



HAL
open science

Background Error Covariance Iterative Updating with Invariant Observation Measures for Data Assimilation

Sibo Cheng, Jean-Philippe Argaud, Bertrand Iooss, Didier Lucor, Angélique Ponçot

► **To cite this version:**

Sibo Cheng, Jean-Philippe Argaud, Bertrand Iooss, Didier Lucor, Angélique Ponçot. Background Error Covariance Iterative Updating with Invariant Observation Measures for Data Assimilation. Stochastic Environmental Research and Risk Assessment, 2019, 33, pp.2033-2051. 10.1007/s00477-019-01743-6 . hal-02307657

HAL Id: hal-02307657

<https://hal.science/hal-02307657v1>

Submitted on 11 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Background Error Covariance Iterative Updating with Invariant Observation Measures for Data Assimilation

Sibo Cheng^{1,2}, Jean-Philippe Argaud¹, Bertrand Iooss^{1,3}, Didier Lucor², Angélique Ponçot¹

¹ EDF R&D

² LIMSI, CNRS, Université Paris-Saclay

³ Institut de Mathématiques de Toulouse, Université Paul Sabatier

Abstract

In order to leverage the information embedded in the background state and observations, covariance matrices modelling is a pivotal point in data assimilation algorithms. These matrices are often estimated from an ensemble of observations or forecast differences. Nevertheless, for many industrial applications the modelling still remains empirical based on some form of expertise and physical constraints enforcement in the absence of historical observations or predictions. We have developed two novel robust adaptive assimilation methods named CUTE (Covariance Updating iTerativeE) and PUB (Partially Updating BLUE). These two non-parametric methods are based on different optimization objectives, both capable of sequentially adapting background error covariance matrices in order to improve assimilation results under the assumption of a good knowledge of the observation error covariances. We have compared these two methods with the standard approach using a misspecified background matrix in a shallow water twin experiments framework with a linear observation operator. Numerical experiments have shown that the proposed methods bear a real advantage both in terms of posterior error correlation identification and assimilation accuracy.

1 Introduction

Data assimilation methods are widely used in engineering applications with the objective of state-estimation or/and parameter registration/identification based on the weighted combination of different sources of noisy information. Data assimilation often starts from some initial (i.e. *prior*) knowledge of the quantity of interest, and then produces a subsequent (i.e. *posterior*) estimator of it. Because of the noise alteration, it is most convenient if the method also provides some statistical information about the posterior estimator, at best in the form of its probability distribution. These algorithms are very well known in geosciences and are used as reference methods in the fields of numerical weather prediction ([32]), nuclear safety ([46]), atmospheric chemistry ([39]), hydrologic modelling [28], seismology, glaciology, agronomy, etc. Over decades, these approaches have been applied in the energy industry, for projects involving temperature field reconstruction ([1]) or forecasting in neutronic ([34]) and hydraulic ([20]). More recently, they have also made their way to other fields such as medicine, biomedical applications ([30]) or wildfire front-tracking problems ([36]).

Data assimilation methods are based on prior estimation of the true state (also called background state) and one or several vectors of observations. There exist non-negligible errors in these two quantities. The essential idea is therefore to find a compromise by fusing the information presented in these two quantities ([5]), while accounting for errors, in order to improve the quality of field reconstruction and forecasting by learning from observations. Due to lack of knowledge, the background state is often provided by some experts or approximated, e.g. from a numerical simulation. A remarkable difficulty in the efficiency of these methods is that the prior error covariance matrices are themselves imperfectly known, especially the one of background errors (often noted as matrix \mathbf{B}).

The modelling of \mathbf{B} as well as the observation error covariance matrix \mathbf{R} , remains a very critical point in data assimilation problems because it determines how prior errors spread spatially or temporally (e.g [38]) and this may substantially change the assimilation results ([21]). It also provides an important information of the relationship between observations and forecasts. As mentioned by [16], there exist a wide variety of methods to estimate these matrices. Well known methods among others, are the one of [22], the *NMC* (National Meteorological Center) method ([32]) and ensemble methods ([10]) which are often combined with algebraic operations such as matrix factorization ([24]) or covariance localization ([29]). For many industrial applications, the paucity of historical observations as well as the large dimension and complexity (e.g.[40]) of the system, make the estimation of these covariance matrices unfeasible. A common practice in this case is to impose a standard form for the covariance matrices by empiricism. Certain types of matrices with homogeneous and isotropic characteristics such as diagonal matrices ([23]) or relying on generic covariance kernels: e.g. Matérn kernel ([39]), are often favoured. Other approaches are based on numerical techniques involving convolution operations ([19]) or the resolution of diffusion equations ([45]). The latter methods are sometimes equivalent to the former ones, under simplifying assumptions, as explained in [31]. An *a priori* choice of fixed covariance matrices with

certain regularity properties thus implicitly imposes extra assumptions to the problem, which may lead to supplementary uncertainties. Research efforts are continuously made toward improving the estimation of error covariances. Suffering from problems dimension and complexity, a variety of background matrix computation methods have been proposed in reduced spaces such as the spectral space ([11]) or the wavelet space ([6]). However, these contributions require prior assumptions about the matrix structure which could be difficult to justify in industrial applications.

Recent works of [14] have also investigated model error covariance modelling (often noted as matrix \mathbf{Q}), which can be seen as the main contributor of the background matrix in a dynamical system. Another pathway of research is to make assimilation more robust to this unavoidable lack of knowledge. This challenge applies to both background and observation error covariances. In this work, we focus on the former but both are important.

In this paper, we are interested in iterative algorithms that can be resilient to inconsistent prior background error covariance. More specifically, we look for algorithms that would automatically adjust, thanks to an optimization process, the structure of the error covariance matrices. Our objective is to gain a better knowledge of error correlation which leads to a reduction of *a posteriori* reconstruction errors with limited available data. The meteorology community has been a strong contributor to this topic, and several algorithms and their improved versions have been developed in [13], [12], [7] etc. These methods have been applied world-widely in industrial problems ever since, e.g. [18], [44]. Among them, the Desroziers & Ivanov tuning method consists in finding a fixed point for the assimilated state by regulating the ratio between background and observation covariance matrices magnitude without modifying their structures. This method, for which no statistical estimation of full matrices is required, could have a particular interest for industrial applications with limited prior data. However, it relies on a good knowledge of the correlation of prior errors as shown in [7]. This last condition can be difficult to fulfil without enough historical statistics or in the case of a new application. Another iterative method, based on a diagnostic in the observation space ([12]), aims at estimating the whole covariance matrices (see [25]). However, this method strongly relies on the statistics of either redundant observation data or historical innovation quantity, which are difficult to obtain in our industrial context.

In this work, we develop two novel methods, consisting in repeating several times the assimilation procedure of the state-estimating problem with the same set of observations. A related idea of reusing several times the same observation data set has been carried out by [26] in the "running in place"(RIP) method for the Ensemble Kalman Filtering in order to improve the system spin-up. We provide different approaches which are directly based on static Best Unbiased Linear Estimator (BLUE) and involve an updating of the background covariance matrix at the end of each iteration. We further take into account the covariance between the errors of the updated background vectors and the ones of observations; this covariance appearing due to the iterative process itself. Based on this idea, we propose two iterative algorithms: CUTE (Covariance Updating iTerativeE) method and PUB (Partially Updating BLUE) method. These methods can also be considered as some kind of preliminary step, improving a sequence of dynamical reconstruction with prediction, as they provide a more "consistent" covariance estimate, right after the first assimilation step. For numerical testing, a two-dimensional shallow water model with periodic boundary conditions is used to perform twin experiments for validation purposes. Both iterative methods are studied for a static reconstruction problem and a dynamical data assimilation chain.

The paper is organized as follows. Variational data assimilation is introduced briefly in section 2, with a focus on covariance updating. We then propose two novel iterative methods in section 3, together with a simple illustrative scalar test case. In section 4, these methods are then compared on a two-dimensional fluid mechanics system in a twin experiments framework for both state-independent and state-dependent prior errors.

2 Data assimilation and variational methods framework

2.1 Data assimilation concept

The idea behind data assimilation system is to combine different sources of information in order to provide a more reliable estimation of the system state variables which can be a discretized physical field or a set of parameters (see [27]). We focus on the former where the state is presented by a vector of real entries which could for instance represent a discretized multidimensional physical field (e.g. speed, temperature) at some given coordinates. The true state is denoted by \mathbf{x}_t . In general, the information is split into two parts: an initial state estimation \mathbf{x}_b (so called the background state) and an observation vector \mathbf{y} , related to the state and representing measurements. Both parts are noisy and the observations are often sparse especially for field reconstruction/prediction problems. The observation operator \mathcal{H} from the state space to the observable space is supposed to be known. Both background state and observations are uncertain quantities. Their tolerance, regarding theoretical (or 'true') values, are quantified by ϵ_b and ϵ_y , respectively:

$$\begin{aligned}\epsilon_b &= \mathbf{x}_b - \mathbf{x}_t \\ \epsilon_y &= \mathbf{y} - \mathcal{H}(\mathbf{x}_t).\end{aligned}\tag{1}$$

The transformation operator \mathcal{H} and the true state \mathbf{x}_t are assumed to be deterministic quantities, except in the case of a dynamical data assimilation chain, which will not be discussed in great details in this paper.

Following unbiased Gaussian distributions with covariance matrices \mathbf{B} and \mathbf{R} , the background error ϵ_b and the observation error ϵ_y are supposed to be uncorrelated i.e

$$\begin{aligned}\epsilon_b &\sim \mathcal{N}(0, \mathbf{B}) \\ \epsilon_y &\sim \mathcal{N}(0, \mathbf{R}) \\ COV(\epsilon_b, \epsilon_y) &= 0.\end{aligned}\tag{2}$$

Simply speaking, the inverse of these covariance matrices (i.e. $\mathbf{B}^{-1}, \mathbf{R}^{-1}$) acts as some “weights” given to the different information sources. In fact, these covariances not only describe the variation of estimation/instrument errors but also how they are correlated. These correlations may depend on the spatial distance, time scale or other physical quantities between two state variables or measure points.

Things become slightly more complicated, due to the iterative approach proposed in this work to finely tune covariances. Indeed, some error correlation between *updated* state variables and observations may be induced by the iterative process. Therefore, it is crucial to account for this modified covariance, which will be discussed in full details later, in particular in section 3.

This approach is applied in a large variety of scientific domains, such as weather prediction, geophysical problems, signal processing, control theory etc. The mathematical handlings of data assimilation are mainly two-fold: Kalman filter-type methods based on estimation theory and variational methods related to control theory. Certain equivalences exist between these two families, especially when the transformation operator \mathcal{H} is linear. Both approaches can be derived from Bayes’ theorem (see [5]) where the state estimation \mathbf{x}_a provided by the data assimilation procedure may be apprehended as a compromise between the information of background estimation and the ones of observations. In practice, dealing with nonlinear problems of large dimension via Bayesian approaches remains a computationally expensive task. In this paper, we focus on the framework of linearized variational methods. However, the analysis and algorithms developed later in this paper can be directly applied at each updating of Kalman filter-type methods.

2.2 Variational formulation

In order to better focus on the study of background covariance matrix computation, we suppose in this paper that \mathcal{H} is linear and perfectly known, represented in matrix form by \mathbf{H} from now on. For this reason, we refer to instrument errors when computing the observation matrix \mathbf{R} . In the case of field reconstruction, the observation \mathbf{y} is only supposed to provide a partial information on the true state.

As mentioned in section 2.1, the key idea in variational methods is to find a balance between the background and the observations ([4]) according to the weights represented by the inverse of \mathbf{B} and \mathbf{R} . This leads to the loss function:

$$J_{3D-VAR}(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + \frac{1}{2}(\mathbf{y} - \mathcal{H}(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y} - \mathcal{H}(\mathbf{x}))\tag{3}$$

$$= \frac{1}{2}\|\mathbf{x} - \mathbf{x}_b\|_{\mathbf{B}^{-1}}^2 + \frac{1}{2}\|\mathbf{y} - \mathcal{H}(\mathbf{x})\|_{\mathbf{R}^{-1}}^2.\tag{4}$$

The optimisation problem defined by the objective function of Eq. (4) is called three-dimensional variational method (*3D-VAR*), which can also be considered as the general equation of variational methods without considering the transition model error (i.e. except weak-constraint data assimilation) (see.[5]).

2.3 Best Linear Unbiased Estimator (BLUE)

Given some observed datasets (in our case both \mathbf{x}_b and \mathbf{y}) and the associated error variance, the Best Linear Unbiased Estimator (BLUE) combines both source of information to produce an unbiased linear estimator with minimum posterior variance (see [2]). When $\mathcal{H} = \mathbf{H}$ is linear, the *optimal* solution provided by the variational formulation (Eq.4) is identical to the one obtained by a BLUE under the assumption of independence between \mathbf{x}_b and \mathbf{y} in terms of prior estimation errors. This approach is unbiased and minimises optimally the error variance, *assuming the error covariance matrices are perfectly known*. It also coincides with the maximum likelihood estimator when prior errors of both background state and observations are normally distributed. In this case, the analysed state \mathbf{x}_a can be updated explicitly as:

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}_b)\tag{5}$$

where the Kalman gain matrix \mathbf{K} is defined as:

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}.\tag{6}$$

Under the assumption of linearity, the covariance of analysis error $\epsilon_a = (\mathbf{x}_a - \mathbf{x}_t)$ takes an exact explicit form:

$$\begin{aligned} \mathbf{A} &= \text{COV}(\mathbf{x}_a - \mathbf{x}_t) \\ &= (\mathbf{I} - \mathbf{KH})\mathbf{B}(\mathbf{I} - \mathbf{KH})^T + \mathbf{KRK}^T \\ &= (\mathbf{I} - \mathbf{KH})\mathbf{B}. \end{aligned} \tag{7}$$

Under the assumption that both background and observation errors follow centered Gaussian distributions, it is easy to justify that:

$$\epsilon_a = \mathbf{x}_a - \mathbf{x}_t \sim \mathcal{N}(0, \mathbf{A}). \tag{8}$$

Eq.8 holds when \mathbf{H} is linear. We recall that the uncorrelatedness between ϵ_b and ϵ_y is a crucial assumption for Eq.5-8. Furthermore, the assumption of Gaussianity on prior errors ensures the complete knowledge of posterior errors distribution as a Gaussian vector can be fully represented by its expectation and covariance. Nonetheless, the estimation of posterior covariance in Eq.7, as well as in the iterative tuning methods proposed in this work remains valid as long as the prior information (both \mathbf{x}_b and \mathbf{y}) is unbiased, regardless of the nature of prior distributions. A more general form of BLUE is presented later in section 3.4. Non-Gaussianity in data assimilation problems, for example due to the nonlinearity of \mathcal{H} , has been discussed (e.g [41]).

However, we emphasize that when prior covariances are not well known, the estimation provided by Eq.7 could be very different from the exact¹ output error covariance (later noted as \mathbf{A}_E). It is therefore of highest importance to differentiate between well or loosely known prior covariance matrices. This aspect will be investigated further in section 3.

2.4 Misspecification of \mathbf{B} matrix

From here, we follow the notations given in [15], where \mathbf{B}_E designates the unknown exact background error covariance while \mathbf{B}_A stands for the *assumed* (or guessed) matrix which can be considered as a parametric quantity within data assimilation algorithms. In this section, we focus on the impact on the output error covariance and its estimation given by the misspecification of matrix \mathbf{B}_E (mismatch between \mathbf{B}_E and \mathbf{B}_A). Following the current notation, the standard estimation of output error covariance \mathbf{A}_A (here we have kept subscript A to indicate that this form is obtained from \mathbf{B}_A) provided by the plain-vanilla *3D-VAR* method in Eq. (7) becomes:

$$\mathbf{A}_A = (\mathbf{I} - \mathbf{K}(\mathbf{B}_A)\mathbf{H})\mathbf{B}_A, \tag{9}$$

which is different from the exact output error covariance when the unknown \mathbf{B}_E has merely been approximated by \mathbf{B}_A . The gain matrix \mathbf{K} remains a function of the assumed background covariance matrix \mathbf{B}_A and the one of the observation, \mathbf{R} . The latter is supposed to be perfectly known. In fact, the exact output error covariance \mathbf{A}_E depends on the prior error and the parameters of the algorithm. Therefore, as described by [15], \mathbf{A}_E is in function of \mathbf{B}_E and \mathbf{B}_A :

$$\mathbf{A}_E = (\mathbf{I} - \mathbf{K}(\mathbf{B}_A)\mathbf{H})\mathbf{B}_E(\mathbf{I} - \mathbf{K}(\mathbf{B}_A)\mathbf{H})^T + \mathbf{K}(\mathbf{B}_A)\mathbf{R}\mathbf{K}(\mathbf{B}_A)^T. \tag{10}$$

As we have mentioned before, the final analysis of the assimilation procedure is very much dependent on the specification of the weights given to background and observations, through the error covariances. In fact, when the background matrix is perfectly specified, i.e. $\mathbf{B}_A = \mathbf{B}_E$, the obtained Kalman gain matrix $\mathbf{K}(\mathbf{B}_E)$ is a so called *optimal* gain matrix, which ensures that the trace of \mathbf{A}_E is minimal. Because the covariance of these errors are not well known, it is natural to turn to methods producing *a posteriori* diagnoses of the misspecification of the *a priori* errors, in order to (sequentially) adapt them. For instance, the Desroziers tuning algorithms ([13]) allow the adjustment of the multiplicative ratio (i.e. the total variance) between matrices \mathbf{B}_A and \mathbf{R} , in order to improve the quality of the analysis.

Our goal is somewhat different from the Desroziers tuning algorithm, as we wish to gain a better knowledge about the error correlation pattern/structure of the output analysis. Indeed, the knowledge of error correlation is crucial for posterior analysis and provides a finer information than the error variance alone. We remind in general how a covariance matrix \mathbf{Cov} is related to its correlation matrix \mathbf{Cor} :

$$\mathbf{Cov} = \mathbf{D}^{\frac{1}{2}}\mathbf{Cor}\mathbf{D}^{\frac{1}{2}} \tag{11}$$

where \mathbf{D} is a diagonal matrix with identical diagonal elements of \mathbf{Cov} and thus $\mathbf{D}^{\frac{1}{2}}$ represents the standard deviations.

¹Here, by the term ‘‘exact’’, we refer to the covariance truly corresponding to the remaining errors present in the analysed state, no matter the level of optimality of the chosen assimilation scheme.

2.5 Data assimilation for dynamical systems

Data assimilation algorithms could be applied to dynamical systems thanks to a sequential application of variational methods using a transition operator (from discretized time t^k to t^{k+1}) $\mathcal{M}_{t^k \rightarrow t^{k+1}}$, where

$$\mathbf{x}_{t^{k+1}} = \mathcal{M}_{t^k \rightarrow t^{k+1}}(\mathbf{x}_{t^k}). \quad (12)$$

The forecasting in data assimilation thus relies on the knowledge of transition operator $\mathcal{M}_{t^k \rightarrow t^{k+1}}$ and the corrected state at the current time \mathbf{x}_{a,t^k} . The state correction could be carried out at each time step $t = t^k$ with current observation \mathbf{y}_{t^k} . Typically, the background state is often provided by the forecasting from the previous step, i.e.

$$\mathbf{x}_{b,t^k} = \mathcal{M}_{t^{k-1} \rightarrow t^k}(\mathbf{x}_{a,t^{k-1}}). \quad (13)$$

It is known that as long as the transformation operator \mathcal{H} and the transition operator \mathcal{M} are linear, the analysis based on the variational method (*4D-VAR*) and the Kalman filter leads to the same forecasting result ([17]). As in both cases the approximation of \mathcal{M} may bring extra noises which may probably lead to a nonlinear error propagation, we think an error covariance diagnostic/correction at different time step could be helpful. Not relying on the dynamic of the system, the iterative tuning methods proposed in this paper could be applied at any step in a data assimilation chain.

3 Iterative variational methods with advanced covariance updating

For interpolation of complex industrial applications, the model error, due to the approximation of the transition model \mathcal{M} , is often integrated as a part of the background error. This modelling choice usually leads to a less precise knowledge about the background covariance matrix \mathbf{B} relative to the observation covariance matrix \mathbf{R} . Therefore, we consider that the background errors are dominant over observation errors with a noise-free transformation operator \mathcal{H} , but the exact ratio between them is difficult to estimate. It was pointed out in [15] that an *overestimation* of covariance \mathbf{B} will introduce a significant risk of mis-calculating the output error covariances. As a consequence, the main idea of our iterative methods is to iterate the data assimilation procedure for a better posterior state estimation and error covariance specification, avoiding overestimation of \mathbf{B} . Therefore, the adjustment of the state variables and its covariance associated will take place progressively.

3.1 Naive approach

In practice, the data assimilation procedure can be reapplied several times making use of the same observations, in order to balance the weight between background states and observations. This naive approach may be summarised as:

$$\mathbf{x}_{b,n+1} \leftarrow \mathbf{x}_{a,n} = \mathbf{x}_{b,n} + \mathbf{K}_n(\mathbf{y} - \mathbf{H}\mathbf{x}_{b,n}) \quad (14)$$

$$\mathbf{B}_{A,n+1} \leftarrow \mathbf{A}_{A,n} = (\mathbf{I} - \mathbf{K}_n\mathbf{H})\mathbf{B}_{A,n}, \quad (15)$$

where n refers to the iteration number, and:

$$\mathbf{K}_n = \mathbf{B}_{A,n} \mathbf{H}^T (\mathbf{H}\mathbf{B}_{A,n} \mathbf{H}^T + \mathbf{R})^{-1}, \quad (16)$$

is the iterated Kalman gain matrix.

3.2 Mis-calculation of updated covariances

The updating of error covariances is incorrect because of the state-observation error correlation emerging due to the iterative process. In fact, the evolution of the exact analysed/background error covariance $\mathbf{A}_{E,n}/\mathbf{B}_{E,n}$ can be expressed as a function of $\mathbf{B}_{A,n}$ and \mathbf{K}_n :

$$\begin{aligned} \mathbf{B}_{E,n+1} = \mathbf{A}_{E,n} &= (\mathbf{I} - \mathbf{K}_n\mathbf{H})\mathbf{B}_{E,n}(\mathbf{I} - \mathbf{K}_n\mathbf{H})^T + (\mathbf{I} - \mathbf{K}_n\mathbf{H})\mathbf{Cov}(\epsilon_{b,n}, \epsilon_y)\mathbf{K}_n^T \\ &+ \mathbf{K}_n\mathbf{Cov}(\epsilon_y, \epsilon_{b,n})(\mathbf{I} - \mathbf{K}_n\mathbf{H})^T + \mathbf{K}_n\mathbf{R}\mathbf{K}_n^T, \end{aligned} \quad (17)$$

where $\mathbf{Cov}(\epsilon_{b,n}, \epsilon_y) = \mathbf{Cov}(\epsilon_y, \epsilon_{b,n})^T$ represents the error covariance of $\mathbf{x}_{b,n}$ and y . Indeed, the state errors are no longer uncorrelated to the observation ones after the first iteration, i.e.

$$\mathbf{Cov}(\epsilon_{b,n}, \epsilon_y) \neq 0 \quad \text{for } \forall n \geq 1.$$

As a result, the exact analysis error covariance $\mathbf{A}_{E,n}$ tends to be under-estimated by $\mathbf{A}_{A,n}$ in Eq.15 throughout the iterations.

This is an important drawback that we next attempt to illustrate in a straightforward scalar case, where we assume:

$$B_A, R \in \mathbb{R}^+ \setminus \{0\}, H \in \mathbb{R} \setminus \{0\}. \quad (18)$$

Here, we keep the covariance matrix denomination for notation coherence but they only reflect scalar variances. In this case,

$$B_{A,n+1} \leftarrow \left(1 - \frac{B_{A,n}H^2}{B_{A,n}H^2 + R}\right)B_{A,n} = \frac{B_{A,n}R}{B_{A,n}H^2 + R}. \quad (19)$$

In fact, one may see that the assumed error covariance (scalar variance in this case) $\mathbf{B}_{A,n \rightarrow \infty}$ provided by the naive iterations converges to zero, therefore falsely suggesting a reasonable estimator. This convergence can be easily proved by studying the fixed-point and monotonicity of the function f in \mathbb{R}^+ defined as:

$$f(x) = \frac{xR}{xH^2 + R}, \quad \text{where } R \in \mathbb{R}^+ \setminus \{0\}, H \in \mathbb{R} \setminus \{0\}. \quad (20)$$

Zero is obviously the only fixed-point of f . On the other hand,

$$\forall x \in \mathbb{R}^+ \setminus \{0\}, \quad f(x) < x, \quad (21)$$

and $f^{(n)}(x)$ ($f^{(n)}(x) = f(f^{(n-1)}(x))$) is a decreasing sequence with a lower bound zero, thus it is convergent. Because zero is the only fixed-point of f , we can conclude that $B_{A,n \rightarrow \infty} \rightarrow 0$ for any initial value $B_A \in \mathbb{R} \setminus \{0\}$. This theoretical result is numerically confirmed in Fig 1, (solid green line). The distribution of the exact covariance, consistent with the updating loop (Eq. (17)), is depicted by the dashed green line and remains positive and non-zero.

Based on this idea, we propose two different algorithms named CUTE and PUB, aiming at a better control of the output error correlation, and consequently a reduction of assimilation error. From now on, for the simplicity of analysis, we make further hypothesis about the error covariance matrix \mathbf{R} of observations to be well known.

3.3 CUTE (Covariance Updating iTerativeE) method

As pointed out in section 3.2, the state-observation covariance $COV(\epsilon_{b,n}, \epsilon_y)$ must be taken care of in the covariance updating.

Algorithm

As we have mentioned in the previous sections, when $\mathcal{H} = \mathbf{H}$ is a linear operator, the reconstructed state \mathbf{x}_a can be expressed as a linear combination of \mathbf{x}_b and \mathbf{y} . Therefore, the covariance of updated background state and observations can be estimated sequentially as:

$$\mathbf{Cov}(\epsilon_{b,n}, \epsilon_y) = \mathbf{Cov}(\epsilon_y, \epsilon_{b,n})^T = \mathbf{Cov}\left([\mathbf{I} - \mathbf{K}_{n-1}\mathbf{H}]\epsilon_{b,n-1} + \mathbf{K}_{n-1}\epsilon_y, \epsilon_y\right) \quad (22)$$

$$= (\mathbf{I} - \mathbf{K}_{n-1}\mathbf{H})\mathbf{Cov}(\epsilon_{b,n-1}, \epsilon_y) + \mathbf{K}_{n-1}\mathbf{R}, \quad (23)$$

with

$$\mathbf{Cov}(\epsilon_{b,0}, \epsilon_y) = 0_{\dim(x_b) \times \dim(y)}. \quad (24)$$

In practice, especially in the case of a poor quality of matrix specification *a priori*, we have found that it is helpful to control the trace of matrices $\mathbf{B}_{A,n}$ at each iteration in order to balance the weight of background state and observations. Indeed, if no care is taken of that, the norm of the updated covariance $\mathbf{B}_{A,n}$ may reduce too quickly after the first iteration, thereby causing a neglect of the observation data during the next iterations. Therefore a scaling is introduced through a coefficient $\alpha \in (0, 1)$ related to the confidence level of prior matrix estimation, in order to control the trace of the updating matrix $Tr(\mathbf{B}_{A,n+1})$. The latter, representing the posterior covariance estimation, is introduced in Eq.27.

The complete update of the state and the background covariance matrix is therefore written as:

$$\mathbf{x}_{b,n+1} \leftarrow \mathbf{x}_{a,n}, \quad (25)$$

$$\begin{aligned} \mathbf{A}_{A,n} &= (\mathbf{I} - \mathbf{K}_n\mathbf{H})\mathbf{B}_{A,n} + (\mathbf{I} - \mathbf{K}_n\mathbf{H})\mathbf{Cov}(\epsilon_{b,n}, \epsilon_y)\mathbf{K}_n^T \\ &\quad + \mathbf{K}_n\mathbf{Cov}(\epsilon_y, \epsilon_{b,n})(\mathbf{I} - \mathbf{K}_n\mathbf{H})^T, \end{aligned} \quad (26)$$

$$\mathbf{B}_{A,n+1} \leftarrow \frac{(1 - \alpha)Tr(\mathbf{B}_{A,n}) + \alpha Tr(\mathbf{A}_{A,n})}{Tr(\mathbf{A}_{A,n})}\mathbf{A}_{A,n} \quad (27)$$

where \mathbf{K}_n is expressed in Eq. (16). The more confident we are in the initial guess $\mathbf{B}_{A,0}$, the higher level of α should be set. In the extreme case where the initial background matrix is set arbitrarily (which is not rare in industrial applications), setting $\alpha = 0$ is suggested which means the trace of $\mathbf{B}_{A,n}$ will be kept constant in the iterative process.

Analysis

It should be mentioned that, despite our effort on taking the background-observation covariance into account, the evolution of the error covariance can not be perfectly known due to the misspecification of \mathbf{B} at the first iteration. The evolution of exact analysis/background error covariance $\mathbf{B}_{E,n+1}$, which depend on the set up of $\mathbf{B}_{A,n}$, can be expressed as:

$$\begin{aligned} \mathbf{B}_{E,n+1} \leftarrow \mathbf{A}_n = & (\mathbf{I} - \mathbf{K}_n \mathbf{H}) \mathbf{B}_{E,n} (\mathbf{I} - \mathbf{K}_n \mathbf{H})^T + (\mathbf{I} - \mathbf{K}_n \mathbf{H}) \mathbf{Cov}_E(\epsilon_{b,n}, \epsilon_y) \mathbf{K}_n^T \\ & + \mathbf{K}_n \mathbf{Cov}_E(\epsilon_y, \epsilon_{b,n}) (\mathbf{I} - \mathbf{K}_n \mathbf{H})^T + \mathbf{K}_n \mathbf{R} \mathbf{K}_n^T. \end{aligned} \quad (28)$$

We remind that the term $\mathbf{Cov}_E(\epsilon_{b,n}, \epsilon_y)$, which represents the exact background-observation covariances, is calculated as done in Eq. (23) but using the exact updated covariance matrix $\mathbf{B}_{E,n}$ in the expression of \mathbf{K} .

Estimating the covariance introduced between $\mathbf{x}_{b,n}$ and \mathbf{y} , at each iteration, allows for a more "consistent" update, in the sense that if the estimation of \mathbf{B} and \mathbf{R} becomes asymptotically accurate, the iterative process will not add extra errors to the posterior covariance estimate. However, since the covariance between \mathbf{x}_b and \mathbf{y} emerges, the optimality of a 3D-VAR formula in a loop (Eq. (4)) may be questioned. Therefore, under the assumption of linearity, we propose another formulation that relies directly on the BLUE estimator in an extended space.

3.4 PUB (Partially Updating Blue) method

With the CUTE formulation, we have taken $COV(\epsilon_{b,n}, \epsilon_y)$ into account in the covariance updating but they are not considered in the optimization loss function (Eq. (4)). To overcome this shortage, our idea is to merge the background and observations in a broader space of larger dimension (as shown in [43]) with a partial updating dealing only concerning the part of the background state and its associated covariance. By merging the state and the observation space, the cross-covariances $COV(\epsilon_b, \epsilon_y)$ could be taken into account in the iterative applications of minimisation problems using BLUE-type formulation.

Algorithm

In general, the BLUE estimator consists of constructing an unbiased estimate with minimum of variance from a true state θ , an observation z , a transformation operator $\tilde{\mathbf{H}}$ from the state to the observation space and the observation error covariance matrix C , under the assumption that:

$$\mathbf{z} = \tilde{\mathbf{H}}\theta + \mathbf{w}, \quad (29)$$

where \mathbf{w} is a white noise.

The minimization of state minus observation under the norm defined by the error covariance C is:

$$J(\mathbf{x}) = \frac{1}{2} \|\mathbf{z} - \tilde{\mathbf{H}}\theta\|_{C^{-1}}, \quad (30)$$

and yields the BLUE $\hat{\theta}$ and its output covariance estimation $\mathbf{C}_{\hat{\theta}}$:

$$\begin{aligned} \hat{\theta} &= (\tilde{\mathbf{H}}^T \mathbf{C}^{-1} \tilde{\mathbf{H}})^{-1} \tilde{\mathbf{H}}^T \mathbf{C}^{-1} \mathbf{z}, \\ \mathbf{C}_{\hat{\theta}} &= (\tilde{\mathbf{H}}^T \mathbf{C}^{-1} \tilde{\mathbf{H}})^{-1}. \end{aligned} \quad (31)$$

Here we refer to the general form of the BLUE without any extra assumptions, for example, the uncorrelation between ϵ_b and ϵ_y as in section 2.3. Furthermore, under the assumption of linearity of \mathbf{H} , Eq.31 is equivalent to the Maximum Likelihood Estimator. In order to be well adapted to the general framework of the BLUE estimator, we redefine the system (5-7) by simply combining the observation and the background spaces:

$$\theta \equiv \mathbf{x}_t, \quad \mathbf{z} \equiv \begin{pmatrix} \mathbf{x}_b \\ \mathbf{y} \end{pmatrix}, \quad \tilde{\mathbf{H}} \equiv \begin{pmatrix} \mathbf{I} \\ \mathbf{H} \end{pmatrix}, \quad \mathbf{w} \equiv \begin{pmatrix} \epsilon_b \\ \epsilon_y \end{pmatrix}, \quad \mathbf{C} \equiv \begin{pmatrix} \mathbf{B} & 0 \\ 0 & \mathbf{R} \end{pmatrix}. \quad (32)$$

Similarly to the previous algorithms, we suppose that the matrix \mathbf{B} is misspecified, which yields also a misspecification of the matrix \mathbf{C} in Eq. (33). The assumed covariance matrix in the extended space is denoted

as \mathbf{C}_A ($\mathbf{C}_{A,0}$ latter in the iterative method) with no initial covariance between the background state and the observations, which can be written as:

$$\mathbf{C}_A \equiv \begin{pmatrix} \mathbf{B}_A & 0 \\ 0 & \mathbf{R} \end{pmatrix}. \quad (33)$$

As for the CUTE method, we aim to adjust the structure of error covariance matrix \mathbf{C}_A by taking into account the covariance between the updated background and the observation, which yields the updating loop of PUB method:

$$\mathbf{x}_{b,n+1} \leftarrow \mathbf{x}_{a,n} = (\tilde{\mathbf{H}}^T \mathbf{C}_{A,n}^{-1} \tilde{\mathbf{H}})^{-1} \tilde{\mathbf{H}}^T \mathbf{C}_{A,n}^{-1} \mathbf{z}_n, \quad (34)$$

$$\mathbf{z}_{n+1} = \begin{pmatrix} \mathbf{x}_{b,n+1} \\ \mathbf{y} \end{pmatrix} \quad (35)$$

$$\mathbf{A}_{A,n} = (\tilde{\mathbf{H}}^T \mathbf{C}_{A,n}^{-1} \tilde{\mathbf{H}})^{-1} \quad (36)$$

$$\mathbf{B}_{A,n+1} \leftarrow \frac{(1-\alpha)Tr(\mathbf{B}_{A,n}) + \alpha Tr(\mathbf{A}_{A,n})}{Tr(\mathbf{A}_{A,n})} \mathbf{A}_{A,n} \quad (37)$$

$$\mathbf{Cov}(\epsilon_{b,n+1}, \epsilon_y) = (\tilde{\mathbf{H}}^T \mathbf{C}_{A,n}^{-1} \tilde{\mathbf{H}})^{-1} \tilde{\mathbf{H}}^T \mathbf{C}_{A,n}^{-1} \begin{pmatrix} \mathbf{Cov}(\epsilon_{b,n}, \epsilon_y) \\ \mathbf{R} \end{pmatrix} \quad (38)$$

$$\mathbf{C}_{A,n+1} = \begin{pmatrix} \mathbf{B}_{A,n+1} & \mathbf{Cov}(\epsilon_{b,n+1}, \epsilon_y) \\ \mathbf{Cov}(\epsilon_{b,n+1}, \epsilon_y)^T & \mathbf{R} \end{pmatrix} \quad (39)$$

where $\mathbf{C}_{A,n}$ is the *assumed* error covariance matrix in the combined space of background state and observations. Similar to CUTE, the coefficient α is introduced to balance the ratio between *assumed* covariances in $\mathbf{C}_{A,n}$

Analysis

Let $\mathbf{C}_{E,n}$ denotes the exact iterated error covariance with $\mathbf{Cov}_E(\epsilon_{b,n}, \epsilon_y)$ representing the exact covariance between $\mathbf{x}_{b,n}$ and \mathbf{y} in this extended space, i.e.

$$\mathbf{C}_{E,n} = \begin{pmatrix} \mathbf{B}_{E,n} & \mathbf{Cov}_E(\epsilon_{b,n}, \epsilon_y) \\ \mathbf{Cov}_E(\epsilon_{b,n}, \epsilon_y)^T & \mathbf{R} \end{pmatrix}, \quad (40)$$

where, as for the CUTE method, we can also express the exact background error covariance evolution as:

$$\mathbf{B}_{E,n+1} = (\tilde{\mathbf{H}}^T \mathbf{C}_{A,n}^{-1} \tilde{\mathbf{H}})^{-1} \tilde{\mathbf{H}}^T \mathbf{C}_{A,n}^{-1} \mathbf{C}_{E,n} \left((\tilde{\mathbf{H}}^T \mathbf{C}_{A,n}^{-1} \tilde{\mathbf{H}})^{-1} \tilde{\mathbf{H}}^T \mathbf{C}_{A,n}^{-1} \right)^T. \quad (41)$$

We note that this method does not only take into account the updated variances but also modify the optimisation formula (30) in the extended space. This effect could make the PUB method more robust and less sensitive to prior assumptions, which will be shown later in section 4. However, the implementation of the algorithm, especially the matrix conditioning could be a bit more sophisticated due to the vector space of a larger dimension.

3.5 Comparison of these methods using an illustrative simple scalar case

As we explained earlier, the objective of the proposed iterative methods is to obtain a better knowledge of the covariance (amplitude and/or correlation, depending on the application and prior knowledge) of output errors which can be crucial for future predictions in a data assimilation chain.

Going back to the simple numerical illustration of a scalar case introduced in section 3 and depicted in Fig. (1), we monitor the behaviour of our iterative algorithms for ten steps. Here, we display the evolution of the successive analysed covariance matrices (in fact, we simply look at variances due to the scalar variables). The dashed lines represent the evolution of exact error variance for the different methods (i.e. $\mathbf{B}_{E,n}^{\text{Naive}}$, $\mathbf{B}_{E,n}^{\text{CUTE}}$, $\mathbf{B}_{E,n}^{\text{PUB}}$), that we are capable of computing thanks to our perfect knowledge of the exact prior background variance. The solid lines represent their associated estimators (i.e. $\mathbf{B}_{A,n}^{\text{Naive}}$, $\mathbf{B}_{A,n}^{\text{CUTE}}$, $\mathbf{B}_{A,n}^{\text{PUB}}$). In the left figure, all algorithms start from a perfect knowledge of the prior background variance (i.e. $\mathbf{B}_E = \mathbf{B}_A$). Since we are dealing with a scalar problem, there is no need to adjust the matrix trace, while α is set to be one in all applications of CUTE/PUB (i.e. $\mathbf{B}_{A,n+1} = \mathbf{A}_{A,n}$).

In this case, the estimated variances provided by CUTE and PUB coincide with the evolution of the exact variance. Meanwhile the estimated variance of the naive approach converges to zero, which leads to a significant under-estimation. We remind the reader that the first step of these three iterative methods are the same. In the right figure, we voluntarily under-estimate the exact background error variance at the beginning. We notice that $\mathbf{B}_{A,n}^{\text{CUTE}}$ and $\mathbf{B}_{A,n}^{\text{PUB}}$ are stable after some iterations. This behaviour was verified (not displayed here) no matter the choice of the initial variance. Moreover, for CUTE method, we notice that the estimation of error variance becomes consistent with the exact error variance and they both converge to the observation error variance.

Meanwhile despite being under-estimated by its estimator (solid red line), the exact error variance of PUB (dashed red line) remains inferior to the one of CUTE and 3D-VAR (dashed green line). In both situations, a simple naive iteration of the variational method (green solid lines) leads to an important under-estimation (green curves) of the posterior error variance.

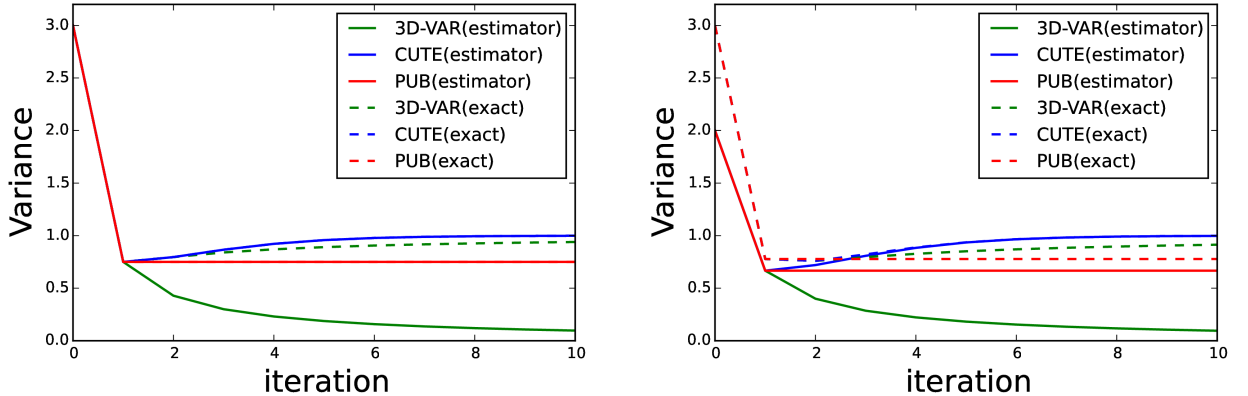


Figure 1: Analysis of the evolution of the exact updated background error variance \mathbf{B}_n (dashed curves) vs its estimation provided by data assimilation algorithms $\mathbf{B}_{A,n}$ (solid curves). On the left side, the prior error variance is perfectly known (i.e. $\mathbf{B}_A = \mathbf{B}_E = 3$) at the initial step. On the right side, the background variance is voluntarily under-estimated: ($\mathbf{B}_A = 2, \mathbf{B}_E = 3$). We remind that the updating of $\mathbf{B}_{A,n}$ is independent of the exact covariance evolution $\mathbf{B}_{E,i=1,\dots,n}$; however, $\mathbf{B}_{E,n}$ is a function of the recurrence $\mathbf{B}_{A,i=1,\dots,n-1}$. The observation error variance is fixed at $R = 1$, perfectly known for both solid and dashed lines.

Unlike our illustration of the scalar case, in a space of larger dimension, these iterative methods may not reach a convergence in terms of error covariance and analysis state. We will discuss later how to define the stopping criteria outside the framework of twin experiments. Under the assumption of lower noise observation level, one well-known quantity that has to be monitored is the innovation quantity: $(\mathbf{y} - \mathbf{H}(\mathbf{x}_{b,n}))$ which will be displayed in the following numerical tests (e.g. Fig. 8, Table. 1).

4 Numerical experiments

Numerical experiments in twin experiments framework are carried out in order to compare the performance of the different methods. This principle is illustrated in Fig. 2 where the background states and the observations are obtained from a chosen true state by adding a *known* artificial noise (dashed line). The objective is to estimate how close is the estimated output to the true state. In this work, it is quantified by computing the expectation of the assimilation error $\mathbb{E}(\|\mathbf{x}_t - \mathbf{x}_a\|_{L^2})$ over the support of the *a priori* noise level, relying on Monte Carlo tests with different realizations of $(\mathbf{x}_b, \mathbf{y})$. More precisely, the original background state $\mathbf{x}_{b,n=0}$ (n stands for the number of iterations in CUTE, PUB) and observation \mathbf{y} (via \mathbf{H}) are first constructed from a chosen true state \mathbf{x}_t , thanks to the exact knowledge of \mathbf{B} and \mathbf{R} . In our experiments, \mathbf{x}_t is obtained by a reference simulation.

4.1 Description of the system

In the following twin experiments, we consider a standard shallow-water fluid mechanics system which is frequently used for evaluating the performance of data assimilation algorithms (as in [42], [9]). The wave-propagation problem is nonlinear and time-dependent. The initial condition is chosen in the form of a cylinder of water of a certain radius that is released at $t = 0$. We assume that the horizontal length scale is more important than the vertical one and we also neglect the Coriolis force. They lead to the Saint-Venant equations ([37]) coupling the fluid velocity and height as shown in Eq.42 where (u, v) are the two components of the two dimensional fluid velocity (in $0.1m/s$) and h stands for the fluid height (in millimeter). The earth gravity constant g is thus scaled to 1 and the dynamical system is defined in a non-conservative form.

The initial values of u and v are set to zero for the whole velocity field and the height of the water cylinder is set to be $h_{t=0}^{CYL} = 0.1mm$ high above the one of the still water as shown in Fig. 3. The domain of size $(L_x \times L_y) = (100mm \times 100mm)$ is discretized with a regular structured grid of size (100×100) and the solution of Eq. (42) is approximated thanks to a finite difference method of first order. The time-integration is also first-order with a time interval $\delta t = 10^{-6}s$. The system is integrated up to a time $t_f = 1.5 \times 10^{-3}s$ (see Fig. 4(c-d)), and the obtained solution is used as the reference state (\mathbf{x}_t) in 4.2 and 4.3. Our objective is to

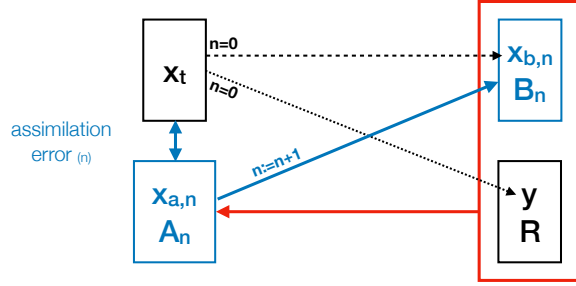


Figure 2: Scheme of a twin experiments data assimilation framework for an iterative method. Quantities in black are kept fixed while iterations are repeated: new assimilated state $\mathbf{x}_{a,n}$ and covariance errors \mathbf{A}_n are injected at the next step in order to update background quantities. The difference (in some norm) between the true state and the output of the algorithm $\|\mathbf{x}_t - \mathbf{x}_{a,n}\|$ is called the error of reconstruction and may be monitored. The entire experiment may be repeated numerous times for different realisations of $\mathbf{x}_{b,n=0}$ to collect statistics of the assimilation results in order to assess the method robustness.

$$\begin{aligned}
 \frac{\partial u}{\partial t} &= -g \frac{\partial}{\partial x}(h) - bu & (42) \\
 \frac{\partial v}{\partial t} &= -g \frac{\partial}{\partial y}(h) - bv \\
 \frac{\partial h}{\partial t} &= -\frac{\partial}{\partial x}(uh) - \frac{\partial}{\partial y}(vh) \\
 u_{t=0} &= 0 \\
 v_{t=0} &= 0.
 \end{aligned}$$

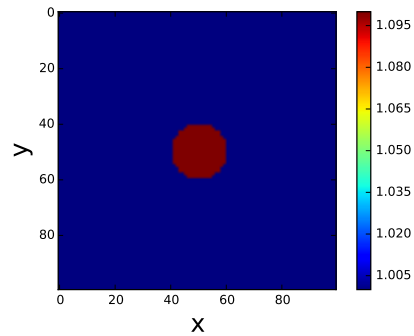


Figure 3: Initial h of shallow water in mm

reconstruct the state $x = (u, v)$ within a non-centered (10×10) subdomain (represented by a red square in Fig. 4 (c) and (d)) from noisy measurements via data assimilation processes. Thanks to an observation operator \mathbf{H} described later, we will use a collection of observations from the subdomain. Therefore, the dimension of the state space (i.e. $\mathbf{x}_t, \mathbf{x}_a, \mathbf{x}_b$) which combined two 2D fields u and v may be algebraically combined in an array of size 200, i.e. $\mathbf{x}_t \equiv \{\mathbf{x}_t(k)\}_{k=1\dots 200} \equiv \{(u(t=t_f), v(t=t_f))\}$. The observation vector \mathbf{y} is of size 100 but with zero elements included as shown in Fig. 4 (b).

In all numerical tests, we keep the assumption of linearity of the transformation operator \mathcal{H} . As we have mentioned in section 2.2, \mathcal{H} could represent a transformation between two physical quantities/fields or even include discretized forecast/model operators. In this work, we wish to remain as general as possible. Therefore, we prefer not to set a particular form of the observation operator, which would promote some space-filling properties or some other type of optimality. With this aim, we decide to model the observation operator with a random matrix \mathbf{H} acting as a binomial selection operator. Each observation will be constructed as a sum of a contribution from a linear combination of a few true state variables randomly collected over the subdomain and some random noise. In order to do so, we introduce the notation for a subset sample $\{\mathbf{x}_t^*(i)\}_{i=1\dots n}$ randomly but homogeneously chosen (with replacement) with probability P among the available data set, i.e. $\{\mathbf{x}_t(k)\}_{k=1\dots 200}$. The subset values \mathbf{x}_t^* are summed up and the process is re-iterated 100 times in order to construct the observations:

$$y(j) = \sum_{i=1}^{n_j} \mathbf{x}_t^*(i) + \epsilon_y, \quad \text{for } j = 1, \dots, 100, \quad (43)$$

where the size n_j of the collected sample used for each j^{th} observation data point $y(j)$ is random and by construction follows a binomial distribution $\mathcal{B}(200, P)$. In the following we choose a sparse representation with $P = 1\%$.

Once \mathbf{H} is randomly chosen, it is kept fixed for a whole set of numerical experiments. This operator \mathbf{H} is shown in Fig. 4 ((c) and (d)). In fact, with this definition of \mathbf{H} , the observed quantities can be apprehended as some sorts of barycenters in the state space. As explained, the number of points in the field associated to each barycenter can thus be seen as a random variable of binomial distribution as shown in Fig. 4 (d). If we increase the probability of success of the selection operator, more points will be selected and combined across the domain, resulting in a more centered barycenters distribution.

We have numerically verified that in general, the results obtained with a transformation operator \mathbf{H} of more

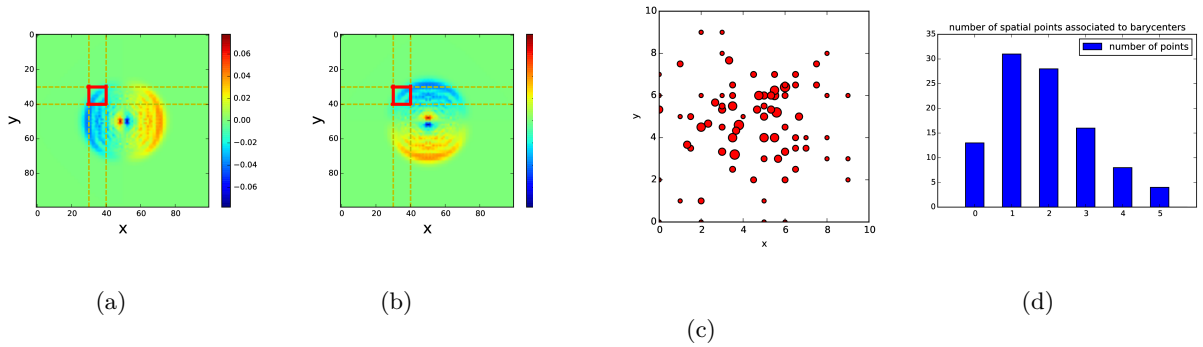


Figure 4: Illustrations of the random observation operator \mathbf{H} and example of 2D flow velocity fields of the shallow water model. Barycenters measured by the linear transformation operator \mathbf{H} are shown in (c) where the symbol radius is proportional to the number of measures associated to each barycenter. The histogram of the number of selected points associated to each barycenter (rows in matrix \mathbf{H}) is shown in (d) and is reminiscent of a binomial distribution. Shallow water 2D velocity fields are represented in (a) and (b) (respectively for u and v in Eq. (42)) at time $t = t_f$. Data assimilations are performed in the red square subdomain.

regular span structure tend to be less optimal than the ones obtained in the case with randomly simulated \mathbf{H} , in terms of output correlation identification. We believe that repeated assimilations based on the same uniform data set of observations may unwillingly put emphasis on certain correlation lengths while missing others, in relation to the structure of \mathbf{H} .

4.2 Experiments with state-independent homogeneous prior errors

For the sake of simplicity, in the analysis of data assimilation algorithms, prior background and observation errors (i.e. ϵ_b and ϵ_y) are often supposed to be independent of their theoretical values (i.e. respectively \mathbf{x}_t and $\mathbf{H}\mathbf{x}_t$). Under this assumption, the assimilation error depends only on the prior errors ϵ_b and ϵ_y , as:

$$\begin{aligned} \mathbf{x}_a - \mathbf{x}_t &= \mathbf{x}_b + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}_b) - \mathbf{x}_t \\ &= \epsilon_b + \mathbf{K}(\epsilon_y - \mathbf{H}\epsilon_b) \\ &= (\mathbf{I} - \mathbf{K}\mathbf{H})\epsilon_b + \mathbf{K}\epsilon_y. \end{aligned} \quad (44)$$

Therefore, the numerical results shown in this section are independent from the choice of the true state, and therefore valid for any (2D) field reconstruction with state-independent prior noise. Here, background states and observations are simulated using chosen error covariances matrices \mathbf{B}_E and \mathbf{R} . Our assumption of higher background error amplitude leads to:

$$Tr(\mathbf{B}_E) > Tr(\mathbf{R}). \quad (45)$$

In our experiments, the average standard deviation of the background error is set to be at least 10 times higher than the observation error. We make further assumption that the correlation pattern of background covariance is poorly known. In order to make numerical tests representative, we make use of homogeneous and isotropic (invariant under rotations and translations) one-dimensional correlation patterns (of spatial euclidean distance $r = \sqrt{\Delta_x^2 + \Delta_y^2}$) for simulating true or initially estimated background errors (i.e. \mathbf{B}_E and $\mathbf{B}_{A,n=0}$). We consider the following correlation function types:

- Exponential type: $\phi(r) = \exp(-\frac{r}{L})$,
- Balgovind type: $\phi(r) = (1 + \frac{r}{L}) \exp(-\frac{r}{L})$,
- Gaussian type: $\phi(r) = \exp(-\frac{r^2}{2L^2})$,

where L is defined as the typical correlation length scale. These correlation functions are part of the Matérn family of covariance function (respectively of order $\nu = 1/2, 3/2$ and ∞) and are often used as imposed structures in background matrix construction (see [39], [34]). For the sake of simplicity, in this section the correlation kernel of the exact background covariance matrices are always chosen to be of Balgovind type with scale length $L = 2$, where observation errors are supposed to be spatially independent (i.e. \mathbf{R} is proportional to an identity matrix). The latter is supposed to be known in the algorithms. Because both the amplitude and the correlation pattern of \mathbf{B}_E are supposed to be poorly specified by \mathbf{B}_A , we choose to set the coefficient of confidence $\alpha = 0$ for the trace operator in all following numerical tests.

In order to verify the robustness of the proposed methods, different scenarios are considered for the correlation pattern of the initial assumed covariance $\mathbf{B}_{A,n=0}$. As mentioned in section 3, the objective of our algorithms is to improve the output error correlation estimation, and in consequence, obtain a reduction of assimilation error. In fact, using Eq. (11), the error correlation matrices associated to $\mathbf{B}_{E,n}$ (exact background error correlation at n^{th} step) and $\mathbf{B}_{A,n}$ (estimation of error correlation at n^{th} step) can be extracted and compared at each iteration. Our objective is therefore to reduce the dissimilarity between these two correlation matrices, the monitoring of this distance taking different forms: – through a simple correlation calibration or – other correlation dissimilarity measure such as the *Affine Invariant Riemannian Metric* (AIRM) ([8]) defined for two semi-positive definite matrices \mathbf{X} and \mathbf{Y} by:

$$D_{\text{AIRM}}(\mathbf{X}, \mathbf{Y}) = \|\log(\mathbf{X}^{-1/2}\mathbf{Y}\mathbf{X}^{-1/2})\|_F \quad (46)$$

where $\|\cdot\|_F$ represent the Frobenius norm of matrices. This similarity measure is widely used as it integrates the knowledge of the manifold structure of the covariance matrices. In our cases, the two semi-positive definite matrices to be compared by AIRM are the assumed background error correlation matrix $\mathbf{Cor}_{\mathbf{B}_{A,n}}$ and the exact background error correlation matrix $\mathbf{Cor}_{\mathbf{B}_{E,n}}$ defined respectively from covariance matrices $\mathbf{B}_{A,n}$ and $\mathbf{B}_{E,n}$ as shown in Eq. (11). According to [33], invariant under linear transformations, AIRM can be seen as a natural choice of metric for symmetric semi-positive definite matrices.

Fig. 5-7 and Table 1 represent the results of twin experiments with different mis-specified (in terms of both amplitude and error correlation) background matrices \mathbf{B}_A , where CUTE and PUB are (arbitrarily) applied for 10 iterations. In each experiment, the true state \mathbf{x}_t is set to be the shallow water solution at $t = t_f$ in an approximation sub-space defined by the finite difference method. We remind that under the assumption of state-independent prior errors, both the output error and its spatial correlation is independent from the choice of the true state. For the Monte-Carlo validations, 10000 background states are simulated independently following a multivariate Gaussian distribution centred at the true state \mathbf{x}_t of fixed background error amplitude with $(\sigma_b = 10 \times \sigma_o = 0.01m/s)$ and imposed correlation kernel (exponential, Balgovind or Gaussian). We show explicitly the evolution of assimilation error as well as posterior error correlation (both the exact correlation kernel and the estimation given by CUTE and PUB).

More specifically, the distribution of background error correlations is shown in sub-figures (a) where the exact original error correlation of \mathbf{B} (black solid line with triangles) and its estimator ($\mathbf{B}_{A,n=0}$, green solid line with circles), both being homogeneous and isotropic, are drawn against spatial distance r (mm). In order to avoid sampling error for large distance, the error correlation is only considered for $r \in (0, 10)$ in a 10×10 grid. The evolution of average background/analysis error $\|\mathbf{x}_t - \mathbf{x}_{b,n}\|$ in CUTE and PUB is shown in sub-figures (b), compared with the analysis error level obtained by a one-shot $3D$ -VAR algorithm (the starred green line) and the results of the same $3D$ -VAR with the *exact* background error covariance matrix (i.e. $\mathbf{B}_A = \mathbf{B}_E$, represented by the dashed black line). The results obtained using the exact background matrix are considered as the optimal target in our study. We observe in (b) (Fig. 5-7) that for both proposed approaches, the average values of the analysed error decrease significantly with algorithm iterations. In fact, the first step of CUTE and PUB is equivalent to a $3D$ -VAR with mis-specified \mathbf{B}_A (starred green line). Then, the experiments show that both assimilation errors of CUTE (blue curve) and PUB (red curve) decrease and remain stable while approaching better the optimal result (dashed black curve) after a sufficient number of iterations.

Standard deviations of the estimators are also displayed with transparent shades. Fig. 5 (c) shows the decrease of the innovation quantity $\|y - \mathbf{H}\mathbf{x}_{b,n}\|$. We consider the innovation quantity, available outside the framework of twin experiments, as an appropriate stopping criteria for CUTE and PUB algorithms, because of its coherence with the assimilation error (Fig. 5 (a)) both in terms of monotonicity and stability.

Despite the fact that output error correlation recognition is significantly improved by CUTE and PUB (as shown in Fig. 5 (d-f) and the correlation mismatches in Table (1), little impact was found on reduction of the output error deviation as shown in the transparent shades in sub-figures (b). The posterior correlation kernels (both the exact one and its estimators), shown in Fig. 5 (d-e) and used to calculate the correlation mismatch in Table 1, are estimated from the data sample by calculating the average correlation value for all pairs of points sharing the same spatial distance in the 2D velocity field of u . Correlation kernels obtained in the velocity field of v are very similar. Compared to the prior scenario, with all three initial guess of prior correlation kernel, the bias of the correlation error estimation is significantly reduced *a posteriori*. This improvement is also very noticeable when examining the L^2 norm of the correlation mismatches as displayed in Table 1. Sub-figure (f) demonstrates that the AIRM criteria decreases significantly for both approaches after several iterations. It is particularly stable for the PUB method but exhibits some asymptotic non-monotonicity for the CUTE method.

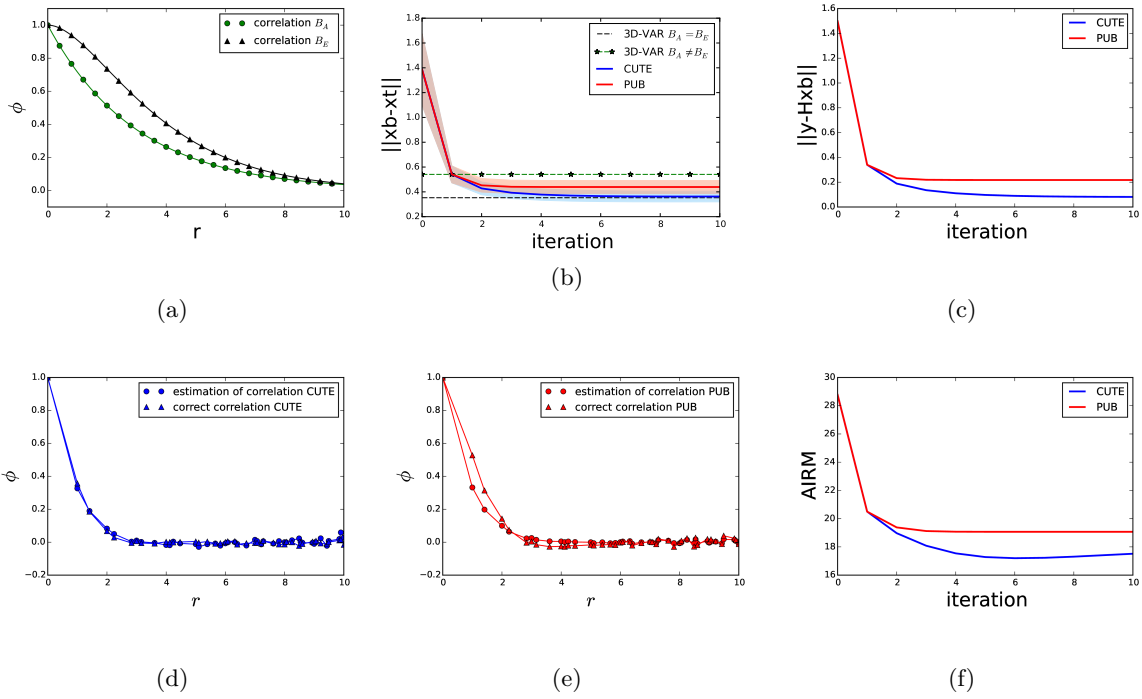


Figure 5: Twin experiments with state-independent homogeneous prior error. Figures on the first line refer to the initial choice of prior correlations (a) and the evolution of assimilation error (b) and innovation quantities (c), while figures on the second line monitor iterated quantities extracted from the errors covariance in the velocity field of u (d-f). In this test, $\mathbf{B}_{A,n=0}$ is chosen to follow an exponential kernel with $L = 3$ (shown by the green curve in (a)).

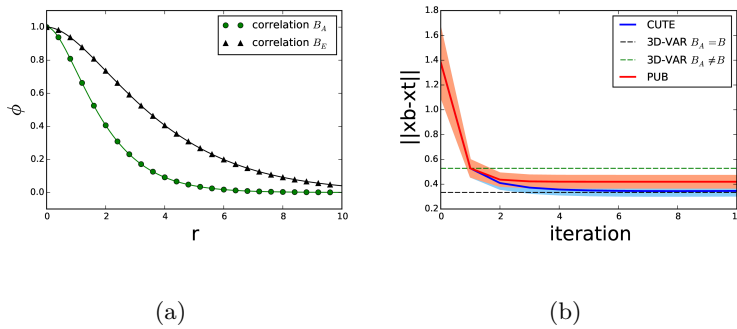


Figure 6: Evolution of assimilation error in twin experiments using same simulated observations as Fig. 5 with different initial background matrix estimation ($\mathbf{B}_{A,n=0}$ is of Balgovind type with scale length $L = 1$).

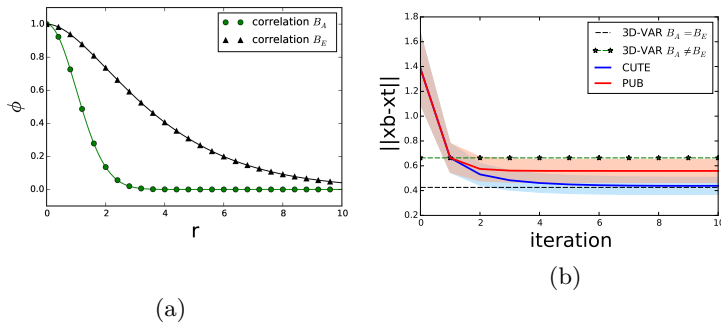


Figure 7: Evolution of assimilation error in twin experiments using same simulated observations as Fig. 5 with different initial background matrix estimation ($\mathbf{B}_{A,n=0}$ is of Gaussian type with length scale $L = 1$).

$\mathbf{B}_{A,n=0}$ kernel choice	Correlation mismatch (u)			AIRM		
	Initial	CUTE	PUB	Initial	CUTE	PUB
Exponential ($L = 3$)	0.667	0.115	0.251	28.772	17.510	19.069
Balgovind ($L = 1$)	1.310	0.140	0.174	23.095	15.607	15.116
Gaussian ($L = 1$)	1.834	0.303	0.660	26.642	19.518	20.957

In conclusion, the iterative approaches improve the assimilation both in terms of reliability of the analyzed error covariance estimate as well as accuracy of the analyzed state. Nevertheless, the final result seems to remain dependent, to some extent, to the level of dissimilarity between the initial guess \mathbf{B}_A and the exact \mathbf{B} . In particular, poorer results are obtained when the prior background error correlation distributions are extremely misspecified (e.g. Gaussian($L = 1$)), regardless of the type of correlation kernel structures considered.

All these numerical results and analyses are obtained under the assumption of a high level of background error variance amplitude, which are here under-estimated by the assumed covariance matrix \mathbf{B}_A (i.e. $Tr(\mathbf{B}_A) < Tr(\mathbf{B}_E)$). This assumption is consistent with the phenomenon of background error inflation as mentioned in 2.2. We remind that as the dimension of the observation space is inferior to the one of the state space, the equation

$$\mathbf{y} = \mathbf{H}\mathbf{x} \quad (47)$$

is underdetermined. It thus defines a hyperplane in the space of \mathbf{x} . Fig. 8 (b) shows that the CUTE method converges to a stable state when the assimilation error $\|\mathbf{x}_{a,n}^{CUTE} - \mathbf{x}_t\|$ is very close to the optimal target $\|\mathbf{x}_a^{\text{optimal}} - \mathbf{x}_t\|$. However, we don't necessarily have

$$\mathbf{x}_{a,n} \rightarrow \mathbf{x}_a^{\text{optimal}}. \quad (48)$$

4.3 Twin experiments with state-dependent prior errors

In this section, we are interested in the performance of our methods in the case of state-dependent errors, i.e. when the assumption of independence between the true state and estimation errors no longer stands (such as the optimal property of the maximum likelihood). State-dependent uncertainties are certainly more complex but it is more realistic for numerous industrial applications. Very recent effort was given along this path in order to improve data assimilation algorithms, e.g. ([3]).

As for the case of homogeneous prior errors, background states and observations are simulated by Gaussian distributions centred around true values (i.e. respectively \mathbf{x}_t and $H\mathbf{x}_t$ for background states and observations). However, the standard deviation at each coordinate is set to be proportional to the magnitude of the true state, while keeping the prior correlation structures as described in 4.2.

In order to better define how state-dependent prior errors are simulated, we denote \mathbf{D}^B (resp. \mathbf{D}^R) as the diagonal of the exact covariance matrix \mathbf{B}_E (resp. \mathbf{R}_E). Above assumption of state-dependent errors leads to:

$$\mathbf{D}_i^B = (\mu_b \times \mathbf{x}_{t,i})^2 \quad (49)$$

$$\mathbf{D}_i^R = (\mu_o \times (H\mathbf{x}_t)_i)^2, \quad (50)$$

where i refers to an index mapping to the two dimensional fields and (μ_b, μ_o) stand for two real coefficients. Combining this state-dependent variance and a homogeneous structure of error correlation, state-dependent error covariance matrices can be written as:

$$\mathbf{B}_E = (\mathbf{D}^B)^{\frac{1}{2}} \mathbf{Cor}_{B,E} (\mathbf{D}^B)^{\frac{1}{2}} \quad (51)$$

$$\mathbf{R}_E = (\mathbf{D}^R)^{\frac{1}{2}} \mathbf{Cor}_{R,E} (\mathbf{D}^R)^{\frac{1}{2}},$$

where the exact prior correlation matrices $\mathbf{Cor}_{B,E}$ and $\mathbf{Cor}_{R,E}$ are still chosen to follow homogeneous and isotropic correlation kernels as in the case of 4.2.

The two velocity fields u and v are supposed to be uncorrelated in terms of prior estimation error. Thus both \mathbf{B}_E and \mathbf{B}_A follow a block diagonal structure. This is obviously a very crude assumption in the context of incompressible fluid mechanics systems. Since observation errors are also state-dependent in this case, the associated observation error covariance cannot be known exactly *a priori*. We introduce the notation R_A for assuming observation error covariance, which is different from the true observation error covariance only in this section (4.3). The assumption of relatively higher background error is also respected by setting 10% standard deviation for background state (i.e. $\mu_b = 10\%$) while 1% for observations (i.e. $\mu_o = 1\%$) in twin experiments.

Monte Carlo twin experiments of 10000 tests with state-dependent prior errors are carried out as presented in Fig. 8-10. We keep homogeneous structure of the assumed covariance matrices \mathbf{B}_A (constructed using correlation kernels) and R_A (set to be the identity matrix) as in section 4.2. We also choose to keep the trace of \mathbf{B}_A and \mathbf{R}_A during iterative processes CUTE/PUB. Results in Fig. 8-10 and Table 2 show that for state-dependent errors, CUTE and PUB iterative methods could also significantly reduce the output errors (sub-figures (b) of Fig. 8-10) compared to the first iteration (standard *3D-VAR* algorithm), as well as the innovation quantity (sub-figures (c)). The latter remains an appropriate candidate for the stopping criteria. Important improvements are obtained in terms of decreasing the bias of error correlation estimation as shown in sub-figures (e) and (f) (comparing with (a)). However, as shown in Fig. 8 (f) and in the last two columns in Table 2, the AIRM

criteria which monitors a global correlation matrix estimation mismatch, reveals a risk of saturation of the use of the same observation data set for the CUTE method after a certain number of iterations. This is due to the imperfect knowledge of observation error covariance. The PUB method is less sensitive and more stable in this case. However, as for the case of state-independent prior errors, CUTE owns a slight advantage over PUB in terms of assimilation error reduction.

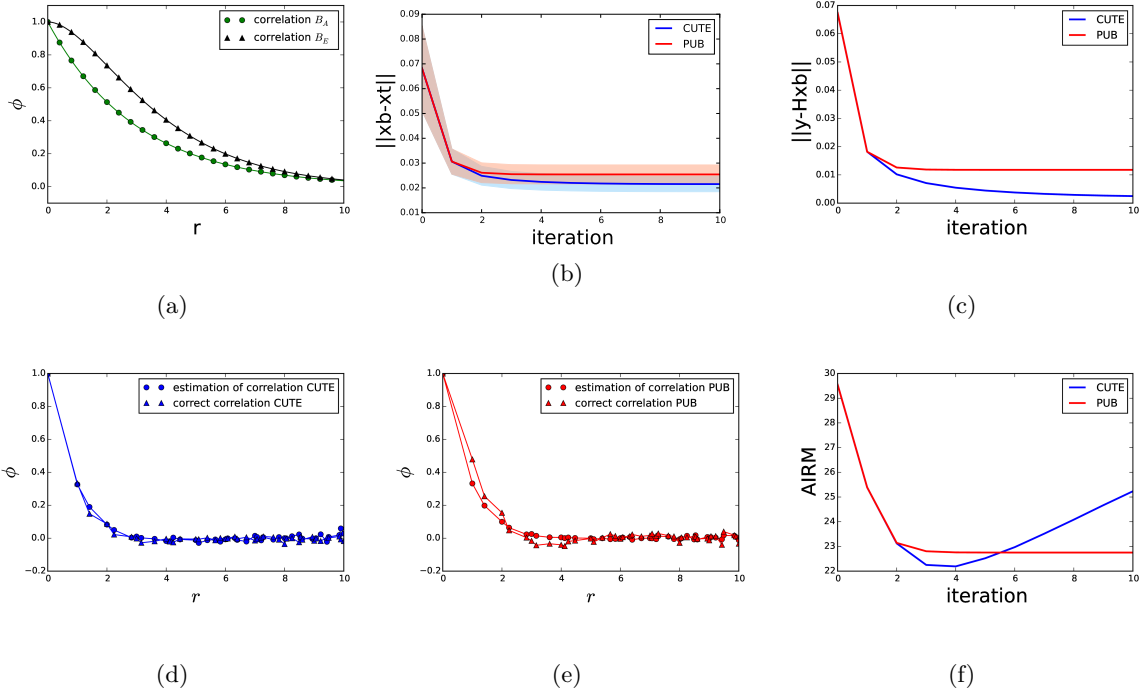


Figure 8: Twin experiments with state-dependent prior error for background state and observations. Figures on the first line refer to the initial choice of prior correlations (a) and the evolution of assimilation error (b) and innovation quantities (c), while figures on the second line monitor iterated quantities extracted from the errors covariance in the velocity field of u (d-f). In this test, the correlation of $\mathbf{B}_{A,n=0}$ is chosen to follow an exponential kernel with $L = 3$ (shown by the green curve in (a)).

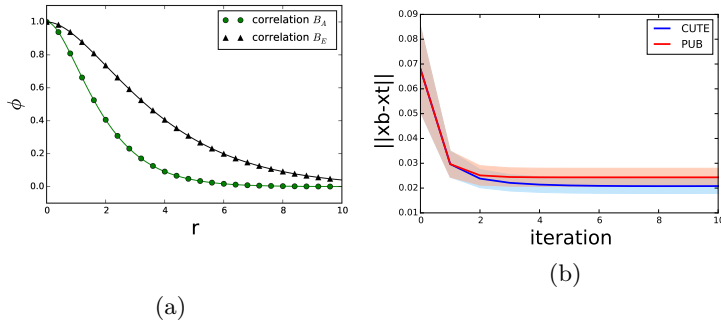


Figure 9: Evolution of assimilation error in twin experiments using same simulated observations as Fig. 8 with different initial background matrix estimation (the correlation kernel of $\mathbf{B}_{A,n=0}$ is of Balgovind type with length scale $L = 1$).

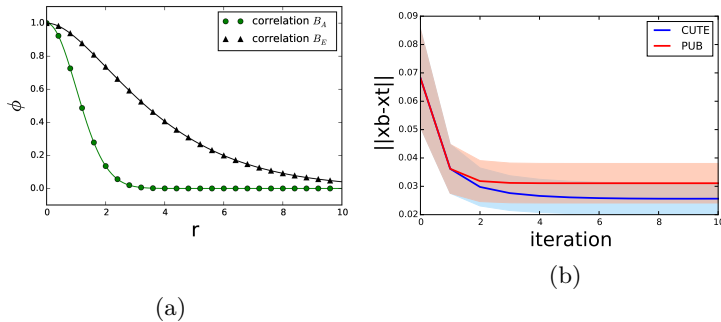


Figure 10: Evolution of assimilation error in twin experiments using same simulated observations as Fig. 8 with different initial background matrix estimation (the correlation kernel of $\mathbf{B}_{A,n=0}$ is of Gaussian type with length scale $L = 1$).

$\mathbf{B}_{A,n=0}$ kernel choice	Correlation mismatch(u)			AIRM		
	Initial	CUTE	PUB	Initial	CUTE	PUB
Exponential ($L = 3$)	0.586	0.147	0.220	29.550	25.234	22.752
Balgovind ($L = 1$)	1.191	0.207	0.180	24.181	24.439	19.785
Gaussian ($L = 1$)	1.733	0.333	0.662	27.495	28.068	23.569

Table 2: Quantification of assimilation results of the CUTE and PUB iterative methods in terms of error correlation identification at the tenth iteration. The prior error covariance is set to be non homogeneous and state-dependent with correlation matrix of Balgovind type, $L = 2$. The mismatch of calibrated correlation functions is calculated with an L^2 norm error between the calibrated one-dimensional correlation curves. The AIRM criteria is also reported.

4.4 Twin experiments in a successive data assimilation process of reconstruction/prediction

The idea of this section is to anticipate on the use of these types of approaches in the wider framework of time-dependent data assimilation problems. Based on the shallow water propagation model introduced in section 4.3, we construct new twin experiments of a dynamical field reconstruction and prediction relying on successive applications of data assimilation algorithm using flow-independent background matrix \mathbf{B}_A . The choice of the test model is made for its simplicity and for better revealing the impact of CUTE and PUB methods. The state dimension remains 200 which is composed of two squarely meshed velocity fields of 10×10 each as for static reconstruction with state-independent prior errors in 4.2. In order to focus on the impact of background error propagation, correct boundary conditions are simulated independently in an error free framework and provided at each reconstruction step for state-transition model to avoid an overlay of model resolution error. In order to observe the impact of CUTE and PUB methods in a long term data assimilation procedure, we choose to apply solely CUTE and PUB at the first reconstruction step of the process for a fixed number of iterations n ($n = 10$ in following numerical tests), following $3D$ -VAR reconstructions every $2 \times 10^{-3}s$. With a significant improvement of assimilation error reduction and error correlation recognition provided by CUTE or PUB, this advantage should be recognised and kept by a standard variational method (in our case, the $3D$ -VAR method) for several further steps in a data assimilation chain. We then compare the results obtained by a standard approach of $3D$ -VAR all the way along. We remind that the difference among the three data assimilation processes shown in Fig. 11 are only the first reconstruction at $t = 10^{-3}s$. The evolution of average assimilation error of 100 independent dynamical simulations is illustrated in Fig. 11 with two different levels of initial errors. We observe a significant improvement due to the iterative process of CUTE and PUB at the first several reconstruction steps. The gaps among the three curves then tend to disappear. In fact, even starting with a high level of noise, a standard successive data assimilation process should be capable of providing a reasonably good long term prediction for both assimilation error reduction and covariance recognition when the information about the state-transition model is accurate enough ([35]), which is the case in our experiments.

The result in Fig. 11 confirms the interest of applying CUTE, PUB methods for a short term prediction. It is also shown that when the assumption of high level or inflated background errors is well respected (Fig. 11 (b)), the "advantage" of an iterative process at the initial step could be kept longer in the dynamical assimilation. However, when the observation error is not sufficiently negligible relatively to the background error, a continuous correction by iterative processes is helpful. The same holds when the information about the dynamical state-transition is not precise, especially in a highly nonlinear system where the misspecification of estimation errors could be enlarged in the successive predictions. Therefore, in these cases an interest can be arisen to apply the

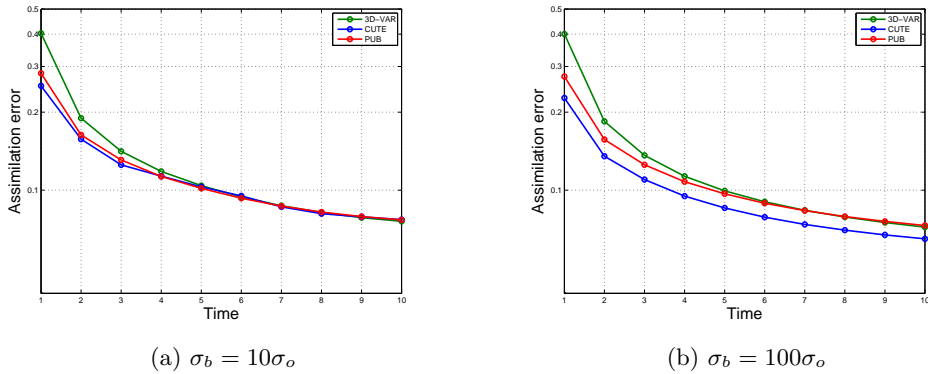


Figure 11: Comparison of standard $3D$ - VAR method with iterative methods in terms of evolution of assimilation error in a dynamical twin experiments framework, where a data assimilation reconstruction takes place every $2 \times 10^{-3}s$. Semi-log grid is used for ordinate coordinates. In these experiments, iterative methods CUTE and PUB are only applied at the first reconstruction of the process, followed later by standard $3D$ - VAR . Simulations are made based on two different level of prior background-observation error: $\sigma_b = 10\sigma_o$ (a), $\sigma_b = 100\sigma_o$ (b).

iterative methods continuously at several different moments. We present in Fig. 12, the same dynamical twin experiments with an implementation of CUTE, PUB at each assimilation step (i.e. every $2 \times 10^{-3}s$) instead of $3D$ - VAR . By construction, the first steps (both (a) and (b)) of Fig. 11 and 12 are equivalent.

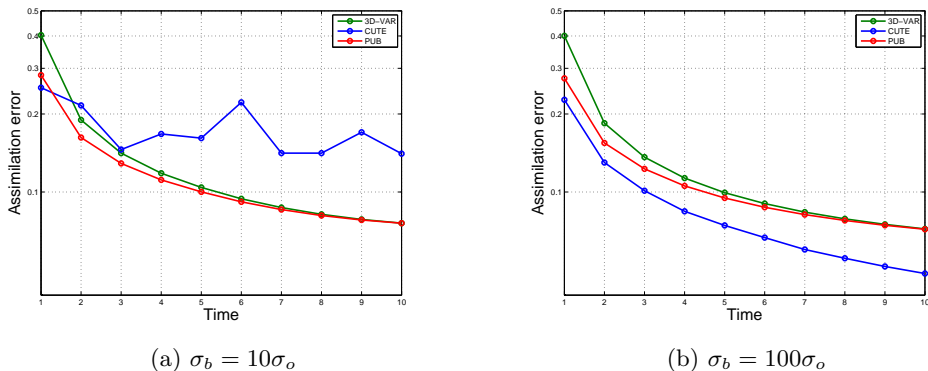


Figure 12: Comparison of *dynamical* twin experiments, with same initial conditions as in Fig. 11, where ten CUTE and PUB sub-iterations are performed at each assimilation step, illustrated by the green and red disk symbols.

We observe clearly from Fig. 12 that, when the observation error is negligible compared to the background error, the implementation of CUTE method at each assimilation step enables a continuous reduction of assimilation error. On the other hand, consistent with previous analysis, the CUTE method is very sensitive to the level of observation errors, especially when being reapplied several times in a dynamical procedure with a good knowledge of the transition model (no inflation of background errors) as presented in Fig. 12 (a). In general, the PUB method remains more robust and less sensitive to the hypothesis of the high level of background error.

5 Conclusion

In this paper we introduced two novel data assimilation iterative methods recycling the observation data for the purpose of damping the detrimental effect of a poor knowledge of the background error covariance. In this framework, we have shown that a naive approach which neglects the background-observation correlation introduced by the iterative process is prone to failure. This indicates that there is a need for a complete covariance updating, as being carried out in the proposed approaches.

Under the assumption of perfect knowledge of the observation error covariance and the transformation operator, we numerically demonstrated that CUTE and PUB methods could noticeably improve output error correlation identification as well as reduce the assimilation error for a variety of initial guesses of the background error covariance matrix, when prior errors are either state-independent or state-dependent. These two methods are different from other iterative methods, in the sense that they not only update the variance of state components but also the background state correlation structure. Other covariance tuning methods, such as the full Desroziers diagnostic used in the observation space, require more data especially for large-scale problem. Originally developed for the purpose of statistical reconstructions or short term predictions, we have shown

that there might be an interest in reapplying the proposed algorithms several times in a dynamical assimilation chain. Limitations of these two methods have also been pointed out in this article, in particular, concerning the risk of straining a redundant observation data set without a careful monitoring of the convergence results.

The difference between CUTE and PUB resides mainly in the minimization function, where the covariance between updated background and observation is only taken into account into the PUB method. This feature makes the PUB method more robust, i.e. less sensitive to the assumption of the trace of the prior errors $Tr(\mathbf{B}_E) > Tr(\mathbf{R}_E)$ and the usage of the same observation data set. Numerically, we have also found that the performance of CUTE can be more optimal when the background error is much underestimated. In fact, the estimation of $COV(\epsilon_b, \epsilon_y)$ is also based on the prior knowledge of matrix \mathbf{B} . In summary, we recommend the utilisation of CUTE when the initial background error covariance matrix is set arbitrarily, especially when it is probably underestimated while the PUB method can be more appropriate when limited data are available for making a rough estimation of \mathbf{B} or its diagonal.

In terms of computational cost, CUTE and PUB methods can be relatively more expensive than the Desroziers approach which only requires *a posteriori* computation of matrix traces. However, when no linearization of the observation operator is needed, the updating process can be done aside once the initial guess for covariance matrices are available and independently from the current background state. This feature promotes a more flexible use of these methods with much lower computational overheads. Future work will investigate along this path of research and will assess their performance in a more realistic/sophisticated industrial application case.

References

- [1] J.-P. Argaud, B. Bouriquet, M. Courtois, and J.-C. Le Roux. Reconstruction by data assimilation of the inner temperature field from outer measurements in a thick pipe. In *Pressure Vessels and Piping Conference, British Columbia, Canada, July 17–21*. ASME, 2016.
- [2] M. Asch, M. Bocquet, and M. Nodet. *Data assimilation: methods, algorithms, and applications*. Fundamentals of Algorithms. SIAM, 2016.
- [3] C. H. Bishop. Data assimilation strategies for state dependent observation error variances. *Quarterly Journal of the Royal Meteorological Society*, 145:217–227, 2019.
- [4] F. Bouttier and P. Courtier. Data assimilation concepts and methods. In *Meteorological Training Course Lecture Series*. ECMWF, 2002.
- [5] A. Carrassi, M. Bocquet, L. Bertino, and G. Evensen. Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *Wiley Interdisciplinary Reviews: Climate Change*, 9(5):e535, 2018.
- [6] V. Chabot, L. Berre, and G. Desroziers. Diagnosis and normalization of gridpoint background-error variances induced by a block-diagonal wavelet covariance matrix. *Quarterly Journal of the Royal Meteorological Society*, 143(704):1268–1279, 2017.
- [7] B. Chapnik, G. Desroziers, F. Rabier, and O. Talagrand. Diagnosis and tuning of observational error in a quasi-operational data assimilation setting. *Quarterly Journal of the Royal Meteorological Society*, 132(615):543–565, 2006.
- [8] A. Cherian, S. Sra, A. Banerjee, and N. Papanikolopoulos. Efficient similarity search for covariance matrices via the Jensen-Bregman LogDet divergence. In *2011 International Conference on Computer Vision*, pages 2399–2406. IEEE, Nov. 2011.
- [9] A. Cioaca and A. Sandu. Low-rank approximations for computing observation impact in 4d-var data assimilation. *Computers & Mathematics with Applications*, 67(12):2112 – 2126, 2014.
- [10] A. M. Clayton, A. C. Lorenc, and D. M. Barker. Operational implementation of a hybrid ensemble/4D-VAR global data assimilation system at the MET office. *Quarterly Journal of the Royal Meteorological Society*, 139(675):1445–1461, 2012.
- [11] P. Courtier, E. Andersson, W. Heckley, D. Vasiljevic, M. Hamrud, A. Hollingsworth, F. Rabier, M. Fisher, and J. Pailleux. The ECMWF implementation of three-dimensional variational assimilation (3D-VAR). i: Formulation. *Quarterly Journal of the Royal Meteorological Society*, 124(550):1783 – 1807, 1998.
- [12] G. Desroziers, L. Berre, B. Chapnik, and P. Poli. Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society*, 131(613):3385 – 3396, 2005.
- [13] G. Desroziers and S. Ivanov. Diagnosis and adaptive tuning of observation-error parameters in a variational assimilation. *Quarterly Journal of the Royal Meteorological Society*, 127(574):1433 – 1452, 2001.

- [14] D. Dreano, P. Tandeo, M. Pulido, B. Ait-El-Fquih, T. Chonavel, and I. Hoteit. Estimating model error covariances in nonlinear state-space models using Kalman smoothing and the expectation-maximisation algorithm. *Quarterly Journal of the Royal Meteorological Society*, 143(705):1877–1885, 2017.
- [15] R. Eyre, J and I. Hilton, F. Sensitivity of analysis error covariance to the mis-specification of background error covariance. *Quarterly Journal of the Royal Meteorological Society*, 139(671):524–533, 2013.
- [16] M. Fisher. Background error covariance modelling. In *Seminar on Recent developments in data assimilation for atmosphere and ocean*, Shinfield Park, Reading, 8-12 September. ECMWF, 2003.
- [17] M. Fisher, M. Leutbecher, and G. A. Kelly. On the equivalence between kalman smoothing and weak-constraint four-dimensional variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 131(613):3235–3246, 2005.
- [18] A. Fitt, J. Norbury, H. Ockendon, and R. Eddie Wilson. *Progress in industrial mathematics at ECMI 2008. Proceedings of the 15th European conference on mathematics for industry, London, UK, June 30 - July 4, 2008*, volume 15. 2010.
- [19] G. Gaspari and E. Cohn, S. Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125(554):723–757, 1999.
- [20] C. Goeury, A. Ponçot, J.-P. Argaud, F. Zaoui, R. Ata, and Y. Audouin. Optimal calibration of TELEMAC-2D models based on a data assimilation algorithm. In *the 14th TELEMAC-MASCARET User Conference, 17 to 20 October 2017, Graz University of Technology, Graz, Austria*, 2017.
- [21] J. Hodges and B. Reich. Adding spatially-correlated errors can mess up the fixed effect you love. *The American Statistician*, 64(4):325–334, 2010.
- [22] A. Hollingsworth and P. Lönnberg. The verification of objective analyses: Diagnostics of analysis system performance. *Meteorology and Atmospheric Physics*, 40:3–27, 1989.
- [23] B. Hunt, E. Kostelich, and I. Szunyogh. Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter. *Physica D: Nonlinear Phenomena*, 230(1):112–126, 2005.
- [24] T. Ishibashi. Tensor formulation of ensemble-based background error covariance matrix factorization. *Monthly Weather Review*, 143(12):4963–4973, 2015.
- [25] T. Janjić, N. Bormann, M. Bocquet, J. A. Carton, S. E. Cohn, S. L. Dance, S. N. Losa, N. K. Nichols, R. Potthast, J. A. Waller, and P. Weston. On the representation error in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 144(713):1257–1278, 2018.
- [26] E. Kalnay and S.-C. Yang. Accelerating the spin-up of Ensemble Kalman Filtering. *Quarterly Journal of the Royal Meteorological Society*, 136(651):1644–1651, 2010.
- [27] M. Leisenring and H. Moradkhani. Snow water equivalent prediction using bayesian data assimilation methods. *Stochastic Environmental Research and Risk Assessment*, 25(2):253–270, 2011.
- [28] W. Li, A. Sankarasubramanian, R. S. Ranjithan, and T. Sinha. Role of multimodel combination and data assimilation in improving streamflow prediction over multiple time scales. *Stochastic Environmental Research and Risk Assessment*, 30(8):2255–2269, 2016.
- [29] C. Liu and M. Xue. Relationships among four-dimensional hybrid ensemble-variational data assimilation algorithms with full and approximate ensemble covariance localization. *Monthly Weather Review*, 144(2):591–606, 2016.
- [30] D. Lucor and O. P. Le Maître. Cardiovascular modeling with adapted parametric inference. *ESAIM: ProcS*, 62:91–107, 2018.
- [31] I. Mirouze. *Régularisation de problèmes inverses à l’aide de l’équation de diffusion, avec application à l’assimilation variationnelle de données océaniques*. PhD thesis, Université de Toulouse, 2010.
- [32] D. F. Parrish and J. C. Derber. The National Meteorological Center’s spectral statistical-interpolation analysis System. *Monthly Weather Review*, 120(8):1747–1763, 1992.
- [33] X. Pennec, P. Fillard, and N. Ayache. A Riemannian framework for tensor computing. *International Journal of Computer Vision*, 66(1):41–66, 2006.
- [34] A. Ponçot, J.-P. Argaud, B. Bouriquet, P. Erhard, S. Gratton, and O. Thual. Variational assimilation for xenon dynamical forecasts in neutronic using advanced background error covariance matrix. *Annals of Nuclear Energy*, 60:39–50, 10 2013.

- [35] F. Rabier. Overview of global data assimilation developments in numerical weather-prediction centres. *Quarterly Journal of the Royal Meteorological Society*, 131(613):3215–3233, 2005.
- [36] M. Rochoux, A. Collin, C. Zhang, A. Trouvé, D. Lucor, and P. Moireau. Front shape similarity measure for shape-oriented sensitivity analysis and data assimilation for Eikonal Equation. *ESAIM: ProcS*, 63:258–279, 2018.
- [37] A. B. Saint-Venant. Théorie du mouvement non permanent des eaux, avec application aux crues des rivières et a l’introduction de marées dans leurs lits. *Comptes rendus de l’Académie des Sciences*, 73:147–154 and 237–240, 1871.
- [38] J. Sénégas, H. Wackernagel, W. Rosenthal, and T. Wolf. Error covariance modeling in sequential data assimilation. *Stochastic Environmental Research and Risk Assessment*, 15(1):65–86, 2001.
- [39] K. Singh, M. Jardak, A. Sandu, K. Bowman, M. Lee, and D. Jones. Construction of non-diagonal background error covariance matrices for global chemical data assimilation. *Geoscientific Model Development*, 4(2):299–316, 2011.
- [40] M. Sinsbeck and D. Tartakovsky. Impact of data assimilation on cost-accuracy tradeoff in multifidelity models. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1):954–968, 2015.
- [41] J. V. T. Sørensen and H. Madsen. Data assimilation in hydrodynamic modelling: on the treatment of non-linearity and bias. *Stochastic Environmental Research and Risk Assessment*, 18(4):228–244, 2004.
- [42] L. M. Stewart, S. L. Dance, and N. K. Nichols. Data assimilation with correlated observation errors: experiments with a 1-d shallow water model. *Tellus A: Dynamic Meteorology and Oceanography*, 65(1):19546, 2013.
- [43] O. Talagrand. A posteriori evaluation and verification of analysis and assimilation algorithms. In *Workshop on Diagnosis of Data Assimilation Systems*, pages 17–28, Shinfield Park, Reading, 1999. ECMWF.
- [44] J. A. Waller, D. Simonin, S. L. Dance, N. K. Nichols, and S. P. Ballard. Diagnosing observation error correlations for doppler radar radial winds in the Met office UKV model using observation-minus-background and observation-minus-analysis statistics. *Monthly Weather Review*, 144(10):3533–3551, 2016.
- [45] A. T. Weaver and I. Mirouze. On the diffusion equation and its application to isotropic and anisotropic correlation modelling in variational assimilation. *Quarterly Journal of the Royal Meteorological Society*, 139(670):242–260, 2013.
- [46] M. Xu, B. Yuan, L. Wang, and L. Zhang. Data assimilation for Fukushima nuclear accident assessments. In *International Conference on Nuclear Engineering, Shanghai, China*, volume 4. ASME, 2017.