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-

tween these errors. Finally, we show that, in contrast to EOFs, the sensitivity of the solution to variation of the regularization parameter is inappreciable.

In conclusion, the novelties of the present work involve the following aspects:

- Sensitivity Analysis: introduction of the B.E.A. for the Variational Data Assimilation problem;
- Conditioning: introduction of the Tikhonov regularization to improve the conditioning of the covariance matrix;
- Regularization Parameter: introduction of an algorithm for computing the regularization parameter based on the Regularization and Perturbation error estimations.

We validate the algorithm on realistic oceanographic data collected on Mediterranean and Caspian seas and provided by the National Institute of Geophysics and Volcanology in Bologna (INGV) and the Space and Atmospheric Physics Group of the Imperial College, London (ICL), respectively.

2. Preliminary definitions

For the sake of completeness, we recall some preliminary definitions that we will use throughout the article concerning linear discrete ill posed problems [15,11,26,30,31,38,43,17].

Let

$$S\mathbf{w} = \mathbf{b}, \quad S \in \Re^{p\times q}, \quad \mathbf{w} \in \Re^{q\times 1}, \quad \mathbf{b} \in \Re^{p\times 1}, \tag{1}$$

where $p \geq q$, be an over-determined linear system. We assume that S has a full (column) rank and that $\mathbf{b}$ is affected by measurement errors, i.e.

$$\mathbf{b} = \mathbf{b}^{ef} + \mathbf{e}, \tag{2}$$

where $\mathbf{b}^{ef} \in \Re^{p\times 1}$ represents the unknown error free right hand side and $\mathbf{e} \in \Re^{p\times 1}$ the additive noise vector.

In presence of linear discrete ill posed problems [38], matrix S is very ill conditioned. In this case, it might make sense to look at the matrix numerical rank [26, p. 72] and Singular Value Decomposition (SVD) enables us to deal with this concept.

Definition 1 *(Singular value decomposition).* Let S be the real matrix as in (1), and let

$$S = U\Sigma W^T$$

be the singular value decomposition (SVD) of S where $U \in \Re^{p\times p}$ and $W \in \Re^{q\times q}$ are orthogonal (or orthonormal) matrices and

$$\Sigma = diag(\sigma_j)_{j=1,\ldots,q}$$

where singular values σ_j appear in decreasing order:

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_q > 0.$$

Definition 2. Let S be as in Definition 1. Given $\epsilon > 0$, the integer

$$r_\epsilon = rank(S, \epsilon),$$

so that

$$\sigma_{r_\epsilon} \geq \epsilon > \sigma_{r_\epsilon + j} \quad j = 1, q - r_\epsilon$$

is said numerical rank of S with respect to ϵ.

Even though the above definition is independent of the behavior of singular values, it is useful only in presence of a significant gap in the singular value spectrum, i.e. if $\sigma_{r_\epsilon} >> \epsilon >> \sigma_{r_\epsilon+1}$. Indeed, in presence of an ill conditioned matrix, the numerical rank allows us to give a precise characterization of a matrix with a *well determined* numerical rank or with an *ill determined* numerical rank, depending on the singular value spectrum. According to this, P.C. Hansen in [31,30] uses the term well determined numerical rank if $\sigma_{r_\epsilon} >> \sigma_{r_\epsilon+1}$.

The most important outcome of Definition 2 is that when the discrete problem is ill posed, with an ill or a well determined numerical rank, it is needed to filter out the contribution to the solution corresponding to the smallest singular values. Filtering can be sharp (e.g. by using the Truncated Singular Value Decomposition) or smooth (e.g. by using the Tikhonov regularization) as given in the following definitions:

Definition 3 *(Truncated singular value decomposition).* Let $S = U\Sigma W^T$ be the SVD of S as in Definition 1. Let $\Phi_{trnc} \in \Re^{q\times q}$ be a matrix such that

$$\Phi_{trnc} = diag(\underbrace{1, 1, 1, \ldots 1}_{trnc}, 0 \ldots, 0), \tag{3}$$

with $1 \leq trnc \leq q$. Then the matrix

$$S^{trnc} := U\Phi_{trnc}\Sigma W^T, \tag{4}$$

is the truncated SVD (TSVD) matrix for S.

Definition 4 *(Tikhonov regularization matrix).* Let $S = U\Sigma W^T$ be the SVD of S as in Definition 1. Let $\Phi_{Tikh(\lambda)} \in \Re^{q\times q}$ be a matrix such that

$$\Phi_{Tikh(\lambda)} = diag\left(\frac{\sigma_j^2}{\sigma_j^2+\lambda}\right)_{j=1,\dots,q}, \tag{5}$$

with $\sqrt{\sigma_q} \leq \lambda \leq \sqrt{\sigma_1}$. Then, the matrix

$$S^{Tikh(\lambda)} := U\Phi_{Tikh(\lambda)}W^T, \tag{6}$$

is the Tikhonov regularization matrix for S.

Given S, let $\mu(S)$ denote the condition number of S.

Remark 2.1. The following results subsist:

$$\mu(S^{Tikh(\lambda)}) = \frac{\sigma_1}{2\sqrt{\lambda}}, \quad \mu(S^{trnc}) = \frac{\sigma_1}{\sigma_{trnc}}.$$

When S has an ill determined numerical rank the TSVD is suitable for preconditioning the linear system if the projection of the right hand side of linear system in (1) on the singular vectors of S decays faster than the singular value, i.e. if the Discrete Picard Condition (DPC) is fulfilled [31,44]:

Definition 5 *(Discrete Picard Condition, DPC).* Let $S = U\Sigma W^T$ be the SVD of S as in Definition 1. Let τ denote the level at which the computed singular values σ_j level off because of rounding errors. The Discrete Picard Condition (DPC) is satisfied if, for all singular values larger than τ, the corresponding Fourier coefficients $|\mathbf{U}_j^T \cdot \mathbf{b}|$ on average, decay faster than the σ_j, with $\mathbf{U}_j^T$ the j-th left singular vector of V and $\cdot$ the scalar product.

Now we give the standard definition of covariance matrix $\mathbf{B}$.

Definition 6 *(Variance-covariance matrix).* Let us assume $\mathbf{X}$ be a matrix of measurements of pv physical variables at spatial location $\mathcal{D} = \{x_j\}_{j=1,\dots,np}$, and at a correlation time window $[0, T_1] = \{\tau_k\}_{k=1,\dots,M}$:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_{NP} \end{bmatrix} \in \Re^{NP\times M} \tag{7}$$

where each of NP row is a time series for a given location and $NP = pv \cdot np$. Let us assume that each row $\mathbf{X}_i$ of $\mathbf{X}$ has mean $E[X_i] = \{m_i\}_{i=1,\dots,NP}$ and let $\mathbf{m} = (m_i)_{i=1,\dots,NP}$. Let

$$V = \mathbf{X} - \mathbf{m} \in \Re^{NP\times M} \tag{8}$$

be the deviation matrix. The variance-covariance matrix $\mathbf{B} \in \Re^{NP\times NP}$ of $\mathbf{X}$ is defined via the expected value[1] of the outer product:

$$\mathbf{B} = VV^T. \tag{9}$$

Remark 2.2. It is worth to underline that, by applying the SVD to the deviation matrix in (8), the eigenvectors of V are the so-called Empirical Orthogonal Functions (EOFs) [28].

[1] If each vector $\mathbf{X}_i$ has a distribution with probability density function P, then the expected value of $\mathbf{X}_i$ is defined by

$$E(\mathbf{X}_i) = \frac{1}{M-1}\sum_{j=1,\dots,M} x_{ij}P(\mathbf{X}_j).$$

We also introduce the following definition of adjoint operator:

Definition 7 *(Adjoint operator).* Let $\mathbf{M} : \mathbf{x} \rightarrow \mathbf{y} = \mathbf{Mx}$ be a linear operator.
The operator $\mathbf{M}^T : \mathbf{y} \rightarrow \mathbf{x} = \mathbf{M}^T\mathbf{y}$ such that

$$\mathbf{y} \cdot \mathbf{Mx} = \mathbf{M}^T\mathbf{y} \cdot \mathbf{x}, \quad \forall \mathbf{x}, \forall \mathbf{y} \tag{10}$$

where $\cdot$ denotes the scalar product, is named Adjoint operator of $\mathbf{M}$.

3. The DA problem and the VarDA formulation

The method we describe here is the most general VarDA method. It is called four-Dimensional (4D-VarDA) because it takes into account observations that are distributed in space (usually, a three dimensional domain) and over an interval of time $[0, T]$ (typically 6 or 12 hours), called the time window. If the time window is reduced to one instant (i.e. the time variable is fixed), the method is called three-Dimensional (3D-VarDA) [10,12,18,34]. Let us give the mathematical settings describing the VarDA problem.

3.1. The DA model: set up and problem definition

If $\Omega \subset \Re^3$ is a spatial three dimensional domain, let:

$$\begin{cases} u(t_2, x) = \mathcal{M}[u(t_1, x)], & \forall x \in \Omega,\ t_1, t_2 \in [0, T],\ (t_2 > t_1 > 0) \\ u(t_0, x) = u_0(x), & t_0 = 0, \quad x \in \Omega, \end{cases} \tag{11}$$

be a symbolic description of the predictive model of interest where

$$u : (t, x) \in [0, T] \times \Omega \mapsto u(t, x) = [u[1](t, x), u[2](t, x), \ldots, u[pv](t, x)] \in \Re^{pv}, \tag{12}$$

is the state function of $\mathcal{M}$ consisting of $pv \in \mathbb{N}$ physical variables, where pv denotes the number of physical variables. Let:

$$v : (t, x) \in [0, T] \times \Omega \mapsto v(t, x) = [v[1](t, x), v[2](t, x), \ldots, v[pov](t, x)] \in \Re^{pov}, \tag{13}$$

be the observations function consisting of $pov \in \mathbb{N}$ physical observed variables, where pov denotes the number of physical observed variables, and

$$\mathcal{H} : \quad u(t, x) \mapsto v(t, x), \qquad \forall (t, x) \in [0, T] \times \Omega \tag{14}$$

denote the non-linear observations mapping. According to the applications of model-based assimilation of observations, we will use the following discrete formulation for the VarDA problem. Given

1. NP points of $\Omega \subset \Re^3$: $\{x_j\}_{j=1,\ldots,NP}$;
2. $nobs$ points of $\Omega \subset \Re^3$, where $nobs << NP$: $\{y_j\}_{j=1,\ldots,nobs}$;
3. N points of $[0, T]$: $\{t_k\}_{k=0,1,\ldots,N-1}$;
4. the vector

$$\mathbf{u}_0 = \{u_0^j\}_{j=1,\ldots,NP} \equiv \{u(t_0, x_j)\}_{j=1,\ldots,NP} \in \Re^{NP};$$

which is the state at time t_0, i.e. the background estimate;
5. the operator

$$\mathbf{M}_{k-1,k} \in \Re^{NP \times NP}, \quad k = 1, \ldots, N,$$

representing a discretization of a linear approximation of $\mathcal{M}$ from t_{k-1} to t_k;
6. the vector

$$\{u_k^j\}_{j=1,\ldots,NP;k=1,\ldots,N-1} \equiv \{u(t_k, x_j)\}_{j=1,\ldots,NP;k=1,\ldots,N-1} \in \Re^{NP \times N-1};$$

representing the solution of $\mathbf{M}_{k-1,k}$ at t_k, for $k = 1, \ldots, N$;
7. the vector

$$\mathbf{v}_k \equiv \{v(t_k, y_j)\}_{j=1,\ldots,nobs} \in \Re^{N \times nobs}$$

consisting of the observations at t_k, for $k = 0, \ldots, N-1$;
8. the linear operator

$$\mathbf{H}_k \in \Re^{nobs \times NP}, \quad k = 0, \ldots, N-1$$

representing a linear approximation of $\mathcal{H}$;

9. a block diagonal matrix $\mathbf{G} \in \Re^{(N \times nobs) \times (NP \times N)}$ such that

$$\mathbf{G} = \begin{cases} diag\,[\mathbf{H}_0, \mathbf{H}_1\mathbf{M}_{0,1}, \ldots, \mathbf{H}_{N-1}\mathbf{M}_{N-2,N-1}] & N > 1; \\ \mathbf{H}_0 & N = 1; \end{cases} \tag{15}$$

10. the measurements error covariance matrix $\mathbf{R} \in \Re^{(N \times nobs) \times (N \times nobs)}$ which describes the probability distribution function (pdf) of measurement errors. Here we assume $\mathbf{R}$ to be defined as follows

$$\mathbf{R} = diag\,(\mathbf{R_k})_{k=0,\ldots,N-1} \quad \mathbf{R_k} := \sigma_0^2\mathbf{I}, \tag{16}$$

with $0 \leq \sigma_0^2 \leq 1$ and $\mathbf{I} \in \Re^{nobs \times nobs}$ be the identical matrix;

11. the background error covariance matrix $\mathbf{B} \in \Re^{NP \times NP}$ which describes the pdf P_B of background errors. Here we assume that $\mathbf{B}$, defined as in Definition 6 where $T_1 > T$, is such that

$$\mathbf{B} = \sigma_b^2\mathbf{C}, \tag{17}$$

where the matrix $\mathbf{C}$ denoting the correlation structure of the background error, is homogeneous, and the correlations depend only on distance between states and not position, i.e.

$$\mathbf{C}_{(NP,h,L)} = (c_{ij}), \quad c_{ij} = \exp\left(-\frac{1}{2}(j-i)^2 \cdot \|x_j - x_{j-1}\|_\infty^2\right), \tag{18}$$

with length scale $L = NP \cdot \|x_j - x_{j-1}\|_\infty$.

Given the DA problem set up, we now define the DA inverse problem.

Definition 8 *(The DA inverse problem).* Given the vectors

$$\mathbf{v} = (\mathbf{v}_k)_{k=0,\ldots,N-1} \in \Re^{N \times nobs}, \quad \mathbf{u}_0 \in \Re^{NP}$$

and the block diagonal matrix

$$\mathbf{G} \in \Re^{(N \times nobs) \times (NP \times N)}$$

a DA problem concerns the computation of

$$\mathbf{u}^{DA} = (\mathbf{u}_k^{DA})_{k=0,\ldots,N-1} \in \Re^{NP \times N},$$

such that

$$\mathbf{v} = \mathbf{G} \cdot \mathbf{u}^{DA}, \tag{19}$$

subject to the constraint that

$$\mathbf{u}_0^{DA} = \mathbf{u}_0. \tag{20}$$

Since $\mathbf{G}$ is typically rank deficient, DA is ill posed [23,39]. In next section we define the variational formulation which leads to an unconstrained least square problem, where the term in (20) ensures the existence of a solution of the (19).

3.1.1. The VarDA model: problem definition and domain decomposition

Definition 9 *(The VarDA problem).* The VarDA problem can be described as following:

$$\mathbf{u}^{DA} = argmin_{\mathbf{u} \in \Re^{NP \times N}} J(\mathbf{u}) \tag{21}$$

with

$$J(\mathbf{u}) = \alpha\|\mathbf{u} - \mathbf{u}_0\|_{\mathbf{B}^{-1}}^2 + \|\mathbf{G}\mathbf{u} - \mathbf{v}\|_{\mathbf{R}^{-1}}^2 \tag{22}$$

where, for any vector $w \in \Re^{NP}$ and $q \in \Re^{N \times nobs}$, $\|w\|_{\mathbf{B}^{-1}} = w^T\mathbf{B}w$ and $\|q\|_{\mathbf{R}^{-1}} = w^T\mathbf{R}w$. Parameter $\alpha > 0$ denotes the regularization parameter. In general, operational DA software assume $\alpha = 1$. By choosing $\alpha = 1$ can be considered as giving the same relative weight to the observations in comparison to the background state.

Following [19], in (22) we let $\alpha = 1$.

In [3,18,19] authors present a reduced order formulation of the VarDA operator based on the problem decomposition approach:

Definition 10 *(DD-VarDA problem).* Let $\Omega = \bigcup_{i=1}^{N_{sub}} \Omega_i$ be an overlapping decomposition of the domain Ω such that $\Omega_i \cap \Omega_j = \Omega_{ij} \neq 0$ if Ω_i and Ω_j are adjacent and Ω_{ij} is called *o*verlapping region. According to this decomposition the DD-VarDA model is a system of N_{sub} problems described in (23)–(24) where J_i in (24) is the restriction of the functional J in (22) on Ω_i.

$$\mathbf{u}^{DA} = \sum_{i=1}^{N_{sub}} \tilde{\mathbf{u}}_i^{DA} \text{ with } \tilde{\mathbf{u}}_i^{DA} = \begin{cases} argmin_{\mathbf{u}_i} J_i(\mathbf{u}_i) & \text{on} \quad \Omega_i \\ 0 & \text{on} \quad \Omega - \Omega_i \end{cases} \tag{23}$$

where

$$J_i(\mathbf{u}_i) = J(\mathbf{u})/\Omega_i + \rho \sum_{k=0}^{N-1} \|(\mathbf{M}/\Omega_i)^k \mathbf{u}_i/\Omega_{ij} - \left(\mathbf{M}/\Omega_j\right)^k \mathbf{u}_j/\Omega_{ij}\|^2_{\mathbf{B_{ij}}^{-1}} \tag{24}$$

and

$$J(\mathbf{u})/\Omega_i = \|\mathbf{u}_i - \mathbf{u}_{i,0}\|^2_{\mathbf{B}_i^{-1}} + \|\mathbf{G}_i\mathbf{u}_i - \mathbf{v}_i\|^2_{\mathbf{R}_i^{-1}} \tag{25}$$

All the quantities in (24) and (25) are the restriction on Ω_i or Ω_{ij} of the variables in Definition 3, $\forall i = 1, \ldots, N_{sub}$. Also, in (24), we have posed $\mathbf{M} = \mathbf{M}_{k-1,k}$ for all $k = 1, \ldots, N$ just for simplify the notations. Parameter ρ in (24) denotes the regularization parameter. Following [19], we let $\rho = 1$.

Since the minimum of the operator J can be obtained by patching together the minimum functions of these reduced functionals J_i, we solve J_i instead of J, without compromising the accuracy of the whole solution (see [19]). Moreover, as we will show in the next section, since the reduced functionals J_i are better conditioned than the global functional J, the problem decomposition approach reduces the VarDA computational complexity as well as the rounding errors propagation.

4. The VarDA sensitivity analysis

Let $i = 1, \ldots, N_{sub}$ be fixed. We analyze how errors affecting $\mathbf{u}_{i,0}$ propagate on $\mathbf{u}_i^{DA}$ computed by Algorithm 1. We will perform SA by obtaining worst-case error bounds with the aid of the following mathematical framework, mainly based on the condition numbers estimation and regularization approaches. In the following we denote as:

- $\mu(J_i)$ the condition number of the minimization problem in (23) used in the Step 4 of Algorithm 1 [20,39];
- $\mu(\mathbf{M}_{0,N-1})$ the condition number of the model integration problem introduced in the Step 2 of Algorithm 1 [24].

Algorithm 1 DD-VarDA on Ω_i.

1. **Given** $\mathbf{u}_{i,0}$
2. **Model Integration**. Compute

$$\mathbf{u}_{i,N} = \mathbf{M}_{0,N-1}[\mathbf{u}_{i,0}]$$

3. **Define** $\mathbf{G_i}$ as in (15), $\mathbf{R_i}$ as in (16) and $\mathbf{B_i}$ as in (9).
4. **Assimilate** $\mathbf{u}_i$ with the observations $\mathbf{v}_i$ minimizing J_i in (24), i.e. compute

$$\mathbf{u}_i^{DA} = \min_{\mathbf{u}_i} J_i(\mathbf{u}_i)$$

end

The following propositions hold.

Proposition 4.1. *Let*

$$\tilde{\mathbf{u}}_{i,0} = \mathbf{u}_{i,0} + \xi_i \tag{26}$$

denote the value of $\mathbf{u}_{i,0}$ *in Step 1 of Algorithm 1, where* ξ_i *is the error given by the finite precision arithmetic system. The Algorithm 1 computes*

$$\tilde{\mathbf{u}}_i^{DA} = \mathbf{u}_i^{DA} + \delta_i \tag{27}$$

where

$$\|\delta_i\|_\infty = \mu(J_i)\mu(\mathbf{M}_{0,N-1})\|\xi_i\|_\infty. \tag{28}$$

Proof. By applying the Backward Error Analysis (B.E.A.) [32], to steps 2 and 4 of Algorithm 1 it follows that, in Step 2, if we let $\tilde{\mathbf{u}}_{i,N}$ denote the computed value of $\mathbf{u}_{i,N}$ using $\tilde{\mathbf{u}}_{i,0}$ as initial value, then

$$\tilde{\mathbf{u}}_{i,N} = \mathbf{u}_{i,N} + \sigma_i, \tag{29}$$

where

$$\|\sigma_i\|_\infty = \mu(\mathbf{M}_{0,N-1})\,\|\xi_i\|_\infty \tag{30}$$

and $\mu(\mathbf{M}_{0,N-1})$ is the condition number of the linear operator $\mathbf{M}_{0,N-1}$.

In Step 4, if $\tilde{\mathbf{u}}_i^{DA} = \mathbf{u}_i^{DA} + \delta_i$ denotes the computed value of $\mathbf{u}_i^{DA}$ obtained from the perturbed data $\tilde{\mathbf{u}}_{i,N}$, then

$$\|\delta_i\|_\infty = \mu(J_i)\|\sigma_i\|_\infty$$

and the (28) follows. □

In the following we consider $\mathbf{M}_{0,N-1}$ given in step 2 of Algorithm 1 as defined by the sequence of the operators $\mathbf{M}_{k-1,k}$ which are the discretization of a linear approximation of $\mathcal{M}$ from t_{k-1} to t_k, $\forall k \in [1, N-1]$ as defined in Section 3, it is [13]:

$$\mathbf{M}_{0,N-1} = \mathbf{M}_{0,1}\mathbf{M}_{1,2}\cdots\mathbf{M}_{N-2,N-1} \tag{31}$$

Proposition 4.2. *Let $\mathbf{M}_{k-1,k}^T$ denote the adjoint operator of $\mathbf{M}_{k-1,k}$ as defined in (10), $\forall k \in [1, N-1]$ and $\mathbf{M}_{0,N-1}$ as in (31). Then*

$$\mu(\mathbf{M}_{0,N-1}) \geq \|\mathbf{M}_{0,N-1}^T\|_\infty^{-1} \tag{32}$$

where

$$\mathbf{M}_{0,N-1}^T = \mathbf{M}_{N-2,N-1}^T\cdots\mathbf{M}_{1,2}^T\mathbf{M}_{0,1}^T. \tag{33}$$

Proof. Let ξ_i and σ_i be the errors on $\mathbf{u}_{i,0}$ and $\mathbf{u}_{i,N}$ respectively as introduced in Proposition 4.1 by (26) and (29). From (31) and the step 2 of Algorithm 1, it is

$$\sigma_i = \mathbf{M}_{0,1}\mathbf{M}_{1,2}\cdots\mathbf{M}_{N-2,N-1}\xi_i.$$

Let $\{\xi_{i,k}\}_{k=1,N-1}$ be a sequence of propagation errors such that $\xi_{i,0} = \xi_i$ and $\xi_{i,k} = \mathbf{M}_{k-1,k}\xi_{i,k-1}$, then $\sigma_i = \mathbf{M}_{N-2,N-1}\xi_{i,N-1}$. From (10) in Definition 7, it is

$$\xi_{i,k}\cdot\mathbf{M}_{k-1,k}\xi_{i,k-1} = \mathbf{M}_{k-1,k}^T\xi_k\cdot\xi_{k-1},\ \forall k \in [1, N-1],$$

which gives

$$\mathbf{M}_{N-2,N-1}^T\cdots\mathbf{M}_{1,2}^T\mathbf{M}_{0,1}^T\sigma_i = \xi_i. \tag{34}$$

By considering the infinity norm $\|\cdot\|_\infty$, from (34) and (33), it is

$$\|\mathbf{M}_{0,N-1}^T\|_\infty\|\sigma_i\|_\infty \geq \|\xi_i\|_\infty \tag{35}$$

then, from (30) and (35), the (32) follows. □

Proposition 4.3. *Let*

$$\mathbf{A} = \mathbf{B}^{-1} + \mathbf{G}^T\mathbf{R}^{-1}\mathbf{G} \tag{36}$$

be the Hessian of the operator J defined in (22). Then

$$\mu(J) = \mu(\mathbf{A}). \tag{37}$$

Proof. The computational kernel of the minimization problem in (23), is the solution of a Quasi Newton (QN) method [19, 21,41]. Then

$$\mathbf{u}^{k+1} = \mathbf{u}^k + \delta\mathbf{u}^k, \tag{38}$$

where k denotes the generic QN iteration and $\delta\mathbf{u}^k$ solution of the linear system

$$\mathbf{A}\delta\mathbf{u}^k = -\mathbf{B}^{-1}(\mathbf{u}^k - \mathbf{u}_0^k) + \mathbf{G}^T R^{-1}(\mathbf{v} - \mathbf{G}^T\mathbf{u}^k) \tag{39}$$

Then, the (37) follows. □

Remark. The computation of the adjoint (in (32) and in (36)) can be obtained using the methodology of the Discrete Adjoint Sensitivity Analysis Procedure (DASAP) [9,13].

In the same way we can prove that

Corollary 4.4. *Let*

$$\mathbf{A}_i = \mathbf{B}_i^{-1} + \mathbf{G}_i^T \mathbf{R}_i^{-1} \mathbf{G}_i \tag{40}$$

be the Hessian of the operator J_i defined in (23). Then

$$\mu(J_i) = \mu(\mathbf{A}_i). \tag{41}$$

These results highlight that computational kernel of global VarDA problem defined in (22) (and of the local VarDA problem, defined in (23)) is the solution of the linear system (39) whose coefficient matrix is the matrix $\mathbf{A}$ in (36) (and $\mathbf{A}_i$ in (40)) (see Algorithm 2, below).

Algorithm 2 DD-VarDA on Ω_i.

1. .....

2. .....

3. .....

4. **Call** the L-BFGS algorithm [41] to minimize $\mathbf{J}_i$:
   k := 0;
   repeat
      k := k + 1;
      Solve the linear system
      $\mathbf{A}_i \delta \mathbf{u}_i^k = -\mathbf{B}_i^{-1}(\mathbf{u}_i^{k-1} - \mathbf{u}_{0,i}) + \mathbf{G}_i^T R^{-1}(\mathbf{v}_i - \mathbf{G}_i^T \mathbf{u}_i^{k-1})$
   until (convergence)
   $\mathbf{u}_i^{DA} := \mathbf{u}_i^k + \delta \mathbf{u}_i^k$

end

We now demonstrate that the condition number of the local functional J_i defined in (23) is smaller than the condition number of the functional J defined in (24).

Proposition 4.5. *It holds that*

$$\mu(\mathbf{A}_i) \leq \mu(\mathbf{A}) \tag{42}$$

Proof. It is[2] [27]

$$\mu(\mathbf{A}) \simeq \mu(\mathbf{B}), \tag{43}$$

and, from the (40), it is

$$\mu(\mathbf{A}_i) \simeq \mu(\mathbf{B}_i). \tag{44}$$

Moreover, from (17) it follows that

$$\mu(\mathbf{B}) = \mu(\mathbf{C}_{(NP,h,L)}), \tag{45}$$

where $\mathbf{C}_{(NP,h,L)}$ is defined in (18). Finally, it is [1]

$$\mu(\mathbf{C}_{(NP,h,L)}) = \exp\left(\frac{1}{2}\left|\left(\frac{\pi}{h}\right)^2 - \left(\frac{\pi}{L}\right)^2\right|\right). \tag{46}$$

Let

$$\mathbf{C}_i := \mathbf{C}_{(NP_{sub},h,L_{sub})}$$

where $NP_{sub} = \frac{NP}{N_{sub}}$ is the number of grid points in Ω_i and $L_{sub} = NP_{sub} \cdot h$, and h is the sampling distance among points of Ω_i. Then from (46), it follows that

$$L_{sub} < L \Rightarrow \mu(\mathbf{C}_i) < \mu(\mathbf{C}_{(NP,h,L)}). \tag{47}$$

Finally, from (43)–(44)–(45)–(47) it is

$$\mu(\mathbf{A}) \simeq \mu(\mathbf{B}) = \mu(\mathbf{C}_{(NP,h,L)}) \geq \mu(\mathbf{C}_i) = \mu(\mathbf{B}_i) \simeq \mu(\mathbf{A}_i). \qquad \square$$

[2] In the following we will use the symbol $a \simeq b$, where a and b are real numbers, to indicate that a and b have the same order of magnitude.

The most popular software, developed in the operative centers, implement the so called incremental formulation of a 3DVar DA model [2,5,16,21,42] described by:

$$J(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + (\mathbf{G}V\mathbf{w} - \mathbf{d})^T\mathbf{R}^{-1}(\mathbf{G}V\mathbf{w} - \mathbf{d}) \tag{48}$$

which uses the covariance matrix as in (9). Similarly to (25), the (48) gives on the subdomains Ω_i [3,18,19]:

$$J_i(\mathbf{w}_i)/\Omega_i = \frac{1}{2}\mathbf{w}_i^T\mathbf{w}_i + (\mathbf{G}_i V_i\mathbf{w}_i - \mathbf{d}_i)^T\mathbf{R}_i^{-1}(\mathbf{G}_i V_i\mathbf{w}_i - \mathbf{d}_i) \tag{49}$$

with $\mathbf{B}_i = V_i V_i^T \in \Re^{NP_{sub}\times NP_{sub}}$, and $V_i \in \Re^{NP_{sub}\times M}$, where $NP_{sub} = \frac{NP}{N_{sub}}$, $M \geq NP_{sub}$ and $\mathbf{w}_i$ such that

$$\mathbf{w}_i = V_i^+(\mathbf{u}_i - \mathbf{u}_{i,0}) \tag{50}$$

with V_i^+ generalized inverse of V_i. Similarly, the (40) becomes

$$\mathbf{A}_i = \mathbf{I}_i + (\mathbf{G}_i V_i)^T\mathbf{R}_i^{-1}\mathbf{G}_i V_i. \tag{51}$$

The condition number of $\mathbf{A}_i$ is bounded above by the condition number of the deviation matrix V_i. Indeed, from the triangle inequality in (51) it is:

$$\mu(\mathbf{A}_i) \leq 1 + \frac{1}{\sigma_o^2}\mu^2(V_i)\mu^2(\mathbf{G}_i).$$

The matrix V_i has an ill determined numerical rank then it is highly ill conditioned [15,28]. In [2,5,21] the conditioning of V_i is reduced by truncating the EOFs of V_i, i.e. by using the TSVD of the V_i matrix. Nevertheless, when the matrix has an ill determined numerical rank the TSVD approach may be suitable only if the projections of the right hand side on the EOFs decay faster than the singular values (see Definition 5). Here we introduce the Tikhonov regularization matrix and we show that it is much more appropriate than the TSVD matrix.

4.1. Tikhonov regularization and error analysis

Here, we consider the generic local linear system

$$V_i\mathbf{w} = (\mathbf{u} - \mathbf{u}_0), \tag{52}$$

which follows from (50), where we omit the index i for simplicity of notations. In particular, we consider the regularized systems

$$V^{Tikh(\lambda)}\mathbf{w}_\lambda = (\mathbf{u} - \mathbf{u}_0), \tag{53}$$

where the matrix $V^{Tikh(\lambda)}$ is defined in Definition 4 (see (6)) and

$$V^{trnc}\mathbf{w}_{trnc} = (\mathbf{u} - \mathbf{u}_0), \tag{54}$$

where the matrix V^{trnc} is defined in Definition 3 (see (4)).

The key point of using $V^{Tikh(\lambda)}$ is a proper choice of the parameter λ that balances the approximation error (the so-called Regularization error) introduced by $V^{Tikh(\lambda)}$ while controlling its condition number (the so-called Perturbation error).

Here, we are interested in demonstrating the gain obtained by means of the Tikhonov regularization of V compared to the TSVD of V. Hence, in order to compute the most appropriate value of the regularization parameter, we use a data-driven parameter choice method, which is a convenient way to compare the solution $\mathbf{w}_\lambda$ of (53) with our reference solution given by the system (54). We note that since the operative software truncates the EOFs at the value $trnc = j \in \mathcal{N}$, with $1 \leq j \leq NP_{sub}$, in this section, we use the notation $EOF(j)$ instead of $trnc$. So we give the following

Definition 11 *(Regularization error).* Let $\mathbf{w}_{EOF(j)}$ be a reference solution of the system in (54). Let $V^{Tikh(\lambda)}$ be defined as in (6). Let

$$\Phi_{EOF(j)} = diag(f_1, \ldots, f_{NP_{sub}}), \quad f_k \neq 0, \quad k = 1, \ldots, NP_{sub}.$$

Then

$$E_\lambda^r := \left[W\left(\Phi_{EOF(j)} - \Phi_{Tikh(\lambda)}\right)W^T\right]\mathbf{w}_{EOF(j)} \tag{55}$$

is the regularization error introduced by $V^{Tikh(\lambda)}$ on $\mathbf{w}_{EOF(j)}$.

Definition 12 *(Perturbation error).* Let $\mathbf{w}_{EOF(j)}$ be a reference solution of the system in (54). Let $V^{Tikh(\lambda)}$ be defined as in (6). Then

$$E^p_\lambda := W\Phi_{Tikh(\lambda)}\Sigma^{-1}U^T(\mathbf{w}_\lambda - \mathbf{w}_{EOF(j)}) \tag{56}$$

is the perturbation error, introduced by $V^{Tikh(\lambda)}$ on $\mathbf{w}_{EOF(j)}$.

The following result holds.

Proposition 4.6. *It is*

$$\lim_{\lambda\to 0}\|E^r_\lambda\|_\infty = +\infty, \quad \lim_{\lambda\to\infty}\|E^r_\lambda\|_\infty = const \neq 0 \tag{57}$$

and

$$\lim_{\lambda\to 0}\|E^p_\lambda\|_\infty = 0, \quad \lim_{\lambda\to\infty}\|E^p_\lambda\|_\infty = +\infty \tag{58}$$

Proof. It descends immediately from the matrices structures in (3)–(4)–(5)–(6) and the limits properties. □

It follows that an estimation of the most appropriate value of the parameter λ, let us say λ_{opt}, which gives the best trade off between these two errors, may be obtained by requiring that

$$\lambda_{opt}: \quad \|E^r_{\lambda_{opt}}\|_\infty = \|E^p_{\lambda_{opt}}\|_\infty \tag{59}$$

as it is described in Algorithm 3.

Algorithm 3 Computation of λ_{opt}.

1. **Given** $\mathbf{w}_{ref}$ as the reference (least square) solution of (52) where the matrix V_i is replaced by the matrix $V_i^{EOF(j)}$ and

$$\Phi_{EOF(j)} = diag(f_1,\ldots,f_{NP_{sub}}), \quad f_k \neq 0;$$

3. **Compute** $\mathbf{w}_\lambda$ as a solution of (53);
4. **Compute** the regularization error, i.e. compute

$$E^r_\lambda = \left[W\left(\Phi_{EOF(j)} - \Phi_{Tikh(\lambda)}\right)W^T\right]\mathbf{w}_{EOF(j)};$$

5. **Compute** $\mathbf{e}_\lambda = \mathbf{w}_\lambda - \mathbf{w}_{EOF(j)}$;
6. **Compute** the perturbation error, i.e. compute

$$E^p_\lambda = \left(W\Phi_{Tikh(\lambda)}\Sigma^{-1}U^T\right)\mathbf{e}_\lambda;$$

7. **Provide**

$$\lambda_{opt}: \|E^r_{\lambda_{opt}}\|_\infty = \|E^p_{\lambda_{opt}}\|_\infty$$

end

Remark 4.7. As is shown by the performance analysis in [19], the SVD computation is the most time consuming step of the DA algorithm. By applying the SVD to the local functionals of the DD-VarDA model the time complexity resulting from computing the SVDs of the N_{sub} local deviation matrices V_i is

$$O(N_{sub}\times NP^3_{sub}) = O\left(\frac{NP^3}{N^2_{sub}}\right)$$

instead of $O(NP^3)$, i.e. if the SVD is applied to the global deviation matrix V. Then, the Scale Up factor in [19], which depends on the time complexity, does not change.

5. Oceanographic applications

Before describing the validation analysis we need to give preliminary details about the DA problem configuration. The data we used for the Mediterranean sea have been provided by INGV and have been produced by the software NEMO [40]. The data we used for the Caspian sea have been provided by the Space and Atmospheric Physics Group which work at

the Imperial College of London (ICL) and have been produced by the software ROMS [42]. It means that for the Mediterranean sea we are using as predictive model $\mathcal{M}$ the general circulation model of ocean currents underlying the software NEMO, while for the Caspian sea we use the general circulation model of ocean currents underlying the software ROMS. Furthermore, the INGV provided values produced by the NEMO software for temperature and salinity of the Mediterranean sea, then (in (13)) $pv = 2$. The ICL provided values produced by the ROMS software for the sea surface temperature, then $pv = 1$.

According to the DD-VarDA problem, the experiments set-up consists on the specification of

- $\mathcal{D}$, the spatial domain in which data are collected, as given in Definition 6. In our application, we consider:
$$\mathcal{D} = \{x_j\}_{j=1,\dots,np} = \{(lat, lon, depth)_j\}_{j=1,\dots,np}$$
where the variables *lat* and *lon* refer to the physical coordinates on the Globe and *depth* denotes the value of deepness with respect the sea level.
- $NP_{sub} = pv \times np$, the size of the domain $\mathcal{D}$.
- $[0, T_1]$, the correlation time window and the number M of points in $[0, T_1]$ as given in Definition 6;
- V, the local deviation matrix of the background error covariance matrix $\mathbf{B}$.

Validation is aimed to compare the (least square) solution of the linear system in (53) with the (least square) solution of the linear system in (54) which is used by the operative software for assimilating the data we considered. Validation is carried out by performing the following steps:

1.) characterization of the linear system (52) by evaluating:
   - ⋆ an estimate of the degree k of ill posedness [33] of V, i.e. such that:
   $$\sigma_j = O(j^{-k});$$
   - ⋆ the condition number of V;
   - ⋆ the Discrete Picard Condition (DPC) associated to the linear system in (52), i.e. the behavior of p_j where:
   $$p_j = \frac{|\mathbf{U}_j^T \cdot (\mathbf{u} - \mathbf{u}_0)|}{\sigma_j} \tag{60}$$
   with $\mathbf{U}_j^T$ the j-th left singular vector of V and $\cdot$ the scalar product;
2.) analysis of the linear systems in (53) and in (54). The analysis is performed by considering:
   - ⋆ the condition number
   $$\mu^{Tikh(\lambda)} := \frac{\sigma_1}{2\sqrt{\lambda}}$$
   of $V^{Tikh(\lambda)}$ and the condition number
   $$\mu^{EOF(j)} := \frac{\sigma_1}{\sigma_j}$$
   of $V^{EOF(j)}$;
   - ⋆ the (relative) residual of the linear system in (53):
   $$r^{Tikh(\lambda)} := \frac{\|V^{Tikh(\lambda)}\mathbf{w}_\lambda - (\mathbf{u} - \mathbf{u}_0)\|_\infty}{\|\mathbf{u} - \mathbf{u}_0\|_\infty} \tag{61}$$
   and the (relative) residual of the linear system in (54):
   $$r^{EOF(j)} := \frac{\|V^{EOF(j)}\mathbf{w}_{EOF(j)} - (\mathbf{u} - \mathbf{u}_0)\|_\infty}{\|\mathbf{u} - \mathbf{u}_0\|_\infty}; \tag{62}$$
3.) estimate of the value of regularization parameter λ_{opt} computed by means of Algorithm 3.

5.1. Experimental set-up

In order to construct the deviation matrix V of the background covariance matrix $\mathbf{B}$, according to the Definition 5, we consider

- ⋆ for the Mediterranean sea:
  - $\mathcal{D} = \{(lat = 360°, log = 145°, depth = h)\}_{h=0,\dots,56}$;
  - $NP_{sub} = pv \times np = 2 \times 57 = 114$;

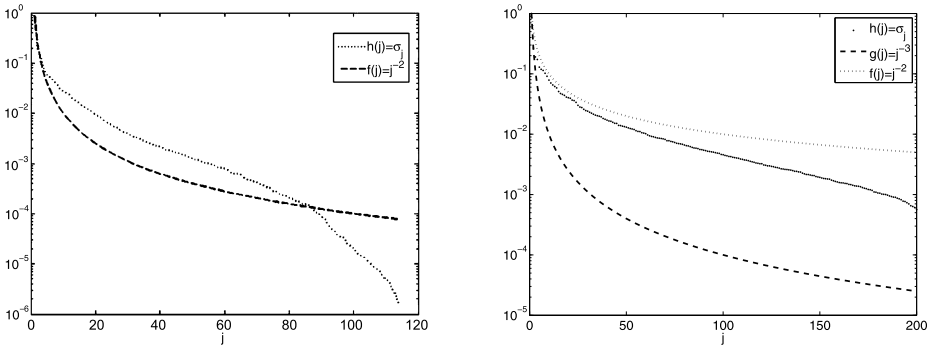

Fig. 1. Singular value spectrum of the matrix V_{Med} (on the left) and the matrix V_{Cas} (on the right) compared with the values $f(j) = j^{-2}$ and $g(j) = j^{-3}$.

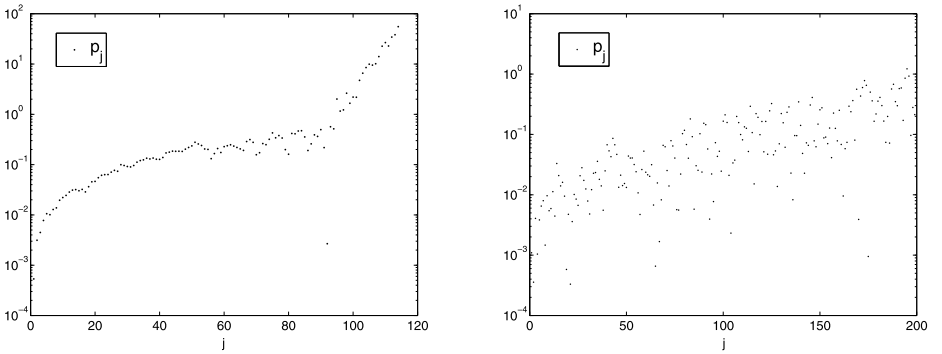

Fig. 2. DPC for the Mediterranean sea data (on the left) and for the Caspian sea data (on the right).

- $M = 14508$, time series in March (2012);
- $V := V_{Med} \in \Re^{114\times 14508}$;
- $V_{Med}^{EOF(j)} \in \Re^{j\times 14508}$ and $V_{Med}^{Tikh(\lambda)} \in \Re^{114\times 14508}$;

⋆ for the Caspian sea:

- $\mathcal{D} = \{(lat = i, log = j, depth = 0)\}_{50^\circ \le i \le 60^\circ,\ 160^\circ \le j \le 180^\circ}$;
- $NP_{sub} = pv \times np = 1 \times np = 200$;
- $M = 403$ time series obtained in January (1948);
- $V := V_{Cas} \in \Re^{200\times 403}$;
- $V_{Cas}^{Tikh(\lambda)} \in \Re^{200\times 403}$ and $V_{Cas}^{EOF(j)} \in \Re^{j\times 403}$.

5.2. Validation

1.) Characterization of the linear system in (52).

The matrices V_{Med} and V_{Cas} have condition number $\mu(V_{Med}) = 1.09 \times 10^5$ and $\mu(V_{Cas}) = 1.70 \times 10^3$. Both matrices have degree of ill posedness $k = 2$, as Fig. 1 shows. The behavior of the values p_j, defined in (60) in Fig. 2, shows that the DPC is violated, meaning that a stable solution of (52) cannot be obtained by using the EOFs truncation.

2.) Analysis of the linear systems in (53) and in (54).

As shown in Fig. 3, the difference between the orders of magnitude of the condition numbers $\mu^{EOF(j)}$ and $\mu^{Tikh(\lambda)}$ increases as $EOF(j) \simeq NP_{sub}$ and $\lambda \simeq 0$. This behavior is in agreement with the expected result that $\lambda \simeq 0 \Rightarrow V^{Tikh(\lambda)} \simeq V$ and that $EOF(j) \simeq NP_{sub} \Rightarrow V^{EOF(j)} \simeq V$. For instance, for the Caspian sea it holds that

$$\frac{\mu^{EOF(j)}(V_{Cas}^{EOF(j)})}{\mu^{Tikh(\lambda)}(V_{Cas}^{Tikh(\lambda)})} \in [10^1, 10^2]; \tag{63}$$

and

$$\frac{\mu^{EOF(j)}(V_{Cas}^{EOF(j)})}{\mu^{Tikh(\lambda)}(V_{Cas}^{Tikh(\lambda)})} \simeq 10^2, \quad j \simeq NP_{sub} = 200 \quad \lambda \simeq \sqrt{\sigma}_{NP_{sub}} = 10^{-2}; \tag{64}$$

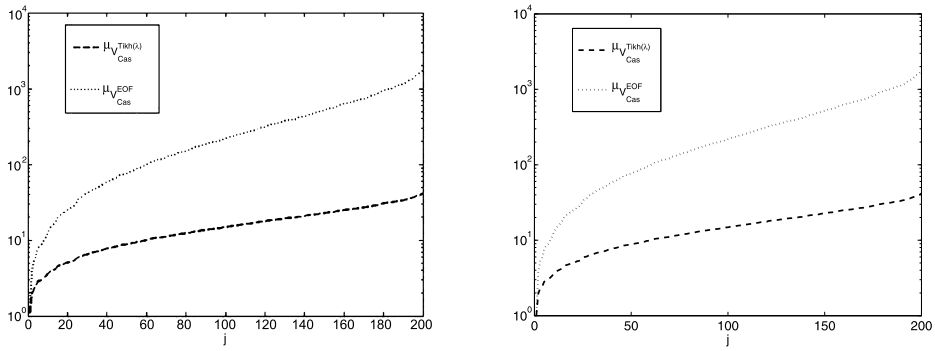

Fig. 3. On the left are the condition number values of the matrices $V_{Med}^{Tikh(\lambda)}$ and $V_{Med}^{EOF(j)}$ for $1.0 \times 10^{-2} = \sqrt{\sigma}_{NP_{sub}} \leq \lambda \leq \sqrt{\sigma}_1 = 25$ and $1 \leq j \leq 114$. On the right are the condition number values $\mu^{Tikh(\lambda)}$ and $\mu^{EOF(j)}$ of the matrices $V_{Cas}^{Tikh(\lambda)}$ and $V_{Cas}^{EOF(j)}$, for $1.0 \times 10^{-2} = \sqrt{\sigma}_{NP_{sub}} \leq \lambda \leq \sqrt{\sigma}_1 = 13$ and $1 \leq j \leq 200$.

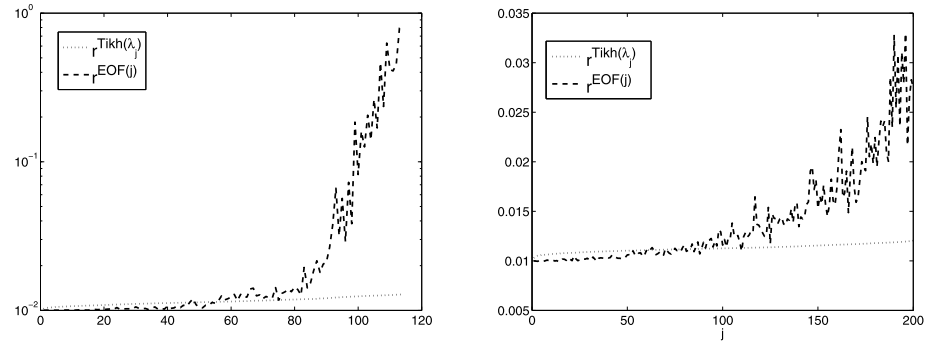

Fig. 4. On the left: residuals $r^{Tikh(\lambda)}$, for $1.0 \times 10^{-2} = \sqrt{\sigma}_{NP_{sub}} \leq \lambda \leq \sqrt{\sigma}_1 = 25$, and $r^{EOF(j)}$ for $1 \leq j \leq 114$ for the Mediterranean sea data. On the right: $r^{Tikh(\lambda)}$, for $1.0 \times 10^{-2} = \sqrt{\sigma}_{NP_{sub}} \leq \lambda \leq \sqrt{\sigma}_1 = 13$, and $r^{EOF(j)}$ for $1 \leq j \leq 200$ for the Caspian sea data.

while for the Mediterranean sea it holds that

$$\frac{\mu^{EOF(j)}(V_{Med}^{EOF(j)})}{\mu^{Tikh(\lambda)}(V_{Med}^{Tikh(\lambda)})} \in [10^2, 10^5], \tag{65}$$

and

$$\frac{\mu^{EOF(j)}(V_{Med}^{EOF(j)})}{\mu^{Tikh(\lambda)}(V_{Med}^{Tikh(\lambda)})} \simeq 10^5, \quad j \simeq NP_{sub} = 114, \quad \lambda \simeq \sqrt{\sigma}_{NP_{sub}} = 10^{-2}. \tag{66}$$

It means that the accuracy of $\mathbf{w}_{EOF(j)}$ exhibits a more severe sensibility to the variation of the value of the truncation parameter j than to the variation of the regularization parameter λ in the solution $\mathbf{w}_\lambda$.

The sensibility to the variation of the value of the truncation parameter j clearly appears on the values of $r^{EOF(j)}$ particularly for

$$j > \frac{NP_{sub}}{2}$$

where $\frac{NP_{sub}}{2} = 57$ for the Mediterranean sea data and $\frac{NP_{sub}}{2} = 100$ for the Caspian sea data, as shown in Fig. 4. This is in agreement with the result that the DPC is not satisfied.

Since the relative error in the solution is bounded above by the product of the residual times the condition number [14], we may conclude that the relative error in $\mathbf{w}_\lambda$ is smaller than in $\mathbf{w}_{EOF(j)}$.

3.) Estimate of the regularization parameter.

In Fig. 5 we show the Regularization and Perturbation errors defined in (55) and (56). Algorithm 3 computes the estimate of the optimal regularization value $\lambda_{opt} = 4.2$ for the Mediterranean sea data and $\lambda = 6.08$ for the Caspian sea data. These values are right now used by the INGV and the ICL for the implementation of the Tikhonov regularization in the operative software for the assimilation of these type of data.

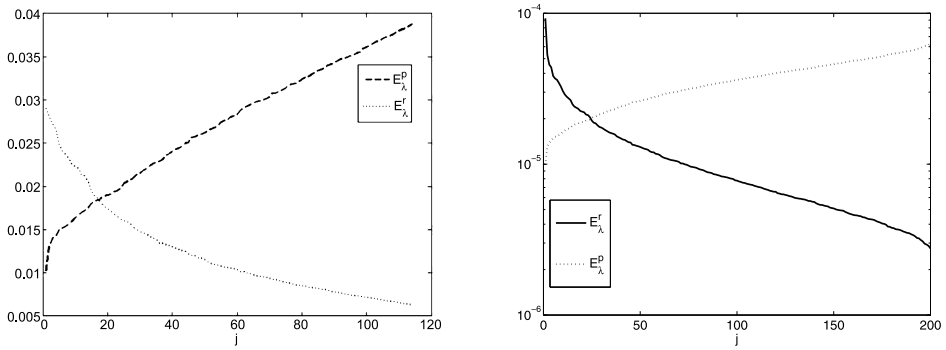

Fig. 5. Regularization error $\|E^r_\lambda\|_\infty$ and perturbation error $\|E^p_\lambda\|_\infty$ versus λ where $1.0 \times 10^{-2} = \sqrt{\sigma}_{NP_{sub}} \leq \lambda \leq \sqrt{\sigma}_1 = 25$, for Mediterranean sea data (on the left) and $1.0 \times 10^{-2} = \sqrt{\sigma}_{NP_{sub}} \leq \lambda \leq \sqrt{\sigma}_1 = 13$ for the Caspian sea data (on the right). Intersection occurs at $\lambda = 4.2$ and $\lambda = 6.08$, respectively.

6. Conclusions

The computational kernel at the heart of the variational data assimilation problem is the solution of a linear system which is highly ill conditioned. The accuracy of the solution of this linear system heavily entails that one of the data assimilation problem. We proposed a computing approach to solve this kernel which reveals to be much more appropriate than the standard approach which is usually implemented in the most operative software for Oceanographic Data Assimilation. In particular, we proposed to employ the Tikhonov regularization for filtering the singular values of the deviation matrix related to the background covariance matrix, which reveals to be more appropriate than the truncation of the EOFs. We applied this approach to the reduced functional in [19] so that the computational cost of the Tikhonov regularization only depends on the local problem size, which can be computationally smaller than the original.

We validated the algorithm on realistic oceanographic data collected on the Mediterranean and the Caspian seas and provided by the National Institute of Geophysics and Volcanology in Bologna (INGV) and the Space and Atmospheric Physics Group of the Imperial College, London (ICL), respectively.

Acknowledgements

This work was developed within the research activity of the H2020-MSCA-RISE-2016 NASDAC Project N. 691184. The computing architectures are located at the University of Naples Federico II, Naples, Italy.

References

[1] R. Ababou, A.C. Bagtzoglou, E.F. Wood, On the condition number of covariance matrices in kriging, estimation, and simulation of random fields, Math. Geol. 26 (1) (1994) 99–133.
[2] E. Andersson, J. Haseler, P. Undén, P. Courtier, G. Kelly, D. Vasiljevic, C. Brancovic, C. Cardinali, C. Gaffard, A. Hollingsworth, C. Jakob, P. Janssen, E. Klinker, A. Lanzinger, M. Miller, F. Rabier, A. Simmons, B. Strauss, J-N. Thepaut, P. Viterbo, The ECMWF implementation of three dimensional variational assimilation (3DVar). Part III: experimental results, Q. J. R. Meteorol. Soc. 124 (550) (1998) 1831–1860.
[3] R. Arcucci, L. D'Amore, L. Carracciuolo, On the problem decomposition of scalable 4D-Var data assimilation models, in: HPCS-IEEE, 2015, pp. 589–594, 978-1-4673-7812-3.
[4] R. Arcucci, L. D'Amore, L. Carracciuolo, A scalable numerical algorithm for solving Tikhonov regularization problems, in: Lecture Notes in Computer Science, vol. 9574, 2016, pp. 45–54.
[5] D.M. Baker, W. Huang, Y.R. Guo, J. Bourgeois, Q.N. Xiao, Three-dimensional variational data assimilation system for MM5: implementation and initial results, Mon. Weather Rev. 132 (2004) 897–914.
[6] F. Bauera, M.A. Lukas, Comparing parameter choice methods for regularization of ill-posed problems, Math. Comput. Simul. 81 (9) (2011) 1795–1841.
[7] D.G. Cacuci, Sensitivity theory for nonlinear systems. I. Nonlinear functional analysis approach, J. Math. Phys. 22 (1981) 2794.
[8] D.G. Cacuci, Sensitivity and Uncertainty Analysis, Chapman & Hall/CRC, NY, 2003.
[9] D.G. Cacuci, M. Ionescu-Bujor, I.M. Navon, Sensitivity and Uncertainty Analysis, Applications to Large-Scale Systems, vol. II, CRC Press, 2005.
[10] D.G. Cacuci, I.M. Navon, M. Ionescu-Bujor, Computational Methods for Data Evaluation and Assimilation, CRC Press, 2013.
[11] R. Campagna, L. D'Amore, A. Murli, An efficient algorithm for regularization of Laplace transform inversion in real case, J. Comput. Appl. Math. 210 (1–2) (2007) 84–98.
[12] P. Courtier, A strategy for operational implementation of 4D-VAR, using an incremental approach, Q. J. R. Meteorol. Soc. 120 (519) (1994) 1367–1387.
[13] D.N. Daescu, I.M. Navon, Sensitivity analysis in nonlinear variational data assimilation: theoretical aspects and applications, in: Istvan Farago, Zahari Zlatev (Eds.), Advanced Numerical Methods for Complex Environmental Models: Needs and Availability, Bentham Science Publishers, December 2013.
[14] G. Dahlquist, A. Bjorck, Numerical Methods, Prentice–Hall, Englewood Cliffs, N.J., 1974.
[15] L. D'Amore, A. Murli, Regularization of a Fourier series method for the Laplace transform inversion with real data, Inverse Probl. 18 (4) (2002) 1185–1205.
[16] L. D'Amore, R. Arcucci, L. Marcellino, A. Murli, A parallel three-dimensional variational data assimilation scheme, AIP Conf. Proc. 1389 (2011) 1829–1831.
[17] L. D'Amore, R. Campagna, A. Galletti, L. Marcellino, A. Murli, A smoothing spline that approximates Laplace transform functions only known on measurements on the real axis, Inverse Probl. 28 (2) (2012) 025007 (37 pages).
[18] L. D'Amore, R. Arcucci, L. Carracciuolo, A. Murli, DD-OceanVar: a domain decomposition fully parallel data assimilation software in Mediterranean sea, Proc. Comput. Sci. 18 (2013) 1235–1244.

[19] L. D'Amore, R. Arcucci, L. Carracciuolo, A. Murli, A scalable variational data assimilation, J. Sci. Comput. 61 (2014) 239–257.
[20] J.E. Dennis Jr., R.B. Schnabel, Numerical Methods for Unconstrained Optimization and Nonlinear Equation, SIAM, 1996.
[21] S. Dobricic, N. Pinardi, An oceanographic three-dimensional variational data assimilation scheme, Ocean Model. 22 (2008) 89–105.
[22] M.P. Ekstrom, A spectral characterization of the ill conditioning in numerical deconvolution, IEEE Trans. Audio Electroacoust. AU-21 (4) (1973) 344–348.
[23] H.K. Engl, M. Hanke, A. Neubauer, Regularization of Inverse Problems, Kluwer, 1996.
[24] R.M. Errico, What is an adjoint model, Bull. Am. Meteorol. Soc. 78 (1997) 2577–2591.
[25] G.H. Golub, M. Heath, G. Wahba, Generalized cross validation as a method for choosing a good ridge parameter, Technometrics 21 (2) (1979) 215–223.
[26] G.H. Golub, C.F. Van Loan, Matrix Computations, John Hopkins University Press, 1996.
[27] S.A. Haben, Conditioning and Preconditioning of the Minimisation Problem in Variational Data Assimilation, PhD. thesis, The University of Reading, 2011.
[28] A. Hannachi, A Primer for EOF Analysis of Climate Data, Department of Meteorology, University of Reading, UK, Reading, 2004.
[29] A. Hannachi, I.T. Jolliffe, D.B. Stephenson, Empirical orthogonal functions and related techniques in atmospheric science: a review, Int. J. Climatol. 1152 (May 2007) 1119–1152.
[30] C. Hansen, Rank-Deficient and Discrete Ill-Posed Problems, Numerical Aspects of Linear Inversion, SIAM, 1998.
[31] P.C. Hansen, J.G. Nagy, D.P. O'Leary, Deblurring Images: Matrices, Spectra, and Filtering, SIAM, 2006.
[32] N. Higham, Accuracy and Stability of Numerical Algorithms, second ed., SIAM, 2002.
[33] B. Hofmann, Regularization for Applied Inverse and Ill-Posed Problems, Teubner, Leipzig, 1986.
[34] E. Kalnay, Atmospheric Modeling, Data Assimilation and Predictability, Cambridge, 2003.
[35] A.B. Kostinski, A.C. Koivunen, On the condition number of Gaussian sample covariance matrices, IEEE Trans. Geosci. Remote Sens. 38 (2000) 329–332.
[36] E.N. Lorenz, Empirical Orthogonal Functions and Statistical Weather Prediction, Sci. Rep., vol. 1, Statistical Forecasting Project, M.I.T., Cambridge, MA, 1956.
[37] V.A. Morozov, Methods for Solving Incorrectly Posed Problems, translation ed.: M.Z. Nashed, Springer, New York, 1984.
[38] J.G. Nagy, D.P. O'Leary, Image Deblurring: I can see clearly now, Comput. Sci. Eng. (2003) 82–84.
[39] N. Nichols, Mathematical concepts in data assimilation, in: W. Lahoz, et al. (Eds.), Data Assimilation, Springer, 2010.
[40] NEMO Web page, www.nemo-ocean.eu.
[41] J. Nocedal, R.H. Byrd, P. Lu, C. Zhu, L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization, ACM Trans. Math. Softw. 23 (4) (1997) 550–560.
[42] ROMS Web page, www.myroms.org.
[43] V.I. Rotar, Probability and Stochastic Modeling, Chapman and Hall/CRC, 2012.
[44] Y. Wang, I.M. Navon, X. Wang, Y. Cheng, 2D Burgers equation with large Reynolds number using POD/DEIM and calibration, Int. J. Numer. Methods Fluids 82 (12) (2016) 909–931, http://dx.doi.org/10.1002/fld.4249.