Procedia Computer Science
Volume 80, 2016, Pages 496–506
ICCS 2016. The International Conference on Computational Science

An Ensemble Approach to Weak-Constraint Four-Dimensional Variational Data Assimilation

Jeremy A. Shaw[1] and Dacian N. Daescu[2]

[1] Portland State University, PO Box 751, Portland, OR, 97207, USA.
shaw@pdx.edu
[2] Portland State University, PO Box 751, Portland, OR, 97207, USA.
daescu@pdx.edu

Abstract
This article presents a framework for performing ensemble and hybrid data assimilation in a weak-constraint four-dimensional variational data assimilation system (w4D-Var). A practical approach is considered that relies on an ensemble of w4D-Var systems solved by the incremental algorithm to obtain flow-dependent estimates to the model error statistics. A proof-of-concept is presented in an idealized context using the Lorenz multi-scale model. A comparative analysis is performed between the weak- and strong-constraint ensemble-based methods. The importance of the weight coefficients assigned to the static and ensemble-based components of the error covariances is also investigated. Our preliminary numerical experiments indicate that an ensemble-based model error covariance specification may significantly improve the quality of the analysis.

Keywords: model error; weak constraint; variational data assimilation; ensemble methods; error covariance; error bias

1 Introduction

Four-dimensional variational data assimilation (4D-Var) provides an estimate to the state of a dynamical system through the minimization of a cost functional that measures the distance to a prior state (background) estimate and observations [20] over a time window $[t_0, t_N]$. The analysis fit to each information input component is determined by the specification of the error covariance matrices in the data assimilation system (DAS).

Unlike the extended Kalman filter, error covariances are typically not updated between 4D-Var assimilation cycles. A practical approach to improve the quality of the analysis is to include the "errors of the day" by using an ensemble-based estimate to the background error covariance matrix. Evensen [11] introduces this Monte Carlo alternative as the ensemble Kalman filter (EnKF) and it has since been implemented in various studies, e.g. [15], [17], [18]. Lorenc [21]

Selection and peer-review under responsibility of the Scientific Programme Committee of ICCS 2016

doi:10.1016/j.procs.2016.05.329

and Fairbairn et al. [12] investigate the potential use of EnKF for numerical weather prediction (NWP) applications and its analysis performance, as compared with 4D-Var.

For large-scale dynamical systems, the number of ensemble forecasts is much smaller as compared to the dimension of the discrete state vector. Therefore, an ensemble-based representation to the background error covariance matrix is of low rank and corrupted by sampling errors. To alleviate these issues, several approaches have been considered for practical implementation including covariance localization [15, 16] and the formulation of hybrid methods that aim to synergistically combine the merits of variational and ensemble-based DA [1, 2].

Weak-constraint 4D-Var (w4D-Var) provides a theoretical framework to account for modeling errors in the analysis scheme. Trémolet [25] investigates some possible implementations of w4D-Var. In addition to the specification of the background error covariance (**B**) matrix, the w4D-Var formulation requires information on the model error statistics and specification of the model error covariance. Up to now, the increased computational cost associated with w4D-Var has prevented its practical implementation. Various simplifications to reduce the computational burden have been considered, including writing the model error covariance as a scalar multiple of the background error covariance (see [8] for example) and modeling the model error [14, 26, 27]. Research to implement an ensemble data assimilation approach to model error covariance estimation in w4D-Var is at an incipient stage. Mitchell and Carrassi [24] use ensembles to account for model error, but in the context of the ensemble transform Kalman filter. Desroziers et al. [9] investigate a possible implementation of an ensemble 4D-Var using a four-dimensional ensemble covariance.

Ensemble data assimilation can estimate not only the model error covariance matrices, but also bias. Traditionally, an assumption is made that the errors in data assimilation are unbiased to simplify the computational cost or because the information about error biases is not available. Bias in data assimilation has been explored in the works by Dee [4], Dee and Da Silva [5], and Dee and Todling [6], where it is noted that errors in models and the data are often systematic rather than random. Attempts to correct for error bias have been made in the form of bias detection and correction methods and "bias-aware" data assimilation methods, including bias correction in variational data assimilation [7], but not in the context of w4D-Var. Bias-aware Kalman filters have been explored by Drécourt et al. [10].

This work investigates novel applications of ensemble and hybrid techniques to estimate the model error statistics in a w4D-Var DAS. The implementation of ensemble-based DA for w4D-Var is presented. A proof-of-concept and comparison of w4D-Var to the ensemble data assimilation and hybrid assimilation schemes is made in numerical experiments. The terminology and notation follow closely to that of Ide et al. [19] and Trémolet [25].

2 Ensemble Data Assimilation

Weak-constraint 4D-Var provides a sequence of time-distributed analyses $\mathbf{x}_i^a \in \mathbb{R}^n$ that estimate the true state $\mathbf{x}_i^t$ of a dynamical system at time t_i of the data assimilation interval $[t_0, t_N]$. The nonlinear cost functional associated with w4D-Var is defined as

$$J(\mathbf{x}_0, \ldots, \mathbf{x}_N) = \frac{1}{2}[\mathbf{x}_0 - \mathbf{x}_0^b]^{\mathrm{T}}\mathbf{B}^{-1}[\mathbf{x}_0 - \mathbf{x}_0^b] + \frac{1}{2}\sum_{i=0}^{N}[\mathbf{y}_i - \mathbf{h}_i(\mathbf{x}_i)]^{\mathrm{T}}\mathbf{R}_i^{-1}[\mathbf{y}_i - \mathbf{h}_i(\mathbf{x}_i)] + \frac{1}{2}\sum_{i=1}^{N}[\boldsymbol{\eta}_i - \mathbf{q}_i]^{\mathrm{T}}\mathbf{Q}_i^{-1}[\boldsymbol{\eta}_i - \mathbf{q}_i] \tag{1}$$