Journal of Computational Physics 335 (2017) 311–326

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

# On the variational data assimilation problem solving and sensitivity analysis

Rossella Arcucci [a,b,c], Luisa D'Amore [a,b,*], Jenny Pistoia [d], Ralf Toumi [c], Almerico Murli [e,b]

[a] University of Naples Federico II, Naples, Italy
[b] Euro-Mediterranean Center on Climate Change, Lecce, Italy
[c] Imperial College London, London, UK
[d] National Institute of Geophysics and Volcanology, Bologna, Italy
[e] Southern Partnership for Advanced Computational Infrastructures, Italy

ARTICLE INFO

*Article history:*
Received 13 June 2016
Received in revised form 16 January 2017
Accepted 18 January 2017
Available online 24 January 2017

*Keywords:*
Data Assimilation
Sensitivity analysis
Inverse Problem

ABSTRACT

We consider the Variational Data Assimilation (VarDA) problem in an operational framework, namely, as it results when it is employed for the analysis of temperature and salinity variations of data collected in closed and semi closed seas. We present a computing approach to solve the main computational kernel at the heart of the VarDA problem, which outperforms the technique nowadays employed by the oceanographic operative software. The new approach is obtained by means of Tikhonov regularization. We provide the sensitivity analysis of this approach and we also study its performance in terms of the accuracy gain on the computed solution. We provide validations on two realistic oceanographic data sets.



## 0. Introduction and motivation

The present work is placed in the context of the design of reliable algorithms for solving large scale Variational Data Assimilation (VarDA) applications. We start from considering two concrete scenarios: the analysis of temperature and salinity variations of data collected in closed and semi closed seas, namely the Mediterranean sea and the Caspian sea. The software system used in the Mediterranean sea by the Institute of Geophysics and Volcanology (INGV) is named OceanVar [21]. This is used within the Mediterranean Forecasting System (MFS) to assimilate observational data with results (the so called backgrounds) produced by an high resolution general circulation model of ocean currents named NEMO (Nucleus for European Modeling of the Ocean) [40]. The software system used for the Caspian sea by the Imperial College London (ICL) is ROMS (Regional Ocean Modeling System) [42].

The VarDA functional which is at the heart of these operative software, is highly ill conditioned [27,39], requiring the use of suitable computation approaches aimed to mitigate the effects on the solution of perturbations in the input data, without compromising its accuracy.

In this work, we employ the algorithm in [19] which splits the VarDA functional into several VarDA functionals. As a consequence, instead of solving one larger and worse conditioned DA problem (let us say, the global problem) we solve several smaller and better conditioned DA problems reproducing the DA problem at smaller dimensions (let us say, the local

* Corresponding author.
*E-mail address:* luisa.damore@unina.it (L. D'Amore).

http://dx.doi.org/10.1016/j.jcp.2017.01.034
0021-9991/

problems). The computational kernel of each local problem is the solution of a linear system [4]. Caused by the background error covariance matrices these systems are still ill conditioned [27,39].

As the accuracy of the solution of these linear systems heavily entails that one of DA problem, here we show how to reduce condition number of local VarDA problems without compromising the accuracy of global VarDa solution. To this aim, we use singular values analysis of local deviation matrices related to local background error covariance matrices. We provide sensitivity analysis of this approach and we compare the accuracy with that one obtained by using existing techniques.

The article is organized as follows. In section 1, contribution of the present work with respect to related works is discussed. Section 2 provides mathematical settings and preliminary definitions. Section 3 describes VarDA problem while in section 4 we perform its sensitivity analysis in terms of the propagation of condition error. Finally, in section 5 we apply results on data arising from the Mediterranean sea and the Caspian sea while in section 6 conclusions are summarized.

## 1. Related work and contribution of the present work

Sensitivity Analysis (SA) refers to the determination of the contributions of individual uncertainty on data to the uncertainty in the solution [8]. The first step of SA is to understand the errors that arise at the different stages of the solution process, namely, the uncertainty in the mathematical model, in the model's solution and in the measurements. These are the errors intrinsic to the DA inverse problem. Moreover, there are the approximation errors introduced by the linearization, the discretization, the model reduction. These errors incur when infinite-dimensional equations are replaced by a finite-dimensional system (that is, the process of discretization), or when simpler approximations to the equations are developed (e.g., by model reduction). Finally, given the numerical problem, an algorithm is developed and implemented as a mathematical software. At this stage, the inevitable rounding errors introduced by working in finite-precision arithmetic occur. The first historical approach to SA is known as the local approach [14]. The impact of small input perturbations on the model output is studied. These small perturbations occur around nominal values. This approach consists in calculating or estimating the partial derivatives of the model at a specific point. For models with a large number of input variables, adjoint-based methods are used [9]. Such approach is commonly used in solving large environmental systems as in climate modeling, oceanography, hydrology, etc. [7]. Sensitivity of the four-dimensional VarDA model has been studied in [13] where an Adjoint modeling is used to obtain first- and second-order derivative information and a reduced-order approach is formulated to alleviate the computational cost associated with the sensitivity estimation. This method makes rerunning less expensive, the parameters must still be selected a priori, and, consequently, important sensitivities may be missed [10].

In the present work, instead of introducing simplified models, we apply a SA to the reduced functional in [19] so that the computational cost only depends on local problem size, which can be computationally much smaller than the original. We perform a sensitivity analysis based on the Backward Error Analysis (B.E.A.) [14] of the VarDA function which emphasizes the relationship between the error propagation and the condition number of the covariance matrices.

The inherent ill conditioning of covariance matrices was investigated in the literature in different applications [22,35]. In DA applications the behavior of the condition number with respect to sampling distance, number of data points, domain size, for Gaussian-type covariances has been studied in [27,39]. In [21], and in most relevant DA operative software as well (see for example [2,5]), a variable transformation is performed on the variational functional to reduce the computational cost needed for computing the covariance matrix explicitly; moreover, to improve the conditioning, only Empirical Orthogonal Functions (EOFs) of the first largest eigenvalues of the error covariance matrix are considered.

Since its introduction to meteorology by Edward Lorenz [36], EOFs analysis has become a fundamental tool in atmosphere, ocean, and climate science for data diagnostics and dynamical mode reduction. Each of these applications basically exploits the fact that EOFs allow a decomposition of a data function into a set of orthogonal functions, which are designed so that only a few of these functions are needed in lower-dimensional approximations [29]. Furthermore, since EOFs are the eigenvectors of the error covariance matrix [28], its condition number is reduced as well.

Nevertheless, the accuracy of the solution obtained by truncating EOFs exhibits a severe sensibility to the variation of the value of the truncation parameter, so that a suitably choice of the number of EOFs is strongly recommended. This issue introduces a severe drawback to the reliability of EOFs truncation, hence to the usability of the operative software in different scenarios [28].

In the present work we propose to employ Tikhonov regularization which reveals to be more appropriate than truncation of EOFs and we face the selection of the regularization parameter. In the literature, many methods for choosing this parameter have been proposed. Basically, there are three types of parameter choice methods: a-priori methods, such as the Morozov's discrepancy method; a-posteriori method, such as the Generalized Cross Validation and the L-curve criterion; and data-driven methods [31,37,25]. A-priori methods are not really practical because they need information on the noise affecting data. A-posteriori methods seek to determine the value of the regularization parameter providing the optimal trade-off between the size of the regularized solution and the quality of its fit to the data. In other words, these methods aim to quantify the amount of regularization affecting the regularized solution, as a function of the regularization parameter. By contrast, the methods of the last type, i.e. data-driven methods, are a convenient way to compare the computed solution with a so-called reference solution. For this reason, these methods are sometimes called heuristic or pragmatic methods [6].

For a proper choice of the regularization parameter, in the present work we provide an estimate of Regularization and Perturbation errors which allows us to get a computable value of the regularization parameter giving the right trade off be-