Ocean Modelling 128 (2018) 67–86

Contents lists available at ScienceDirect

Ocean Modelling

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

Extending an oceanographic variational scheme to allow for affordable hybrid and four-dimensional data assimilation

Andrea Storto[*,a,b], Paolo Oddo[b], Andrea Cipollone[a], Isabelle Mirouze[a], Benedicte Lemieux-Dudon[a]

[a] Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC), Viale Carlo Berti Pichat, 6/2, I-40127 Bologna, Italy
[b] NATO Centre for Maritime Research and Experimentation (CMRE), Viale San Bartolomeo 400, La Spezia I-19126, Italy

ARTICLE INFO

Keywords:
4DVAR
Ocean reanalysis
Ocean forecasts
Tangent-linear model

ABSTRACT

The traditional formulation of three-dimensional variational (3DVAR) data assimilation schemes for oceanographic applications neglects the temporal evolution of background errors within and across assimilation temporal windows. Such a simplification may be limiting for many climate (e.g. reanalyses) and operational (e.g. medium-range forecast) applications. This work explores possible extensions of the OceanVar data assimilation code aiming at overcoming these limitations. General formulations are proposed and implemented in order to extend the 3DVAR scheme of OceanVar into a simplified hybrid (ensemble-variational) four-dimensional variational (4DVAR) assimilation scheme, where (i) background-error covariances combine stationary and flow-dependent components through an augmented control vector and (ii) a simplified tangent-linear and adjoint model, which assumes that only temperature and salinity are independent variables. These extensions are shown to allow the background-error covariances to follow the time-varying structure typical of climate modes like ENSO, and to shape the analysis increments in agreement with the underlying ocean circulation, respectively. The two extensions are cross-compared in terms of computational time cost and accuracy and further combined together into a hybrid 4DVAR scheme. The hybrid formulation provides in general largely positive impact at short forecast ranges, while 4DVAR at long ones. The hybrid 4DVAR scheme improves the verification skill scores in most cases.

1. Introduction

Owing to the requirements of long-term climate prediction systems (e.g. seasonal forecasts) and activities that rely on operational ocean forecasting (e.g. search-and-rescue, route optimization, oil spill, etc.), oceanographic data assimilation schemes are a crucial component of environmental monitoring. This is testified by the growing attention devoted to them by international programs during the last decade, such as in Europe the Copernicus Marine Environment Monitoring Service (CMEMS) supported by the European Union through the European Commission Directorate on Enterprise and Industry. A number of data assimilation systems have therefore been developed during the last two decades for both operational and reanalysis applications (e.g. Martin et al., 2015; Masina et al., 2017).

New hybrid data assimilation algorithms that merge advantages of ensemble and variational schemes have been mostly developed to improve the accuracy of operational Numerical Weather Prediction (NWP), as suggested for instance by the review article of Bannister (2017). Oceanographic data assimilation has some delay in following meteorological data assimilation advances. Several reasons concur to this; NWP shows a much larger impact than ocean forecasts on diverse aspects of our daily life that can potentially affect the economy and the safety of human activities on a short-time scale. Furthermore, on a technical ground, the much less dense observing network that has implications on the feasibility of the schemes, e.g. preventing the reliability of ensemble derived error statistics (Panteleev et al., 2015).

While atmospheric data assimilation systems in operational centers mostly rely on four-dimensional data assimilation, recently upgraded to hybrid ensemble-variational formulation, e.g. at the European Centre for Medium-range Weather Forecasts (ECMWF)(Bonavita et al., 2012), at the MetOffice (Clayton et al., 2013), and at Météo-France (Raynaud et al., 2011), the extension of three-dimensional data assimilation systems to hybrid four-dimensional is more limited and recent in time within the oceanic forecasting community.

Potential advantages of 4DVAR with respect to 3DVAR reside in the

* Corresponding author at: Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC), Viale Carlo Berti Pichat, 6/2, I-40127 Bologna, Italy.
E-mail addresses: andrea.storto@cmre.nato.int, andrea.storto@cmcc.it (A. Storto).

https://doi.org/10.1016/j.ocemod.2018.06.005
Received 10 December 2017; Received in revised form 28 June 2018; Accepted 29 June 2018
Available online 30 June 2018

4DVAR ability to implicitly evolve covariances within the assimilation time-window through the tangent-linear model approximation, thus shaping the analysis increments consistently with the actual circulation (Lorenc, 2003a). The tangent-linear assumption required in the 4DVAR formulation limits its time-window length, although the weak-constraint formulation, where the tangent-linear model is not assumed to be perfect, may alleviate this weakness (Fisher et al., 2011). 4DVAR has however very large costs, not only in terms of computational demand, due to the tangent-linear and adjoint model integrations at every minimization iteration, but also in terms of human resources needed for software coding and maintenance. Software engineering behind tangent-linear and adjoint model is indeed non-trivial, and even when automatic differentiation tools are adopted, manual intervention is still required (Elizondo et al., 2002). Such a limitation has fostered the development of adjoint-free 4DVAR formulations that in different ways exploit the information about model-error temporal covariances implicitly contained in ensemble systems (e.g. Yaremchuk et al., 2009; Bishop et al., 2017; Yaremchuk et al., 2017).

Only a few ocean data assimilation systems support 4DVAR. For instance, the Regional Ocean Modelling System (ROMS) (Moore et al., 2011) allows for strong- and weak- constraint four-dimensional data assimilation and it is used for regional applications (both reanalyses and operational oceanography). Still for regional applications, the U.S. Naval Research Laboratory has developed a 4DVAR system coupled with the Navy Coastal Ocean Model (Ngodock and Carrier, 2014), recently extended to allow weak-constraint data assimilation (Ngodock et al., 2016). For global applications, the MIT OGCM has been recently complemented with a 4DVAR data assimilation capability (Fenty et al., 2017); the NEMOVAR variational code (Weaver et al., 2003) developed by the NEMOVAR consortium implements the possibility of running 4DVAR when it is coupled to the NEMOTAM package (i.e. the tangent-linear and adjoint version of the NEMO ocean model, Vidard et al., 2015), although all real-world implementations for reanalysis or forecast applications still use a three-dimensional formulation (e.g. Waters et al., 2015). It is thus clear that the oceanographic community lacks detailed assessments of the potential benefits of four-dimensional versus three-dimensional data assimilation, which is among the objectives of the present work. Furthermore, comparisons between variational and ensemble methods have been outlined several times within global NWP applications (Lorenc, 2003b; Kalnay et al., 2007; Fairbairn et al., 2014), but never with ocean applications.

Unlike variational data assimilation methods, ensemble-based filters include a flow-dependent definition of error covariances, implicit in the time-evolving ensemble-based cross-covariances. A step forward towards more accurate error characterization might reside in the use of hybrid static-ensemble background-error covariances in the variational scheme. This relatively simple extension relaxes the assumption of stationarity of the background errors. Thus, hybrid 4DVAR schemes overcome the 3DVAR limitation of stationarity of covariances either across (through flow-dependent covariance component) or within (through the implicit 4DVAR propagation of covariances) assimilation time-windows.

Hybrid covariance methods have been introduced by Hamill and Snyder (2000), who demonstrated how variational methods can benefit from incorporating information about the *error-of-the-day*. Conversely, ensemble methods may benefit from the variational solution scheme and from the incorporation of stationary covariances that can limit problems arising from limited-size ensemble systems. Since then, hybrid methods have become popular in NWP and are now implemented in many short- and medium- range operational forecast systems in Europe (e.g. Bonavita et al., 2012; Clayton et al., 2013) and at the U.S. National Centers for Environmental Prediction (NCEP) (Wang et al., 2013). Hybrid covariance formulation of data assimilation systems for ocean applications is emerging recently, testified by only a few works at global (Penny et al., 2015) and regional (Oddo et al., 2016) scales, although growing attention is being devoted to include ensemble-based background-error covariances in many ocean variational data assimilation schemes (e.g. in NEMOVAR, Weaver et al., 2018).

Theoretical justification of the use of hybrid covariances has been recently highlighted. The work of Bishop and Satterfield (2013) and Bishop et al. (2013) on hidden variances shows that features such as unknown sources of model error, finite ensemble size and ensemble covariance localization may limit the optimality of the mere use of ensemble-based covariances. Furthermore, they provide a mathematical framework to combine stationary and ensemble-based covariances. In the context of impulsive synchronization perspective, Penny (2017) showed that hybrid data assimilation methods are able to recover the lost stability found when ensemble methods are implemented with limited ensemble size and cannot represent the unstable modes. The work of Ménétrier and Auligné (2015) has also a similar aim: the authors apply a linear filtering framework to sample covariances in order to simultaneously optimize hybridization weights and localization parameters. However, it should be noted that it is customary in real-world applications to perform sensitivity tests to identify the optimal hybrid weight that maximizes certain skill scores.

In this work, we summarize the latest developments in the OceanVar data assimilation code, originally developed by Fondazione CMCC (Centro Euro-Mediterraneo sui Cambiamenti Climatici), – see Dobricic and Pinardi (2008) – and used for both global and regional reanalysis and operational oceanography applications (e.g. Adani et al., 2011; Storto and Masina, 2016), in the context of CMEMS. The developments include i) the support of hybrid formulation for background-error covariances (either vertical only or both vertical and horizontal covariances) and ii) a new four-dimensional variational formulation that required the development of a simplified tangent-linear and adjoint model. These approaches are compared to the original 3DVAR formulation in terms of derivation, computational costs and accuracy.

The article is structured as follows: Section 2 introduces the original OceanVar formulation and the experimental setup; Section 3 describes the new hybrid and 4DVAR formulation and developments, showing case studies to assess their potential benefits. Section 4 cross-compares the new developments with respect to the original 3DVAR, in terms of both computational demand and accuracy, while Section 5 discusses the main achievements and future plans.

2. Original formulation of OceanVar

The assimilation scheme presented here is called OceanVar. It was originally developed for the Mediterranean Sea Forecasting system (Dobricic and Pinardi, 2008) and later adapted to global ocean configurations (Storto et al., 2011), mostly for reanalysis applications (Storto et al., 2016; Storto and Masina, 2016). OceanVar implements the tangent-linear approximation to the observational term of the cost function. The tangent linear assumption considers a Taylor expansion for the model equivalents in observation space. Considering the ocean state $\mathbf{x}$ and the vector of observations $\mathbf{y}$, their difference is approximated with the following rule chain, where $\mathbf{x}^b$ is the ocean background state:

$$\mathcal{H}(\mathcal{M}(\mathbf{x})) - \mathbf{y} \simeq [\mathcal{H}(\mathcal{M}(\mathbf{x}^b)) + \mathbf{HM}\delta\mathbf{x}] - \mathbf{y} = \mathbf{HM}\delta\mathbf{x} - \mathbf{d}. \quad (1)$$

where $\mathcal{H}()$ is the observation function, mapping the ocean state in model space into observation space and $\mathcal{M}()$ is the non-linear model function, propagating forward in time the ocean state. $\mathbf{H}$ and $\mathbf{M}$ are their respective tangent-linear operators formally defined as $\mathbf{H} = \partial\mathcal{H}(\mathbf{x})/\partial\mathbf{x}|_{\mathcal{M}(\mathbf{x}^b)}$ and $\mathbf{M} = \partial\mathcal{M}(\mathbf{x})/\partial\mathbf{x}|_{\mathcal{M}(\mathbf{x}^b)}$. $\mathbf{d} = \mathbf{y} - \mathcal{H}(\mathcal{M}(\mathbf{x}^b))$ is the innovation vector and it may be calculated online during the model integration.

In the traditional 3DVAR formulation of OceanVar, with the First Guess at Appropriate Time (FGAT), the temporal evolution of the ocean state within the assimilation window is neglected for the analysis solution ($\mathbf{M} = \mathbf{I}$) and