}, h = \{1, \cdots, r\}, \tag{12}$$

with ε being the size of the perturbation. The reduced dynamics operator $\widetilde{M}_i$ can now be computed by premultiplying the above formulae by P^T:

$$\widetilde{\mathbf{M}}_i = P^T(\frac{\partial M_i}{\partial X_b(t_i)}p_1, \cdots, \frac{\partial M_i}{\partial X_b(t_i)}p_r). \tag{13}$$

Notice also that only the original model simulations are needed here.The reduced model requires less computational time as it simulates a reduced state within the dimension r instead of the original dimension n where $r < n$. The dimension on which the reduced model operates is $(r + n^p) \times (r + n^p)$ with n^p being the number of estimated parameters.

Collection of the snapshots and POD basis

The POD method is used here to obtain an approximate low-order formulation of the original tangent linear model. POD is an optimal technique of finding a basis which spans an ensemble of data (snapshots) collected from an experiment or a numerical simulation of a dynamical system. The POD modes are optimal at approximating a given dataset. Since the reduced model is used here to estimate uncertain parameters (depth D and manning coefficient c_m), the snapshots should be able to represent the behaviour of the system for these parameters. Therefore the snapshot vectors $e_i \in \Re^s$ are obtained from the perturbations $\frac{\partial M_i}{\partial \gamma_k}$ along each estimated parameter γ_k to get a matrix

$$E = \{e_1, \cdots, e_s\}; i = \{1, 2, \cdots, s\}. \tag{14}$$

The dimension of this ensemble matrix E is $s = n^p \times n^s$, where n^s is the number of snapshot collected for each individual parameter γ_k. The covariance matrix Q can be constructed from the ensemble E of the snapshots by taking the outer product

$$Q = EE^T \tag{15}$$

The projection matrix P used in the previous section is based on the dominant eigenvectors (POD modes) of this covariance matrix. This covariance matrix is usually huge as in the current application with state vector of dimension $\sim 3 \times 10^6$, so direct solution of eigenvalue problem is not feasible. To shorten the calculation time necessary for solving the eigenvalue problem for this high dimensional covariance matrix, we define a covariance matrix G as an inner product

$$G = E^t E \tag{16}$$

In the method of snapshots (Sirovich (1987)), one then solves the $s \times s$ eigenvalue problem

$$G\mathbf{z}_i = E^t E\mathbf{z}_i = \lambda_i \mathbf{z}_i, i \in \{1, 2, \cdots, s\} \tag{17}$$

where λ_i are the eigenvalues of the above eigenvalue problem. The eigenvectors $\mathbf{z}_i$ may be chosen to be orthonormal and the POD modes P are then given by:

$$\mathbf{p}_i = E\mathbf{z}_i/\sqrt{\lambda_i} \tag{18}$$

A physical explanation of POD modes is that they maximise the average energy in the projection of data onto subspace spanned by the modes. We define a measure ψ_i for the relative information to choose a low dimensional basis by neglecting modes corresponding to the small eigenvalues:

$$\psi_i = \frac{\lambda_i}{\sum_{l=1}^{s} \lambda_l} 100\%, i = \{1, 2, \cdots, s\} \tag{19}$$

We collect p_r $(r < s)$ modes such that $\psi_1 > \psi_2 > \ldots > \psi_r$ and they totally explain at least the required variance ψ^e,

$$\psi^e = \sum_{l=1}^{r} \psi_l \tag{20}$$

The total number of eigenmodes r in the POD basis P depends on the required accuracy of the reduced model.

Approximate objective function and its adjoint

In reduced model approach, we look for an optimal solution of the (1) to minimise the approximate objective function $\hat{J}$ in an incremental way:

$$\hat{J}(\Delta\gamma) = \sum_i [\{Y(t_i) - \mathbf{H}(X_b(t_i))\} - \hat{\mathbf{H}}\xi(t_i, \Delta\gamma)]^T \mathbf{R}_i^{-1} [\{Y(t_i) - \mathbf{H}(X_b(t_i))\} - \hat{\mathbf{H}}\xi(t_i, \Delta\gamma)]. \tag{21}$$

The value of the approximate objective function $\hat{J}$ is obtained by correcting the observations $Y(t_i)$ for background state $X_b(t_i)$ which is mapped on the observational space through a mapping $\mathbf{H}$ and for the reduced model state $\xi(t_i, \Delta\gamma)$ which is mapped to the observational space through mapping $\hat{\mathbf{H}}$ with $\hat{\mathbf{H}} = \mathbf{H}P$.

Since the reduced model has linear characteristics, it is easy to build an approximate adjoint model for the computation of gradient of the approximate objective function (21). The gradient of $\hat{J}$ with respect to $\Delta\gamma$ is given by

$$\frac{\partial \hat{J}}{\partial(\Delta\gamma)} = \sum_i -[\hat{\nu}(t_{i+1})]^T \frac{\partial \xi(t_{i+1})}{\partial(\Delta\gamma)}, \tag{22}$$

where $\hat{\nu}(t_{i+1})$ is the reduced adjoint state variable (see Appendix A). Once the gradient has been computed, the process of minimizing the approximate objective function $\hat{J}$ is done along the direction of the gradient vector in the reduced space.

After the minimization process the initial parameters γ are updated and new set of updated parameters γ^{up} is obtained:

$$\gamma^{up} = \gamma + \Delta\gamma. \tag{23}$$

This process of minimization is repeated several times by constructing new POD model with new set of updated parameters γ^{up} to get optimal parameters.

Workflow with POD algorithm

In order to perform the whole parameter estimation process, the following steps are executed.

1. Outer Iteration:
   (a) Generate an ensemble of forward model simulations using initial parameters γ^b.
   (b) Solve eigenvalue problem to get dominant eigenmodes p_i
   (c) Establish a POD reduced model and its adjoint model using eigenmodes p_i.
2. Inner Iteration:
   (a) Perform optimisation iterations in the reduced space to obtain the optimal solution of the approximate objective function $\hat{J}$.
   (b) Update the initial parameters γ^b after the minimisation process obtain new set of updated parameters γ^{up}.
3. Return to step 1 with new set of updated parameters γ^{up} until optimality condition is achieved.

Convergence criterion for inner and outer iterations

The minimisation is performed using a quasi-Newton optimisation algorithm where the Hessian of the objective function is updated using the limited Broyden-Fletcher-Goldfarb-Shanno (LBFGS) method. The minimisation algorithm requires convergence criteria to terminate. We have defined two criterions, one is for inner iterations and one is for outer iterations of the optimisation process. We stop the present inner iteration α and switch to a new outer iteration β with updated parameters γ^{up} by criterion μ, which is defined as

$$\mu = \frac{\mid \hat{J}_{\alpha_{i+1}} - \hat{J}_{\alpha_i} \mid}{max\{\mid \hat{J}_{\alpha_{i+1}} \mid, 1\}} < \epsilon, \tag{24}$$

where α_i represents the i^{th} inner iteration. The value of the ϵ is chosen such that the approximate objective function $\hat{J}$ stops to change, i.e. $\epsilon = 10^{-4}$ (see Oliver *et al.* (2008)). The outer iteration cycle is aborted when the terminal value of ρ is obtained

$$\rho = \frac{\mid J_{\beta_i} - J_{\beta_{i-1}} \mid}{\mid J_{\beta_i} \mid} \leq \kappa, \tag{25}$$

where β_i stands for the i^{th} outer iteration, κ is the terminal value.

Computational efficiency of the algorithm

The computational efficiency of the model-reduced approach is influenced by three factors.

1. Ensemble generation: The computational costs of the reduced model approach are dominated by the generation of the ensemble of forward model simulations. If the dynamics of the system does not change significantly during the course of simulation then a smaller simulation period can be chosen for the generation of ensemble Altaf *et al.* (2009). Using this ensemble the optimisation problem can then be solved over the whole period of model simulation.

   To achieve convergence, the POD method needs to be updated in each outer iteration β, so the ensemble E of snapshot vectors is required in each β. Instead of defining a new model subspace of the leading eigenvectors in each β by generating a new ensemble of the forward model simulations, it is possible to obtain the reduced model by projecting the original model with updated parameters onto the same subspace.

2. Ensemble size: The efficiency of optimisation process is also influenced by the ensemble size. A large ensemble size leads to a huge eigenvalue problem. On the other hand, since the ensemble gives the representation of the model behaviour with respect to each γ_k, it is important that the number of snapshot vectors included in the ensemble must give this representation. So the quality of ensemble is crucial for a reduced-order procedure to be effective. It is possible to include only those snapshots in the ensemble for the period where data is available.

3. Outer iteration: The convergence criterion ρ should be carefully chosen. It should not be chosen too small as this causes jumping of the updated parameters γ^{up} around the optimal global solution Vermeulen and Heemink (2006).

0.3.2 Example Module

Introduction of the method including usage, background and references.

Description

Detailed description of the method.

Workflow

procedural flow of the method and if possible give the pseudo code here

Computational Costs

Bibliography

Altaf, M. U., Heemink, A. W., and Verlaan, M. (2009). Inverse shallow-water flow modelling using model reduction. *International Journal for Multiscale Computational Engineering*, **7**, 577–596.

Altaf, M. U., Verlaan, M., and Heemink, A. W. (2012). Efficient identification of uncertain parameters in a large scale tidal model of european continental shelf by proper orthogonal decomposition. *Int. J. Numer. Meth. Fluids*, **68**, 422–450.

Anderson, J. L. (2001). An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903.

Antoulas, A. C. (2005). *Approximation of large-scale Dynamical Systems*. USA: SIAM.

Bengtsson, T., Bickel, P., and Li, B. (2008). Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems. *IMS Collections Probability and Statistics: Essays in Honor of David A. Freedman*, **2**, 316–334.

Bishop, C. H., Etherton, B. J., and Majumdar, S. J. (2001). Adaptive sampling with ensemble transform Kalman filter. Part I: theoretical aspects. *Mon. Wea. Rev.*, **129**, 420–436.

Burgers, G., van Leeuwen, P. J., and Evensen, G. (1998). On the analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719–1724.

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

Cohn, S. and Todling, R. (1996). Approximate data assimilation schemes for stable and unstable dynamics. *J. Meteor. Soc. Japan*, **74**, 63–75.

Courtier, P. and Talagrand, O. (1990). Variational assimilation of meteorological observations with direct and adjoint shallow water equations. *Tellus*, **42A**, 531–549.

Delay, R. (1991). *Atmospheric data analysis*. Cambridge Unicersity Press.

Elbern, H., Schmidt, H., and Ebel, A. (1997). Variational data assimilation for tropospheric chemistry modeling. *J. Goephys. Res.*, **102**, 15967–15985.

Evensen, G. (1994a). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10143–10162.

Evensen, G. (1994b). Sequential data assimilation with a nonlinear quasi-geostropic model using monte carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10,143–10,162.

Evensen, G. (2003). The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dyn.*, **53**, 343–367.

Evensen, G. and van Leeuwen, P. J. (1996). Assimilation of geosat altimeter data for the aghulas current using the ensemble Kalman filter with a quasi-geostrophic model. *Mon. Wea. Rev.*, **124**, 85–96.

Francis, B. A. (1987). *A Course in H_∞ Control Theory*. Springer-Verlag.

Gunzburger, M. D. (2004). Reduced-order modeling, data compression and the design of experiments. In *Second DOE workshop on multiscale Mathematics*, Broomfield, Colorado.

Heemink, A. W. and Kloosterhuis, H. (1990). Data assimilation for non-linear tidal models. *International Journal for Numerical Methods in Fluids*, **11**, 1097–1112.

Heemink, A. W., Mouthaan, E. E. A., and Roest, M. R. T. (2002). Inverse 3D shallow water flow modeling of the continental shelf. *Continental Shelf Research*, **22**, 465–484.

Hoteit, I., Pham, D. T., and Blum, J. (2001). A semi-evolutive partially local filer for data assimilation. *Marine Pollution Bulletin*, **43**, 164–174.

Hoteit, I., Pham, D. T., and Blum, J. (2002). A simplified reduced order Kalman filtering and application to altimetric data assimilation in Tropical Pacific. *Journal of Marine Systems*, **36**, 101–127.

Houtekamer, P. L. and Mitchell, H. L. (1998). Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811.

Kalman, R. (1960). A new approach to linear filtering and prediction problems. *Trans. ASME, Ser. D, J. Basic Eng.*, **82**, 35–45.

Kaminski, T., Giering, R., and Scholze, M. (2003). An example of an automatic differentiation-based modeling system. *Lecture Notes Comput. Sci.*, **2668**, 5–104.

Karhunen, K. (1946). Zur spektral theorie stochasticher prozsee. *Ann. Acad. Sci. Fenicae*, **34 (A1)**, 1–7.

Lardner, R. W., Al-Rabeh, A. H., and Gunay, N. (1993). Optimal estimation of parameters for a two dimensional hydrodynamical model of the arabian gulf. *J. Geophys. Res. Oceans*, **98**, 229–242.

Loeve, M. (1946). Functions aleatoire de second ordre. *Revue Science*, **84**, 195–206.

Lumley, J. L. (1967). The structure of inhomogeneous turbulence. In *A. M. Yaglom and V. I. Tatarski (eds)*, pages 166–178, Nauka, Moscow.

Luo, X. and Hoteit, I. (2011). Robust ensemble filtering and its relation to covariance inflation in the ensemble Kalman filter. *Mon. Wea. Rev.*, **139**, 3938–3953.

Nerger, L., Hiller, L., and Schröter, J. (2005). A comparison of error subspace Kalman filters. *Tellus*, **57A**, 715–735.

Oliver, D. S., Reynolds, A. C., and Liu, N. (2008). *Inverse theory for petroleum reservoir characterization and history matching*. UK: Cambridge.

Pearson, K. (1901). On lines and planes of closest fit to points in space. *Phil. Mag.*, **2**(6), 559–572.

Pham, D. T., Verron, J., and Roubaud, M. C. (1998). A singular evolutive extended Kalman filter for data assimilation in oceanography. *Journal of Marine Systems*, **16**, 323–340.

Segers, A., Heemink, A. W., Verlaan, M., and van Loon, M. (2000). A modified rrsqrt-filter for assimilating data in atmospheric chemistry models. *Environmental Modeling and Software*, **15**, 663–671.

Simon, D. (2006). *Optimal State Estimation: Kalman, H-Infinity, and Nonlinear Approaches*. Wiley-Interscience.

Sirovich, L. (1987). choatic dynamics of coherent structures. *Physica D*, **37**, 126–145.

Talagrand, O. (1997). Assimilation of observations, an introduction. *Journal of the Meteorological Society of Japan*, **75**, 191–209.

Ten-Brummelhuis, P. G. J., Heemink, A. W., and van den Boogard, H. F. P. (1993). Identification of shallow sea models. *Int. J. for Num. Met. Fluids*, **17**, 637–665.

Le Dimet, F. X. and Talagrand, O. (1986). Variational algorithms for analysis and assimilation of meteorological observations: Theoratical aspects. *Tellus*, **38**, 97–110.

Thacker, W. C. and Long, R. B. (1988). Fitting models to inadequate data by enforcing spatial and temporal smoothness. *J. Geophys. Res.*, **93**, 10655–10664.

Tziperman, E., Thacker, W. C., and Long, R. B. (1992). Oceanic data analysis using a general circulation model, part2: A north atlantic model. *J. Phys. Oceanography*, **22**, 1458–1485.

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

Van Leeuwen, P. J. (2009). Particle filtering in geophysical systems. *Mon. Wea. Rev.*, **137**, 4089–4114.

Verlaan, M. and Heemink, A. W. (1996). Data assimilation schemes for non-linear shallow water flow models. *Adv. Fluid Mechanics*, **96**, 277–286.

Verlaan, M. and Heemink, A. W. (1997). Tidal flow forecasting using reduced rank square root filters. *Stochastic Hydrology and Hydraulics*, **11**, 349 – 368.

Vermeulen, P. T. M. and Heemink, A. W. (2006). Model-reduced variational data assimilation. *Mon. Wea. Rev.*, **134**, 2888–2899.

Whitaker, J. S. and Hamill, T. M. (2002). Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924.