Ocean Modelling 117 (2017) 41–51

Contents lists available at ScienceDirect

Ocean Modelling

journal homepage: www.elsevier.com/locate/ocemod

A hybrid approach to generating search subspaces in dynamically constrained 4-dimensional data assimilation

Max Yaremchuk[a,*], Paul Martin[a], Christopher Beattie[b]

[a] Naval Research Laboratory at Stennis Space Center, USA
[b] Department of Mathematics, Virginia Tech, USA

ARTICLE INFO

Article history:
Received 16 November 2016
Revised 26 July 2017
Accepted 2 August 2017
Available online 3 August 2017

Keywords:
Data assimilation
Adjoint analysis
Regional modeling

ABSTRACT

Development and maintenance of the linearized and adjoint code for advanced circulation models is a challenging issue, requiring a significant proportion of total effort in operational data assimilation (DA). The ensemble-based DA techniques provide a derivative-free alternative, which appears to be competitive with variational methods in many practical applications. This article proposes a hybrid scheme for generating the search subspaces in the adjoint-free 4-dimensional DA method (a4dVar) that does not use a predefined ensemble. The method resembles 4dVar in that the optimal solution is strongly constrained by model dynamics and search directions are supplied iteratively using information from the current and previous model trajectories generated in the process of optimization. In contrast to 4dVar, which produces a single search direction from exact gradient information, a4dVar employs an ensemble of directions to form a subspace in order to proceed. In the earlier versions of a4dVar, search subspaces were built using the leading EOFs of either the model trajectory or the projections of the model-data misfits onto the range of the background error covariance (BEC) matrix at the current iteration. In the present study, we blend both approaches and explore a hybrid scheme of ensemble generation in order to improve the performance and flexibility of the algorithm. In addition, we introduce balance constraints into the BEC structure and periodically augment the search ensemble with BEC eigenvectors to avoid repeating minimization over already explored subspaces. Performance of the proposed hybrid a4dVar (ha4dVar) method is compared with that of standard 4dVar in a realistic regional configuration assimilating real data into the Navy Coastal Ocean Model (NCOM). It is shown that the ha4dVar converges faster than a4dVar and can be potentially competitive with 4dvar both in terms of the required computational time and the forecast skill.

Published by Elsevier Ltd.

1. Introduction

The ongoing trend toward massive parallelization in computer technologies facilitates the use of ensemble techniques in geophysical data assimilation. The ensemble approach becomes attractive not only because of its favorable parallelization properties (Isaksen, 2011; Desroziers and Berre, 2012). It also brings in more flexibility and realism in representing the background error covariances (e.g., Romine et al., 2014; Ménétrier et al., 2014; Descombes et al., 2015) and appears to be less vulnerable to instabilities associated with model linearization employed by the standard 4dVar technique. Besides, the ensemble approach allows to avoid costly development and maintenance of the linearized models and their adjoints which beyond being costly may impose certain limits on versatility in applying dynamical constraints within a particular adjoint-based assimilation system.

In the last decade, the use of ensembles in DA has been under extensive development in several directions. Apart from improvements in the BEC modeling, major efforts have been made to combine the benefits of the 4dVar and the ensemble methods. In particular, Buehner et al. (2010) have shown that the 4dVar system with the ensemble-generated BEC outperforms the standard 4dVar in the global forecast model. Similar results were obtained by Kuhl et al. (2013) who investigated the performance of the atmospheric DA system (Rosmond and Xu, 2006) with the hybrid BEC formulation. Coupling the regional 4dVar and ensemble KF systems (Zhang and Zhang, 2012; Barker et al., 2012) resulted in a significant reduction of errors for the forecast lead times up to 2.5 days. All these observations underscore the decisive role played by the flow-dependent BECs delivered by ensembles in improving the forecast skill.

* Corresponding author.
E-mail address: max.yaremchuk@nrlssc.navy.mil (M. Yaremchuk).

http://dx.doi.org/10.1016/j.ocemod.2017.08.003
1463-5003/Published by Elsevier Ltd.

Another extensive field of development is related to the so-called 4dEnVar algorithms (Liu et al., 2008; 2009; Fairbairn et al., 2014) which introduce ensembles into the very fabric of the 4dVar optimization. In contrast to 4dVar which implicitly propagates the BEC, these ensemble methods leverage the power of massively parallel computers and explicitly approximate BEC evolution on the model grid. The major issue with this approach is a computationally efficient localization of the raw ensemble-generated BECs which generally suffer from sampling errors caused by the limited number of ensemble members. In their recent studies, Desroziers et al. (2014) and Liu and Xue (2016) established useful relationships between the 4dVar and 4dEnVar variants with different preconditioners and covariance localization schemes. As a result of these developments, hybrid 4dVar and 4dEnVar methods were implemented operationally in the European (Clayton et al., 2013) and Canadian (Buehner et al., 2013; 2015) weather prediction facilities.

In practice, the 4dEnVar technique is formulated as a search for optimal corrections to the control variables which is performed in the range of the background error covariance $\mathbf{B}$. For that reason preconditioning is often made by the square root of $\mathbf{B}$ and the variational optimization problem is considered in the dual (observation space) formulation which usually has much smaller dimension than the original control space formulation and therefore will be more efficient computationally. In particular, this approach has been adopted in the NAVDAS-AR atmospheric DA system (Rosmond and Xu, 2006).

In the ocean, observations are less abundant than in the atmosphere and the ensemble-based BEC estimates which constitute the backbone of 4dEnVar technique tend to be much less accurate. For that reason, one has to rely on heuristic BEC approximations (e.g. Yaremchuk et al., 2013; Weaver et al., 2015). Development of an efficient a4dVar DA method also becomes more problematic as one has to select a few reliable ensemble perturbations with more care. Early predecessors of practical a4dvar algorithms limited optimization to predetermined low-dimensional subspaces spanned either by the reduced-order approximations of the model Green functions (Stammer and Wunsch, 1996; Menemenlis and Wunsch, 1997), or by the dominant principal orthogonal vectors (EOFs) associated with the model statistics (e.g., Robert et al., 2005; Qui et al., 2007; Hoteit, 2008). The 4dEnVar technique proposed by Liu et al. (2008); 2009), generalizes this approach by representing the search subspace by the Schur products of the ensemble members with the eigenvectors of the reduced-order representation of the localization matrix.

In the present paper, we further develop an iterative ensemble-based 4dVar technique (Yaremchuk et al., 2009) which appears to be competitive with 4dVar in oceanographic applications (Panteleev et al., 2015, Yaremchuk et al., 2016a, hereinafter Y16). A distinctive feature of the technique is its self-sufficiency: in contrast to many ensemble estimation methods which employ a given well-trained ensemble to optimize the control variables within a given time window, the a4dvar sequentially generates search subspaces (bundles of search directions) entirely from the statistics of the model trajectories and/or the respective model-data misfits obtained in the course of optimization. In that respect, the a4dVar technique resembles the 4dVar, which uses the adjoint code to generate a new search direction, whereas in a4dVar that direction is replaced by a search subspace spanned by the ensemble of search directions.

In the previously considered versions of the method search subspaces were built using the leading EOFs of either the model trajectory or the projections of the model-data misfits onto the range of $\mathbf{B}$ at the current iteration inheriting information from either dynamical constraints or modeling errors respectively. The present study blends both approaches in an attempt to improve a4dVar performance and flexibility. In addition, search subspaces are explicitly confined to the range of $\mathbf{B}$, whose structure is constrained by the balance operator, which facilitates searches in hydrostatically and geostrophically balanced directions. To avoid searches in the directions nearly parallel to the ones already explored on the previous iterations, the descent process is restarted by augmenting the search subspaces with the eigenvectors of the background error covariance. It is shown that all these modifications result in a significant improvement in the performance of the algorithm.

The paper is organized as follows. In the next section we briefly describe the basics of 4dvar methodology and its ensemble-based (4dEnVar) variants, outline the a4dvar method, and describe considerations in support of the proposed hybrid methodology of selecting the search subspaces. In Section 3, performance of the a4dVar technique is analyzed using NCOM configuration in the Adriatic sea with a particular focus on the impact of balance constraints on the forecast skill and of the new restart procedure on the convergence rate. Summary and discussion of the results conclude the paper.

2. Variational optimization methodologies

We follow the terminological convention proposed by Lorenc (2013), and refer to “4dEnVar” for the adjoint-free optimization algorithms that recover the gradient information from predetermined ensembles which are intended to capture the dominant features of the BEC structure. The a4dVar algorithm being tested here is designed to perform without a given ensemble: Instead of the BEC model derived from the ensemble, we use a heuristic BEC model, which we believe contributes to a more robust strategy in the face of sparse data. We then iteratively retrieve ensemble members (search directions) either from a model trajectory on current iteration, or from dominant spectral modes of the BEC matrix computed off-line.

2.1. 4dVar

In order to better illuminate connections between the 4dVar framework and what follows, the 4dVar approach in this section is formulated as a linear discrete least-squares problem constrained by model dynamics in a small vicinity of the model's background trajectory $\mathbf{x}_b^n$:

$$J = \frac{1}{2}\left[\mathbf{x}^{0\top}\mathbf{B}^{-1}\mathbf{x}^0 + \sum_{n=1}^{N}(\mathbf{H}_n\mathbf{x}^n - \mathbf{d}^n)^\top\mathbf{R}_n^{-1}(\mathbf{H}_n\mathbf{x}^n - \mathbf{d}^n)\right] \to \min_{\mathbf{x}^0} \quad (1)$$

where $\mathbf{x}^n$ are the deviations of the model state from $\mathbf{x}_b^n$ at time t_n, n enumerates observation times, $\mathbf{B}$ is the BEC matrix of $\mathbf{x}_b^n$ which describes the (Gaussian) statistics of the model state at $n = 0$, $\mathbf{H}_n$ are the model-data projection operators, $\mathbf{d}^n$ are the discrepancies $\mathbf{d}_n^* - \mathbf{H}_n\mathbf{x}_b^n$ between observations $\mathbf{d}_n^*$ and the corresponding background model values, $\mathbf{R}_n$ are the observation error covariances, and $^\top$ denotes transposition. If $\mathbf{B}$ is rank-deficient, $\mathbf{B}^{-1}$ is to be understood as a Moore–Penrose pseudoinverse. We will denote the dimension of the discretized model state vector $\mathbf{x}$ by M and the number of observations available at time t_n by L_n.

The correction vectors, $\mathbf{x}^n$, are governed by the recursive relationship

$$\mathbf{x}^n = \mathbf{M}_n\mathbf{x}^{n-1}, \quad (2)$$

where $\mathbf{M}_n$ is the dynamical operator of the model linearized in the vicinity of the background trajectory $\mathbf{x}_b^n$ at the time interval (t_{n-1}, t_n), so that

$$\mathbf{x}^n = \mathbf{M}_n\mathbf{M}_{n-1}\ldots\mathbf{M}_2\mathbf{M}_1\mathbf{x}^0. \quad (3)$$

Introduce the preconditioned variable $\mathbf{c} = \mathbf{B}^{-1/2}\mathbf{x}^0$ for the control vector, where $\mathbf{B}^{-1/2}$ is the square root of $\mathbf{B}^{-1}$, and denote the aggregated n-step propagator as $\mathbf{M}^n \equiv \mathbf{M}_n\ldots\mathbf{M}_2\mathbf{M}_1$. Define (briefly)