Error covariance specification and localization in data assimilation with industrial application

Sibo Cheng

université PARIS-SACLAY

Error covariance specification and localization in data assimilation with industrial application

Thèse de doctorat de l'université Paris-Saclay

École doctorale n°579 Sciences Mécaniques et Energétiques, Matériaux et Géosciences (SMEMAG)
Spécialité de doctorat: énergétique
Domaine Scientifique : mathématiques et leurs interactions
Unité de recherche : Université Paris-Saclay, CNRS, LIMSI, 91400, Orsay, France
Référent : Faculté des sciences d'Orsay

Thèse présentée et soutenue en visioconférence totale, le 18/11/2020, par

Sibo Cheng

Composition du Jury

Caroline Nore Professeure, LIMSI, CNRS, Université Paris-Saclay	Présidente
Sarah Dance Professeure, University of Reading	Rapporteur & Examinatrice
Eric Blayo Professeur, Université Grenoble Alpes	Rapporteur & Examinateur
Pierre Tandeo Maître de conférences, IMT Atlantique	Examinateur
Didier Lucor Directeur de recherche, LIMSI, CNRS, Université Paris-Saclay	Directeur de thèse
Jean-Philippe Argaud Chercheur expert, EDF R&D	Co-Encadrant & Examinateur
Angélique Ponçot Ingénieure chercheuse, EDF R&D	Co-Encadrante
Bertrand Iooss Chercheur senior, EDF R&D	Co-Encadrant
Delphine Sinoquet Ingénieure de recherche, IFPEN	Invitée

Thèse de doctorat

NNT : 2020UPAST067

Remerciement/Acknowledgements

Je voudrais dans un premier temps remercier, mon directeur de thèse Didier Lucor et mes encradrants à EDF: Jean-Philippe Argaud, Bertrand Iooss et Angélique Ponçot, pour leur patience, leur judicieux conseils, et surtout le temps qu'ils m'ont consacré pour l'avancement de cette thèse. Les travaux présentés ici n'auraient jamais été abouti sans leur encadrement. Je leur remercie également pour leur accueil chaleureux à chaque fois que j'ai sollicité leur aide, notamment lors de préparation de congrés et de publications.

I would like to thank sincerely all the members of the jury: Professor Sarah Dance and Professor Eric Blayo for having conscientiously read my work and having spent your time to write the constructive reports. My thanks also go to Professor Caroline Nore, Doctor Pierre Tandeo, Doctor Delphine Sinoquet for your examination and your interest in my thesis. Thanks also go to Doctor christian Tenaud for the mid-term examination of my thesis.

Je remercie également toute l'équipe I23 (Olivier, Frank, Bruno, Romain, Gérald, Soizic, François,Laura, Lucie, Alexei, Alexandre, Helin, Morgane et al.) du département PERICLES, pour avoir eu la patience de répondre à mes innombrables questions aux techniques scientifiques ou à l'administration. Merci de m'avoir aussi bien intégrée à l'équipe et aux discussions. Merci en

particulier à Dimitri, avec qui je partage un bureau pendant ces trois ans, pour sa bonne humeur et ses encouragements.

J'adresse mes sincères remerciements aux chercheurs et aux camarades de doctorat à LIMSI et L'Universite Paris-Saclay. J'ai beaucoup apprécié les discussions fructueux avec vous sur des thèmes de recherche communs.

Very special thanks to all my friends in Europe and in China, your support has been very important to me.

Last, but definitely not least, I would like to thank my parents Caiyuan and Yi as well as the rest of my family for the continuous encouragement and unlimited support.

Résumé

Les méthodes d'assimilation de données et plus particulièrement les méthodes variationnelles sont mises à profit dans le domaine industriel pour deux grands types d'applications que sont la reconstruction de champ physique et le recalage de paramètres. Ces méthodes consistent à trouver un compromis entre différents types d'information : une estimation a priori, appelée ébauche, du champ ou des paramètres, et un ensemble de mesures de l'état du système appelées observations. Ce compromis est établi à partir de la confiance entourant ces informations par des matrices de covariance d'erreur $\mathbf{B}$ et $\mathbf{R}$, qui représentent respectivement les covariances établies *a priori* des erreurs d'ébauche et d'observation. Une des difficultés de mise en œuvre des algorithmes d'assimilation est que la structure de ces matrices, surtout celle de la matrice $\mathbf{B}$, n'est souvent pas ou mal connue. De plus, l'absence de données statistiques nous empêche de l'estimer directement. Pour mettre en place ces méthodes d'assimilation, les ingénieurs choisissent souvent une matrice de pondération "standard" par empirisme. Cependant beaucoup d'études et d'expériences montrent que la qualité des reconstruction/prévision est sensible à cette méconnaissance des covariances. Dans cette thèse, on s'intéresse à la spécification et la localisation de matrices de covariance dans des systèmes multivariés et multidimensionnels, et dans un cadre industriel. Dans un premier temps, on cherche à adapter/améliorer notre connaissance sur les covariances d'analyse à l'aide d'un processus itératif. Dans ce but nous avons développé deux nouvelles méthodes itératives pour la construction de matrices de covariance d'erreur d'ébauche sous l'hypothèse d'avoir une bonne connaissance des covariances d'erreur d'observation : CUTE (Covariance Updating iTerative mEthod) et PUB (Partially Updating BLUE method). On applique d'abord ces méthodes dans un modèle du shallow water pour vérification. L'efficacité de ces méthodes est montrée numériquement avec des erreurs indépendantes ou relatives aux états vrais. On propose ensuite un nouveau concept de localisation pour le diagnostic et l'amélioration des covariances des erreurs. Au lieu de s'appuyer sur une distance spatiale, cette localisation est établie exclusivement à partir de liens entre les variables d'état et les observations. En appliquant cette stratégie de localisation, on arrive à réduire le coût de calcul et gagner une paramétrisation plus flexible. Finalement, on applique une combinaison de ces nouvelles approches et de méthodes plus classiques existantes, pour un modèle hydrologique multivarié développé à EDF. L'assimilation de données est mise en œuvre pour corriger la quantité de précipitation observée afin d'obtenir une meilleure prévision du débit d'une rivière en un point donné. Un schéma "optimal" qui combine les différentes méthodes est également proposé pour lequel on observe une amélioration significative de la prévision des débits. Pour cette application, on constate que CUTE/PUB

a la meilleure performance parmi les algorithmes qui corrige la structure de covariances.

Abstract

Data assimilation techniques, more precisely variational methods, are widely applied in industrial problems of field reconstruction or parameter identification. It consists of finding an "optimal" compromise between prior simulations and real-time observations where the weights are given by the error covariance matrices **B** and **R**. These matrices represent not only prior error variance but also how these errors are correlated spatially or temporally. These matrices are often difficult to specify due to the lack of knowledge about state dynamics. To apply variational methods, engineers often choose a "standard" weighting matrix by empiricism. Several studies show that the assimilation precision, both in terms of state estimation and output covariance identification, can be sensitive to the mis-specification of the prior covariances. In this thesis, we are interested in the specification and localization of covariance matrices in multivariate and multidimensional systems in an industrial context. In this thesis, we propose to improve the covariance specification by iterative processes. Hence, we develop two new iterative methods: CUTE (Covariance Updating iTerative mEthod) and PUB (Partially Updating BLUE method) for **B** matrix recognition based on the assumption of good knowledge on observation covariance. We first apply these methods to the shallow water model for validation. The strength of these methods is demonstrated numerically with independent errors or relative errors. We then propose a new concept of localization and apply it for error covariance tuning. Instead of relying on spatial distance, this localization is established purely on links between state variables and observations. We put two state variables in the same subspace if they are mainly measured by the same group of observations. By using this localization strategy in covariance tuning algorithms, we can reduce the computational cost and gain a more flexible parameterization. Finally, we apply these new approaches, together with other classical methods for comparison, to a multivariate hydrological model developed at EDF. Variational assimilation is implemented to correct the observed precipitation in order to obtain a better river flow forecast. An "optimal" algorithm scheme which combines the different covariance specification methods is also proposed in this work. A significant improvement in short-range flow forecasting, compared to the original model, is achieved. In addition, we note that, in this application, CUTE / PUB has the best performance among all the tested algorithms that correct the covariance structures.

Symbols and Abbreviations

$\mathbf{A}$	analyzed error covariance
$\mathbf{A}_{\mathrm{A}}$	assumed analyzed error covariance
$\mathbf{A}_{\mathrm{E}}$	exact analyzed error covariance
$\mathbf{B}$	background error covariance
$\mathbf{B}_{\mathrm{A}}$	assumed background error covariance
$\mathbf{B}_{\mathrm{E}}$	exact background error covariance
$\mathbf{Cov}$	covariance matrix
$\mathbf{Cor}$	correlation matrix
$\mathbf{D}$	diagonal matrix of error variance
$\mathbb{E}[.]$	expectation operator
$\|\|.\|\|_F$	the Frobenius norm
ϵ_b	background error
ϵ_y	observation error
$\mathcal{G}_{\mathcal{S}}$	observation-based state graph
$\mathbf{A}_{\mathcal{G}_{\mathcal{S}}}$	adjacency matrix associated to $\mathcal{G}_{\mathcal{S}}$
$\mathcal{H}$	non-linear transformation operator
$\mathbf{H}$	linearized transformation operator
$\mathcal{J}$	cost function
$\mathcal{J}_b$	background cost function
$\mathcal{J}_o$	observation cost function
L	correlation scale
$\mathcal{M}$	transition model
$\mathbf{I}$	Identity matrix
$\mathbf{K}$	Kalman gain matrix
$\mathbf{Q}$	model error covariance
$\mathbf{R}$	observation error covariance
$\mathbf{R}_\infty$	fixed-point of observation matrix in Desroziers iteration

T	size of assimilation window (in the hydrological model)
$\sigma_{b,E}$	exact background error amplitude
$\sigma_{b,A}$	assumed background error amplitude
$\sigma_{o,E}$	exact observation error amplitude
$\sigma_{o,A}$	assumed observation error amplitude
$\mathbf{x}_t$	true state
$\mathbf{x}_b$	background state
$\mathbf{x}^i$	i^{th} community of state variables
$\mathbf{y}$	observation vector
$\mathbf{y}^i$	i^{th} community of observations

BLUE	best linear unbiased estimator
DI01	Desroziers & Ivanov tuning method
D05	Desroziers iterative method
CUTE	Covariance Updating iTerativE
PUB	Partially Updating BLUE
3D-Var	3D Variational

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

1.1.1 What is data assimilation

The main objective of data assimilation (DA) algorithms could be summarized as constructing an optimally weighted combination of different information sources. The essential idea of data assimilation is to combine observable information and simulation results while taking the associated uncertainties into account. Based on observed and simulated data, DA algorithms provide not only a history matching in dynamical systems, but more importantly they correct the current states to initialize numerical forecast models. DA methods have been originally developed in meteorological and environmental science. In fact, over the past decades, numerical weather prediction (NWP) has been significantly improved, both in terms of forecast accuracy as well as forecast range thanks to the development of data assimilation techniques, among other reasons.

Because of its capacity of dealing with high dimensional data (e.g. 10^9 state variables in NWP and geophysic problems) and of integrating real-time observations, data assimilation has been applied in a wide variety of industrial domains, including geophysical modeling ([Carrassi et al., 2018]), computational fluid mechanics, chemical engineering ([Sandu and Chai, 2011]), etc. The main purpose of data assimilation approaches can be split into two parts: physical field reconstruction or parameter identification. The former often aims to estimate a multidimensional physical field, such as temperature, velocity, concentration, while the latter corrects the current physical state via an update of some model parameters. In both cases, DA methods rely on a prior estimation (also known as the background state) of the true state and one or several vectors of observation, both subject to prior errors. Therefore, the basic idea of data assimilation approaches is to find a compromise by merging the information presented in these two quantities

to improve the history matching or forecasting.

1.1.2 Applications at EDF

EDF (Électricité de France) is a French electric utility company which stands for one of the world's largest producer of electricity. EDF produces its electricity mainly from nuclear power (58 active nuclear reactors) and renewable energy, including hydroelectricity. The research of EDF covers almost all trades and activities of the energy sector. Then, DA techniques are also adopted in engineering problems of EDF on a consistent basis. The applications include not only classical DA problems such as weather forecast but also some specific problems in nuclear, electrical or civil engineering. For example, data assimilation can used for analyzing measures or estimating high dimensional neutron flux with mesh-modeling. This estimation can be carried out via a field reconstruction or via parameter identification with observations provided by thermal/neutron sensors ([Argaud et al., 2018], [Argaud et al., 2017]). Other applications at EDF include, for example, concrete creep material laws calibration where observation data can be found via deformation sensors or measurement on test pieces. EDF R&D has developed its own DA solver ADAO, integrated into the $Salome$ open-source study platform ([Argaud, 2019]), which supplies a broad range of state of the art DA approaches. In general, for industrial applications at EDF, the state dimension usually ranges from 10^1 to 10^3 in calibration problems and from 10^4 to 10^7 in interpolation problems, considered as somewhat medium regarding the problem size in NWP and geoscience. This relatively small problem dimension leads to the possibility of implementing refined covariance tuning algorithms.

Another use of data assimilation at EDF concerns hydrology studies, for hydroelectric plants and pressurized water reactors. The latter uses a large amount of water from the sea, lake or river to cool the reactor core via a secondary steam circuit. Therefore, the forecast of floods and droughts is crucial for the management of general water resources and in particular for cooling water of power plants. To this end, EDF develops a precipitation-flow simulator, called MORDOR. This simulator takes spatially distributed geophysical data, e.g. rainfall, snowfall, temperature, to provide a simulation of river flow, which is also spatially distributed. Since the last decade of the twentieth century, MORDOR has been broadly used for flow forecast in a considerable number of study areas, mostly in France, such as the Loire valley ([Rouhier et al., 2017]) or the Durance river ([Paquet, E., 2004]). Continuous effort has been given in uncertainty quantification, sensitivity analysis and data assimilation to improve the forecast accuracy of MORDOR ([Rouhier et al., 2017]). For this purpose, the modeling of input error covariance is critical.

1.2 Motivation

History matching (reanalysis) and forecast are essential problems in many industrial applications, other than NWP or oceanography. These problems are well studied in crude oil production or space object tracking problems. Data assimilation has always been an important tool to improve the reanalysis and forecast accuracy by combining the prior simulation with some observable quantities. The weight of different information sources in data assimilation is determined by background and observation error covariances, respectively denoted as $\mathbf{B}$ and $\mathbf{R}$. In a dynamical DA chain, the former can be deduced from the initial errors and the model error covariance $\mathbf{Q}$ associated to the uncertainty of prior knowledge on the dynamical model. Thus there exists a strong link between the estimation of $\mathbf{B}$ and $\mathbf{Q}$ as, for example, shown in the Kalman filter. Since this study is more about variational approaches, we focus on the specification of $\mathbf{B}$ and $\mathbf{R}$. These matrices make a significant impact on the assimilation accuracy ([Bannister, 2008]). In fact, both background and observation errors in data assimilation can be considered as an overlapping of different sources of uncertainties, such as the measurement error, the resolution error or the representation error ([Janjić et al., 2018]), leading to extra difficulties in covariance estimation. It is mentioned and numerically demonstrated in [Tandeo et al., 2018] and [R. Eyre and I. Hilton, 2013] that DA algorithms are very sensitive to the mis-specification of $\mathbf{B}$ and $\mathbf{R}$, concerning both $||\mathbf{B}||/||\mathbf{R}||$ ratio and their structures. It has also been shown in industrial applications that well specified error covariance could be helpful, e.g. for estimating xenon dynamics in nuclear engineering ([Ponçot et al., 2013]), identifying element concentration in chemical engineering ([Singh et al., 2011]) or predicting disease spreading ([Cobb et al., 2014]).

Actually, the precision of error covariance estimation impacts not only the assimilation accuracy but also the specification of analyzed (output) error distribution. The latter is crucial in dynamical DA or filtering problems, such as NWP or signal processing. Several families of algorithms have been developed to improve the error covariance specification, which will be discussed later in chapter 2 and chapter 3. However, most of these methods require either a sufficiently long data assimilation chain or a relatively precise knowledge of the dynamical operator while these conditions could be difficult to fulfill in some industrial applications. For example, in some industrial problems, very little information is available about the dynamical model, especially the associated model error distribution. Another significant barrier is about the high dimension of the problem, making the refined covariance tuning method computationally difficult, if not infeasible. Common solutions include, for example, localization techniques or data compression approaches. The former is often used in ensemble-based DA approaches where the probability density function (pdf) of the current state is represented by a set of sampling background trajectories. Covariance localization attempts to avoid long-distance sampling error ([Gaspari and Cohn, 1999]). On the other hand, the objective of domain localization is to break global assimilation problems into appropriate subspaces to reduce the computational cost. Localization techniques are widely adopted in multivariate geophysical systems ([Carrassi et al., 2018]). But most of these methods depend on assumptions on prior error covariance, such as the cut-off correlation radius or the pre-defined covariance kernel. Being pointed out by recent works of [van Leeuwen, 2019] and [Waller et al., 2017], the incoherence between the assumption and the re-

ality may result in a less optimal covariance tuning and thus less accurate assimilation corrections.

To overcome these difficulties, this thesis aims to develop and implement error covariance specification algorithms adapting industrial conditions with efficient computation strategy. Our ultimate goal is to improve the error covariance tuning and as a consequence, the assimilation accuracy in multidimensional and multivariate industrial problems with a reasonable computational cost ([Arcucci et al., 2018]). We remind that, although the assumption is often made in data assimilation for prior errors to be zero-mean Gaussian and state-independent, the error bias, the non-Gaussianity and the state-dependent errors are sometimes discussed in real-world applications. Continuous effort has been devoted to quantifying the impact of these properties on the error covariance computation ([Bocquet et al., 2010], [Bishop, 2019]). These tasks are not included in the objectives of this thesis, except that the state-dependent error is tested numerically in the twin experiments in chapter 3.

1.3 Overview of the thesis

1.3.1 Covariance specification and diagnosis

While the observation error is considered as independent from the DA process in general ([Janjić et al., 2018]), the specification of the background matrix, both in terms of error amplitude and correlation structure, is sometimes more sophisticated [Fisher, 2003] in real applications. Current methods, such as the ensemble-based methods [Evensen, 1994] or the National Meteorological Center (NMC) method [Parrish and Derber, 1992], often require a sufficiently long dynamics for the convergence of background error estimation. In this thesis, we aim to develop non-parametric algorithms that make full use of good knowledge on the observation matrix (compared to the background matrix) to improve the assimilation accuracy, as well as the estimation of output error covariance in the state space. Since we intend to ameliorate the short-range forecast with limited prior data (often subjected by industrial condition), these methods should not require either a large ensemble of prior data or a long DA chain, regarding existed approaches. Two novel iterative assimilation methods named as CUTE (Covariance Updating iTerativE) and PUB (Partially Updating BLUE) are introduced in this thesis. These methods iterate the current analysis state using invariant observation data, taking into account the background-observation covariance emerged from the iterative process itself. More precisely, CUTE only considers the state-observation covariance in the covariance updating while PUB take this covariance directly into account in the cost function. With different optimization objectives, both methods are capable of sequentially adapting background error covariance matrices in order to improve assimilation results in terms of state estimation and output covariance specification. CUTE and PUB are first tested in a 2D shallow water framework. Under the assumption of a higher level of background error (relative to the observation), numerical results show the strength of CUTE and PUB with both state independent and state dependent prior error. Starting with an initial guess of the $\mathbf{B}$ matrix,

the estimated analysis error decreases significantly against the algorithm iteration. Furthermore, we gain a more accurate estimation of the output error correlation as well. These results are numerically robust while changing the correlation kernel of $\mathbf{B}$ and $\mathbf{R}$.

1.3.2 Graph-based localization techniques

To ensure the computational efficiency, we also focus on the combination of localization techniques and covariance tuning algorithms. The key idea consists of finding an optimal decomposition of the state space where each subspace, considered homogeneous in terms of error amplitude, is connected to one specific group of observations in the ideal case. Localization techniques can be roughly split into two families: covariance localization methods and domain localization methods, both based on prior assumptions on the correlation scale [Gaspari and Cohn, 1999]. Very recently, the concept of domain localization was introduced in posterior covariance diagnostic methods [Waller et al., 2017] for reducing the computational cost. However, it is shown in [Waller et al., 2017] that flawed prior assumptions on the correlation scale may lead to significant error in posterior error analysis. In this work, we seek for unsupervised learning algorithms to perform localization methods in automatically detected subspaces. Expected achievements are two-fold: firstly avoiding the conflict between the prior assumptions and the reality, compared to traditional localization approaches; secondly getting a more flexible parameterization of error covariance. Here the localization term refers to the idea of breaking up global assimilation into subproblems. Instead of relying on the spatial distance, we introduce a novel concept of localization by performing clustering algorithms on graphs ([Parés et al., 2017], [Gueuning et al., 2019]) which connect the state variables. The graph connection among state variables is solely decided by the linearized state-observation transformation operator. Briefly speaking, we group the state variables measured by the same class of observations, if existed. The variances and covariances associated to these states are more likely to be improved jointly by posterior tuning algorithms, as proved numerically using synthetic data. We also introduce two strategies to deal with group overlapping while introducing the concept of graph-based localization and its application in covariance tuning. Numerical experiments are performed using synthetic data with DI01 tuning algorithm ([Desroziers and Ivanov, 2001]).

1.3.3 The hydrological industrial application

Alongside theoretical developments, a hydrological model based on MORDOR, issued from an industrial problem of EDF, is also studied in this thesis for comparing different covariance specification approaches. We are interested in the study area around the Tarn river, located in the south of France. Variational assimilation algorithms are implemented to correct the daily-measured precipitation with available observations of river flow in 9 catchments. We aim to improve the short-range forecast of river flow, with a particular interest in flood periods. An essential challenge in this hydrological model is the lack of knowledge on prior error covariance, both spatial and temporal. Starting with a localized error amplitude tuning, we apply the novel iterative

methods CUTE and PUB, combining with other posterior covariance specification methods in a variational DA framework. Hundreds of assimilation windows are studied and analyzed. As proved by numerical results, the DA correction with an advanced covariance tuning improves the short-range flow forecast significantly with a reduction of around 30% to 60% observation-prediction difference in average. These results are highly more optimal than using arbitrarily set covariance matrices. The improvement is most significant in flood periods, except for some extremely high points. The hydrological application is introduced in detail in chapter 5, including objectives, modeling and constraints. We then perform the advanced covariance tuning methods along with CUTE and PUB, following with an analysis of the short-range numerical flow forecast.

1.4 Outline

In chapter 2, we introduce briefly the concept of variational data assimilation and the notation/definition used throughout this thesis, with specific attention on covariance tuning and localization methods. The rest of the thesis is divided into three main parts, each relies on a published or submitted journal paper:

- chapter 3: CUTE and PUB algorithms, together with the twin experiment results using two dimensional shallow water model, published in [Cheng et al., 2019].
- chapter 4: Graph-based localization approach combing with DI01 algorithm and numerical results using synthetic data, given in the submitted paper [Cheng et al., 2020b].
- chapter 5: Application of covariance specification methods on an industrial hydrological model for improving flow prediction and reanalysis.

Finally, the conclusion and future work are addressed in chapter 6. The dependency of different chapters is shown in Fig. 1.1.

Publications and communications

Published/submitted works and conference presentations related to the content of this thesis are listed herebelow.

- Argaud, J.-P., Cheng, S., Iooss, B., Lucor, D., and Ponçot, A. (2018). Methods for improving background error covariance matrix rebuild in data assimilation in *The 11th International Conference of the ERCIM WG on Computational and Methodological Statistics* (CMStatistics), 14-16 December 2018, Pisa, Italy.

- Cheng, S., Argaud, J.-P., Iooss, B., Lucor, D., and Ponçot, A. (2019). Improvement of error covariance matrix computation in variational methods in *2019 MASCOT-NUM annual conference* , 18-20, March 2019, Rueil-Malmaison, France.

- Argaud, J.-P., Cheng, S., Iooss, B., Lucor, D., and Ponçot, A. (2019). Iterative methods for improving error covariance matrices modelling in data assimilation in *3rd International Conference on Uncertainty Quantification in Computational Sciences and Engineering* (UNCECOMP), 24-26 June 2019, Crete, Greece.

- Cheng, S., Argaud, J.-P., Iooss, B., Lucor, D., and Ponçot, A. (2019). Background error covariance iterative updating with invariant observation measures for data assimilation. *Stochastic Environmental Research and Risk Assessment*, 33(11):2033–2051.

- Cheng, S., Argaud, J.-P., Iooss, B., Ponçot, A., and Lucor, D. (2020). A graph clustering approach to localization for adaptive covariance tuning in data assimilation based on state-observation mapping, submitted to *Mathematical Geoscience*

- Cheng, S., Argaud, J.-P., Iooss, B., Lucor, D. and Ponçot, A. (2020). Error covariance tuning in variational data assimilation: application to an operating hydrological model, accepted for publication in *Stochastic Environmental Research and Risk Assessment*

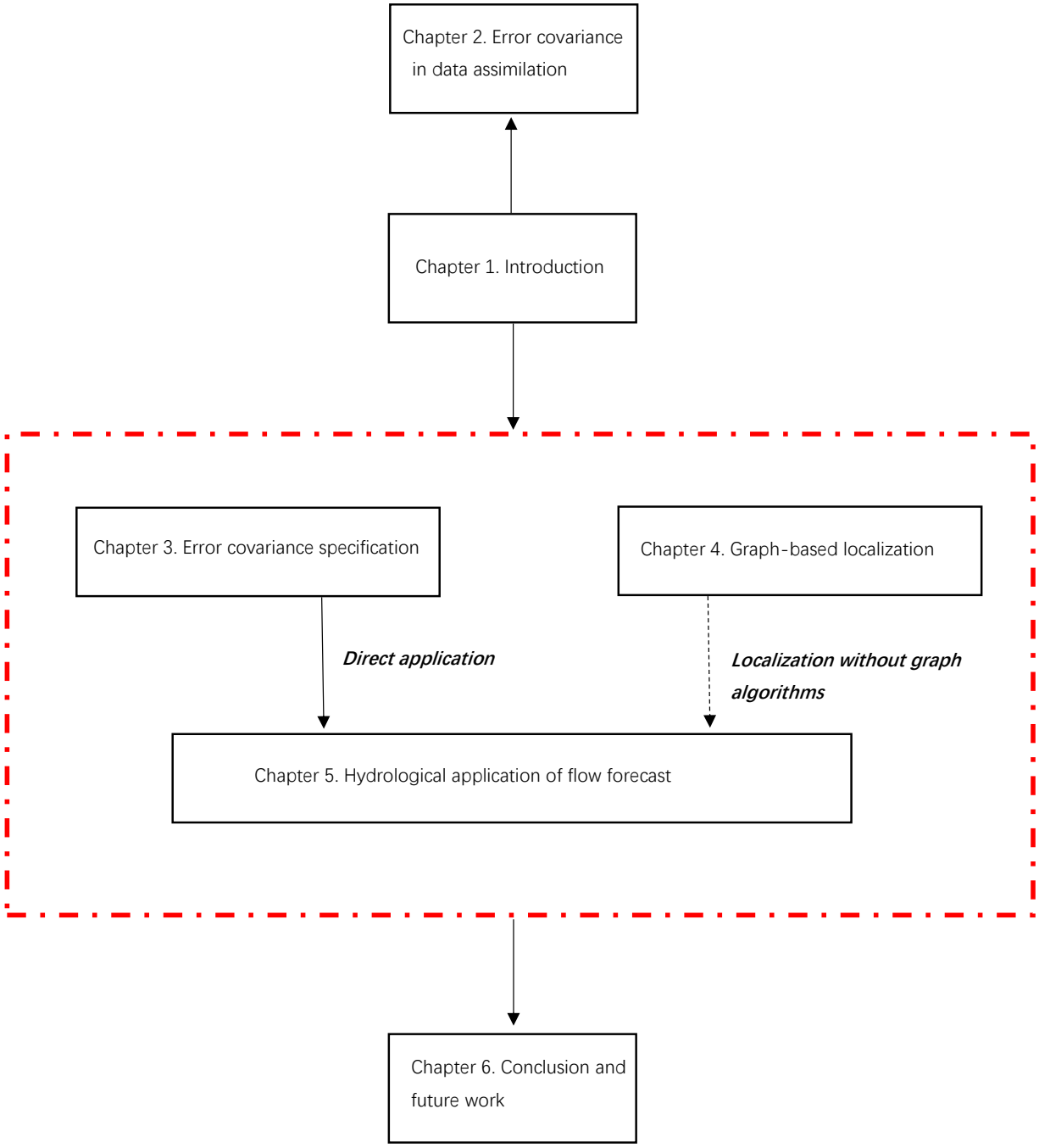


Figure 1.1: Flowchart of dependency

Chapter 2

Error covariance in data assimilation

In this chapter, we introduce the concept and notations of variational assimilation algorithms. The challenging problems of covariance specification and localization is also explained in short with a brief review of existed methods. More detailed descriptions could be found in the introduction of chapter 3, 4, 5.

2.1 Brief introduction to data assimilation

The objective of data assimilation is the estimation or the forecast of the state **x**, which could be a physical field or a set of parameters, relying a some real function H with

$$\mathbf{y} = \mathcal{H}(\mathbf{x}) \tag{2.1}$$

where **y** is a vector of observable quantities. Eq. 2.1 corresponds to the classical formulation of inverse problems where the goal is to find numerical values of **x** which fits best the observation vector **y**. Both **x** and **y** are with some uncertainties. Besides data assimilation, common solutions of inverse problem include surrogate models ([Arcucci et al., 2017]) or Bayesian approaches. Among these methods, data assimilation techniques own an advantage when dealing with high dimensional dynamical systems thanks to be assumption of unbiased Gaussianity which will be explained later.

More precisely, data assimilation relies on a prior state estimation, also known as the background state,

$$\mathbf{x}_b = \mathbf{x}_t + \epsilon_b \tag{2.2}$$

where $\mathbf{x}_t$ stands for the true state and ϵ_b is a stochastic additive term, representing the background error. Noisy observations are available to correct the prior estimation. These observations are represented by:

$$\mathbf{y} = \mathcal{H}(\mathbf{x}_t) + \epsilon_y \tag{2.3}$$

where ϵ_y is the observation error, supposed to be uncorrelated to ϵ_b and the state to observation transformation function $\mathcal{H}$ is supposed to be known. Due to the high dimensionality of DA problems, $\mathbf{x}_b$ and $\mathbf{y}$ are often supposed to be Gaussian, i.e.,

$$\epsilon_b \sim \mathcal{N}(0, \mathbf{B}), \quad \epsilon_y \sim \mathcal{N}(0, \mathbf{R}) \tag{2.4}$$

where $\mathbf{B}$ and $\mathbf{R}$ represent respectively the background and the observation error covariance matrices. These matrices reflect not only the prior error amplitude but also how these errors are connected spatially or temporally. Thanks to the assumption of unbiased Gaussian distribution, the prior error probability density function (pdf) relies solely on these matrices.

DA approaches are mainly split into two families: the variational assimilation, originally from the control theory; and the sequential assimilation, originally from the probability estimation theory. Both being widely applied in industrial problems, the former relies on the optimization of some objective function $\mathcal{J}$ to find the optimal state or trajectory while the latter aims to assimilate sequentially the observations where the current background state is obtained via propagation of the previous analysis. When the transformation operator $\mathcal{H}$ is linear, the static variational assimilation (also known as $3D-Var$) and the current step of a sequential assimilation lead to the same estimation, both being equivalent to the Best Linear Unbiased Estimator (BLUE). For more general problems (e.g. non-linear, time-dependent), continuous effort ([Fisher et al., 2005], [Buehner et al., 2010]) has been devoted to study the equivalence between the variational assimilation (e.g. $4D-Var$) and the sequential assimilation (typically Kalman-based filters). Variational assimilation methods and the BLUE are presented in details in the introduction of chapter 3. As for sequential assimilation, since it is not essential in this thesis, interested readers are referred to the book of [Chui and Chen, 1991].

Both variational and sequential DA algorithms can be performed at any past (smoothing), present (filtering) or future (predicting) time in a discretized dynamical system,

$$\mathbf{x}_{k+1} = \mathcal{M}(\mathbf{x}_k) + \eta_k \tag{2.5}$$

where $\mathcal{M}$ stands for the transition operator and the additive term η_k, known as the model error, is often supposed to follow a Gaussian distribution $\mathcal{N}(0, \mathbf{Q})$.

In dynamical data assimilation problems, despite that sometimes the model error is treated separately, the model error covariance $\mathbf{Q}$ is often integrated into the $\mathbf{B}$ matrix while updating the

assimilation results.

2.2 Error covariance

The error covariance $\mathbf{B}$ and $\mathbf{R}$ play an essential role in DA algorithms. Their associated inverse matrices, $\mathbf{B}^{-1}, \mathbf{R}^{-1}$ decide the weight of different information sources in the optimization function (see 3.1 for details) in variational assimilation.

2.2.1 Covariance specification

In statistics, the covariance matrix of a random vector is often specified via empirical estimation. A sufficient number of simultaneous samplings is required for the empirical estimation ([Wishart, 1928]). Typically, when the sampling number is inferior to the problem dimension, the estimated covariance will be rank deficient. In DA framework, these samplings can be obtained, for example, via an ensemble of simulations for the background covariance or via different measuring instruments for the observation covariance ([Daget, 2008]). However, as mentioned in chapter 1, the high problem dimensionality and the lack of simultaneous data stand for significant obstacles of covariance computation in data assimilation. To overcome these difficulties, we often rely on some generic correlation kernels, often with homogeneous and isotropic characteristics, together with balanced operators for multivariate systems ([Derber and Rosati, 1989]). Among these correlation kernels, the family of Matérn functions, including the Exponential kernel (Matérn $1/2$), the Balgovind kernel (Matérn $3/2$, also known as second order auto-regressive (SOAR) function) and the Gaussian kernel (Matérn $5/2$), is frequently used for covariance computing thanks to its capability to capture physical processes ([Stein, 1999]). Different Matérn kernels are continuously studied and compared in DA applications (e.g. [Stewart et al., 2013], [Weston et al., 2014]), in terms of smoothness, differentiability and so on. Other stationary covariance models involve, for example, diffusion-based operators ([Weaver and Courtier, 2001], [Mirouze and Weaver, 2010]) or convolution formulation ([Gaspari and Cohn, 1999]), both contribute to an efficient storage of covariance matrices. In fact, very tight links could be found between the Matérn kernel and the diffusion/convolution modeling. We refer to the thesis of [Mirouze, 2010] for the equivalence between diffusion-based modelings and the Matérn kernel (4.A), and the thesis of [Daget, 2008] (appendix B) for the equivalence between convolution and diffusion operators. One can also read the work of [Melkumyan and Ramos, 2011] for the link between Matérn kernels and convolution operator. We illustrate these equivalencies in Fig. 2.1. All covariance modeling approaches mentioned in this section until now, depend strongly on the prior assumption of the covariance structure. The associated parameters, such as the Matérn parameter or the diffusion coefficient, are often determined via a calibration process using real data. However, problems may occur when the supposed prior covariance structure does not reflect, at least approximately, the true error distribution.

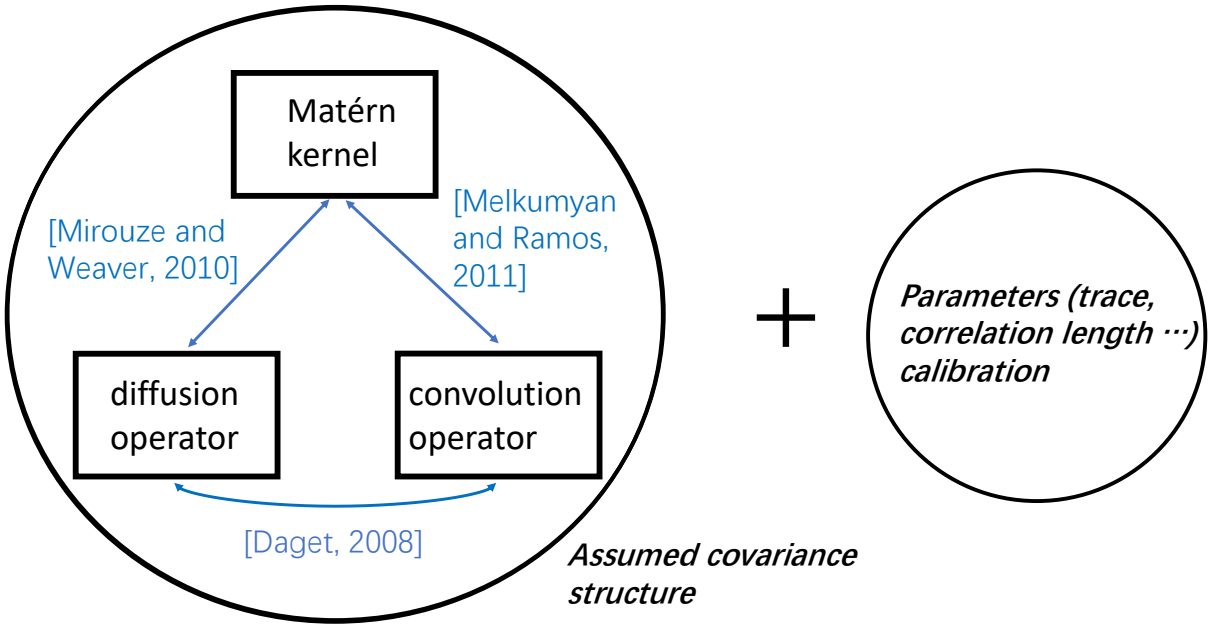


Figure 2.1: equivalence between different methods for computing error covariance structure

Some classical statistical methods, such as [Hollingsworth and Lönnberg, 1986], [Parrish and Derber, 1992] (NMC) and [Evensen, 1994] (EnKF) are developed to provide a non-parametric covariance estimation, especially for the $\mathbf{B}$ matrix. Some of these methods depend on the propagation of an ensemble of simulated trajectories initialized by adding some artificially set perturbations on the current state. According to [Fisher, 2003], ensemble-based methods require firstly careful attention on the set up of the initial noise, and secondly a sufficiently long assimilation time (smoothing period) for the generated trajectories to approach a true background ensemble. Our interest here is to improve the covariance estimation using less observation and sequential time.

To gain a more clear insight of covariance evolution, several methods of posterior diagnosis are developed based on the analysis of innovation quantities i.e., $\mathbf{y} - \mathcal{H}(\mathbf{x}_b)$ or $\mathbf{y} - \mathcal{H}(\mathbf{x}_a)$. Being a strong contributor to this topic, the meteorology community invented several well-known posterior diagnosis and their improved versions ([Desroziers and Ivanov, 2001], [Chapnik et al., 2004], [Desroziers et al., 2005]) to adjust the $||\mathbf{B}||/||\mathbf{R}||$ ratio, correlation scales or the full covariance structure in the observation space. Some iterative processes, based on the fixed-point theory, have also been proposed for error covariance tuning. Recent works of [Ménard, 2016] and [Bathmann, 2018] have proved its convergence in the ideal case. However, it is also mentioned by [Bathmann, 2018] that a regularization step is necessary in practice for applying the method of [Desroziers et al., 2005] and the convergence of the regularized iterations remains an open question. These iterative methods are explained in detail in chapter 3, 4, 5 and applied in the

hydrological model in chapter 5. To deal with time-varying systems, lag-innovation statistics are also used for error covariance estimation ([Daley, 1992]). The basic idea is to build a secondary Kalman-filtering process for adjusting **Q** and **R** using time-shifted innovation vectors. For more details of innovation-based methods, we refer to the overview of [Tandeo et al., 2018] which also covers some other estimation methods, such as the family of likelihood-based approaches.

2.2.2 Localization techniques

Localization techniques are widely applied in data assimilation to reduce the computational cost and avoid sampling error, leading to the possibility of performing refined covariance tuning algorithms. The concept of localization in data assimilation relies on the exponential decrease of the error correlation against the spatial distance in geophysical problems ([Carrassi et al., 2018]). It is commonly applied in ensemble-based methods. There are two main families of localization approaches: the covariance localization and the domain localization. The former aims to reduce the sampling error by building an element-wise (Schur) product of the sample covariance matrix and a fixed smooth matrix ([Gaspari and Cohn, 1999]). From this fact, all long-distance correlations appearing in the covariance matrices are eliminated as they are considered as spurious sampling errors. On the other hand, the domain localization intends to break global DA problems into several small ones, usually also depending on the spatial distance among state variables, to reduce the computational cost. Instead of performing diagnosis in the entire space, [Waller et al., 2017] apply the Desroziers iterative method in local subspaces, determined by setting a fixed distance scale around each state variable. These subspaces, so called the "region of influence", is decided purely by spatial distance. [Waller et al., 2017] show that the localized approach could significantly make the tuning algorithm less costly. However, they also mentioned that the estimation of error covariance could be troublesome when the "region of influence" does not reflect the reality, i.e., an analyzed state can be impacted by an observation outside the "region of influence". To solve this problem, it might be beneficial to develop new localization methods with a more flexible definition of the "region of influence", according to how state variables are connected to the observations.

Chapter 3

Background error covariance iterative updating with invariant observation measures for data assimilation

published as Cheng, S., Argaud, J.-P., Iooss, B., Lucor, D., and Ponçot, A. (2019). Background error covariance iterative updating with invariant observation measures for data assimilation. *Stochastic Environmental Research and Risk Assessment*, 33(11):2033–2051.

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 iTerativE) 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.

3.1 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 ([Parrish and Derber, 1992]), nuclear safety ([Xu et al., 2017]), atmospheric chemistry ([Singh et al., 2011]), hydrologic modelling [Li et al., 2016], seismology, glaciology, agronomy, etc. Over decades, these approaches have been applied in the energy industry, for projects involving temperature field reconstruction ([Argaud et al., 2016]) or forecasting in neutronic ([Ponçot et al., 2013]) and hydraulic ([Goeury et al., 2017]). More recently, they have also made their way to other fields such as medicine, biomedical applications ([Lucor and Le Maître, 2018]) or wildfire front-tracking problems ([Rochoux et al., 2018]).

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 ([Carrassi et al., 2018]), 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 [Sénégas et al., 2001]) and this may substantially change the assimilation results ([S. Hodges and J. Reich, 2010]). It also provides an important information of the relationship between observations and forecasts. As mentioned by [Fisher, 2003], there exist a wide variety of methods to estimate these matrices. Well known methods among others, are the one of [Hollingsworth and Lönnberg, 1989], the *NMC* (National Meteorological Center) method ([Parrish and Derber, 1992]) and ensemble methods ([Clayton et al., 2012]) which are often combined with algebraic operations such as matrix factorization ([Ishibashi, 2015]) or covariance localization ([Liu and Xue, 2016]). For many industrial applications, the paucity of historical observations as well as the large dimension and complexity (e.g.[Sinsbeck and Tartakovsky, 2015]) 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 ([Hunt et al., 2005]) or relying on generic covariance kernels: e.g. Matérn kernel ([Singh et al., 2011]), are often favoured. Other approaches are based on numerical techniques involving convolution operations ([Gaspari and Cohn, 1999]) or the resolution of diffusion equations ([Weaver and Mirouze, 2013]). The latter methods are sometimes equivalent to the former ones, under simplifying assumptions,

as explained in [Mirouze, 2010]. 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 ([Courtier et al., 1998]) or the wavelet space ([Chabot et al., 2017]). However, these contributions require prior assumptions about the matrix structure which could be difficult to justify in industrial applications.

Recent works of [Dreano et al., 2017] 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 [Desroziers and Ivanov, 2001], [Desroziers et al., 2005], [Chapnik et al., 2006] etc. These methods have been applied world-widely in industrial problems ever since, e.g. [Fitt et al., 2010], [Waller et al., 2016]. 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 [Chapnik et al., 2006]. 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 ([Desroziers et al., 2005]), aims at estimating the whole covariance matrices (see [Janjić et al., 2018]). 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 [Kalnay and Yang, 2010] 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 iTerativE) 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 3.2, with a focus on covariance updating. We then propose two novel iterative methods in section 3.3, together with a simple illustrative scalar test case. In section 3.4, 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.

3.2 Data assimilation and variational methods framework

3.2.1 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 [Leisenring and Moradkhani, 2011]). 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:

$$\begin{aligned} \epsilon_b &= \mathbf{x}_b - \mathbf{x}_t \\ \epsilon_y &= \mathbf{y} - \mathcal{H}(\mathbf{x}_t). \end{aligned} \tag{3.1}$$

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

$$\begin{aligned} \epsilon_b &\sim \mathcal{N}(0, \mathbf{B}) \\ \epsilon_y &\sim \mathcal{N}(0, \mathbf{R}) \\ Cov(\epsilon_b, \epsilon_y) &= 0. \end{aligned} \tag{3.2}$$

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 3.3.

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 [Carrassi et al., 2018]) 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.

3.2.2 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 3.2.1, the key idea in variational methods is to find a balance between the background and the observations ([Bouttier and Courtier, 2002]) according to the weights

represented by the inverse of $\mathbf{B}$ and $\mathbf{R}$. This leads to the loss function:

$$J_{\text{3D-VAR}}(\mathbf{x}) = \frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) + \frac{1}{2}(\mathbf{y}-\mathcal{H}(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{y}-\mathcal{H}(\mathbf{x})) \tag{3.3}$$

$$= \frac{1}{2}||\mathbf{x}-\mathbf{x}_b||^2_{\mathbf{B}^{-1}} + \frac{1}{2}||\mathbf{y}-\mathcal{H}(\mathbf{x})||^2_{\mathbf{R}^{-1}}. \tag{3.4}$$

The optimisation problem defined by the objective function of Eq. (3.4) 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.[Carrassi et al., 2018]).

3.2.3 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 [Asch et al., 2016]). When $\mathcal{H} = \mathbf{H}$ is linear, the *optimal* solution provided by the variational formulation (Eq.3.4) 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:

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y}-\mathbf{H}\mathbf{x}_b) \tag{3.5}$$

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

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}. \tag{3.6}$$

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

$$\begin{aligned}\mathbf{A} &= Cov(\mathbf{x}_a - \mathbf{x}_t) \\ &= (\mathbf{I}-\mathbf{K}\mathbf{H})\mathbf{B}(\mathbf{I}-\mathbf{K}\mathbf{H})^T + \mathbf{K}\mathbf{R}\mathbf{K}^T \\ &= (\mathbf{I}-\mathbf{K}\mathbf{H})\mathbf{B}.\end{aligned} \tag{3.7}$$

Under the assumption that both background and observation errors follow centered Gaussian

distributions, it is easy to justify that:

$$\epsilon_a = \mathbf{x}_a - \mathbf{x}_t \sim \mathcal{N}(0, \mathbf{A}). \tag{3.8}$$

Eq.3.8 holds when $\mathbf{H}$ is linear. We recall that the uncorrelatedness between ϵ_b and ϵ_y is a crucial assumption for Eq.3.5-3.8. 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.3.7, 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 latter in section 3.3.4. Non-Gaussianity in data assimilation problems, for example due to the nonlinearity of $\mathcal{H}$, has been discussed (e.g [Sørensen and Madsen, 2004]).
However, we emphasize that when prior covariances are not well known, the estimation provided by Eq.3.7 could be very different from the exact[1] output error covariance (later noted as $\mathbf{A}_{\mathrm{E}}$). It is therefore of highest importance to differentiate between well or loosely known prior covariance matrices. This aspect will be investigated further in section 3.3.

3.2.4 Misspecification of B matrix

From here, we follow the notations given in [R. Eyre and I. Hilton, 2013], where $\mathbf{B}_{\mathrm{E}}$ designates the unknown exact background error covariance while $\mathbf{B}_{\mathrm{A}}$ 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 $\mathbf{B}_{\mathrm{E}}$ (mismatch between $\mathbf{B}_{\mathrm{E}}$ and $\mathbf{B}_{\mathrm{A}}$). Following the current notation, the standard estimation of output error covariance $\mathbf{A}_{\mathrm{A}}$ (here we have kept subscript A to indicate that this form is obtained from $\mathbf{B}_{\mathrm{A}}$) provided by the plain-vanilla *3D-VAR* method in Eq. (3.7) becomes:

$$\mathbf{A}_{\mathrm{A}} = (\mathbf{I} - \mathbf{K}_{\mathrm{A}}\mathbf{H})\mathbf{B}_{\mathrm{A}}, \tag{3.9}$$

where

$$\mathbf{K}_{\mathrm{A}} = \mathbf{B}_{\mathrm{A}}\mathbf{H}^T(\mathbf{H}\mathbf{B}_{\mathrm{A}}.\mathbf{H}^T + \mathbf{R})^{-1}, \tag{3.10}$$

$\mathbf{A}_{\mathrm{A}}$ is different from the exact output error covariance when the unknown $\mathbf{B}_{\mathrm{E}}$ has merely been approximated by $\mathbf{B}_{\mathrm{A}}$. The gain matrix $\mathbf{K}$ remains a function of the assumed background covariance matrix $\mathbf{B}_{\mathrm{A}}$ and the one of the observation, $\mathbf{R}$. The latter is supposed to be perfectly known.

[1]Here, 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.

In fact, the exact output error covariance $\mathbf{A}_{\mathrm{E}}$ depends on the prior error and the parameters of the algorithm. Therefore, as described by [R. Eyre and I. Hilton, 2013], $\mathbf{A}_{\mathrm{E}}$ is in function of $\mathbf{B}_{\mathrm{E}}$ and $\mathbf{B}_{\mathrm{A}}$:

$$\mathbf{A}_{\mathrm{E}} = (\mathbf{I} - \mathbf{K}_{\mathrm{A}}\mathbf{H})\mathbf{B}_{\mathrm{E}}(\mathbf{I} - \mathbf{K}_{\mathrm{A}}\mathbf{H})^T + \mathbf{K}_{\mathrm{A}}\mathbf{R}\mathbf{K}_{\mathrm{A}}^T. \tag{3.11}$$

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. $\mathbf{B}_{\mathrm{A}} = \mathbf{B}_{\mathrm{E}}$, the obtained Kalman gain matrix $\mathbf{K}(\mathbf{B}_{\mathrm{E}})$ is a so called *optimal* gain matrix, which ensures that the trace of $\mathbf{A}_{\mathrm{E}}$ 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 ([Desroziers and Ivanov, 2001]) allow the adjustment of the multiplicative ratio (i.e. the total variance) between matrices $\mathbf{B}_{\mathrm{A}}$ 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 **Cov** is related to its correlation matrix **Cor**:

$$\mathbf{Cov} = \mathbf{D}^{\frac{1}{2}}\mathbf{Cor}\mathbf{D}^{\frac{1}{2}} \tag{3.12}$$

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

3.2.5 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}) $\mathcal{M}_{t^k \to t^{k+1}}$, where

$$\mathbf{x}_{t^{k+1}} = \mathcal{M}_{t^k \to t^{k+1}}(\mathbf{x}_{t^k}). \tag{3.13}$$

The forecasting in data assimilation thus relies on the knowledge of transition operator $\mathcal{M}_{t^k \to t^{k+1}}$ 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.

$$\mathbf{x}_{b,t^k} = \mathcal{M}_{t^{k-1}\to t^k}(\mathbf{x}_{a,t^{k-1}}). \tag{3.14}$$

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 ([Fisher et al., 2005]). 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.

3.3 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 [R. Eyre and I. Hilton, 2013] 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.

3.3.1 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:

$$\mathbf{x}_{b,n+1} \leftarrow \mathbf{x}_{a,n} = \mathbf{x}_{b,n} + \mathbf{K}_n(\mathbf{y} - \mathbf{H}\mathbf{x}_{b,n}) \tag{3.15}$$

$$\mathbf{B}_{\mathrm{A},n+1} \leftarrow \mathbf{A}_{\mathrm{A},n} = (\mathbf{I} - \mathbf{K}_n\mathbf{H})\mathbf{B}_{\mathrm{A},n}, \tag{3.16}$$

where n refers to the iteration number, and:

$$\mathbf{K}_n = \mathbf{B}_{\mathrm{A},n}\,\mathbf{H}^T(\mathbf{H}\mathbf{B}_{\mathrm{A},n}\mathbf{H}^T + \mathbf{R})^{-1}, \tag{3.17}$$

is the iterated Kalman gain matrix.

3.3.2 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 $\mathbf{B}_{\mathrm{A},n}$ and $\mathbf{K}_n$:

$$\begin{aligned}\mathbf{B}_{\mathrm{E},n+1} = \mathbf{A}_{\mathrm{E},n} = (\mathbf{I} - \mathbf{K}_n\mathbf{H})\mathbf{B}_{\mathrm{E},n}(\mathbf{I} - \mathbf{K}_n\mathbf{H})^T + (\mathbf{I} - \mathbf{K}_n\mathbf{H})\mathbf{Cov}(\epsilon_{b,n}, \epsilon_y)\mathbf{K}_n^T \\ + \mathbf{K}_n\mathbf{Cov}(\epsilon_y, \epsilon_{b,n},)(\mathbf{I} - \mathbf{K}_n\mathbf{H})^T + \mathbf{K}_n\mathbf{R}\mathbf{K}_n^T,\end{aligned} \tag{3.18}$$

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.

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

As a result, the exact analysis error covariance $\mathbf{A}_{\mathrm{E},n}$ tends to be under-estimated by $\mathbf{A}_{\mathrm{A},n}$ in Eq.3.16 throughout the iterations.

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

$$B_{\mathrm{A}}, R \in \mathbb{R}^+ \setminus \{0\}, H \in \mathbb{R} \setminus \{0\}. \tag{3.19}$$

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

$$B_{\mathrm{A},n+1} \leftarrow \left(1 - \frac{B_{\mathrm{A},n}H^2}{B_{\mathrm{A},n}H^2 + R}\right)B_{\mathrm{A},n} = \frac{B_{\mathrm{A},n}R}{B_{\mathrm{A},n}H^2 + R}. \tag{3.20}$$

In fact, one may see that the assumed error covariance (scalar variance in this case) $\mathbf{B}_{\mathrm{A},n\to\infty}$ 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:

$$f(x) = \frac{xR}{xH^2 + R}, \quad \text{where} \quad R \in \mathbb{R}^+ \setminus \{0\}, H \in \mathbb{R} \setminus \{0\}. \tag{3.21}$$

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

$$\forall x \in \mathbb{R}^+ \setminus \{0\}, \quad f(x) < x, \tag{3.22}$$

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 $B_{\mathrm{A},n\to\infty} \to 0$ for any initial value $B_{\mathrm{A}} \in \mathbb{R} \setminus \{0\}$. This theoretical result is numerically confirmed in Fig 3.1, (solid green line). The distribution of the exact covariance, consistent with the updating loop (Eq. (3.18)), 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 **R** of observations to be well known.

3.3.3 CUTE (Covariance Updating iTerativE) method

As pointed out in section 3.3.2, 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:

$$\mathbf{Cov}(\epsilon_{b,n}, \epsilon_y) = \mathbf{Cov}(\epsilon_y, \epsilon_{b,n})^T = \mathbf{Cov}\Big(\Big[(\mathbf{I} - \mathbf{K}_{n-1}\mathbf{H})\epsilon_{b,n-1} + \mathbf{K}_{n-1}\epsilon_y\Big], \epsilon_y\Big) \tag{3.23}$$

$$= (\mathbf{I} - \mathbf{K}_{n-1}\mathbf{H})\mathbf{Cov}(\epsilon_{b,n-1}, \epsilon_y) + \mathbf{K}_{n-1}\mathbf{R}, \tag{3.24}$$

with

$$\mathbf{Cov}(\epsilon_{b,0}, \epsilon_y) = 0_{\dim(x_b)\times\dim(y)}. \tag{3.25}$$

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 $\mathbf{B}_{\mathrm{A},n}$ 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 $\mathbf{B}_{\mathrm{A},n}$ 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 $Tr(\mathbf{B}_{\mathrm{A},n+1})$. The latter, representing the posterior covariance estimation, is introduced in Eq.3.28.

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

$$\mathbf{x}_{b,n+1} \leftarrow \mathbf{x}_{a,n}, \tag{3.26}$$

$$\begin{aligned}\mathbf{A}_{\mathrm{A},n} = &(\mathbf{I}-\mathbf{K}_n\mathbf{H})\mathbf{B}_{\mathrm{A},n} + (\mathbf{I}-\mathbf{K}_n\mathbf{H})\mathbf{Cov}(\epsilon_{b,n},\epsilon_y)\mathbf{K}_n^T \\ &+ \mathbf{K}_n\mathbf{Cov}(\epsilon_y,\epsilon_{b,n})(\mathbf{I}-\mathbf{K}_n\mathbf{H})^T,\end{aligned} \tag{3.27}$$

$$\mathbf{B}_{\mathrm{A},n+1} \leftarrow \frac{(1-\alpha)Tr(\mathbf{B}_{\mathrm{A},n}) + \alpha Tr(\mathbf{A}_{\mathrm{A},n})}{Tr(\mathbf{A}_{\mathrm{A},n})}\mathbf{A}_{\mathrm{A},n} \tag{3.28}$$

where $\mathbf{K}_n$ is expressed in Eq. (3.17). The more confident we are in the initial guess $\mathbf{B}_{\mathrm{A},0}$, 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 $\mathbf{B}_{\mathrm{A},n}$ 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 $\mathbf{B}_{\mathrm{E},n+1}$, which depend on the set up of $\mathbf{B}_{\mathrm{A},n}$, can be expressed as:

$$\begin{aligned}\mathbf{B}_{\mathrm{E},n+1} \leftarrow \mathbf{A}_n = &(\mathbf{I}-\mathbf{K}_n\mathbf{H})\mathbf{B}_{\mathrm{E},n}(\mathbf{I}-\mathbf{K}_n\mathbf{H})^T + (\mathbf{I}-\mathbf{K}_n\mathbf{H})\mathbf{Cov}_{\mathrm{E}}(\epsilon_{b,n},\epsilon_y)\mathbf{K}_n^T \\ &+ \mathbf{K}_n\mathbf{Cov}_{\mathrm{E}}(\epsilon_y,\epsilon_{b,n})(\mathbf{I}-\mathbf{K}_n\mathbf{H})^T + \mathbf{K}_n\mathbf{R}\mathbf{K}_n^T.\end{aligned} \tag{3.29}$$

We remind that the term $\mathbf{Cov}_{\mathrm{E}},(\epsilon_{b,n},\epsilon_y)$, which represents the exact background-observation covariances, is calculated as done in Eq. (3.24) but using the exact updated covariance matrix $\mathbf{B}_{\mathrm{E},n}$ 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. (3.4)) may be questioned. Therefore, under the assumption of linearity, we propose

another formulation that relies directly on the BLUE estimator in an extended space.

3.3.4 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. (3.4)). To overcome this shortage, our idea is to merge the background and observations in a broader space of larger dimension (as shown in [Talagrand, 1998]) 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:

$$\mathbf{z} = \tilde{\mathbf{H}}\theta + \mathbf{w}, \tag{3.30}$$

where **w** is a white noise.

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

$$J(\mathbf{x}) = \frac{1}{2}||\mathbf{z} - \tilde{\mathbf{H}}\theta||_{C^{-1}}, \tag{3.31}$$

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

$$\begin{aligned} \hat{\theta} &= (\tilde{\mathbf{H}}^T\mathbf{C}^{-1}\tilde{\mathbf{H}})^{-1}\tilde{\mathbf{H}}^T\mathbf{C}^{-1}\mathbf{z}, \\ \mathbf{C}_{\hat{\theta}} &= (\tilde{\mathbf{H}}^T\mathbf{C}^{-1}\tilde{\mathbf{H}})^{-1}. \end{aligned} \tag{3.32}$$

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 3.2.3. Furthermore, under the assumption of linearity of **H** and under the assumption of the Gaussian distributions, Eq.3.32 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 (3.5-3.7) by simply combining the observation and the background spaces:

$$\theta \equiv \mathbf{x}_t, \quad \mathbf{z} \equiv \begin{pmatrix} \mathbf{x}_b \\ \mathbf{y} \end{pmatrix}, \quad \tilde{\mathbf{H}} \equiv \begin{pmatrix} \mathbf{I} \\ \mathbf{H} \end{pmatrix}, \quad \mathbf{w} \equiv \begin{pmatrix} \epsilon_b \\ \epsilon_y \end{pmatrix}, \quad \mathbf{C} \equiv \begin{pmatrix} \mathbf{B} & 0 \\ 0 & \mathbf{R} \end{pmatrix}. \tag{3.33}$$

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. (3.33). The assumed covariance matrix in the extended space is denoted as $\mathbf{C}_{\mathrm{A}}$ ($\mathbf{C}_{\mathrm{A},0}$ latter in the iterative method) with no initial covariance between the background state and the observations, which can be written as:

$$\mathbf{C}_{\mathrm{A}} \equiv \begin{pmatrix} \mathbf{B}_{\mathrm{A}} & 0 \\ 0 & \mathbf{R} \end{pmatrix}. \tag{3.34}$$

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

$$\mathbf{x}_{b,n+1} \leftarrow \mathbf{x}_{a,n} = (\tilde{\mathbf{H}}^T \mathbf{C}_{\mathrm{A},n}^{-1} \tilde{\mathbf{H}})^{-1} \tilde{\mathbf{H}}^T \mathbf{C}_{\mathrm{A},n}^{-1} \mathbf{z}_n, \tag{3.35}$$

$$\mathbf{z}_{n+1} = \begin{pmatrix} \mathbf{x}_{b,n+1} \\ \mathbf{y} \end{pmatrix} \tag{3.36}$$

$$\mathbf{A}_{\mathrm{A},n} = (\tilde{\mathbf{H}}^T \mathbf{C}_{\mathrm{A},n}^{-1} \tilde{\mathbf{H}})^{-1} \tag{3.37}$$

$$\mathbf{B}_{\mathrm{A},n+1} \leftarrow \frac{(1-\alpha) Tr(\mathbf{B}_{\mathrm{A},n}) + \alpha Tr(\mathbf{A}_{\mathrm{A},n})}{Tr(\mathbf{A}_{\mathrm{A},n})} \mathbf{A}_{\mathrm{A},n} \tag{3.38}$$

$$\mathbf{Cov}(\epsilon_{b,n+1}, \epsilon_y) = (\tilde{\mathbf{H}}^T \mathbf{C}_{\mathrm{A},n}^{-1} \tilde{\mathbf{H}})^{-1} \tilde{\mathbf{H}}^T \mathbf{C}_{\mathrm{A},n}^{-1} \begin{pmatrix} \mathbf{Cov}(\epsilon_{b,n}, \epsilon_y) \\ R \end{pmatrix} \tag{3.39}$$

$$\mathbf{C}_{\mathrm{A},n+1} = \begin{pmatrix} \mathbf{B}_{\mathrm{A},n+1} & \mathbf{Cov}(\epsilon_{b,n+1}, \epsilon_y) \\ \mathbf{Cov}(\epsilon_{b,n+1}, \epsilon_y)^T & \mathbf{R} \end{pmatrix} \tag{3.40}$$

where $\mathbf{C}_{\mathrm{A},n}$ 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 $\mathbf{C}_{\mathrm{A},n}$

Analysis

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

$$\mathbf{C}_{\mathrm{E},n} = \begin{pmatrix} \mathbf{B}_{\mathrm{E},n} & \mathbf{Cov}_{\mathrm{E}}(\epsilon_{b,n}, \epsilon_y) \\ \mathbf{Cov}_{\mathrm{E}}(\epsilon_{b,n}, \epsilon_y)^T & \mathbf{R} \end{pmatrix}, \tag{3.41}$$

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

$$\mathbf{B}_{\mathrm{E},n+1} = (\tilde{\mathbf{H}}^T \mathbf{C}_{\mathrm{A},n}^{-1} \tilde{\mathbf{H}})^{-1} \tilde{\mathbf{H}}^T \mathbf{C}_{\mathrm{A},n}^{-1} \, \mathbf{C}_{\mathrm{E},n} \left((\tilde{\mathbf{H}}^T \mathbf{C}_{\mathrm{A},n}^{-1} \tilde{\mathbf{H}})^{-1} \tilde{\mathbf{H}}^T \mathbf{C}_{\mathrm{A},n}^{-1} \right)^T. \tag{3.42}$$

We note that this method does not only take into account the updated variances but also modify the optimisation formula (3.31) 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 3.4. 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.

3.3.5 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 3.3 and depicted in Fig. (3.1), 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. $\mathbf{B}_{\mathrm{E},n}^{\mathsf{Naive}}$, $\mathbf{B}_{\mathrm{E},n}^{\mathrm{CUTE}}$, $\mathbf{B}_{\mathrm{E},n}^{\mathrm{PUB}}$), 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. $\mathbf{B}_{\mathrm{A},n}^{\mathsf{Naive}}$, $\mathbf{B}_{\mathrm{A},n}^{\mathrm{CUTE}}$, $\mathbf{B}_{\mathrm{A},n}^{\mathrm{PUB}}$). In the left figure, all algorithms start from a perfect knowledge of the prior background variance (i.e. $\mathbf{B}_{\mathrm{E}} = \mathbf{B}_{\mathrm{A}}$). 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. $\mathbf{B}_{\mathrm{A},n+1} = \mathbf{A}_{\mathrm{A},n}$).

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 $\mathbf{B}_{\mathrm{A},n}^{\mathrm{CUTE}}$ and $\mathbf{B}_{\mathrm{A},n}^{\mathrm{PUB}}$ 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.

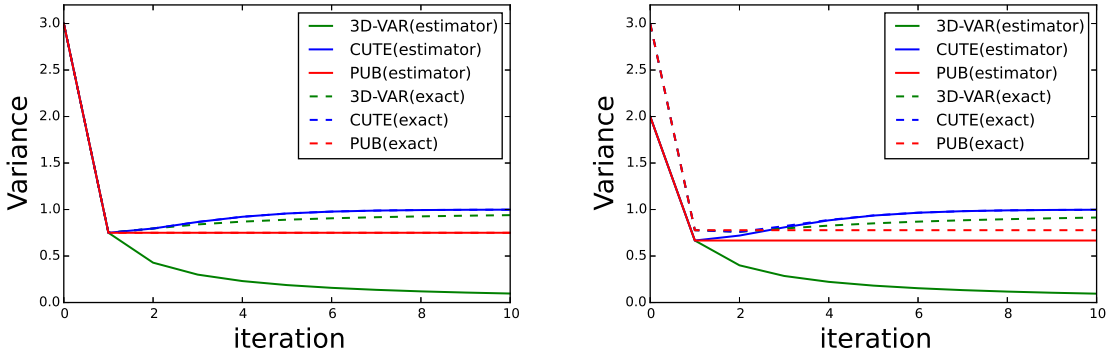


Figure 3.1: Analysis of the evolution of the exact updated background error variance $\mathbf{B}_n$ (dashed curves) vs its estimation provided by data assimilation algorithms $\mathbf{B}_{\mathrm{A},n}$ (solid curves). On the left side, the prior error variance is perfectly known (i.e. $\mathbf{B}_{\mathrm{A}} = \mathbf{B}_{\mathrm{E}} = 3$) at the initial step. On the right side, the background variance is voluntarily under-estimated: $(\mathbf{B}_{\mathrm{A}} = 2, \mathbf{B}_{\mathrm{E}} = 3)$. We remind that the updating of $\mathbf{B}_{\mathrm{A},n}$ is independent of the exact covariance evolution $\mathbf{B}_{\mathrm{E},i=1,\ldots,n}$; however, $\mathbf{B}_{\mathrm{E},n}$ is a function of the recurrence $\mathbf{B}_{\mathrm{A},i=1,\ldots,n-1}$.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. 3.8, Table. 3.1).

3.4 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. 3.2 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.

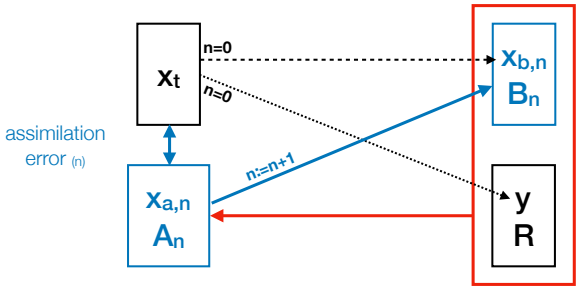


Figure 3.2: Scheme of a twin experiments data assimilation framework for an iterative method. Quantities in black are kept fixed while iterations are repeated: new assimilated state $\mathbf{x}_{a,n}$ and covariance errors $\mathbf{A}_n$ are injected at the next step in order to update background quantities. 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.

3.4.1 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 [Stewart et al., 2013], [Cioaca and Sandu, 2014]). 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 ([Saint-Venant, 1871]) coupling the fluid velocity and height,

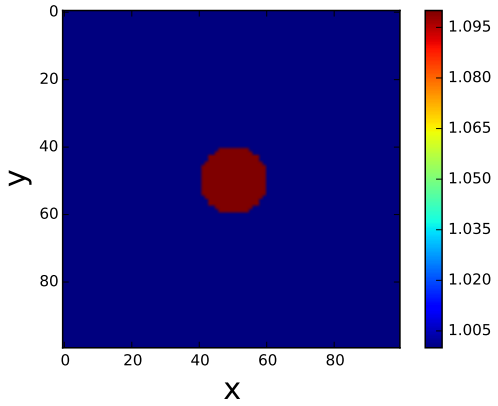


Figure 3.3: Initial h of shallow water in mm

$$
\begin{aligned}
\frac{\partial u}{\partial t} &= -g\frac{\partial}{\partial x}(h) - bu \\
\frac{\partial v}{\partial t} &= -g\frac{\partial}{\partial y}(h) - bv \\
\frac{\partial h}{\partial t} &= -\frac{\partial}{\partial x}(uh) - \frac{\partial}{\partial y}(vh) \\
u_{t=0} &= 0 \\
v_{t=0} &= 0.
\end{aligned}
\tag{3.43}
$$

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 $h^{\mathrm{CYL}}_{t=0} = 0.1mm$ high above the one of the still water as shown in Fig. 3.3. 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. (3.43) 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. 3.4(a-b)), and the obtained solution is used as the reference state ($\mathbf{x}_t$) in 3.4.2 and 3.4.3. Our objective is to reconstruct the state $x = (u, v)$ within a non-centered (10×10) subdomain (represented by a red square in Fig. 3.4 (c) and (d)) from noisy measurements via data assimilation processes. Thanks to an observation operator **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. $\mathrm{x}_t \equiv \{\mathrm{x}_t(k)\}_{k=1\ldots200} \equiv \{(u(t = t_f), v(t = t_f))\}$. The observation vector **y** is of size 100 but with zero elements included as shown in Fig. 3.4 (d).

In all numerical tests, we keep the assumption of linearity of the transformation operator $\mathcal{H}$. As we have mentioned in section 3.2.2, $\mathcal{H}$ 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 **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 $\{\mathrm{x}_t^*(i)\}_{i=1\ldots n}$ randomly but homogeneously chosen (with replacement) with probability P among the available data set, i.e. $\{\mathrm{x}_t(k)\}_{k=1\ldots 200}$. The subset values x_t^* are summed up and the process is re-iterated 100 times in order to construct the observations:

$$y(j) = \sum_{i=1}^{n_j} \mathrm{x}_t^*(i) + \epsilon_y, \quad \text{for } j = 1, \ldots, 100, \tag{3.44}$$

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 **H** is randomly chosen, it is kept fixed for a whole set of numerical experiments. This operator **H** is shown in Fig. 3.4 ((c) and (d)). In fact, with this definition of **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. 3.4 (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 **H** of more regular span structure tend to be less optimal than the ones obtained in the case with randomly simulated **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 **H**.

3.4.2 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

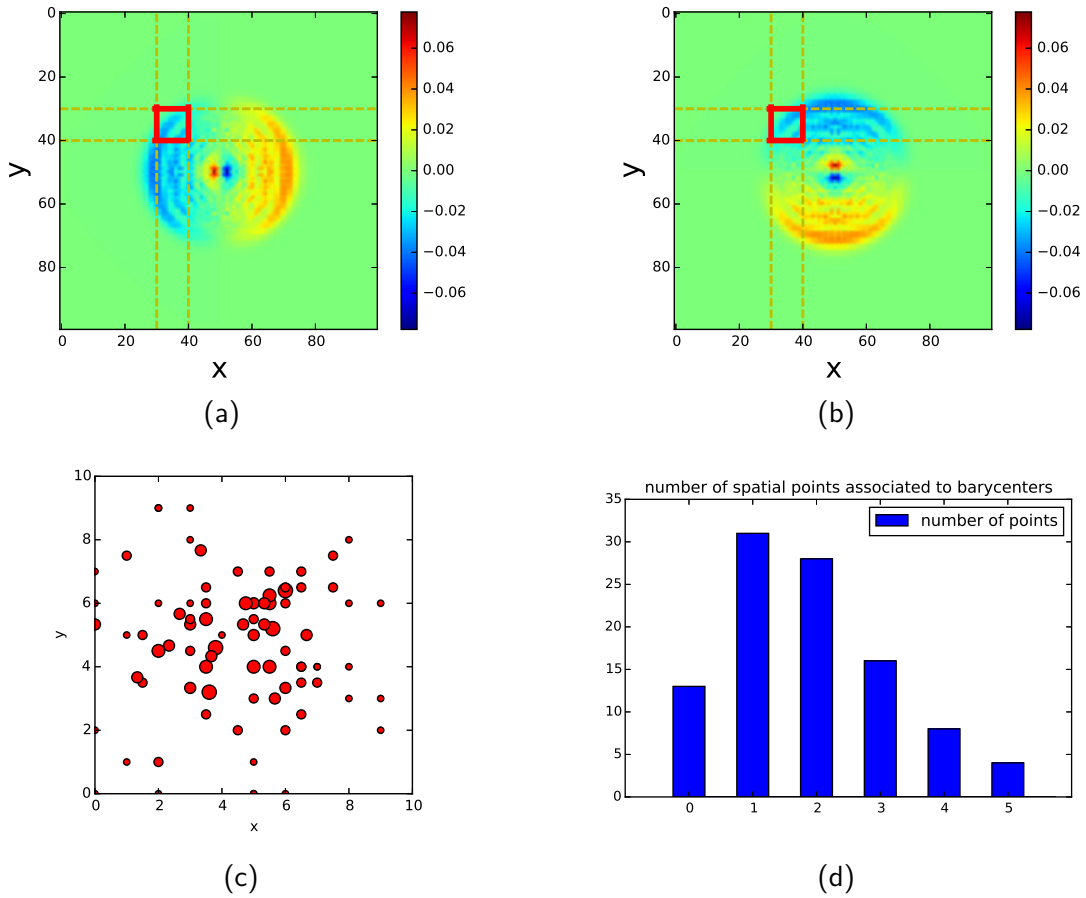


Figure 3.4: Illustrations of the random observation operator $\mathbf{H}$ and example of 2D flow velocity fields of the shallow water model. Barycenters measured by the linear transformation operator $\mathbf{H}$ are shown in (c) where the symbol radius is proportional to the number of measures associated 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. (3.43)) at time $t = t_f$. Data assimilations are performed in the red square subdomain.

prior errors ϵ_b and ϵ_y, as:

$$\begin{aligned} \mathbf{x}_a - \mathbf{x}_t &= \mathbf{x}_b + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}_b) - \mathbf{x}_t \\ &= \epsilon_b + \mathbf{K}(\epsilon_y - \mathbf{H}\epsilon_b) \\ &= (\mathbf{I} - \mathbf{K}\mathbf{H})\epsilon_b + \mathbf{K}\epsilon_y. \end{aligned} \tag{3.45}$$

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 $\mathbf{B}_{\mathrm{E}}$ and $\mathbf{R}$. Our assumption of higher background error amplitude leads to:

$$Tr(\mathbf{B}_{\mathrm{E}}) > Tr(\mathbf{R}). \tag{3.46}$$

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. $\mathbf{B}_{\mathrm{E}}$ and $\mathbf{B}_{\mathrm{A},n=0}$). 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 [Singh et al., 2011], [Ponçot et al., 2013]). 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 $\mathbf{B}_{\mathrm{E}}$ are supposed to be poorly specified by $\mathbf{B}_{\mathrm{A}}$, 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 $\mathbf{B}_{\mathrm{A},n=0}$. As mentioned in section 3.3, 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. (3.12), the error correlation matrices associated to $\mathbf{B}_{\mathrm{E},n}$ (exact background error correlation at n^{th} step) and $\mathbf{B}_{\mathrm{A},n}$ (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) ([Cherian et al., 2011]) defined for two semi-positive definite matrices $\mathbf{X}$ and $\mathbf{Y}$ by:

$$D_{\mathrm{AIRM}}(\mathbf{X}, \mathbf{Y}) = ||\log(\mathbf{X}^{-1/2}\mathbf{Y}\mathbf{X}^{-1/2})||_F \tag{3.47}$$

where $||.||_F$ 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 $\mathbf{Cor}_{\mathbf{B}_{\mathrm{A},n}}$ and the exact background error correlation matrix $\mathbf{Cor}_{\mathbf{B}_{\mathrm{E},n}}$ defined respectively from covariance matrices $\mathbf{B}_{\mathrm{A},n}$ and $\mathbf{B}_{\mathrm{E},n}$ as shown in Eq. (3.12). According to [Pennec et al., 2006], invariant under linear transformations, AIRM can be seen as a natural choice of metric for symmetric semi-positive definite matrices.

Fig. 3.5 -3.7 and Table 3.1 represent the results of twin experiments with different mis-specified (in terms of both amplitude and error correlation) background matrices $\mathbf{B}_{\mathrm{A}}$, 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 ($\mathbf{B}_{\mathrm{A},n=0}$, 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 stared green line) and the results of the same *3D-VAR* with the *exact* background error covariance matrix (i.e. $\mathbf{B}_{\mathrm{A}} = \mathbf{B}_{\mathrm{E}}$, 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. 3.5-3.7) 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 $\mathbf{B}_{\mathrm{A}}$ (stared 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. 3.5 (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. 3.5 (c)) 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. 3.5 (d-f) and the correlation mismatches in Table (3.1), 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. 3.5 (d-e) and used to calculate the correlation mismatch in Table 3.1, 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 3.1. 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-motonicity for the CUTE method.

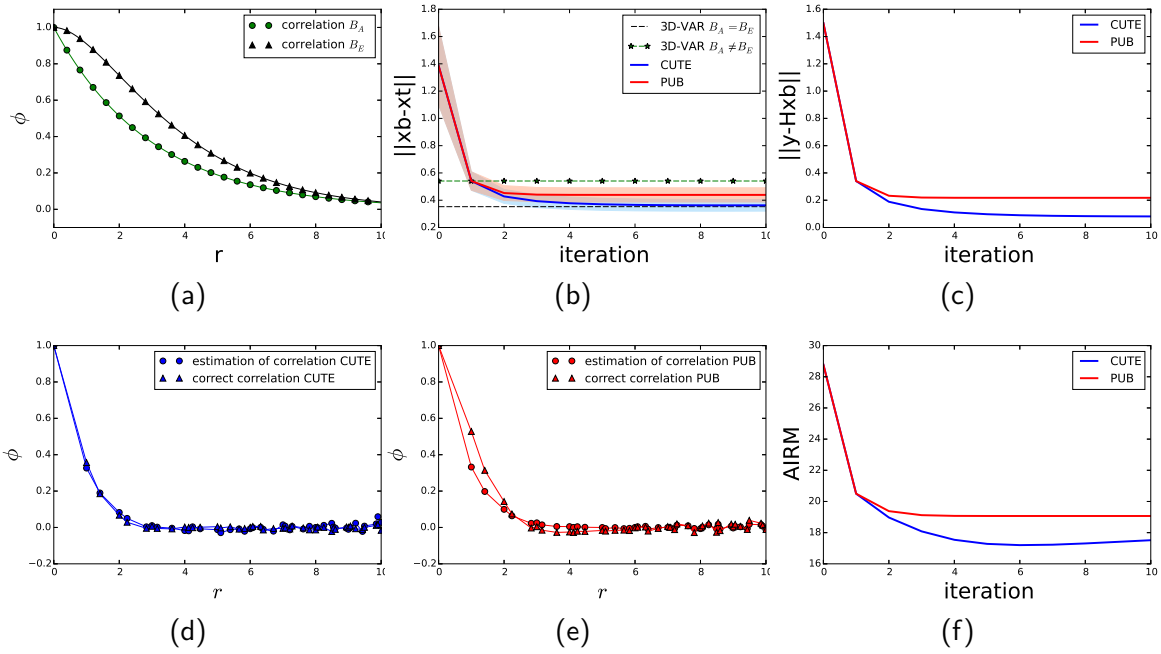
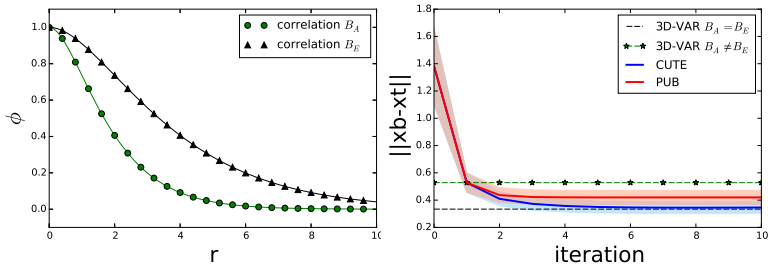
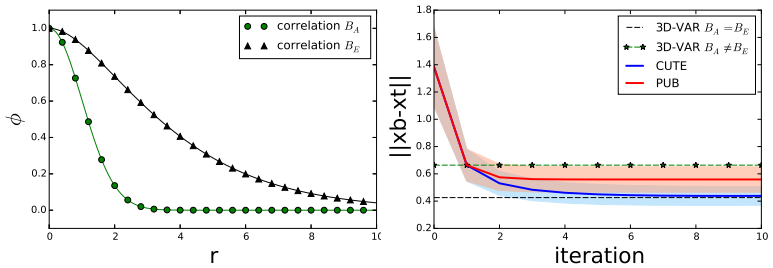


Figure 3.5: Twin experiments with state-independent homogeneous prior error. Figures on the first line refer to the initial choice of prior correlations (a) and the evolution of assimilation error (b) and innovation quantities (c), while figures on the second line monitor iterated quantities extracted from the errors covariance in the velocity field of u (d-f). In this test, $\mathbf{B}_{\mathrm{A},n=0}$ is chosen to follow an exponential kernel with $L = 3$ (shown by the green curve in (a)).



(a) (b)

Figure 3.6: Evolution of assimilation error in twin experiments using same simulated observations as Fig. 3.5 with different initial background matrix estimation ($\mathbf{B}_{\mathrm{A},n=0}$ is of Balgovind type with scale length $L = 1$).



(a) (b)

Figure 3.7: Evolution of assimilation error in twin experiments using same simulated observations as Fig. 3.5 with different initial background matrix estimation ($\mathbf{B}_{\mathrm{A},n=0}$ is of Gaussian type with length scale $L = 1$).

$\mathbf{B}_{\mathrm{A},n=0}$ 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	0.305	0.660	26.642	19.518	20.957

Table 3.1: Quantification of results of the CUTE and PUB iterative methods in terms of error correlation identification at the tenth iteration. The prior error covariance $\mathbf{B}_{\mathrm{E}}$ is set to be of Balgovind type with correlation length $L = 2$, homogeneous and state-independent. The mismatch of calibrated correlation functions is calculated with an L^2 norm error between the one-dimensional correlation curves. The AIRM criteria is also reported.

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 $\mathbf{B}_{\mathrm{A}}$ 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 $\mathbf{B}_{\mathrm{A}}$ (i.e. $Tr(\mathbf{B}_{\mathrm{A}}) < Tr(\mathbf{B}_{\mathrm{E}})$). This assumption is consistent with the phenomenon of background error inflation as mentioned in 3.2.2. We remind that as the dimension of the observation space is inferior to the one of the state space, the equation

$$\mathbf{y} = \mathbf{H}\mathbf{x} \tag{3.48}$$

is underdetermined. It thus defines a hyperplane in the space of $\mathbf{x}$. Fig. 3.8 (b) shows that the CUTE method converges to a stable state when the assimilation error $||\mathbf{x}_{a,n}^{\mathrm{CUTE}} - \mathbf{x}_t||$ is very close to the optimal target $||\mathbf{x}_a^{\mathrm{optimal}} - \mathbf{x}_t||$. However, we don't necessarily have

$$\mathbf{x}_{a,n} \rightarrow \mathbf{x}_a^{\mathrm{optimal}}. \tag{3.49}$$

3.4.3 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. ([Bishop, 2019]).

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 3.4.2.
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 $\mathbf{B}_{\mathrm{E}}$ (resp. $\mathbf{R}_{\mathrm{E}}$). Above assumption of

state-dependent errors leads to:

$$\mathbf{D}_i^B = (\mu_b \times \mathbf{x}_{t,i})^2 \tag{3.50}$$
$$\mathbf{D}_i^R = (\mu_o \times (\mathbf{H}\mathbf{x}_t)_i)^2, \tag{3.51}$$

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:

$$\mathbf{B}_{\mathrm{E}} = (\mathbf{D}^B)^{\frac{1}{2}}\mathbf{Cor}_{B,\mathrm{E}}(\mathbf{D}^B)^{\frac{1}{2}} \tag{3.52}$$
$$\mathbf{R}_{\mathrm{E}} = (\mathbf{D}^R)^{\frac{1}{2}}\mathbf{Cor}_{R,\mathrm{E}}(\mathbf{D}^R)^{\frac{1}{2}},$$

where the exact prior correlation matrices $\mathbf{Cor}_{B,\mathrm{E}}$ and $\mathbf{Cor}_{R,\mathrm{E}}$ are still chosen to follow homogeneous and isotropic correlation kernels as in the case of 3.4.2.

The two velocity fields u and v are supposed to be uncorrelated in terms of prior estimation error. Thus both $\mathbf{B}_{\mathrm{E}}$ and $\mathbf{B}_{\mathrm{A}}$ 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 $\mathbf{R}_{\mathrm{A}}$ for assuming observation error covariance, which is different from the true observation error covariance only in this section (3.4.3). 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. 3.8-3.10. We keep homogeneous structure of the assumed covariance matrices $\mathbf{B}_{\mathrm{A}}$ (constructed using correlation kernels) and $\mathbf{R}_{\mathrm{A}}$ (set to be the identity matrix) as in section 3.4.2. We also choose to keep the trace of $\mathbf{B}_{\mathrm{A}}$ and $\mathbf{R}_{\mathrm{A}}$ during iterative processes CUTE/PUB. Results in Fig. 3.8-3.10 and Table 3.2 show that for state-dependent errors, CUTE and PUB iterative methods could also significantly reduce the output errors (sub-figures (b) of Fig. 3.8-3.10) 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. 3.8 (f) and in the last two columns in Table 3.2, 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.

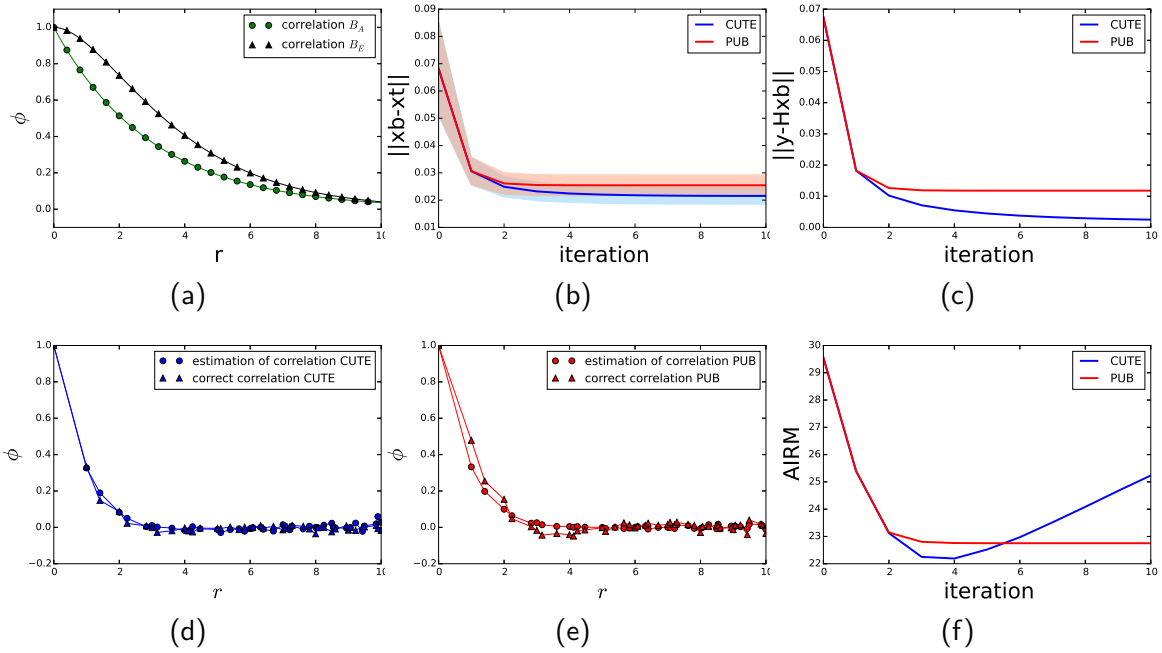


Figure 3.8: Twin experiments with state-dependent prior error for background state and observations. Figures on the first line refer to the initial choice of prior correlations (a) and the evolution of assimilation error (b) and innovation quantities (c), while figures on the second line monitor iterated quantities extracted from the errors covariance in the velocity field of u(d-f). In this test, the correlation of $\mathbf{B}_{\mathrm{A},n=0}$ is chosen to follow an exponential kernel with $L = 3$ (shown by the green curve in (a)).

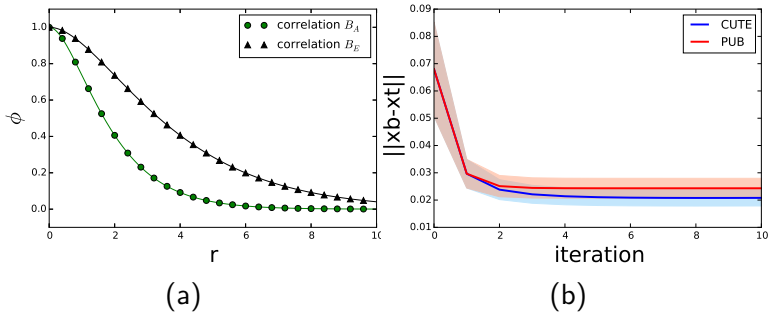


Figure 3.9: Evolution of assimilation error in twin experiments using same simulated observations as Fig. 3.8 with different initial background matrix estimation (the correlation kernel of $\mathbf{B}_{\mathrm{A},n=0}$ is of Balgovind type with length scale $L = 1$).

$\mathbf{B}_{\mathrm{A},n=0}$ 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 3.2: 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, $L = 2$.

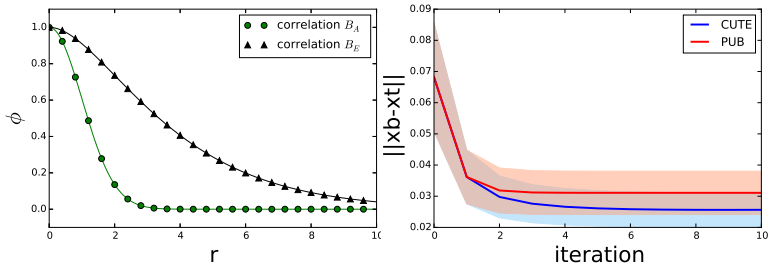
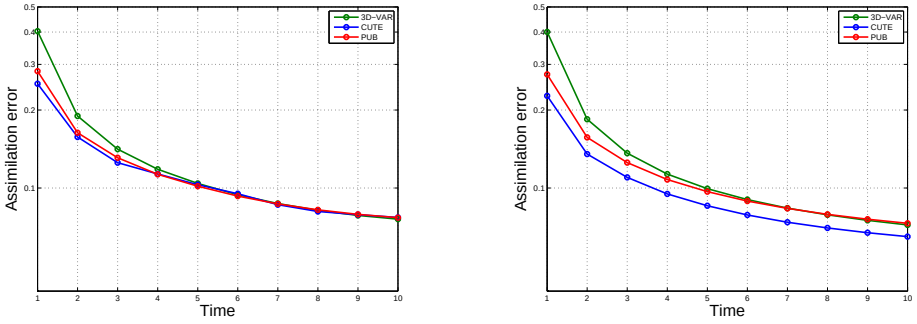


Figure 3.10: Evolution of assimilation error in twin experiments using same simulated observations as Fig. 3.8 with different initial background matrix estimation (the correlation kernel of $\mathbf{B}_{\mathrm{A},n=0}$ is of Gaussian type with length scale $L = 1$).

3.4.4 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 3.4.3, 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 $\mathbf{B}_{\mathrm{A}}$. 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 3.4.2. 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. 3.11 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. 3.11 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 ([Rabier, 2005]), which is the case in our experiments.

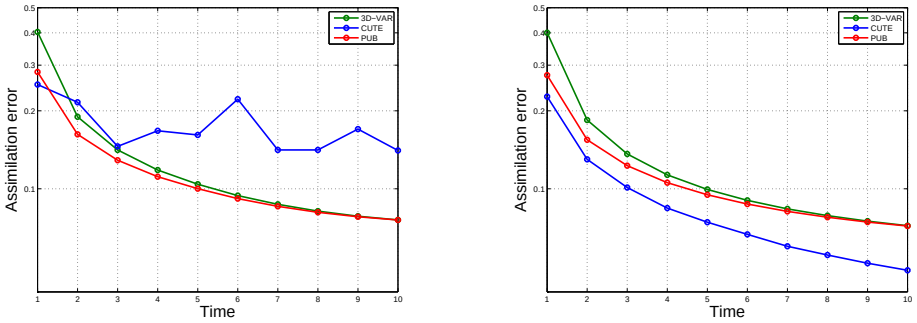


(a) $\sigma_b = 10\sigma_o$ (b) $\sigma_b = 100\sigma_o$

Figure 3.11: 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).

The result in Fig. 3.11 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. 3.11 (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 iterative methods continuously at several different moments. We present in Fig. 3.12, 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. 3.11 and 3.12 are equivalent.

We observe clearly from Fig. 3.12 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. 3.12 (a). In general, the PUB method remains more robust and less sensitive to the hypothesis of the high level of background error.



(a) $\sigma_b = 10\sigma_o$ (b) $\sigma_b = 100\sigma_o$

Figure 3.12: Comparison of *dynamical* twin experiments, with same initial conditions as in Fig. 3.11, where ten CUTE and PUB sub-iterations are performed at each assimilation step, illustrated by the green and red disk symbols.

3.5 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 $Tr(\mathbf{B}_{\mathrm{E}}) > Tr(\mathbf{R}_{\mathrm{E}})$ 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 access their performance in a more realistic/sophisticated industrial application case.

Chapter 4

A graph clustering approach to localization for adaptive covariance tuning in data assimilation based on state-observation mapping

submitted for publication to *Mathematical Geoscience* as
Cheng, S., Argaud, J.-P., Iooss, B., Ponçot, A., and Lucor, D. (2020). A graph clustering approach to localization for adaptive covariance tuning in data assimilation based on state-observation mapping

Abstract

An original graph clustering approach to efficient localization of error covariances is proposed within an ensemble-variational data assimilation framework. Here the localization term is very generic and refers to the idea of breaking up a global assimilation into subproblems. This unsupervised localization technique based on a linearized state-observation measure is general and does not rely on any prior information such as relevant spatial scales, empirical cut-off radius or homogeneity assumptions. The localization is performed thanks to graph theory, a branch of mathematics emerging as a powerful approach to capture complex and highly interconnected Earth and environmental systems in computational geosciences. The novel approach automatically segregates the state and observation variables in an optimal number of clusters, more amenable to scalable data assimilation. The application of this method does not require underlying block-diagonal structures of prior covariance matrices. In order to deal with inter-cluster connectivity, two alternative data adaptations are proposed. Once the localization is completed, a covariance diagnosis and tuning is performed within each cluster, which contribution is sequentially integrated into the entire covariance matrices. Numerical twin-experiment tests show that this approach is less costly and more flexible than a global covariance tuning, and most often brings more accurate results both for observation- and background-error parameters tuning.

4.1 Introduction

Data assimilation techniques, originally developed for numerical weather prediction (NWP), have been widely applied in the field of geosciences ([Blayo et al., 2012], [Carrassi et al., 2018]) for instance for reconstruction of physical fields (e.g temperature or velocity fields) or parameter identification. The applications, often complex, multi-physics and nonlinear with different model resolutions and prediction time horizons, vary from reservoir modelling ([Kumar, 2018]), heat transfer problems ([Jiang et al., 2020]), to geological feature prediction ([Vo and Durlofsky, 2014]), to operational oceanography. The goal of data assimilation is to reduce the uncertainty in prediction that arise due to uncertainties in input variables such as parameters and state variables by combining the information embedded in a prior estimation (also known as the background state) and real time observations or measures. Unfortunately, the gigantic size, i.e. $\mathcal{O}(10^{6-9})$ for multi-dimensional problems of geosciences data assimilation problems makes a full Bayesian approach computationally unaffordable. Instead, a variational approach weighs these two information sources, thanks to the background state $\mathbf{x}_b$ and the observation $\mathbf{y}$ with their associated error covariances, represented by the matrix $\mathbf{B}$ and $\mathbf{R}$ respectively.

These prior covariance matrices can be estimated with the help of a correlation kernel (e.g. [Stewart et al., 2013], [Gong et al., 2020b]) or a diffusion operator (e.g. [Weaver and Courtier, 2001]). The computation of these covariances may also be performed/improved by ensemble methods ([Evensen, 1994]), or some iterative methods for which some features of $\mathbf{B}$ and/or $\mathbf{R}$ are supposed to be known e.g. [Desroziers and Ivanov, 2001], [Desroziers et al., 2005], [Cheng et al., 2019]. These approaches quite often rely on converged state ensemble statistics, noiseless dynamical system or assumption of error amplitude ([Talagrand, 1998],[Cheng et al., 2019]). These conditions are usually difficult to be satisfied for high-dimensional geophysical systems.

When the state ensemble size is too small compared to the problem dimension, sampling errors may very well induce spurious long-distance error correlations resulting in poor conditioning of $\mathbf{B}$ and $\mathbf{R}$. An important ingredient used to make data assimilation more efficient and robust, follows the idea of *localization*. It relies on the intuitive idea that "distant" states of the system are more likely to be independent, at least for sufficiently short time scales. For applications where system variables depend on spatial coordinates, such as NWP, it is possible to *spatially* localize the analysis. For other systems, e.g. the interchannel radiance observation ([Garand et al., 2007]) or problems of parameter identification ([Schirber et al., 2013]), the correlation between different ranges/scales of the state or observation variables may not be directly interpreted in terms of spatial distances and the assumption of weak long-distance correlations might be less relevant. In this paper, we will refer to the more generic "long-range correlation" expression instead. Also, there might be situations for which a prior covariance structure has limited spatial extent, that is smaller than the support of the observation operator that maps state to observations spaces. In this case, non-local observations, i.e. observations that cannot be really allocated to one specific spatial location, because they may result from spatial averages of linear or non-linear functions of the system variables can have a large influence on the assimilation, cf. the work of [van Leeuwen, 2019].

Existing localization methods are mainly two kinds: covariance localization and domain localization. The first family of localization methods is implicit and works on a regularization of the

covariance matrix that is operated using a Schur matrix product with certain short-range predefined correlation matrices ([Gaspari and E. Cohn, 1999]), which ensures the (semi)definitiveness of the new matrix and therefore avoids the introduction of spurious long-range correlation. These methods have been widely improved, e.g. ensemble-based Kalman filters (EnKF) ([Farchi and Bocquet, 2019]) where the covariance localization is crucial to produce more accurate analyses especially when covariance inflation take place ([Hamill et al., 2001]). The second class of families (domain localization) is explicit and performs data assimilation for each state variable by using only a local subset of available observations, typically within a fixed range of this point. In this case, a relevant localization length must be carefully chosen. This is the main disadvantage of the approach: if this length is chosen too small, some important short- to medium-range correlation will be falsely neglected.

Recent works have shown that a local diagnosis/correction of error covariance computation could be helpful for improving the forecast quality of the global system, e.g. [Waller et al., 2017], as well as reducing the computational cost. From the point of view of an observation, it introduces the concepts of – *domain of dependence*, i.e. the set of elements of the model state that are used to predict the model equivalent of this observation; and of – the *region of influence*, i.e. the set of analysis states that are updated in the assimilation using this observation. According to [Waller et al., 2017], difficulties appear with the domain localization when the region of influence is far offset from the domain of dependence. In fact, the former which represents the set of analysis states that are updated in the assimilation using same observations may be imposed based on prior assumptions while the later is obtained from the linearized transformation operator, which depicts how the state variables are "connected" via the observations. Nevertheless, relying purely on imposed cut-off radius for localization may deteriorate this connection, resulting in a less optimal posterior estimation especially when long-range error covariance is present, as illustrated in the numerical experiments of [Waller et al., 2017]. Empirical choice of cut-off or distance thresholds may result in removal of true physical long-range correlations, thus inducing imbalance in the analysis ([Greybush et al., 2011]). This conclusion points to the relevance of more efficient and less arbitrary segregation operators. The spatial dependence between state variables and observations stands for an essential problem in inversion of nonlinear problems, such as subsurface flows. The probability conditioning method (PCM) ([Jafarpour and Khodabakhshi, 2011]), is another class of data assimilation using probability maps of state variables from an ensemble of updated models and assimilating the probability with multipoint statistical techniques for generating geological patterns. This allows for the representation of realistic natural formations with non-Gaussian statistics. Performing domain localization based on state-observation mapping may improve the quality of these probability maps, contributing overall to the algorithmic efficiency and training process of such approaches.

In practice, data assimilation often deals with non-uniform error fields, containing underlying structure due to the heterogeneity of the data, which calls for *unsupervised* localization schemes. One of the main objectives of unsupervised learning is to split a data set into two or more classes based on a similarity measure over the data, without resorting to any *a priori* information on how it should be done (see [Hastie et al., 2001], section 14). Fig. 4.1 illustrates with a very simple schematic the class of problems which could benefit from such an approach. It depicts

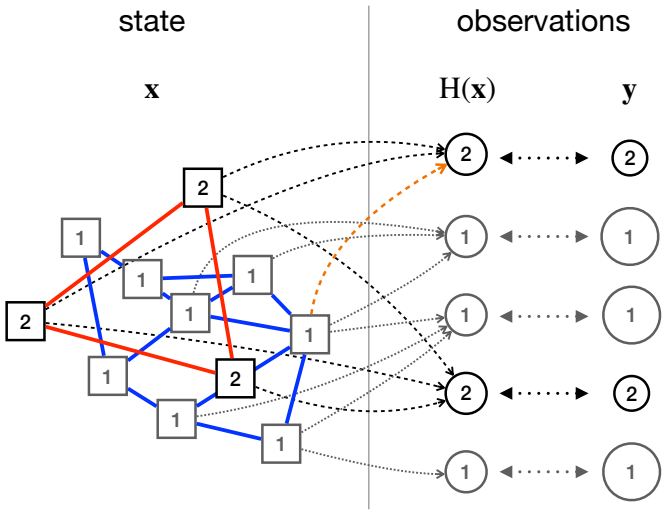


Figure 4.1: Simple sketch illustrating the type of relations between state variables and observations considered for data assimilation in this paper. The observation operator $\mathcal{H}$ maps some state variables $\mathbf{x}$ to the space of observations so that they can be compared with the experimental measurements $\mathbf{y}$. A graph clustering approach is put to use as a *localizer* to reveal unknown state variable/observation communities.

the type of relations between state variables and observations considered for data assimilation. The observation operator $\mathcal{H}$ maps some state variables $\mathbf{x}$ to the space of observations so that they can be compared with the experimental measurements $\mathbf{y}$. Despite the various contributions, the mapping is quite exclusive as some variables do not contribute to some observations, i.e., observations of 2-type depend on a certain group of variables, while the observations of 1-type inherit some values from another group of variables[1]. For illustration, one may apprehend the two groups of state variables in terms of spatial scales. This situation may arise for instance if two classes of sensors of different precision (illustrated by the circles size) and span are used to collect the data. A key ingredient of our data assimilation approach will be to automatically and correctly *localize* these state variables/observations *clusters* (otherwise named as subspaces or communities), for instance to be able to reveal inner-cluster networks.
In this study, we choose to segregate the state variables directly based on the information provided by the state-observation mapping. This unifying approach avoids potential conflicts between the region of influence and the domain of dependence of the localized assimilation. In this study, we choose to segregate the state variables directly based on the information provided by the state-observation mapping for a more flexible and efficient covariance tuning. This unifying approach

[1]In Fig. 4.1, only a single observation contributes from both groups of state variables, cf. orange arrow.

avoids potential conflicts between the region of influence and the domain of dependence of the localized assimilation.

A first original idea of our work, is to turn to efficient localization strategies based on *graph clustering* theory, which are able to automatically detect several clusters or "communities" (we will also refer to them as "subspaces" in the state and observation space) of state variables and corresponding observations. This clustering of variables will allow more local assimilation, likely to be more flexible and efficient than a standard global assimilation technique. In recent years, graph theory has been introduced in geosciences for a large range of utilities, such as: quantifying complex network properties, e.g., similarity, centrality and clustering or identifying special graph structures, e.g., small-world or scale-free networks. These graph-based techniques are very useful for improving the computational efficiency of geophysical problems, as well as bringing more insight into the quantification of feature interactions ([Phillips et al., 2015], [Qian et al., 2019]).

In a more general framework, graphical models are used in data assimilation problems of geoscience for representing both spatial and temporal dependencies of variables which reveals potential links among states and observations. More precisely, a data assimilation chain could be modeled as a hidden Markov process where the state variables are unobserved/hidden ([T. Ihler et al., 2005]). In this circumstance, graphical models could be considered as a variable dependency based localization methods. Another advantage of graphical representations, as pointed out by [T. Ihler et al., 2005], is introducing sparsity to the covariance structures which makes the covariance specification/modification more tractable. In this paper, we take one step further by applying directly a graph localization approach based on variable dependencies for covariance tuning. In summary, a similarity measure is evaluated for each state variables pair regarding their sensitivity to common observation points, which forms subsequently a graph/network structure. Community detection algorithms are then deployed in this network in order to provide subspaces segmentation. This network, called an observation-based state network, will only depend on the linearized transformation operator $\mathbf{H}$ between state variables and observations. More precisely, our objective is to classify the state variables represented by the same observation to the same subspace, regardless of their spatial distance.

Once the graph clustering approach has been efficiently applied for localizing several state communities, the next step is to take advantage of it in order to improve the prior state/observation errors covariance. Our approach proposes to perform a fine tuning of the entire matrices by sequentially updating the covariances thanks to the correction contribution coming from each cluster. In particular, we wish to improve the error covariance tuning *without* deteriorating prior error correlation knowledge. Therefore, it is crucial to rely on an appropriate posterior covariance tuning strategy, while appropriately assigning subset of observations to each community of state variables. We will show how different modeling and computational approaches are possible along those lines.

As mentioned previously, remarkable efforts have been made on posterior diagnosing and iterative adjustment of error covariance quantification, especially by the meteorology community. (e.g. [Desroziers and Ivanov, 2001], [Desroziers et al., 2005]). Among these tuning methods, the

one of Desroziers and Ivanov (also known as $\mathrm{DI01}$), which consists of finding a fixed point for the assimilated state by adjusting the ratio between background and observation covariance matrices amplitude without modifying their correlation structures, is well received in NWP. This approach presents the flexibility to be implemented either in a static or at any step of a dynamical data assimilation process for both variational methods and Kalman-type filtering, even with limited background/observation data. A different approach with full covariance estimation/diagnosis based on large ensembles, is for instance proposed in [Desroziers et al., 2005]. The later is based on statistics of prior and posterior innovation quantities. In fact, the deployment of $\mathrm{DI01}$ in subspaces has already been introduced in [Chapnik et al., 2004] for block diagonal structures of $\mathbf{B}$ and $\mathbf{R}$. In this paper, we adopt a $\mathrm{DI01}$ approach that we extend to a more general approach, where the block diagonal structure of the covariances matrix is no longer required, but covariance between extra-diagonal blocks remains accounted for.

The paper is organized as follows. The standard formulation of data assimilation is introduced, as well as its resolution in the case of a linearized Jacobian matrix, in section 4.2. We then explain how this Jacobian matrix, considered as a state-observation mapping, can be used to build an observation-based state network. The subspaces decomposition is carried out by applying graph-based community detection algorithms. The localized version of $\mathrm{DI01}$ is then introduced (section 4.4) and investigated in a twin experiments framework (section 4.5). We close the paper with a discussion (section 4.6).

4.2 Data assimilation framework

The goal of data assimilation algorithms is to correct the state $\mathbf{x}$ of a a dynamical system with the help of a prior estimation $\mathbf{x}_b$ and an observation vector $\mathbf{y}$, the former being often provided by expertise or a numerical simulation code. This correction brings the state vector closer to its true value denoted by $\mathbf{x}_t$, also known as the true state. In this paper, each state component $\mathbf{x}_i$ is called a state variable and $\mathbf{y}_j$ is called an observation where i, j represent the vector indices. The principle of data assimilation algorithms is to find an optimally weighted combination of $\mathbf{x}_b$ and $\mathbf{y}$ by optimizing the minimum cost of a cost function J defined as

$$
\begin{aligned}
J(\mathbf{x}) &= \frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) \\
&\quad + \frac{1}{2}(\mathbf{y}-\mathcal{H}(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{y}-\mathcal{H}(\mathbf{x})) \qquad (4.1)\\
&= J_b(\mathbf{x}) + J_o(\mathbf{x}) \qquad (4.2)
\end{aligned}
$$

where the observation operator $\mathcal{H}$ denotes the mapping from the state space to the one of observations. $\mathbf{B}$ and $\mathbf{R}$ are the associated error covariance matrices, i.e.

$$\mathbf{B} = \mathbf{cov}(\epsilon_b, \epsilon_b), \tag{4.3}$$
$$\mathbf{R} = \mathbf{cov}(\epsilon_y, \epsilon_y), \tag{4.4}$$

where

$$\epsilon_b = \mathbf{x}_b - \mathbf{x}_t, \tag{4.5}$$
$$\epsilon_y = \mathcal{H}(\mathbf{x}_t) - \mathbf{y}. \tag{4.6}$$

Their inverse matrices, $\mathbf{B}^{-1}$ and $\mathbf{R}^{-1}$, represent the weights of these two information sources in the objective function. Prior errors, ϵ_b and ϵ_y, are supposed to be centered Gaussian variables in this paper, thus they can be perfectly characterized by the covariance matrices, i.e.

$$\epsilon_b \sim \mathcal{N}(0, \mathbf{B}), \tag{4.7}$$
$$\epsilon_y \sim \mathcal{N}(0, \mathbf{R}). \tag{4.8}$$

The two covariance matrices $\mathbf{B}$ and $\mathbf{R}$, which are difficult to know perfectly *a priori*, play essential roles in data assimilation. The state-observation mapping $\mathcal{H}$ is possibly nonlinear in real applications. However, for the sake of simplicity, a linearization of $\mathcal{H}$ is often required to evaluate the posterior state and its covariance. The linearized operator $\mathbf{H}$, often known as the Jacobian matrix of $\mathcal{H}$ in data assimilation, can be seen as a mapping from the state space to the one of observation.

In the case where $\mathcal{H} = \mathbf{H}$ is linear and the covariances matrices $\mathbf{B}$ and $\mathbf{R}$ are well known, the optimization problem (Eq. 4.1) can be perfectly solved by linear formulation of the best linear unbiased estimator (BLUE)

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}_b) \tag{4.9}$$

which is also equivalent to a maximum *a posteriori* estimator. The Kalman gain matrix $\mathbf{K}$ is defined as

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}. \tag{4.10}$$

Several diagnosis or tuning methods, such as the ones of [Desroziers et al., 2005], [Desroziers and Ivanov, 2001], [Dreano et al., 2017] have been developed to improve the quality of covariance estimation/construction. Much effort has also been devoted to apply these methods in subspaces (e.g. [Waller et al., 2017], [Sandu and Cheng, 2015]). The subspaces are often divided by the physical nature of state variables or their spatial distance. The prior estimation errors are often considered as uncorrelated among different subspaces. A significant disadvantage of this approach

is that the cut-off correlation radius remains difficult to determine and the hypothesis of no error correlation among distant state variables is not always relevant depending on the application.

4.3 State-observation localization based on graph clustering methods

For the purpose of the simplicity of implementation, representing state variables/observations by block diagonal matrices is sometimes used in data assimilation (for example, see [Chabot et al., 2015]). In this case, only uncorrelated state variables can be separated. In this work, we are interested in applying covariance diagnosis methods in subspaces identified from the state-observation mapping, and we wish to make no assumption of block diagonal structures of the covariance.

The state subspaces will be detected thanks to an unsupervised graph clustering learning technique. Here, the graph will be formed by a set of vertices (i.e. the state discrete nodes) and a set of edges (based on a similarity measure over the state variables-observations mapping) connecting pairs of vertices. The graph clustering will automatically group the vertices of the graph into clusters taking into consideration the edge structure of the graph in such a way that there should be many edges within each cluster and relatively few between the clusters.

4.3.1 State space decomposition via graph clustering algorithms

Principles

The idea is to perform a localization by segregating the state vector $\mathbf{x} \in \mathbb{R}^{n_\mathbf{x}}$ (we drop the background or analysed subscript for the ease of notation) into a partition $\mathcal{C}$ of subvectors: $\mathcal{C} = \{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^p\}$, each $\mathbf{x}^i$ being non-empty. We will call later $\mathcal{C}$ a *clustering* and the elements $\mathbf{x}^i$ *clusters*. Similarly to the standard localization approach, for each identified subset of state variables, it will then be necessary to identify an associated subset of observations: $\{\mathbf{y}^1, \mathbf{y}^2, \ldots, \mathbf{y}^p\}$.

In the work of [Waller et al., 2016], a threshold of spatial distance $\tilde{r}$ is *arbitrarily* imposed *a priori* to define local subsets of state variables influenced by each observation during the data assimilation updating. In other words, each observation component $\mathbf{y}_i$, of the complete vector $\mathbf{y}$, is only supposed to influence the updating of a subset of state variables within the spatial range of $\tilde{r}$. This subset of state variables $\mathcal{R}_{\text{influence}}(\mathbf{y}_i) = \{\mathbf{x}_k : \phi(\mathbf{y}_i, \mathbf{x}_k) \leq \tilde{r}\}$, where ϕ measures some spatial distance, is called the *region of influence* of $\mathbf{y}_i$.

However, that method faces a significant difficulty when the Jacobian matrix $\mathbf{H}$ of $\mathcal{H}$ is dense or non local, i.e. the updating of state variables depends on observations out of the region of influence. In fact, the non-locality of matrix $\mathbf{H}$ may contain terms that will induce a "connection" between state variables and observations beyond the critical spatial range $\tilde{r}$. The

domain of dependence defined as

$$\mathcal{D}_{\text{dependence}}(\mathbf{y}_i) = \{\mathbf{x}_k : \mathbf{H}_{i,k} \neq 0\}, \tag{4.11}$$

is introduced to quantify the range of this state-observation connection which is purely decided by $\mathbf{H}$ instead of the spatial distance. [Waller et al., 2016] have shown that problems may occur in the covariance diagnosis when $\mathcal{R}_{\text{influence}}(\mathbf{y}_i)$ and $\mathcal{D}_{\text{dependence}}(\mathbf{y}_i)$ do not overlap. This incoherence not only impacts the assimilation accuracy but also the posterior covariance estimation. This phenomenon is also highlighted and studied in the work of [van Leeuwen, 2019] where the author proposes an extra step to assimilate observations outside the region of influence.

Observation-based state connections

Rather than considering the region of influence, our proposed approach uses a clustering strategy directly based on the domain of dependence, i.e. taking advantage of the particular structure of the transformation function $\mathcal{H}$ (or its linearized version $\mathbf{H}$). The main idea is to separate the ensemble of state variables into several subsets regarding their occurrence in the domains of dependence of different observations. In order to do so, we introduce the notion of observation-based connection between two state variables $\mathbf{x}_i$ and $\mathbf{x}_j$ when they appear in the domain of dependence of the same observation $\mathbf{y}_k$, i.e.

$$\exists k, \quad \text{such that} \quad \frac{\partial[\mathcal{H}(\mathbf{x})]_k}{\partial \mathbf{x}_i} \neq 0, \qquad \frac{\partial[\mathcal{H}(\mathbf{x})]_k}{\partial \mathbf{x}_j} \neq 0, \tag{4.12}$$

where $[\mathcal{H}(\mathbf{x})]_k$ stands for the k^{th} element in the reconstruction, referring to model equivalent observation $\mathbf{y}_k$. In this paper, we consider time-invariant mappings because they lead to invariant domains of dependence, which is beneficial from a computational point of view. For a linearized state-observation operator $\mathbf{H}$, it simply becomes

$$\exists k, \quad \text{such that} \quad \mathbf{H}_{k,i} \neq 0, \qquad \mathbf{H}_{k,j} \neq 0. \tag{4.13}$$

Our goal is to determine if we can group the state variables which are strongly connected based on the observations, regardless of their spatial distance. In order to do so, we define the strength of this connection for each pair of state variables,

$$\mathcal{S} : \mathbb{R}^{n_{\mathbf{x}}} \times \mathbb{R}^{n_{\mathbf{x}}} \mapsto \mathbb{R}^{+} \tag{4.14}$$

$$\text{if } i \neq j : \tag{4.15}$$

$$\mathcal{S}(\mathbf{x}_i, \mathbf{x}_j) \equiv \mathcal{S}_{i,j} = \sum_{k, i \neq j} \left| \frac{\partial [\mathcal{H}(\mathbf{x})]_k}{\partial \mathbf{x}_i} \right| \left| \frac{\partial [\mathcal{H}(\mathbf{x})]_k}{\partial \mathbf{x}_j} \right|$$

$$= \sum_{k} |\mathbf{H}|_{k,i} |\mathbf{H}|_{k,j} \quad \text{if} \quad \mathcal{H} \equiv \mathbf{H} \quad \text{is linear},$$

$$\text{if } i = j : \tag{4.16}$$

$$\mathcal{S}(\mathbf{x}_i, \mathbf{x}_j) = 0, \tag{4.17}$$

where $|\cdot|$ represents the absolute value (symmetric) function on the whole matrix (i.e $|\mathbf{H}|_{k,i} = |\mathbf{H}_{k,i}|$). The formulation is proposed for general problems, but in case of linearity of $\mathbf{H}$ the graph-clustering identification becomes easier, especially when the data assimilation problem is of large dimension with a sparse observation operator. In fact, several data assimilation algorithms already require a linearization of $\mathcal{H}$. In these cases, little computational overhead is added for graph computing. When the operator is fully nonlinear, careful attention has to be given to the evaluation of the partial derivatives.
In the rest of this paper, we will assume that $\mathcal{H} \equiv \mathbf{H}$ is linear.

Moreover, we will assume that the function is null when measuring the connection strength of one state variable with itself. In case $|\mathbf{H}|$ exhibits extremely large values, extra smoothing (e.g. of sigmoid type) could be applied on $|\mathbf{H}|_{k,i}|\mathbf{H}|_{k,j}$ in order to appropriately balance the graph weight. Finally, in case of data assimilation of multi-variate problems, care has to be taken of inhomogeneous $\mathbf{H}$ matrix, which would result in perturbations for graph clustering. We propose to either deal with each variable type individually (i.e performing graph-clustering localization for each type of state variables) or introduce some kind of normalization to balance the structure of $\mathbf{H}$. For example, the sum of each column in $|\mathbf{H}|$ could be set as a fixed value.

We now consider an undirected graph $\mathcal{G}$ that is a pair of sets $\mathcal{G} = (\mathbf{x}, \mathbf{E})$, where $\mathbf{x}$ plays the role of the set of vertices (the number of vertices $n_{\mathbf{x}}$ is the order of the graph) and the set $\mathbf{E}$ contains the edges of the graph (the edge cardinality, i.e. $|\mathbf{E}| = m$ represents the size of the graph). Each edge is an unordered pair of endpoints $\{\mathbf{x}_k, \mathbf{x}_l\}$. We are going to use our measure $\mathcal{S}$ as a weight function to define the weighted version of the graph $\mathcal{G}_{\mathcal{S}} = (\mathbf{x}, \mathbf{E}, \mathcal{S})$. This translates into the weighted adjacency matrix $\mathbf{A}_{\mathcal{G}_{\mathcal{S}}}$ of the graph, that is a $n_{\mathbf{x}} \times n_{\mathbf{x}}$ matrix $\mathbf{A}_{\mathcal{G}_{\mathcal{S}}} = (a^{\mathcal{G}_{\mathcal{S}}}_{\mathbf{x}_i, \mathbf{x}_j})$,

$$a^{\mathcal{G}_{\mathcal{S}}}_{\mathbf{x}_i, \mathbf{x}_j} = \begin{cases} \mathcal{S}_{i,j} \text{ if } \{\mathbf{x}_i, \mathbf{x}_j\} \in \mathbf{E}, \\ 0 \quad \text{otherwise.} \end{cases} \tag{4.18}$$

This matrix will be useful to perform the graph clustering.

Each edge of the graph thus represents the connection strength between two state variables. For some problems, it is possible to organize the graph into clusters, with many edges joining vertices of the same cluster and comparatively few edges joining vertices of different clusters. We have the partition $\mathcal{C} = \{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^p\}$ of $\mathbf{x}$, and we identify a cluster $\mathbf{x}^i$ with a node-induced subgraph of $\mathcal{G}_\mathcal{S}$, i.e. the subgraph $\mathcal{G}_\mathcal{S}[\mathbf{x}^i] := \left(\mathbf{x}^i, \mathbf{E}(\mathbf{x}^i), \mathcal{S}_{|\mathbf{E}(\mathbf{x}^i)}\right)$, where $\mathbf{E}(\mathbf{x}^i) := \{\{\mathbf{x}_k, \mathbf{x}_l\} \in \mathbf{E} : \mathbf{x}_k, \mathbf{x}_l \in \mathbf{x}^i\}$. So $\mathbf{E}(\mathcal{C}) := \bigcup_{i=1}^{p} \mathbf{E}(\mathbf{x}^i)$ is the set of intra-cluster edges and $\mathbf{E}\backslash\mathbf{E}(\mathcal{C})$ is the set of inter-cluster edges of cluster $\mathbf{x}^i$ respectively, with $|\mathbf{E}(\mathcal{C})| = m(\mathcal{C})$ and $|\mathbf{E}\backslash\mathbf{E}(\mathcal{C})| = \bar{m}(\mathcal{C})$, while $\mathbf{E}(\mathbf{x}^i, \mathbf{x}^j)$ denotes the set of edges connecting nodes in $\mathbf{x}^i$ to nodes in $\mathbf{x}^j$. We denote $\bar{m}^c(\mathcal{C}) = p(p-1) - \bar{m}(\mathcal{C})$, representing the number of non-connecting inter-cluster pairs of vertices. It is important to stress that the identification of structural clusters is made easier if graphs are *sparse*, i. e. if the number of edges m is of the order of the number of nodes $n_\mathbf{x}$ of the graph ([Fortunato, 2010]).

Clustering algorithms

One of the main paradigms of clustering is to find groups/clusters which ensure both intra-cluster density and inter-cluster sparsity ([Cheng et al., 2017]). Despite the fact that many problems related to clustering are NP-hard problems, there exist many approximation methods for graph-based community detection, such as the Louvain algorithm ([Blondel et al., 2008]) and the Fluid community algorithm ([Parés et al., 2017]). These methods are mostly based on random walks ([Gueuning et al., 2019]) or centrality measures in a network with the advantage of low computational cost. The use of graph theory in numerical simulation problems such as the Cuthill–McKee algorithm ([Cuthill and McKee, 1969]) already exists, for instance for sorting multidimensional grid points in a more efficient way (in terms of reducing the matrix band). In this paper, we introduce a different approach with the objective of identifying observation-based state variable communities which will be later considered as state subsets in covariance tuning. The community detection is performed on the observation-based state network, regardless of the algorithms chosen. Considering the computational cost, the Fluid community detection algorithm proposed by [Parés et al., 2017] could be an appropriate choice for sparse transformation matrix because its complexity is *linear* to the number of edges in the network, i.e. $\mathcal{O}(|\mathbf{E}|)$. When the state dimension is very large, the computation of $\mathcal{G}$ may be numerically infeasible. Whereas, researches in graph theory have shown that if the jacobian matrix is sparse, community detection algorithms could be performed without the computation of the full adjacency matrix (i.e. $|\mathbf{H}||\mathbf{H}|^T$), for example via a k-means method applied directly on $|\mathbf{H}|$, as shown in [Browet and Van Dooren, 2014].

In real applications of graph theory, the number of optimal cluster p is often not known in advance. Finding appropriate cluster number remains a popular research topic. Several methods have been developed in order to propose some objective functions with notion of optimal coverage, performance or inter-cluster conductance, e.g. the Elbow method ([Ketchen and Shook, 1996]) or the Gap statistic method ([Tibshirani et al., 2001]). For instance the following performance

metric will be used later for the experiments in section 4.5.2,

$$\text{performance} := \frac{m(\mathcal{C}) + \bar{m}^c(\mathcal{C})}{\frac{1}{2} n_{\mathbf{x}}(n_{\mathbf{x}} - 1)}. \tag{4.19}$$

It represents the fraction of node pairs that are clustered correctly, i.e. those connected node pairs that are in the same cluster and those non-connected node pairs that are separated by the clustering.

A simple example of state-observation graph clustering

For illustration purpose, inspired by the pedagogical approach of [Waller et al., 2017], we consider the following simple system with $\mathbf{x} \in \mathbb{R}^{n_{\mathbf{x}}=9}$ and $\mathbf{y} \in \mathbb{R}^{n_{\mathbf{y}}=4}$,

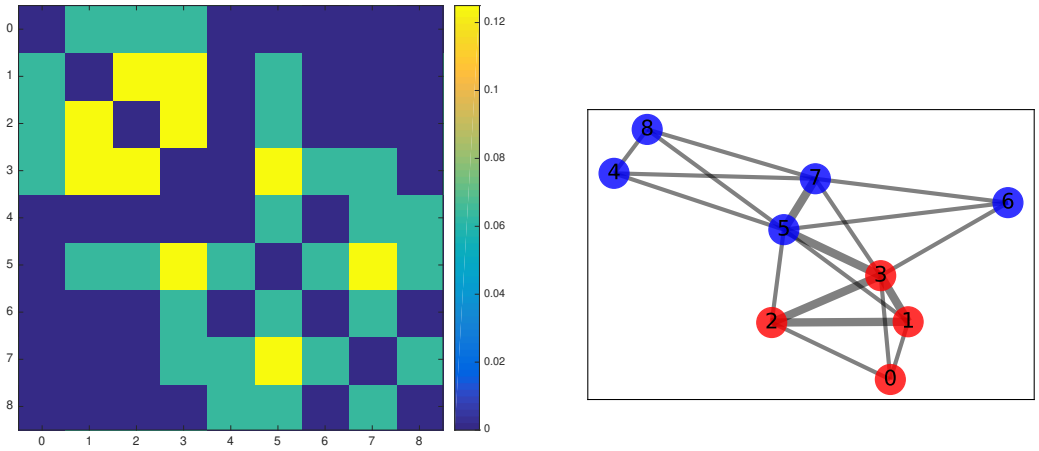
$$\mathbf{H} = 0.25 \times \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \end{bmatrix}, \tag{4.20}$$

where the magnitude of non-zero $\mathbf{H}$ entries is assumed constant for simplicity, which leads to the associated state-observation transformation function

$$\begin{array}{rcl} 0.25(\mathbf{x}_0 + \mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3) & = & \mathbf{y}_0 \\ 0.25(\mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3 + \mathbf{x}_5) & = & \mathbf{y}_1 \\ 0.25(\mathbf{x}_3 + \mathbf{x}_5 + \mathbf{x}_6 + \mathbf{x}_7) & = & \mathbf{y}_2 \\ 0.25(\mathbf{x}_4 + \mathbf{x}_5 + \mathbf{x}_7 + \mathbf{x}_8) & = & \mathbf{y}_3. \end{array} \tag{4.21}$$

The obtained observation-based adjacency matrix is represented in Fig. 4.2(a) and is quite sparse with only $m = 19$ edges. The clustering result obtained by the Fluid community detection algorithm is illustrated in Fig. 4.2(b). Two communities (red and blue colors) of state variables could be identified, where points in each community are tightly connected. In particular some intra-cluster nodes with strong connections (eg. $\{\mathbf{x}_1, \mathbf{x}_2\}$ or $\{\mathbf{x}_5, \mathbf{x}_7\}$) are well identified by the algorithm, in accordance with the large values of the adjacency matrix. However, connections across clusters can also be found, for example the connection $\{\mathbf{x}_3, \mathbf{x}_5\}$. These inter-cluster connections, are still managed by the algorithm. In fact, an output partition of perfect (noise free) subsets can hardly be obtained in real application problems.

After the identification of the state clusters, we need to associate each one with an ensemble of observations. As discussed previously, difficulty appears for observations with domains of dependence spanning across multiple clusters. In this case, it is necessary to operate some data preprocessing. For instance, the assignment of $\{\mathbf{y}_0\}$ and $\{\mathbf{y}_3\}$ respectively to the first (red) and



(a) (b)

Figure 4.2: (a) Observation-based state adjacency matrix obtained from the transformation operator **H** in Eq. (4.20). (b) Corresponding network identified by the community detection algorithm. The graph edge weights (measure of the strength of observation-based state connections) are represented by their widths.

second (blue) state community is without ambiguity while both $\{\mathbf{y}_1\}$ and $\{\mathbf{y}_2\}$ are overlapped by the two communities. Dealing with this type of overlapping in the observation partition is therefore crucial for the covariance tuning.

4.3.2 Dealing with inter-cluster observation region of dependence for assimilation

Assuming that a p-cluster structure $\mathcal{C} = \{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^p\}$ is provided by the community detection algorithm, we should assign, for each cluster $\mathbf{x}^i$, an associated observation subset $\mathbf{y}^i$, in order to perform local covariance tuning later on. As we will see in the following, while the partition $\mathcal{C} = \{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^p\}$ of $\mathbf{x}$ will remain the same, the partition of the observations $\{\mathbf{y}^1, \mathbf{y}^2, \ldots, \mathbf{y}^p\}$ will be constructed on a subvector of observations $\tilde{\mathbf{y}} \in \mathbb{R}^{n_{\tilde{\mathbf{y}}} \leq n_{\mathbf{y}}}$ or on a modified vector of observations $\hat{\mathbf{y}} \in \mathbb{R}^{n_{\mathbf{y}}}$. In this work, we propose two alternative methods, named "observation reduction" and "observation adjustment", providing appropriate observation subsets associated with each state cluster.

Observation reduction

Applying this strategy, the observation components $\mathbf{y}_k$ (thus $[\mathcal{H}(\mathbf{x})]_k$) with connections to several state variable clusters must be identified and canceled, i.e. all observations such that,

$$\frac{\partial[\mathcal{H}(\mathbf{x})]_{k=0,\ldots,n_{\mathbf{y}}-1}}{\partial \mathbf{x}^i_{l=1,\ldots,n^i_{\mathbf{x}},i=1,\ldots,p}} \neq 0, \tag{4.22}$$

for more than a single cluster, must be withdrawn from the assimilation procedure.

Nevertheless, we emphasize that these observation data can still be used later on for evaluating the posterior estimation $\mathbf{x}_a$ in the data assimilation procedure. Back to Eq. (4.21), the observations $\{\mathbf{y}_1\}$ and $\{\mathbf{y}_2\}$ are voluntarily excluded to perform the covariance correction, i.e. the tuning will be performed with only two clusters of subvectors,

$$\begin{array}{rcl} \mathbf{x}^1 & = & \{\mathbf{x}_{k=0,\ldots,3}\}, \quad \tilde{\mathbf{y}}^1 = \{\mathbf{y}_0\}, \\ \mathbf{x}^2 & = & \{\mathbf{x}_{k=4,\ldots,8}\}, \quad \tilde{\mathbf{y}}^2 = \{\mathbf{y}_3\}. \end{array} \tag{4.23}$$

The reduced global state-observation operator $\tilde{\mathbf{H}}$ thus becomes

$$\tilde{\mathbf{H}} = 0.25 \times \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \end{bmatrix}. \tag{4.24}$$

Observation adjustment

Here, the idea is to modify the observation data dependent on multiple clusters, in order to simply keep its strongest dependence to a single cluster. This way, each observation will be assigned to only one subset of state variables based on the state-observation mapping. This is done by substracting from the original observation value, the contribution of the surplus quantity related to the other clusters. We rely on the values of the background states to evaluate those surpluses. If more than one background state sample is available (that will be the case in the next section), the expected value of the background ensemble is used instead.

For example, if $\{\mathbf{y}_l\}$ has stronger ties to $\mathbf{x}^j$, then it should be readjusted as

$$\hat{\mathbf{y}}_l = \mathbf{y}_l - \sum_{i=1,\ldots,p, i\neq j} \sum_{k \mid \mathbf{x}_k \in \mathbf{x}^i} \mathbf{H}_{l,k} \mathbb{E}_b[\mathbf{x}_k], \tag{4.25}$$

where $\mathbb{E}_b[.]$ denotes the *empirical* expected value based on the prior background ensemble at hand. This approach leads to an adjusted Jacobian matrix $\hat{\mathbf{H}}$ that induces adjacency matrix with no overlapped domains. This is obviously an approximation due to the averaged operator. In fact, there are two error sources, a main one coming from the prior backgroup measure and another one due to the sampling error. We will see examples in section 4.5.2

Applied to the example, $\{\mathbf{y}_1\}$ and $\{\mathbf{y}_2\}$ can be respectively adjusted to belong to the first and

the second cluster. With the help of background state $\mathbf{x}_b$, Eq. (4.21) can be adjusted to:

$$\begin{aligned}
0.25(\mathbf{x}_0 + \mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3) &= \hat{\mathbf{y}}_0 = \mathbf{y}_0 \\
0.25(\mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3) &= \hat{\mathbf{y}}_1 = \mathbf{y}_1 - 0.25\mathbb{E}_b[\mathbf{x}_5] \\
0.25(\mathbf{x}_5 + \mathbf{x}_6 + \mathbf{x}_7) &= \hat{\mathbf{y}}_2 = \mathbf{y}_2 - 0.25\,\mathbb{E}_b[\mathbf{x}_3] \\
0.25(\mathbf{x}_4 + \mathbf{x}_5 + \mathbf{x}_7 + \mathbf{x}_8) &= \hat{\mathbf{y}}_3 = \mathbf{y}_3.
\end{aligned} \tag{4.26}$$

Thus the new operator can be written as

$$\hat{\mathbf{H}} = 0.25 \times \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \end{bmatrix}. \tag{4.27}$$

For real applications, one may envision a mixture of these two approaches.

4.4 Localized error covariance tuning

Now that we have localized our system based on the state-observation linearized measure, and thanks to graph clustering methods, we next explain how we take advantage of the localization in order to improve the error covariance tuning.

4.4.1 Desroziers & Ivanov diagnosis and tuning approach

The [Desroziers and Ivanov, 2001] tuning algorithm (DI01) was first proposed and applied in the meteorological science at the beginning of the 21st century. This method is based on the diagnosis and verification of innovation quantities and has been widely applied in geoscience (e.g. [Hoffman et al., 2013]) and meteorology. Consecutive works have been carried out to improve its performance and feasibility in problems of large dimension such as the study of [Chapnik et al., 2004]. Without modifying error correlation structures, the DI01 algorithm adjusts the observation-error weighting parameters by applying an iterative fixed-point procedure.

It was proven in [Talagrand, 1998] and [Desroziers and Ivanov, 2001] that, under the assumption of perfect knowledge of the covariance matrices $\mathbf{B}$ and $\mathbf{R}$, the following equalities are

perfectly satisfied in a 3D-VAR assimilation system:

$$\begin{aligned}\mathbb{E}\left[J_b(\mathbf{x}_a)\right] &= \frac{1}{2}\mathbb{E}\left[(\mathbf{x}_a-\mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x}_a-\mathbf{x}_b)\right] \\ &= \frac{1}{2}\mathrm{Tr}(\mathbf{KH}),\end{aligned} \tag{4.28}$$

$$\mathbb{E}\left[J_o(\mathbf{x}_a)\right] = \frac{1}{2}\mathbb{E}\left[(\mathbf{y}-\mathbf{H}\mathbf{x}_a)^T\mathbf{R}^{-1}(\mathbf{y}-\mathbf{H}\mathbf{x}_a)\right] \tag{4.29}$$

$$= \frac{1}{2}\mathrm{Tr}(\mathbf{I}-\mathbf{HK}), \tag{4.30}$$

where $\mathbf{x}_a$ is the output of a 3D-VAR algorithm with a linear observation operator $\mathbf{H}$. Eqs. 4.29 and 4.30 are seldomly satisfied in practice, in the sense that the accurate knowledge of prior error covariances is often out of reach for real data assimilation applications. Nonetheless, if we assume that the *correlation* structures of these matrices are well known, then it is possible to iteratively correct their magnitudes. Using the two indicators

$$s_{b,q} = \frac{2J_b(\mathbf{x}_a)}{\mathrm{Tr}(\mathbf{K}_q\mathbf{H})}, \tag{4.31}$$

$$s_{o,q} = \frac{2J_o(\mathbf{x}_a)}{\mathrm{Tr}(\mathbf{I}-\mathbf{H}\mathbf{K}_q)}, \tag{4.32}$$

where q is the current iteration, the objective of the $\mathrm{DI01}$ tuning method is to adjust the ratio between the weighting of $\mathbf{B}^{-1}$ and $\mathbf{R}^{-1}$ without modifying their correlation structure,

$$\mathbf{B}_{q+1} = s_{b,q}\mathbf{B}_q, \quad \mathbf{R}_{q+1} = s_{o,q}\mathbf{R}_q. \tag{4.33}$$

These two indicators act as scaling coefficients, modifying the error variance magnitude. We remind that both the reconstructed state $\mathbf{x}_a$ and the gain matrix $\mathbf{K}_q$ depend on $\mathbf{B}_q$, $\mathbf{R}_q$ and thus on the iterative coefficients $s_{b,q}$, $s_{o,q}$. The application of this method in subspaces where matrices $\mathbf{B}$ and $\mathbf{R}$ follow block-diagonal structures has also been discussed in [Desroziers and Ivanov, 2001].

Compared to other posterior diagnosis or iterative methods, e.g. [Desroziers et al., 2005], [Cheng et al., 2019], no estimation of full matrices is needed and only the estimation of two scalar values (J_b, J_o) is required in $\mathrm{DI01}$. Therefore, this method could be more suitable when the available data is limited. Another advantage relates to the computational cost of this method as DI01 requires only the computation of matrices trace which can be evaluated in efficient ways.

In practice, a stopping criteria of $\mathrm{DI01}$ could be designed by choosing a minimum threshold

of **max**$(||s_{b,q}-1||, ||s_{o,q}-1||)$. According to [Chapnik et al., 2004], the convergence of s_b and s_o can be very fast, especially in the ideal case where the correlation patterns of **B** and **R** are perfectly known. Under this assumption, [Chapnik et al., 2004] proved DI01 is equivalent to a maximum-likelihood parameter tuning. In addition, large iteration number is not required as the first iteration could already provide a reasonably good estimation of the final result. For this particular method, since the covariance matrices are tuned only based on current ensemble of $\mathbf{x}_b$ and **y**, it is not appropriate to apply DI01 in a dynamical data assimilation chain. In this case, we propose to perform a global tuning by averaging the s_b, s_o obtained at different time stamps.

4.4.2 Adaptation of the DI01 algorithm to localized subspaces

The application of data assimilation algorithms, as well as the full observation matrix diagnosis has been discussed in [Waller et al., 2017]. Following the notation of their paper, we introduce the binary selection matrix Φ_x^i, Φ_y^i of the i^{th} subvector with

$$\mathbf{x}^i = \Phi_x^i \mathbf{x}, \quad \mathbf{y}^i = \Phi_y^i \mathbf{y} \tag{4.34}$$

where i is the index of the subspace. The data assimilation in the subspace, as well as localized covariance tuning could be easily expressed using the standard formulation with projection operators Φ_x^i and Φ_y^i.

Given the example of the first pair of state and observation subsets in the case of Fig. 4.2, we have

$$\Phi_x^1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{4.35}$$

In the case of data reduction strategy (Eq.4.23),

$$\Phi_{y,\text{reduction}}^1 = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}, \tag{4.36}$$

while for data adjustment strategy,

$$\Phi_{y,\text{adjustment}}^1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}. \tag{4.37}$$

The error covariances matrix $\mathbf{B}^i$ (resp. $\mathbf{R}^i$) associated to $\mathbf{x}_b^i$ (resp. $\mathbf{y}^i$) can be written as

$$\mathbf{B}^i = \Phi_x^i \mathbf{B} \Phi_x^{i,T}, \quad \mathbf{R}^i = \Phi_y^i \mathbf{R} \Phi_y^{i,T}. \tag{4.38}$$

Therefore, the associated analyzed subvector $\mathbf{x}_a^i$ could be obtained by applying data assimilation procedure using $\left(\mathbf{x}_b{}^i, \mathbf{y}^i, \mathbf{B}^i, \mathbf{R}^i\right)$. We remind that, due to the cross-community noises (i.e. the updating of $\mathbf{x}_b{}^i$ may not only depend on $\mathbf{y}_b{}^i$ in the global data assimilation system), we don't necessarily have

$$\mathbf{x}_a^i = \Phi_x^i \mathbf{x}_a. \tag{4.39}$$

For more details of decomposition formulations, the interested readers are referred to [Waller et al., 2017]. Our objective for implementing localized covariance tuning algorithms is to gain a finer diagnosis and correction on the covariance computation. The local DI01 diagnosis in $\left(\mathbf{x}_b{}^i, \mathbf{y}^i\right)$ can be expressed as

$$\begin{aligned}
\mathbb{E}\left[J_b(\mathbf{x}_a^i)\right] &= \mathbb{E}\left[(\mathbf{x}_a^i - \mathbf{x}_b^i)^T (\mathbf{B}^i)^{-1} (\mathbf{x}_a^i - \mathbf{x}_b^i)\right] \\
&= \frac{1}{2}\mathrm{Tr}(\mathbf{K}^i \mathbf{H}^i), \qquad (4.40) \\
\mathbb{E}\left[J_o(\mathbf{x}_a^i)\right] &= \mathbb{E}\left[(\mathbf{y}^i - \mathbf{H}^i \mathbf{x}_b^i)^T (\mathbf{R}^i)^{-1} (\mathbf{y}^i - \mathbf{H}^i \mathbf{x}_b^i)\right] \\
&= \frac{1}{2}\mathrm{Tr}(\mathbf{I}^i - \mathbf{H}^i \mathbf{K}^i), \qquad (4.41)
\end{aligned}$$

where the optimization functions J_b and J_o, as well as the localized gain matrix $\mathbf{K}^i$ have also been adjusted in these subspaces. In our approach, the localized state-observation mapping $\mathbf{H}^i$ is obtained thanks to the graph-based localization, i.e. either identifying to the $\hat{\mathbf{H}}^i$ or $\tilde{\mathbf{H}}^i$ characterization. For simplicity, we will keep the $\mathbf{H}^i$ notation in the following. The identity matrix $\mathbf{I}^i$ is of the same dimension as $\mathbf{B}^i$. We then define the local tuning algorithm as

$$s_{b,q}^i = \frac{2J_b(\mathbf{x}_a^i)}{\mathrm{Tr}(\mathbf{K}_q{}^i \mathbf{H}^i)}, \tag{4.42}$$

$$s_{o,q}^i = \frac{2J_o(\mathbf{x}_a^i)}{\mathrm{Tr}(\mathbf{I}^i - \mathbf{H}^i \mathbf{K}_q{}^i)}, \tag{4.43}$$

$$\mathbf{B}_{q+1}^i = s_{b,q}^i \mathbf{B}_q^i, \tag{4.44}$$

$$\mathbf{R}_{q+1}^i = s_{o,q}^i \mathbf{R}_q^i. \tag{4.45}$$

The iterative process is repeated $q_{\max}^i$ times, based on some *a priori* maximum number of iterations

or some stopping criteria monitoring the rate of change. The approach provides a local correction within each cluster thanks to a multiplicative coefficient. This way, the covariance tuning is more flexible than a global approach relying on two coefficients (s_b, s_o) only.

However, if the updating is performed in each subspace (i.e. correction only on the sub-matrices $\mathbf{B}^i, \mathbf{R}^i$), then the adjusted $\mathbf{B}$ and $\mathbf{R}$ are not guaranteed to be positive-definite and the prior knowledge of covariance structure might be deteriorated. In order to circumvent this problem, we keep the correlation structure of ($\mathbf{C_B}$ and $\mathbf{C_R}$) fixed. We remind that a covariance matrix $\mathbf{Cov}$ (of random vector $\mathbf{x}$), which is by its nature positive semi-definite, can be decomposed into its variance and correlation structures as

$$\mathbf{Cov} = \mathbf{D}^{1/2}\mathbf{C}\,\mathbf{D}^{1/2},$$

where $\mathbf{D}$ is a diagonal matrix of the state error variances, and $\mathbf{C}$ is the correlation matrix. By correcting the variance in each subspace only through the diagonal matrices ($\mathbf{D}^i_\mathbf{B}, \mathbf{D}^i_\mathbf{R}$), the positive definiteness of $\mathbf{B}$ and $\mathbf{R}$ is thus guaranteed as the correlation structure remains invariant, cf. Algorithm 1.

4.4.3 Complexity analysis

Reducing computational cost could be seen as an important vocation of localization techniques, especially for domain localization methods ([Waller et al., 2017]). As an example, for a Kalman-type solver, the complexity mainly comes from the inversion and multiplication of matrices of large size, with typical unit cost of the order $\mathcal{O}(n_\mathbf{x}^\mu)$ where $\mu \in (2, 3)$ depending on the algorithm chosen, e.g. [Coppersmith and Winograd, 1990]. Therefore, the global $\mathrm{DI01}$ covariance tuning, for a given state vector of size $n_\mathbf{x}$, is of computational complexity

$$\mathcal{C}_{\text{global}}(n_\mathbf{x}) = q_{\max} \times n_\mathbf{x}^\mu. \tag{4.46}$$

On the other hand, applying algorithm 1 for p clusters of dimension $n_{\mathbf{x}^1}, ..., n_{\mathbf{x}^p}$ with $\sum_{i=1}^p n_{\mathbf{x}^i} = n_\mathbf{x}$, the complexity of localized covariance tuning $\mathcal{C}_{\text{localized}}$ writes:

$$\mathcal{C}_{\text{localized}}(n_{\mathbf{x}^1}, ..., n_{\mathbf{x}^p}) = \sum_{i=1}^{p} q^i_{\max} \times (n^i_\mathbf{x})^\mu. \tag{4.47}$$

Since the graph computation could be carried out offline as long as the operator $\mathbf{H}$ remains invariant, the cost of graph clustering is not considered here. Under the hypothesis that the

Algorithm 1: Localization and updating of $\mathbf{B}$ and $\mathbf{R}$ with cluster-based implementation of DI01 algorithm.

Inputs:
Background state: $\mathbf{x}_b$
Observation data: $\mathbf{y}$
Initially guessed matrix: $\mathbf{B}, \mathbf{R}$
Jacobian matrix: $\mathbf{H}$

Algorithm: Community detection using $\mathbf{H}$ with given or detected community number p
for i *from* 1 *to* p*:* **do**
  Extraction of subvectors ${\mathbf{x}_b}^i, \mathbf{y}^i$ and associated covariance matrices $\mathbf{B}^i, \mathbf{R}^i$.
  for q *from* 1 *to* q_{max} **do**
    calculation and storing of $\left\{s^i_{b,q}, s^i_{o,q}\right\}$ with $\mathbf{B}^i_q, \mathbf{B}^i_q$ via Eq. (4.42-4.43)
  end
  Updating of full covariance matrices from blockwise tuned covariance in current cluster:

$$\mathbf{B} \leftarrow (\mathbf{D}^i_{\mathbf{B}})^{1/2}\mathbf{B}\,(\mathbf{D}^i_{\mathbf{B}})^{1/2}$$
$$\mathbf{R} \leftarrow (\mathbf{D}^i_{\mathbf{R}})^{1/2}\mathbf{R}\,(\mathbf{D}^i_{\mathbf{R}})^{1/2}$$

  where $\mathbf{D}^i_{\mathbf{B}}$ and $\mathbf{D}^i_{\mathbf{R}}$ are diagonal matrices defined as

$$(\mathbf{D}^i_{\mathbf{B}})_{j,j} = \begin{cases} \prod_{q=1}^{q_{\max}} s^i_{b,q} & \text{if } \{\mathbf{x}_j\} \subset \mathbf{x}^i \\ 1 \quad \text{otherwise} \end{cases}$$
$$(\mathbf{D}^i_{\mathbf{R}})_{l,l} = \begin{cases} \prod_{q=1}^{q_{\max}} s^i_{o,q} & \text{if } \{\mathbf{y}_l\} \subset \mathbf{y}^i \\ 1 \quad \text{otherwise.} \end{cases}$$

end

outputs: Improved error covariances

clusters are of comparable size, Eq. 4.47 could be simplified as

$$\mathcal{C}_{\text{localized}}(n_{\mathbf{x}}) = \left(\sum_{i=1}^{p} q_{\max}^{i}\right) \times \left(\mathcal{O}\left(\left(\frac{n_{\mathbf{x}}}{p}\right)^{\mu}\right)\right) = \frac{\sum_{i=1}^{p} q_{\max}^{i}}{p} \times \frac{\mathcal{O}(n_{\mathbf{x}}^{\mu})}{\mathcal{O}(p^{\mu-1})}. \tag{4.48}$$

Considering the number of $\mathrm{DI01}$ iterations per cluster may be represented by a random integer centered around some mean value $\mathbb{E}\left[q_{\max}^{i}\right]$, the first term of Eq. 4.48 represents its empirical mean $\overline{q_{\max}^{i}}$. Because the clusters fragment the global problem in some simpler smaller problems, in general it is reasonable to assume that $\overline{q_{\max}^{i}} \leq q_{\max}$, we can easily deduce that $p^{\mu-1} \times \mathcal{C}_{\text{local}} \leq \mathcal{C}_{\text{global}}$. Therefore, the graph-based method is at least $\mathcal{O}(p^{\mu-1})$ times faster than the standard approach. This derivation also holds for most posterior covariance tuning methods other than DI01. Notice that data assimilation algorithms are often combined with other techniques, such as adjoint modelling. In these cases, the marginal computation cost of each iteration of $\mathrm{DI01}$ (both in subspaces and the global space) could be reduced further. Nevertheless, the value of μ in Eq. 4.48 will always remain strictly superior than one, regardless the computation strategy chosen.

It is also important to emphasize that the computational strategy could be easily ported to parallel computing, in particular in the case where the clusters do not overlap, lowering even more the computational time.

4.5 Illustration with numerical experiments

4.5.1 Test case description

Similar to the works of [Clifford S. et al., 2009] and [Waller et al., 2017], we illustrate our methodology with numerical experiments relying on synthetic data. Our numerical experiments shed some light on the important steps of our approach: a sparse state-observation mapping chosen to implicitly reflect on the presence of some clusters, an algorithm of community detection and the implementation of the covariance tuning method.

Construction of H

A sparse Jacobian matrix **H** reflecting the clustering of the state-observation mapping is generated; the components of which are then randomly mixed in order to hide any particular structural pattern. The dimension of the state space is set to be 100, $\mathbf{x} \in \mathbb{R}^{n_{\mathbf{x}}=100}$, while the dimension of the observation space is set to be $\mathbf{y} \in \mathbb{R}^{n_{\mathbf{y}}=50}$. We consider a case for which the state-observation mapping **H** reflects community structures. For this reason, we construct *a priori* two (this choice is arbitrary) subsets of observations each relating mainly to only one subset of state variables. In fact, clustering structure of Jacobian matrices could often be found in real-world

applications (see an example of building structure data in Fig. 3 of [Gerke, 2011]) due to its non-homogeneity in the space. In order to be as general as possible, we consider $|\mathbf{x}^1| = |\mathbf{x}^2| = 50$ and $|\mathbf{y}^1| = |\mathbf{y}^2| = 25$. For the sake of simplicity, the observation operator $\mathbf{H}$ (of dimension $[50 \times 100]$) is randomly filled with binary elements, forming a dominant blockwise structure with some extra-block non-zero terms. The latter is done in order to mimic realistic problems, i.e. some perturbations are introduced in the form of cross-communities perturbations, therefore the two communities are not perfectly separable.

The background/observation vectors and Jacobian matrix are then randomly shuffled in a coherent manner in order to hide the cluster structure to the community detection algorithm, as for the adjacency matrix in Fig. 4.4(a). More specifically, the state-observation mapping is constructed as follows: we use a binomial distribution with two levels of success probability,

$$Pr(\mathbf{H}_{i,j} = 1) = \begin{cases} 15\% & \text{if} \quad \mathbf{x}_i \in \mathbf{x}^1 \quad \text{and} \quad y_j \in \mathbf{y}^1 \\ 15\% & \text{if} \quad \mathbf{x}_i \in \mathbf{x}^2 \quad \text{and} \quad y_j \in \mathbf{y}^2 \\ 1\% & \text{otherwise} \quad (\text{perturbations}). \end{cases} \tag{4.49}$$

In the following tests, exact and assumed covariance magnitudes will be changed but we will always keep the same choice of Jacobian $\mathbf{H}$. The community detection, remaining also invariant for all Monte Carlo tests, is provided by the Fluid community-detection algorithm.

As explained previously, there is a particular interest to apply $\mathrm{DI01}$ in the case of limited access to data (i.e. small ensemble size of $(\mathbf{x}_b, \mathbf{y})$).

In these twin experiments, the prior errors are assumed to follow the distribution of correlated Gaussian vectors:

$$\epsilon_b = \mathbf{x}_b - \mathbf{x}_t \sim \mathcal{N}(0^{n_{\mathbf{x}}=100}, \mathbf{B}_{\mathrm{E}}), \tag{4.50}$$

$$\epsilon_y = \mathbf{y} - \mathbf{H}\mathbf{x}_t \sim \mathcal{N}(0^{n_{\mathbf{y}}=50}, \mathbf{R}_{\mathrm{E}}), \tag{4.51}$$

where $\mathbf{B}_E, \mathbf{R}_E$ denote the chosen exact prior error covariances, *hidden from the tuning algorithm*. We remind that under the assumption of state independent error and linearity of $\mathbf{H}$, the posterior assimilation error, as well as the posterior correction of $\mathbf{B}$ and $\mathbf{R}$ via $\mathrm{DI01}$ (regardless of the strategy chosen, i.e. data reduction or data adjustment), is independent of the theoretical value of $\mathbf{x}_t$ but only depends on prior errors (i.e. $\mathbf{x}_t - \mathbf{x}_b$ and $\mathbf{y} - \mathbf{H}\mathbf{x}_t$).

Twin experiments setup

In order to reflect the construction of $\mathbf{H}$, we suppose that the *exact* error deviation, *hidden from the tuning algorithm* (respectively denoted by $\sigma^i_{b,E}, \sigma^i_{o,E}$) are constant in each cluster, so for

instance we have:

$$\text{if} \quad \{\mathbf{x}_u, \mathbf{x}_v\} \subset \mathbf{x}^i, \quad \text{then} \quad \sigma^i_{b,E}(\mathbf{x}_u) = \sigma^i_{b,E}(\mathbf{x}_v).$$

For this numerical experiment, a quite challenging case is chosen with:

$$\begin{aligned} \sigma^{i=1}_{b,E}(\mathbf{x}_u) &= \sigma^{i=2}_{b,E}(\mathbf{x}_v) \\ \sigma^{i=1}_{o,E}(\mathbf{y}_u) &= \text{ratio} \times \sigma^{i=2}_{o,E}(\mathbf{y}_v), \end{aligned}$$

so that the background error is homogeneous while the observation error is different in the two communities with a fixed ratio (in the following, we will choose $ratio = 10$). However, the correlation structures of the covariance matrices are supposed to be known *a priori*, and are assumed to follow a Balgovind structure:

$$(\mathbf{C_B})_{i,j} = (\mathbf{C_R})_{i,j} = \left(1 + \frac{r}{L}\right) \exp^{-\frac{r}{L}}, \tag{4.52}$$

where $r \equiv r(\mathbf{x}_i, \mathbf{x}_j) = r(\mathbf{y}_i, \mathbf{y}_j) = |i - j|$ is a pseudo spatial distance between two state variables, and the correlation scale is fixed ($L = 10$) in the following experiments. The Balgovind structure is also known as the $\nu = 3/2$ Matern kernel, often used in prior error covariance computation in data assimilation (see for example [Ponçot et al., 2013], [Stewart et al., 2013]).

We remind that the output of all DI01 based approaches depend on the available background and observation data set. We compare three different methods described previously in this paper, differentiated by the notation used for their output covariances:

- $(\mathbf{B}, \mathbf{R})$: implementation of $\mathrm{DI01}$ in full space,
- $(\tilde{\mathbf{B}}, \tilde{\mathbf{R}})$: implementation of $\mathrm{DI01}$ with graph clustering localization with *data reduction* strategy,
- $(\hat{\mathbf{B}}, \hat{\mathbf{R}})$: implementation of $\mathrm{DI01}$ with graph clustering localization with *data adjustment* strategy.

The performance of the covariance tuning with localization is evaluated with a simple scalar criteria involving the Frobenius norm, relative to the standard approach. This indicator/gain may be expressed for the background covariance tuning with the reduction strategy as

$$\gamma_{\tilde{\mathbf{B}}} = \left(\Delta_{\mathbf{B}} - \Delta_{\tilde{\mathbf{B}}}\right) / \Delta_{\mathbf{B}}, \tag{4.53}$$

with $\Delta_{\cdot} = \mathbb{E}\Big[\| \cdot - \mathbf{B}_{\mathrm{E}} \|_F\Big]$ representing the expected matrices difference in the Frobenius norm, and similarly for the adjustment strategy and for the observations covariance. The larger the

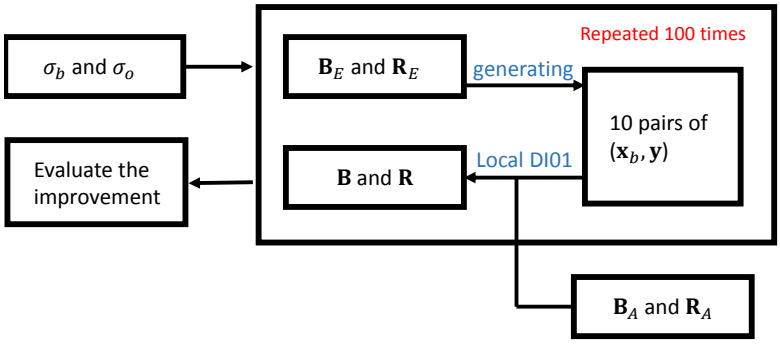


Figure 4.3: Flowchart of Monte-Carlo experiments for adaptive data assimilation with fixed parameters: $\sigma_b, \sigma_o, \mathbf{B}_A, \mathbf{R}_A$.

gain, the more advantage can be expected from the new approach compared to the standard $\mathrm{DI01}$ one.

In the numerical results presented later, empirical expectation of these indicators will be calculated by repeating the tests 100 times, in a Monte Carlo fashion, for each case of standard deviation parameters. In each Monte Carlo simulation, 10 pairs of background state and observation vector are generated to evaluate the coefficients s_b^i and s_o^i necessary for diagnosing and improving $\mathbf{B}$ and $\mathbf{R}$ as shown in Fig. 4.3.

In order to examine the performance of the proposed approach, we choose to always quantify the assumed prior covariances as

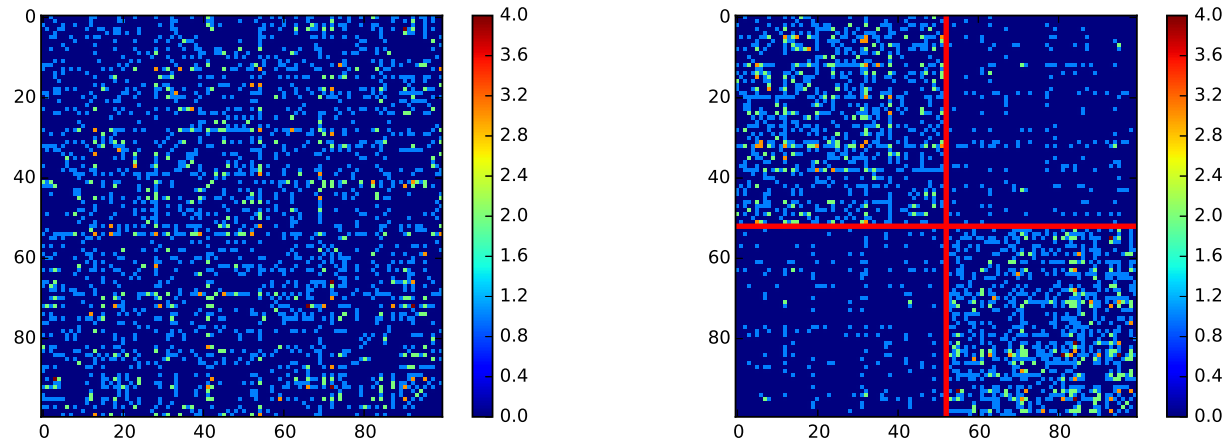
$$\mathbf{B}_A = \sigma_{b,A}^2 \times \mathbf{C}_\mathbf{B},$$
$$\mathbf{R}_A = \sigma_{o,A}^2 \times \mathbf{C}_\mathbf{R},$$

with $\sigma_{b,A} = \sigma_{o,A} = 0.05$.

Meanwhile, the average exact prior error deviation ($\sigma_{b,E}, \sigma_{o,E} = \sqrt{\sigma_{o,1}\sigma_{o,2}}$) varies in the range $([0.025, 0.1])$. In other words, we test a range spanning a domain with over-estimation of 100% of error deviation to an under-estimation of 100%. We remind that the aim of the new approaches is to obtain a more precise estimation of prior covariance structures.

4.5.2 Results

Thanks to the adjacency matrix of the observation-based state network as shown in Fig. 4.4(a), we first apply the Fluid community detection method in the observation-based state network to determine subspaces (communities) in the state space. For real applications, the number of communities is unknown. Here, we apply several times the community detection algorithms with different assumed community number and we evaluate the performance rate ([Fortunato, 2010]) of the obtained partition. It is used as an indicator for finding the optimal community number



(a) (b)

Figure 4.4: (a) Original adjacency matrix of a 100 vertex observation-based state network. (b) Vertex ordering by cluster where the 2-cluster structure is evident thanks to the graph clustering algorithm.

Table 4.1: Quantification of the community detection algorithm results on the observation-based state network followed by the data reduction and data adjustment strategies. The number of communities (i.e. $k = 2$) is set according to the result in Fig. 4.5.

Size of subsets / Strategy chosen	Detected subsets					
	$\lvert\mathbf{x}^1\rvert$	$\lvert\mathbf{x}^2\rvert$	$\lvert\mathbf{x}\rvert$	$\lvert\mathbf{y}^1\rvert$	$\lvert\mathbf{y}^2\rvert$	$\lvert\mathbf{y}\rvert$
Data reduction	52	48	100	12	13	25
Data adjustment	52	48	100	25	25	50

(as shown in Fig. 4.5), which is a standard approach for graph problems.

According to the result presented in Fig. 4.5, we chose to separate the state variables into two subsets, which is the correct number of communities when we simulate the Jacobian matrix **H**. We emphasize that despite the fact that the **H** matrix was generated using two clusters, it was not trivial to rediscover them from the observation-based state network, once the information was shuffled and noisy, cf. from (a) to (b) in Fig. 4.4. The result of graph partition algorithm of two communities is summarized in Table. 4.1 and Fig. 4.4(b). From Table. 4.1, we notice that two state variables from the second subset $\mathbf{x}^2$ are mistakenly assigned to the first one, $\mathbf{x}^1$. The last column of Table. 4.1 shows the total number of observations ($|\mathbf{y}| = |\mathbf{y}^1| + |\mathbf{y}^2|$) used in the covariance tuning. Only half of the observations are considered while applying the strategy of data adjustment.

Fig. 4.6 and Fig. 4.7 collect the results of the Monte-Carlo tests described in 4.5.1 where the ratio of exact error deviation over the assumed one is chosen to vary from 0.5 to 2 for both background and observation errors. The improvement in terms of covariance matrices specification is estimated for the standard DI01 algorithm as well as its localized version with

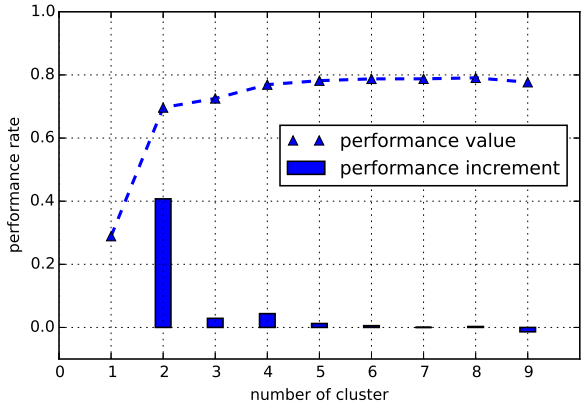


Figure 4.5: The evolution of performance value and its increment against the number of communities chosen.

two strategies for fitting the observation data. We are interested in the potential advantage of the new methods compared to the standard algorithm. The normalized difference of covariance specification error is drawn as mentioned in 4.5.1 where positive values represent an advantage of localized methods. All tuning methods are applied for $q_{\max} = 10$ iterations and we have checked that the sequences $s_{b,q}, s_{o,q}, s^i_{b,q}, s^i_{b,q}$ have been well converged to 1.

Measure of improvement of the localized approaches for the estimation of the background B matrix

From Fig. 4.6, one observes that in this test, the localized $\mathrm{DI01}$ with data reduction always holds a strong advantage (positive value) in terms of matrix **B** estimation, no matter the exact error deviation, compared to the standard approach. The strategy of data adjustment works well for some parameters combinations, but it becomes less optimal when σ_b increases and σ_o decreases. Thus careful attention should be brought on the error level of the background state while applying data adjustment strategy. In fact, when the background error level is high, adjustment of the observation data with background state of large variance will take a considerable risk of polluting the observations both in terms of observation accuracy and the knowledge of error covariances.

Measure of improvement of the localized approaches for the estimation of the observation R matrix

From Fig. 4.7, one observes significant advantages in most cases for both new adaptive approaches. In fact, according to the hypothesis of our experiments, the non-homogeneity of observation errors is completely neglected by a standard $\mathrm{DI01}$. This non-homogeneity could be covered using the graph-based new approach. Similar to the matrix **B**, less optimal results are

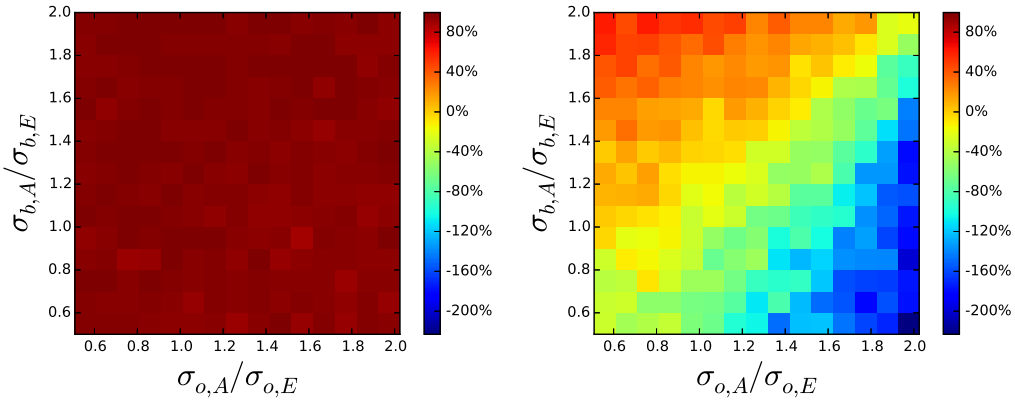


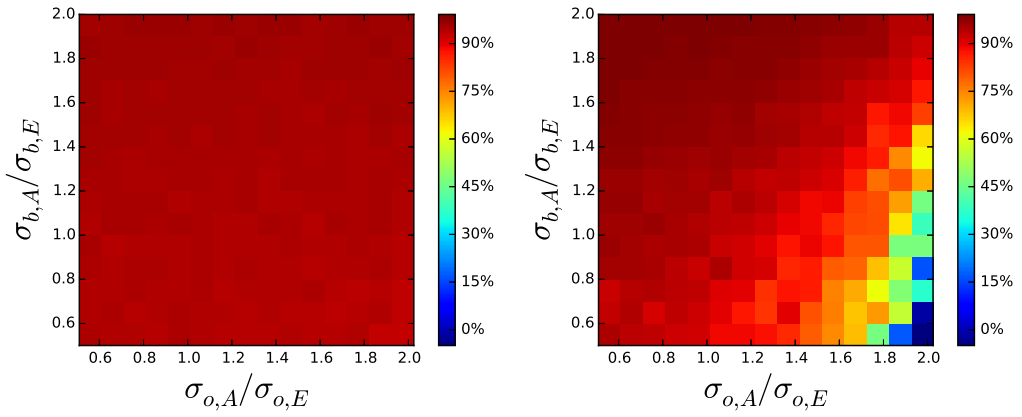
Figure 4.6: Average improvement (in % according to the measures introduced in 4.5.1) of the background error covariance **B** corrected by the proposed localized approach relative to the standard global tuning, (a): with data reduction ($\tilde{\gamma}_{\mathbf{B}}$); (b): with data adjustment ($\hat{\gamma}_{\mathbf{B}}$); A stands for assumed and E for exact values, respectively, with ($\sigma_{b,E}, \sigma_{o,E} = \sqrt{\sigma_{o,1}\sigma_{o,2}}$) both varying in $[0.025, 0.1]$.

found when the background error is considerably higher than the observation one.

In these twin experiments, we may conclude that despite the fact that half of the observations are ignored for the covariance tuning, the strategy of data reduction owns in general an advantage over the one of data adjustment. However, for problems of large dimension, it is possible that most observations are imperfect concerning the correspondence to state communities. Therefore, how to wisely combine these two strategies in real applications for improving the covariance tuning could be a promising topic.

Test case with a larger difference of error deviation across the two observation clusters

Similar experiments are also performed with a more significant difference between the two observation groups in terms of their prior error deviation. The setup of experiments is the same as in section 4.5.1, except that the ratio of $\sigma_{o,E}^{i=1}(\mathbf{y}_u) \Big/ \sigma_{o,E}^{i=2}(\mathbf{y}_v)$ is now set to be 100 instead of 10. The same number of experiments as in the previous case are carried out. The test results are summarized in Table 4.2, according to the cases of under- or over-estimation of prior error amplitude. As expected, due to the larger difference between the two observation groups and thanks to the assumed homogeneous observation matrix $\mathbf{R}_{\mathrm{A}}$, the results of the new approaches are even more impressive over a standard DI01 while keeping the same trends against the variations of $\sigma_{b,\mathrm{A}} \Big/ \sigma_{b,\mathrm{E}}, \sigma_{o,\mathrm{A}} \Big/ \sigma_{o,\mathrm{E}}$ similarly to Fig. 4.6 and Fig. 4.7. On the other hand, while the prior estimation of $\sigma_{b,A}, \sigma_{o,A}$ is of extremely poor quality, for example,$\sigma_{o,\mathrm{A}} \Big/ \sigma_{o,\mathrm{E}} > 100 \quad (\mathrm{or} < 1/100)$, we recommend to consider the standard DI01 in the first place.



(a) (b)

Figure 4.7: Same figure as in Fig. 4.6 for observation error covariance improvement $\tilde{\gamma}_{\mathbf{R}}$ (a) and $\hat{\gamma}_{\mathbf{R}}$ (b).

Table 4.2: Averaged gain improvement of error covariances ($(\mathbf{B}, \mathbf{R})$in $\%$) with $\sigma_{o,E}^{i=1}(\mathbf{y}_u) = 100\sigma_{o,E}^{i=2}(\mathbf{y}_v)$ via two graph clustering localization strategies (observation reduction and observation adjustment). Both $\sigma_{b,E}$ and $\sigma_{o,E}$ vary in $[0.025, 0.1]$.

Improvement of B and R (in $\%$)	$\sigma_{o,A} < \sigma_{o,E}$		$\sigma_{o,A} > \sigma_{o,E}$	
observation reduction	$\overline{\gamma_{\tilde{\mathbf{B}}}}$	$\overline{\gamma_{\tilde{\mathbf{R}}}}$	$\overline{\gamma_{\tilde{\mathbf{B}}}}$	$\overline{\gamma_{\tilde{\mathbf{R}}}}$
$\sigma_{b,A} < \sigma_{b,E}$	98.5%	96.65%	96.41%	96.08%
$\sigma_{b,A} > \sigma_{b,E}$	99.24%	96.08%	97.81%	96.69%
observation adjustment	$\overline{\gamma_{\hat{\mathbf{B}}}}$	$\overline{\gamma_{\hat{\mathbf{R}}}}$	$\overline{\gamma_{\hat{\mathbf{B}}}}$	$\overline{\gamma_{\hat{\mathbf{R}}}}$
$\sigma_{b,A} < \sigma_{b,E}$	34.76%	99.27%	-4.03%	96.94%
$\sigma_{b,A} > \sigma_{b,E}$	68.22%	99.64%	30.83%	98.81%

4.6 Discussion

Localization technique is an important numerical tool which contributes to the success of solving high-dimensional data assimilation problems for which ensemble estimates are unreliable. It is based on the assumption that correlations between dynamical system variables eventually decay with the physical distance. This simple rationale is put to use either to make the assimilation of observations more local (domain localization) or to numerically impose a tapering of distant spurious correlations (covariance localization) and leads to very different implementations and numerical difficulties. Domain localization is interesting because it makes the problem more scalable and the implementation more flexible in the sense that the original global formulation can be broken-up into several smaller subproblems. Nevertheless, the assimilation of *non-local* observations and/or observations from different sources and at different scales (e.g. satellite observations) becomes increasingly challenging.

If we consider as a motivating example the application of data assimilation to hydrological modeling ([Cheng et al., 2020a]), we known that hydrological changes induced by precipitations (in the form of rain or snow) in the various watersheds of a region affect hydraulic conditions and the accompanying flood levels, sediment transport rates, and habitat conditions within various distributed streams. While any particular location along a stream channel may depend on several close or distant watersheds, there are sometimes geographically closed watersheds that do not contribute to the same stream due the position of the ridgeline separating their neighboring drainage basins. Therefore, assimilation of discrete streamflow measurements in order to correct water levels in drainage basin and reservoirs remains challenging due to the complex network structure of the hydrological domain ([Castronova and Goodall, 2014]). For this particular application, a domain localization technique solely based on spatial distance seems therefore to be a poor approach. Similar arguments may apply to modeling of subsurface flow and transport properties involved in groundwater flow and contaminant transport, energy recovery from geothermal and hydrocarbon reservoirs, and geologic storage of CO2 in deep underground formations.

In this work, we propose to generalize the concept of domain localization relying on graph clustering state decomposition techniques. The idea is to automatically detect and segregate the state and observation variables in an optimal number of clusters, more amenable to scalable data assimilation, and use this decomposition to perform efficient adaptive error covariances tuning. Compared to classical domain localization, the novel method is more effective when long-distance observations and error correlation exist, either in $\mathbf{B}$ or $\mathbf{R}$.This unsupervised localization technique based on a *linearized* state-observation measure is general and does not rely on any prior information such as relevant spatial scales, empirical cut-off radius or homogeneity assumptions. In this paper, the Fluid method is chosen for applications because of its computational simplicity, especially for sparse graphs. In terms of covariance diagnosis, the $\mathrm{DI01}$ is chosen because the ratio of available data to problem size is often limited for geosciences applications. Furthermore, the correction of $\mathrm{DI01}$ in subspaces allows a more flexible tuning on error covariances without deteriorating prior knowledge of error correlation. Finally, we have shown that our approach reduces the computational complexity and provides some speedup. It is best suited for problems of intermediate size such as the ones involving transformed data set, mentioned hereinbefore.

In this paper, our methodology is applied to a simple twin experiments data assimilation problem for which the Jacobian matrix of the observation operator is chosen to reflect a dual clustering of the state-observation mapping; the components of which are then randomly mixed in order to hide any particular structural pattern. Simply speaking, there exist two *hidden* communities of state variables, each of them preferably connected to their own observations community. The problem is far from trivial as — the segregation resulting in clustering is not related to let us say spatial separations, — exact background error magnitude is supposed to be homogeneous in our tests but the clusters have different exact observation errors and also because — there exists some inter-connectivity between the clusters. Considering the latter, two simple numerical approaches are proposed in order to handle a data *reduction* or a data *adjustment* strategy. The problem is investigated for a wide range of assumed prior covariances and the graph clustering approach with adaptive covariance tuning is much more efficient than a global adaptive covariance tuning approach, especially in the case of DI01 tuning where $\mathbf{B}$ and $\mathbf{R}$ are jointly corrected.

The graph clustering algorithm uses an adjacency matrix derived from a linearization of the observation operator. Therefore, it seems reasonable to anticipate that the approach will be more appropriate for linear or weakly nonlinear problems. For time-dependent strongly nonlinear problems, one may need to rely on the community detection algorithm multiple times, which could be computationally expensive. In this paper, we choose to operate the graph-based localization on state variables $\mathbf{x}_i$ while this approach could also be applied for $\mathbf{y}_i$. The choice might be determined according to the priority between $\mathbf{B}$ and $\mathbf{R}$ localization, as discussed in [Greybush et al., 2011].

Another critical point relates to the inter-cluster connectivity which materialize the fact that real applications problems will never be fully separable. Here, we have made the choice to circumvent the difficulty by disposing of the troublesome shared observations. Nevertheless, this approach might be impractical for applications with a large number of clusters and overlaps. In this case, alternate strategies will have to be considered, much likely involving a search for optimal ordering of the subspaces covariance tuning.

Finally, our localization approach will perform better if the assimilation problem, represented by a graph, is well separable under our cluster analysis; i.e. in the sense that the data assimilation problem is decomposed into a certain number of subsets problems minimizing the overlap between subsets. This will somewhat depend on the graph cluster analysis algorithm retained but more predominantly on the chosen measure of similarity. For the former, it will be useful to monitor some performance metrics as a function of the number of clusters for a given graph clustering algorithm. In this work, we base the measure of similarity on the linearized observation operator. Complementary to this approach, it may be interesting to combine a measure involving prior knowledge of error covariances with the state-observation mapping, i.e. $|\mathbf{H}|\,\mathbf{B}\,|\mathbf{H}|^T$ instead of $|\mathbf{H}|\,|\mathbf{H}|^T$. This might provide a way to scalable optimization of covariance structures between observations and model variables instead of covariance structures in the prior alone. After this

methodological contribution, future work will consider applying these methods to more challenging real industrial applications.

Appendix

This Appendix is not included in the submitted paper. The proof is synthesised based on the work of [Ménard, 2016], [Chapnik et al., 2004] and [Desroziers and Ivanov, 2001] with some simplifications.

Convergence of $\mathrm{DI01}$ when $\mathbf{HBH}^T$ and $\mathbf{R}$ share same correlation structure

It is found that while the correlation structure of $\mathbf{HBH}^T$ and $\mathbf{R}$ are not sufficiently different, the fixed point of $\mathrm{DI01}$ may not lead to the exact error covariances. The reasoning is given in [Ménard, 2016]. In order to distinguish the exact covariances matrices (represent the true estimation error covariances) and the estimated ones, the former is denoted as $\mathbf{B}_\mathrm{E}, \mathbf{R}_\mathrm{E}$ while $\mathbf{B}_\mathrm{A}, \mathbf{R}_\mathrm{A}$ are used for the assumed covariances. In addition, the error covariance of prior innovation $d = y - \mathcal{H}(x_b)$ is denoted by $\mathbf{D}$. We remind that the analysis is carried out under the assumption of a linear $\mathbf{H}$, where the explicit expression of $\mathbf{D}$ can be found, i.e.

$$\mathbf{D} = \mathbf{HBH}^T + \mathbf{R}. \tag{4.54}$$

Similar to $\mathbf{B}_\mathrm{A}, \mathbf{R}_\mathrm{A}$ the assumed innovation covariance matrix $\mathbf{D}_\mathrm{A}$ is defined as

$$\mathbf{D}_\mathrm{A} = \mathbf{HB}_\mathrm{A}\mathbf{H}^T + \mathbf{R}_\mathrm{A}. \tag{4.55}$$

Following the assumption of [Desroziers and Ivanov, 2001] which says the correlation patterns of these matrices are perfectly identified, for any current iteration q,

$$\mathbf{B}_{\mathrm{A},q} = \beta_q \mathbf{B}_\mathrm{E} \tag{4.56}$$
$$\mathbf{R}_{\mathrm{A},q} = \alpha_q \mathbf{R}_\mathrm{E}. \tag{4.57}$$

where β_q, α_q are real numbers. Hence, the DI01 covariance tuning is equivalent to a tuning of two scalar sequences,

$$\beta_{q+1} = s_q^b \beta_q \tag{4.58}$$
$$\alpha_{q+1} = s_q^o \alpha_q. \tag{4.59}$$

According to [Ménard, 2016], in the ideal case, the multiplicative coefficient updating could be

expressed as:

$$\beta_{q+1} = \beta_q \frac{\mathrm{Tr}\{\mathbf{D}_{\mathrm{A},q}^{-1}\mathbf{D}\mathbf{D}_{\mathrm{A},q}^{-1}\mathbf{H}\mathbf{B}_{\mathrm{A},q}\mathbf{H}^T\}}{\mathrm{Tr}\{\mathbf{D}_{\mathrm{A},q}^{-1}\mathbf{H}\mathbf{B}_{\mathrm{A},q}\mathbf{H}^T\}}. \tag{4.60}$$

$$\alpha_{q+1} = \alpha_q \frac{\mathrm{Tr}\{\mathbf{D}_{\mathrm{A},q}^{-1}\mathbf{D}\mathbf{D}_{\mathrm{A},q}^{-1}\mathbf{R}_{\mathrm{A},q}\}}{\mathrm{Tr}\{\mathbf{D}_{\mathrm{A},q}^{-1}\mathbf{R}_{\mathrm{A},i}\}}. \tag{4.61}$$

It is obvious that while $\mathbf{D}_{\mathrm{A},q} = \mathbf{D}$ (i.e. $\mathbf{H}\mathbf{B}_{\mathrm{A}}\mathbf{H}^T + \mathbf{R}_{\mathrm{A}} = \mathbf{H}\mathbf{B}_{\mathrm{E}}\mathbf{H}^T + \mathbf{R}_{\mathrm{E}}$) both the iterations of α_q and β_q converge numerically. If $\mathbf{H}\mathbf{B}_{\mathrm{A}}\mathbf{H}^T$ and $\mathbf{R}_{\mathrm{A}}$ are of different structures, i.e.

$$\nexists \tau \in \mathbb{R}, \quad \text{such that} \quad \mathbf{H}\mathbf{B}_{\mathrm{A}}\mathbf{H}^T = \tau\mathbf{R}_{\mathrm{A}}, \tag{4.62}$$

then the convergence of (β_q, α_q) is equivalent to $\mathbf{B}_{\mathrm{A}} = \mathbf{B}_{\mathrm{E}}$ and $\mathbf{R}_{\mathrm{A}} = \mathbf{B}_{\mathrm{E}}$. On the other hand when $\mathbf{H}\mathbf{B}_{\mathrm{A}}\mathbf{H}^T$ and $\mathbf{R}_{\mathrm{A}}$ share same correlation structure, fixed-points other than true covariance matrices could be found for DI01.

Maximum likelihood property of DI01

As stated by [Chapnik et al., 2004], $\mathrm{DI01}$ is equivalent to tune a maximum likelihood algorithm to the innovation covariances matrix $\mathbf{D}(s^b, s^o)$ parameterized by the multiplicative coefficients. At each updating step,

$$\mathbf{D}(s^b, s^o) = \mathbf{H}s^b\mathbf{B}\mathbf{H}^T + s^o\mathbf{R}. \tag{4.63}$$

In fact, suppose $\mathbf{s}$ (in our case s^b, s^o) is a parameter vector such that $\mathbf{D} = \mathbf{D}(s)$, the conditional probability density function of the innovation quantity $\mathbf{d}$ could be written as:

$$f(\mathbf{d}|\mathbf{s}) = \frac{1}{\sqrt{(2\pi)^p \det(\mathbf{D}(\mathbf{s}))}} exp\Big(-\frac{1}{2}\mathbf{d}^T\mathbf{D}(\mathbf{s})^{-1}\mathbf{d}\Big) \tag{4.64}$$

where p stands for the dimension of $\mathbf{d}$. We then deduce the log-likelihood function,

$$\mathcal{L}(\mathbf{s}) = -\log(f(\mathbf{d}|\mathbf{s})) = \frac{p}{2}\log(2\pi) + \frac{1}{2}\log[\det(\mathbf{D}(\mathbf{s}))] + \frac{1}{2}\mathbf{d}^T D(\mathbf{s})^{-1}\mathbf{d}. \tag{4.65}$$

The minimum of this function should satisfy

$$\text{For all component } s \text{ of } \mathbf{s}, \frac{\partial \log[\det(\mathbf{D}(\mathbf{s}))]}{\partial s} + \mathbf{d}^T \frac{\mathbf{D}(\mathbf{s})^{-1}}{\partial s}\mathbf{d} = 0. \tag{4.66}$$

Using following algebraic properties,

$$\log[\det(\mathbf{D}(\mathbf{s}))] = \mathrm{Tr}[\log\mathbf{D}(\mathbf{s})], \tag{4.67}$$

$$\frac{\partial \mathrm{Tr}[\log(\mathbf{D}(\mathbf{s}))]}{\partial s} = \mathrm{Tr}\Big[\mathbf{D}(\mathbf{s})^{-1}\frac{\partial \mathbf{D}(\mathbf{s})}{\partial s}\Big], \tag{4.68}$$

$$\frac{\mathbf{D}(\mathbf{s})^{-1}}{\partial s} = \mathbf{D}(\mathbf{s})^{-1}\frac{\mathbf{D}(\mathbf{s})}{\partial s}\mathbf{D}(\mathbf{s})^{-1} \tag{4.69}$$

Eq.4.66 could be simplified. Concerning the first term on the left side,

$$\frac{\partial \log[\det(\mathbf{D}(\mathbf{s}))]}{\partial s} = \frac{\mathrm{Tr}[\partial \log\mathbf{D}(\mathbf{s})]}{\partial s} = \mathrm{Tr}\Big[\mathbf{D}(\mathbf{s})^{-1}\frac{\partial \mathbf{D}(\mathbf{s})}{\partial s}\Big] = \mathrm{Tr}\Big[\frac{\partial \mathbf{D}(\mathbf{s})}{\partial s}\mathbf{D}(\mathbf{s})^{-1}\Big]. \tag{4.70}$$

In our case, $\mathbf{s} = \{s^b, s^o\}$ therefore

$$\frac{\partial \mathbf{D}(\mathbf{s})}{\partial s^b} = \mathbf{H}\mathbf{B}\mathbf{H}^T \tag{4.71}$$

$$\frac{\partial \mathbf{D}(\mathbf{s})}{\partial s^o} = \mathbf{R}. \tag{4.72}$$

Eq.4.66 is then equivalent to the two following marginal formulas,

$$\mathrm{Tr}[\mathbf{H}\mathbf{B}\mathbf{H}^T\mathbf{D}^{-1}] - \mathbf{d}^T\mathbf{D}^{-1}\mathbf{H}\mathbf{B}\mathbf{H}^T\mathbf{D}^{-1}\mathbf{d} = 0 \tag{4.73}$$

$$\mathrm{Tr}[\mathbf{R}\mathbf{D}^{-1}] - \mathbf{d}^T\mathbf{D}^{-1}\mathbf{R}\mathbf{D}^{-1}\mathbf{d} = 0. \tag{4.74}$$

On the other hand, from Eq.4.32, one can deduce

$$2J_b(x_a) = s^b \mathrm{Tr}[\mathbf{KH}] \tag{4.75}$$
$$||\mathbf{K}(\mathbf{y} - \mathbf{H}x_b)||_{\mathbf{B}^{-1}} = s^b \mathrm{Tr}[\mathbf{BH}^T\mathbf{D}^{-1}\mathbf{H}] \tag{4.76}$$
$$(\mathbf{Kd})^T(s^b\mathbf{B})^{-1}\mathbf{Kd} = s^b \mathrm{Tr}[\mathbf{HBH}^T\mathbf{D}^{-1}] \tag{4.77}$$
$$\mathbf{d}^T\mathbf{D}^{-1}\mathbf{H}(s^b\mathbf{B})^T(s^b\mathbf{B})^{-1}(s^b\mathbf{B})\mathbf{H}^T\mathbf{D}^{-1}\mathbf{d} = s^b \mathrm{Tr}[\mathbf{HBH}^T\mathbf{D}^{-1}] \tag{4.78}$$
$$\mathbf{d}^T\mathbf{D}^{-1}\mathbf{HB}^T\mathbf{B}^{-1}\mathbf{BH}^T\mathbf{D}^{-1}\mathbf{d} = \mathrm{Tr}[\mathbf{HBH}^T\mathbf{D}^{-1}]. \tag{4.79}$$

Meanwhile,

$$2J_o(x_a) = s^o \mathrm{Tr}[\mathbf{I} - \mathbf{KH}] \tag{4.80}$$
$$((\mathbf{I} - \mathbf{HK})\mathbf{d})^T\mathbf{R}^{-1}((\mathbf{I} - \mathbf{HK})\mathbf{d}) = s^o \mathrm{Tr}[\mathbf{I} - \mathbf{BH}^T(\mathbf{HBH}^T + \mathbf{R})^{-1}\mathbf{H}] \tag{4.81}$$
$$\mathbf{d}^T(s^o\mathbf{RD}^{-1})^T(s^o\mathbf{R})^{-1}(s^o\mathbf{RD}^{-1})\mathbf{d} = s^o\Big(p - \mathrm{Tr}[\mathbf{HBH}^T(\mathbf{HBH}^T + \mathbf{R})^{-1}]\Big) \tag{4.82}$$
$$s^o\mathbf{d}^T\mathbf{D}^{-1}\mathbf{RD}^{-1}\mathbf{d} = s^o \mathrm{Tr}[\mathbf{R}(\mathbf{HBH}^T + \mathbf{R})^{-1}] \tag{4.83}$$
$$\mathbf{d}^T\mathbf{D}^{-1}\mathbf{RD}^{-1}\mathbf{d} = \mathrm{Tr}[\mathbf{RD}^{-1}]. \tag{4.84}$$

Therefore, the $\mathrm{DI01}$ method is equivalent to a Maximum likelihood estimation of $\mathbf{B}$ and $\mathbf{R}$ parameterized by s^b, s^o.

Chapter 5

Error covariance tuning in variational data assimilation: application to an operating hydrological model

Accepted for publication in *Stochastic Environmental Research and Risk Assessment* as
Cheng, S., Argaud, J.-P., Iooss, B., Lucor, D. and Ponçot, A. (2020). Error covariance tuning in variational data assimilation: application to an operating hydrological model

Abstract

Because the true state of complex physical systems is out of reach for real-world data assimilation problems, error covariances are uncertain and their specification remains very challenging. These error covariances are crucial ingredients for the proper use of data assimilation methods and for an effective quantification of the *a posteriori* errors of the state estimation. Therefore, the estimation of these covariances often involves at first a chosen specification of the matrices, followed by an adaptive tuning to correct their initial structure. In this paper, we propose a flexible combination of existing covariance tuning algorithms, including both online and offline procedures. These algorithms are applied in a specific order such that the required assumption of current tuning algorithms are fulfilled, at least partially, by the application of the ones at the previous steps. We use our procedure to tackle the problem of a multivariate and spatially-distributed hydrological model based on a precipitation-flow simulator with real industrial data. The efficiency of different algorithmic schemes is compared using real data with both quantitative and qualitative analysis. Numerical results show that these proposed algorithmic schemes improve significantly short-range flow forecast. Among the several tuning methods tested, recently developed CUTE and PUB algorithms are in the lead both in terms of history matching and forecast.

5.1 Introduction

In order to improve the estimation of state variables, especially in dynamical systems, data assimilation (DA) techniques, originally developed for Numerical Weather Prediction (NWP) [Parrish and Derber, 1992] and geosciences [Carrassi et al., 2018], have been widely applied in industrial problems, including hydrology [Houser et al., 2012], nuclear engineering [Gong et al., 2020b], biomedical applications [Rochoux et al., 2018], etc. The objectives of DA methods could be mainly divided into two groups: - field reconstruction and - parameter identification. The former aims at improving the estimation/forecast of a physical field of interest (e.g. temperature, velocity, usually multidimensional), while the latter consists in estimating a set of optimal parameters in order to provide a parameterized simulation of the physical field. The goal of DA algorithms could be simply summarized as finding an "optimal" compromise via noisy information embedded in different sources of simulation/observation. The weight of each information source is defined by associated prior error covariance matrices. As DA problems are often of large state dimension (e.g. $10^6 \sim 10^{10}$ in NWP or geosciences), prior errors are usually assumed to be Gaussian in order to simplify the probability distribution of uncertainties. In fact, the Gaussian property ensures that both prior and posterior errors could be fully characterized by the first (mean) and second (covariance) moment. In this paper, the terms "prior" and "posterior" are defined relatively to the assimilation process. As an example, "posterior" data covariance refers to the covariance of assimilated data.

The estimation of error covariances thus plays an essential role in both variational ([Fisher, 2003]) and sequential ([Sénégas et al., 2001]) DA algorithms. It weighs the confidence of different information sources but also describes how prior errors are spatially (or temporally) correlated. The former aspect is mainly decided by the amplitude of covariances matrices while the latter depends on the extra-diagonal elements. The specification of these covariance matrices impacts significantly the accuracy of DA algorithms ([Tandeo et al., 2018]). Major obstacles in estimating covariances for real applications are mainly two folds: - the large size of simulation/observation vectors, and, - the fact that these measures might be evaluated non-simultaneously in a dynamical system. Both reasons make the empirical estimation of covariances extremely difficult, if not infeasible. A common solution in data assimilation is to take *a priori* defined covariance matrices, often with a diagonal structure (e.g. [Argaud et al., 2016]) or a correlation kernel of Matern-type (e.g. [Singh et al., 2011], [Gong et al., 2020b]). Other modeling via diffusion equation ([Mirouze and Weaver, 2010]) or convolution operators ([Gaspari and Cohn, 1999]) have also been proposed for a more efficient covariance computation. However, all these mentioned methods often rely on few parameterized structures, and are thus less flexible to fit the true error covariances in a generic procedure of uncertainty quantification through these matrices.

Continuous effort has been devoted to improve the covariance tuning in general. For example, several iterative methods based on posterior diagnosis have been developed. These methods focus respectively on correcting the amplitude of different sources ([Desroziers and Ivanov, 2001]), the evolved background matrix (associated to a prior state simulation/estimation) structure ([Cheng et al., 2019]), or the observation covariances from non-simultaneous data [Desroziers et al., 2005].

Each of these algorithms has its own assumptions about the knowledge of prior errors, such as correlation structure [Desroziers and Ivanov, 2001], observation covariances [Cheng et al., 2019] or covariance flow-independence [Desroziers et al., 2005]. In this work, numerically we demonstrate that a combination of these methods, rarely applied in industrial problems other than NWP or geosciences, could open a potential general solution to industrial problems, typically the ones with non-simultaneous historical data.

In this paper, we focus on some industrial hydrological applications tackled with the MORDOR-TS software, developed by EDF (Électricité de France, French electric utility company) [Garçon, 1996]. This numerical tool relies on observed precipitation and temperatures as input and predicts simulation of river flows as outputs. The precipitation-flow simulation is carried out via conceptual reservoir systems, which ensures its high computational efficiency. Its high accuracy of flow forecast has been proved numerically by several studies in different hydraulic areas in France (e.g. the Alps [Garavaglia et al., 2017], the Loire valley [Rouhier, 2018]). In this work the study area is the south of France around the Tarn river. The Tarn river, being known for its extreme variability of water-level values and high sensitivity to precipitations (see chap 2 of [Lerat, 2009]), is an ideal benchmark for comparing different DA strategies. The assimilation scheme setup consists in correcting both the reservoir levels at the beginning of the assimilation window and the daily precipitations over the window by assimilating daily flow measurements. For DA solving, we use the ADAO tool [Argaud, 2019], developed by EDF R&D and integrated into the SALOME open-source study platform [CEA/DEN et al., 2020]. The dimension of the problem is considered somewhat intermediate in DA, with the state/observation vector size usually ranging from 30 to 1000, depending on the length of the chosen assimilation window and the chosen variables to estimate.

DA algorithms have already been adopted in hydraulic/hydrological problems for improving history matching or forecast accuracy with applications from flow modeling to soil moisture content (see [Houser et al., 2012]). Being often multivariate and multidimensional, covariance computation is far from trivial in hydrological DA problems. The main challenge in this study is to balance the weight of different information sources (i.e. daily precipitation, initial reservoir levels and observed river flows), which requires a careful computation of prior error covariances. At that time, it is important to point out that the French climate is temperate. The rainfall is spread across the year even if drought can occur in summer. Actually the rainfall distribution can be considered as a random variable and the seasonality of the errors does not exist. Moreover, knowledge of rainfall over an area requires a numerical interpolation model, necessarily imperfect, that reconstructs a rainfall field over the entire domain from a limited number of local rainfall measurement stations. Consequently, the rainfall field given as an input to MORDOR-TS contains unknown errors, that we try to quantify and model through the DA background a priori covariance matrix. In fact, this covariance matrix is itself uncertain and its modelling relies on uncertainty quantification that we choose to build on experimental data, in addition to mathematical modeling and algorithmic adaptation. For posterior covariance diagnosis, a rich historical data of over 20 years (mainly between 1990 and 2010) is available. In order to make

full use of these non simultaneous data, we make the classical assumption of flow-independence of covariance matrices (i.e. being time-invariant per window). A block-diagonal structure is assigned to the background matrix, which is a common practice for multi-variable DA problems. We first apply posterior diagnosis to adjust the error amplitudes of these two diagonal blocks, relative to the observation error amplitude. Both offline (using average-adjusted ratio obtained from historical data) and online (real-time adjusting) approaches are proposed and compared in terms of their improvement on flow forecast/reanalysis. Once the error amplitude is adjusted, we move on to more refined tuning algorithms ([Desroziers et al., 2005, Cheng et al., 2019]) in order to improve the specification of error correlation based on an initial guess of Matern-type or diagonal structure. Numerical results show that significant improvement (over 30% compared to model forecast) of short-range flow forecast is achieved with this covariance tuning.

The paper is organized as follows: the industrial background of the hydrological model and the study area are described in section 5.2. We then recall some standard formulation of variational assimilation, following with the posterior covariance diagnosis in section 5.3. In section 5.4, we describe the DA modeling of the hydrological application, including parametrization, constraints and objective. We analyze detailed results of covariance tuning and flow forecast in sections 5.5 and 5.6, respectively, and we end the paper with a discussion. In the appendix, following the work of [Ménard, 2016] and [Bathmann, 2018], the convergence of Desroziers iterative method is also discussed, with some elements concerning regularization.

5.2 Industrial background

EDF has a particular interest in the study of hydrology, for hydroelectric production or cooling of pressurized water reactors. Moreover, water resource must be managed as a shared resource for agriculture and other water-based human activities. Therefore, the forecast of floods and droughts is crucial for its general management. In this paper, we want to improve river discharge forecast and reanalysis, by correcting historical precipitation and initial reservoir level.

5.2.1 MORDOR: an operating hydrological model

Developed since the early 1990's by EDF and commonly applied in operational applications, MORDOR is a widely applied hydrological model for water management studies, such as scenario impact investigations [Garçon, 1996]. Continuous effort is done to improve MORDOR in order to enhance operational hydrology support, leading to different versions, based in particular on different runoff spatial representations [Garavaglia et al., 2017]. In this work, we concentrate on the last version, named MORDOR-TS, which supports a hydrological mesh-modeling representation of the catchment area.

MORDOR-TS aims at calculating the river flow in a catchment according to weather conditions (precipitations and air temperatures) called forcing. MORDOR-TS is a conceptual hydrological model, relying on the water state of the catchment by reservoirs or stores which feed each other with the help of balance equations. In order to take into account the physical characteristics of the simulated watershed, MORDOR-TS contains several parameters which are calibrated using real flow data. Calibration of MORDOR-TS physical parameters is operated in a separate step and is not the purpose of this paper. For detailed informations about MORDOR-TS calibration, see [Rouhier, 2018].

To get the river flow at a catchment point, water reservoirs have to be transformed in a production term (or runoff). This is done with the help of internal variables among with the reservoirs that define the MORDOR-TS dynamical state.

We list the five types of MORDOR-TS internal storage below:

- Snow storage **S**, which contains both solid and liquid precipitation. The threshold of liquid proportion is fixed at 10% where surplus liquid drops into other reservoirs (i.e. the following **U, L, Z**);
- Surface water storage **U** (with maximum capacity $U_{\max}$) which represents the capacity of water absorption of soil surface. A proportional rate of evaporation is also considered;
- Intermediate water storage **L** (with maximum capacity $L_{\max}$) which accepts water from the storage **U** and plays the role of percolation to deep storage **N**;
- Evaporating storage **Z** (with maximum capacity $Z_{\max}$) which accepts a part of indirect water runoff and contributes to evaporation;
- Deep storage **N** which contributes to baseflow with no limitation of storage capacity.

MORDOR-TS relies on these storage modeling to support the simulation of exchanges among different hydrological components. This modeling is computationally very efficient. For example, a spatially distributed flow simulation of several years may take only a few CPU seconds using MORDOR-TS. The transition among different storage types and the connection to mesh production is detailed in Fig. 1 of [Rouhier et al., 2017].

5.2.2 Study area

The study area is set on the Tarn river catchment, in the south of France. We use operational daily stream flow measurements at 9 gauges of the Tarn catchment. The historical discharge data of over twenty years is provided by EDF and the French water management agencies. Forcing data (rainfall and temperature) are processed in order to get allocated to the 28 mesh cells used by MORDOR-TS as shown in Fig. 5.1. This figure also shows the 9 stream flow gauges, positioned at 9 mesh outlets, which are located for some of them on the Tarn river, but for others on its affluents (Dourbie, Breze, Jonte, Mimente, Tarnon rivers). Due to its position, the Tarn river

outlet at Millau (named here *Tarn at Millau*) is of particular interest in the hydrological study. Located downstream, *Tarn at Millau* receives the flow from other streams, usually with a time delay. For this reason, if a flood takes place somewhere in the studied domain, it is most likely experienced by the *Tarn at Millau* station.

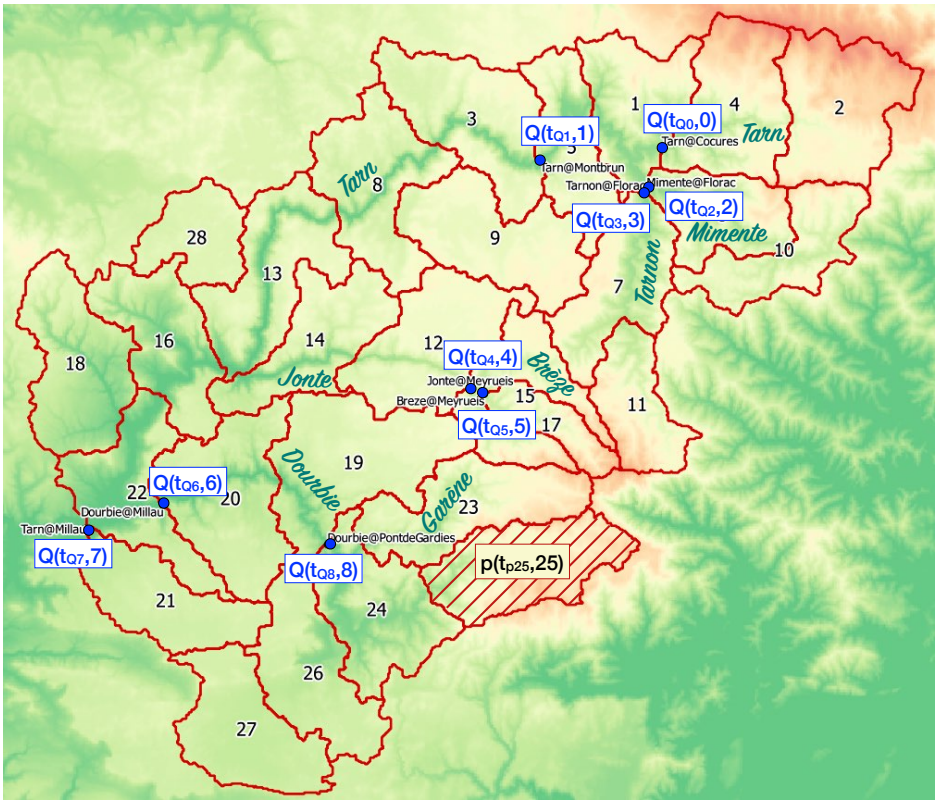


Figure 5.1: Spatial discretization of the Tarn basin with its 28 mesh cells for forcing data (rainfall and temperature) and the locations of the 9 observation gauges (blue dots). The area in darker green represents the elevation of the Tarn river and its affluents.

As an example, we show in Fig. 5.2 the simulated and daily observed Tarn river discharges at Millau, for 3 months in 1990, chosen to illustrate some typical regional hydrological events. The simulation is carried out using observed precipitation and temperature as forcing over the 28 spatial hydrological mesh cells. The averaged precipitation is also included in the figure. We observe that abrupt rainfall over the basin induces floods at *Tarn at Millau* 2 to 3 days later. This delay differs from different gauges, according to their geographical positions. As a preliminary study, we compute the lag correlation between averaged daily precipitation and flow observations at different gauges using data from 1990 to 2010, as shown in Fig. 5.3. The lag correlation is broadly used in the analysis of time series in geosciences (see 3.2 of [Oliver and Webster, 2015]) where the lag refers to the offset length. For example, when lag $= -9$ (first row of Fig. 5.3), we compute the correlation between precipitation and flow observations where the latter is shifted 9 days ahead. As we observe from Fig. 5.3, significant positive correlation exists

between precipitation and observed flow for lag $= 0$ (last column) to -3 days. This range could be extended for *Tarn at Millau*, *Tarn at Montbrun* and *Dourbie at Millau* since they are located downstream. In other words, if the actual precipitation is better characterized, we could expect an improvement of short-range flow forecast (up to 3 to 4 days).

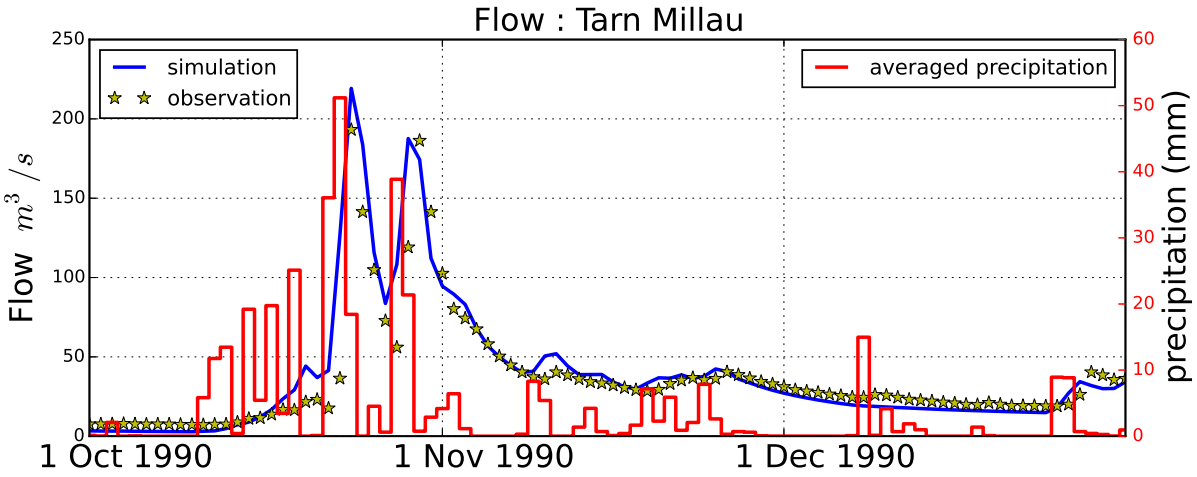


Figure 5.2: Example of simulation predicted by MORDOR-TS using daily precipitation, and observed Tarn discharges at Millau for three months in 1990. Simultaneous observed precipitations are in red bars (with the scale on the right vertical axis).

5.3 Data assimilation and covariance tuning

5.3.1 Variational data assimilation

Data assimilation aims to improve the state estimation $\mathbf{x}$ of a static or dynamical system, thanks to a prior simulation/estimation $\mathbf{x}_b$ and to an observation vector $\mathbf{y}$ (see [Bouttier and Courtier, 2002] for more details). The true value of state variable, usually unknown, is denoted by a vector $\mathbf{x}_t$, also known as the true state. Data assimilation algorithms can be seen as techniques to find an optimal weighted compromise between the prior estimation $\mathbf{x}_b$ and the observation $\mathbf{y}$, minimizing the loss function J defined for a state $\mathbf{x}$ as:

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) + \frac{1}{2}(\mathbf{y}-\mathcal{H}(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{y}-\mathcal{H}(\mathbf{x})) \tag{5.1}$$

$$= \frac{1}{2}||\mathbf{x}-\mathbf{x}_b||^2_{\mathbf{B}^{-1}} + \frac{1}{2}||\mathbf{y}-\mathcal{H}(\mathbf{x})||^2_{\mathbf{R}^{-1}} \tag{5.2}$$

$$= J_b(\mathbf{x}) + J_o(\mathbf{x}), \tag{5.3}$$

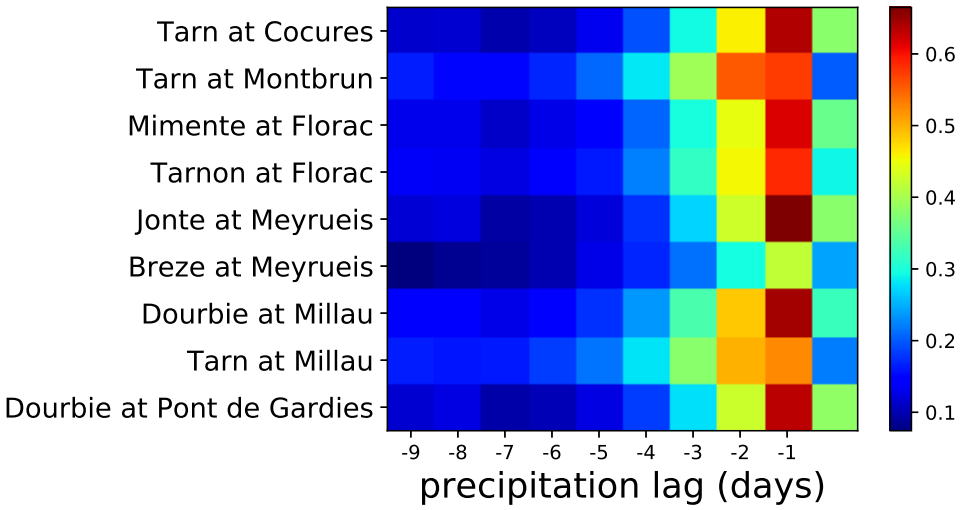


Figure 5.3: The lag correlation between daily precipitation and observed river flow at 9 observation gauges.

where $\mathcal{H}$ denotes the transformation operator from the state space to the one of observations and $J_b(.), J_o(.)$ stand for background/observation cost functions. Matrices $\mathbf{B}$ and $\mathbf{R}$ in the loss function are the associated error covariance matrices defined as:

$$\mathbf{B} = \mathbf{Cov}(\epsilon_b, \epsilon_b) \tag{5.4}$$
$$\mathbf{R} = \mathbf{Cov}(\epsilon_y, \epsilon_y), \tag{5.5}$$

where

$$\epsilon_b = \mathbf{x}_b - \mathbf{x}_t \tag{5.6}$$
$$\epsilon_y = \mathcal{H}(\mathbf{x}_t) - \mathbf{y}, \tag{5.7}$$

represent the background/observation error, respectively. These errors are supposed to be zero-mean Gaussian, being perfectly characterized by their covariance matrices i.e.

$$\epsilon_b \sim \mathcal{N}(0, \mathbf{B}) \tag{5.8}$$
$$\epsilon_y \sim \mathcal{N}(0, \mathbf{R}). \tag{5.9}$$

The inverse matrices of the covariances $\mathbf{B}$ and $\mathbf{R}$ represent the weights of these information sources in the objective function.

The optimization problem of minimizing the functional form given in Eq.5.1, is the so called 3D-Var formulation, and is a generic representation of variational data assimilation. The optimal

output is named analysis and denoted as $\mathbf{x}_a$, i.e.

$$\mathbf{x}_a = \underset{\mathbf{x}}{\operatorname{argmin}}\Big(J(\mathbf{x}) \Big). \tag{5.10}$$

The two covariance matrices **B** and **R** play essential roles in data assimilation. They are difficult to know precisely, both because of their statistical nature in relation to an unknown state $\mathbf{x}_t$ and because of their large size.

The non-linearity of the $\mathcal{H}$ transformation operator increases significantly the computational difficulty of the optimization problem, as both the analyzed state and its associated covariance could not be estimated via linear algebraic formulations. In the present work, the optimization of the loss function of Eq.5.1 is solved using the classical and efficient iterative algorithm called "L-BFGS-B" algorithm [Byrd et al., 1995, Zhu et al., 1997], using the ADAO solver [Argaud, 2019] of the SALOME platform [CEA/DEN et al., 2020]. This methodology has already been applied in various industrial problems, including nuclear and hydrological problems (see for example [Goeury et al., 2017]).

5.3.2 Covariances matrices diagnosis/tuning

The specification of error covariances (both **B** and **R**) impacts crucially the algorithm accuracy [Fisher, 2003]. In this hydrological application, despite a rich available historical database, error covariances could not be directly estimated because of the lack of knowledge of the unknown true precipitation/flow dynamics. This difficulty has also been noticed in many other DA applications with time-varying data (e.g. see [Bannister, 2008]). Innovation-based covariances diagnosis has been widely adopted in numerical weather prevision (NWP) and geosciences. A detailed overview and comparison of these methods could be found in [Tandeo et al., 2018]. Most of these methods are based on an on-line covariance learning process in a data assimilation chain. For our hydrological model, we use a 3D-Var data assimilation framework with a temporal correlation of the background state, which makes the Desroziers-type covariance diagnosis [Desroziers and Ivanov, 2001, Desroziers et al., 2005] suitable for this application. These methods are introduced and discussed in the following.

Desroziers & Ivanov covariance tuning (DI01)

First developed in the meteorological domain, the $\mathrm{DI01}$ tuning algorithm [Desroziers and Ivanov, 2001] consists in adjusting the ratio between **B** and **R** based on a simple parametrization of the traces of covariance matrices. This is done with an iterative fixed-point procedure for two scalar indicators. As no estimation of full matrices is required, unlike ensemble-based methods (e.g., EnKF [Evensen, 1994]), $\mathrm{DI01}$ could be applied with a single real-time assimilation trajectory.

More precisely, the iterative process could be synthesized as:

$$s_{b,n} = \frac{2\mathbb{E}\Big[J_b(\mathbf{x}_a)\Big]}{\mathrm{Tr}(\mathbf{K}_n\mathbf{H})}, \quad s_{o,n} = \frac{2\mathbb{E}\Big[J_o(\mathbf{x}_a)\Big]}{\mathrm{Tr}(\mathbf{I}-\mathbf{H}\mathbf{K}_n)},$$
$$\mathbf{B}_{n+1} = s_{b,n}\mathbf{B}_n, \quad \mathbf{R}_{n+1} = s_{o,n}\mathbf{R}_n, \tag{5.11}$$

where n is the iteration index, $\mathbf{K}_n$ is the Kalman gain matrix (as a function of $\mathbf{B}_n$ and $\mathbf{R}_n$) and $s_{b,n}, s_{o,n}$ represent two real value indicators. The closer to 1 they are, the more likely the ratio $\mathrm{Tr}(\mathbf{B})/\mathrm{Tr}(\mathbf{R})$ is to be well balanced. According to [Chapnik et al., 2004], this iterative tuning method is equivalent to a maximum-likelihood tuning of the matrix traces $Tr\Big(\mathbf{B}(s_b)\Big)$ and $Tr\Big(\mathbf{R}(s_o)\Big)$. The application of this iterative method in sub-spaces has also been discussed in [Desroziers and Ivanov, 2001] for bloc-diagonal prior error covariances. Very recent work of [Cheng et al., 2020b] has extended the application of $\mathrm{DI01}$ to prior error correlated subspaces, where the domain decomposition is performed via the state-observation mapping to get a more flexible parametrization. In this paper, we perform the block-wise DI01 to adjust the state/observation ratio as well as the amplitudes of two groups of state variables ξ_t^p and $\xi^{r,j}$. The marginal error deviations are considered homogeneous inside each group of variables.

Desroziers iterative method (D05) in the observation space

The Desroziers diagnosis (D05) is based on the observation-minus-background (O-B) and observation-minus-analysis (O-A) residuals, also known respectively as prior and posterior innovation quantity, in the observation space. It is proved in [Desroziers et al., 2005] that, in linear data assimilation, with perfectly specified covariances $\mathbf{B}$ and $\mathbf{R}$, the expectation of the analysis state should satisfy:

$$\mathbb{E}\Big([\mathbf{y}-\mathcal{H}(\mathbf{x}_a)][\mathbf{y}-\mathcal{H}(\mathbf{x}_b)]^T\Big) = \mathbf{R}. \tag{5.12}$$

Hence, the difference between the left side and the right side of Eq.5.12, where $||.||_F$ denotes the Frobenius norm:

$$||\mathbf{R} - \mathbb{E}\Big([\mathbf{y}-\mathbf{H}(\mathbf{x}_a)][\mathbf{y}-\mathbf{H}(\mathbf{x}_b)]^T\Big)||_F \tag{5.13}$$

can be used as a validation indicator of $\mathbf{R}$ matrix estimation. Expression in Eq.5.13 is estimated via Kalman-type formulation based on a best linear unbiased estimator (BLUE) [Desroziers et al., 2005].

Being fully established through statistics of residuals, an important advantage of this method is that time variant observation/background data could be used to estimate or diagnose the error covariances. Assuming perfect knowledge of $\mathbf{B}$ matrix, an iterative process has been put into

practice for $\mathbf{R}$ matrix specification [Bathmann, 2018]:

$$\mathbf{R}_{n+1} = \mathbb{E}\Big([\mathbf{y} - \mathcal{H}(\mathbf{x}_{a,n})][\mathbf{y} - \mathcal{H}(\mathbf{x}_b)]^T \Big). \tag{5.14}$$

We remind that the current analysis state $\mathbf{x}_{a,n}$ depends on the specification of $\mathbf{R}_n$ while other prior information (e.g. $\mathbf{x}_b, \mathbf{B}, \mathbf{y}$) remains invariant. It is proved in [Ménard, 2016] and [Bathmann, 2018] that, under the assumption of sufficient observation data and well specified $\mathbf{B}$ matrix, the iterative process of Eq.5.14 converges to the true observation error covariance. However, as mentioned by [Bathmann, 2018], despite the exact error covariance is symmetric positive definite (SPD), the intermediate matrices $\mathbf{R}_n$ could be non-symmetric and possibly contain negative or complex eigenvalues. In these cases, the variational objective function $\mathcal{J}$ could no longer be written as the sum of two metric distances because of the loss of positive definiteness of $\mathbf{R}$. Moreover, we show in the appendix an example that proves the BLUE-type linear solving, which is a key assumption of D05 diagnosis, is no longer valid when $\mathbf{R}$ owns negative eigenvalues.

Pointed out by [Bathmann, 2018], a posterior regularization at each iteration step is necessary to ensure the positive definiteness of $\mathbf{R}_n$ in practice. Regularization (e.g. localization [Gaspari and Cohn, 1999]) of error covariances has been widely studied and applied in ensemble-based DA algorithms. As mentioned in [Bathmann, 2018], the first step of regularization could be forcing the matrix to be symmetric by adding its transpose as:

$$\mathbf{R}_n \longleftarrow \frac{1}{2}(\mathbf{R}_n + \mathbf{R}_n^T). \tag{5.15}$$

The symmetry of $\mathbf{R}_n$ is thus guaranteed. As a consequence, the spectrum of $\mathbf{R}_n$ contains only real numbers but not necessarily positive. A standard approach in ensemble-based DA methods to ensure the positive definiteness is called the "hybrid method", which consists of combining a prior defined covariance matrix $\mathbf{C}$ (often diagonal or Matérn type) with the estimated one. This technique is widely adopted in practice for industrial applications (see [Carrassi et al., 2018]). We thus obtain the formulation of the regularized observation matrix $\mathbf{R}_{\mathrm{r,n}}$:

$$\mathbf{R}_{\mathrm{r,n}} \longleftarrow (1-\mu)\mathbf{R}_n + \mu\mathbf{C}, \tag{5.16}$$

following Eq.5.15 with $\mu \in (0, 1)$. Without extra information on the prior covariances (e.g. correlation scale), the matrix $\mathbf{C}$ is usually chosen to be diagonal as it helps improving the matrix conditioning. However, as mentioned in the discussion of [Bathmann, 2018], the convergence of regularized observation matrices remains an open question. In the appendix 5.8.3 of the present paper, we show that the matrix sequence built according to the regularization formula given in Eq.5.42, does not necessarily converge to the exact observation covariance matrix: a counter-example is given where a fixed point of Eq.5.14 other than the exact observation covariance is

found.

CUTE and PUB iterative covariance tuning methods

Recently proposed iterative methods called CUTE (Covariance Updating iTerativE) and PUB (Partially Updating Blue) [Cheng et al., 2019] consist in re-assimilating several times the observation data in order to reduce the posterior innovation and gains a better knowledge of the output error covariance. These methods are appropriate when the $\mathbf{B}$ matrix is unknown *a priori*, especially when the background error is underestimated. Unlike [Desroziers et al., 2005], the innovation quantity is not taken into account directly in the covariance computation. However [Cheng et al., 2019] mentioned that the posterior innovation could be used as a stopping criteria. CUTE and PUB update not only the current state variables $\mathbf{x}_{b,n}$ but also the associated state covariances by considering the newly emerged state-observation covariances. Because the observation error is supposed to be smaller than the background one in CUTE and PUB, $\mathbf{R}$ and $\mathbf{y}$ remain invariant in these two iterative methods.

CUTE method

More precisely, an iteration of CUTE $(n \rightarrow n+1)$ is based on a classical 3D-Var framework:

$$\mathbf{x}_{b,n+1} = \mathbf{x}_{a,n} = \underset{\mathbf{x}}{\operatorname{argmin}} \left(\frac{1}{2} ||\mathbf{x} - \mathbf{x}_{b,n}||^2_{\mathbf{B}_n^{-1}} + \frac{1}{2} ||\mathbf{y} - \mathcal{H}(\mathbf{x})||^2_{\mathbf{R}^{-1}} \right), \tag{5.17}$$

with a careful attention on the estimation of state-observation error covariances $\mathbf{Cov}(\epsilon_{b,n}, \epsilon_y)$ which emerged due to the iterative process itself. This process could be written as:

$$\begin{aligned} \mathbf{A}_n &= (\mathbf{I} - \mathbf{K}_n\mathbf{H})\mathbf{B}_n + (\mathbf{I} - \mathbf{K}_n\mathbf{H})\mathbf{Cov}(\epsilon_{b,n}, \epsilon_y)\mathbf{K}_n^T + \mathbf{K}_n\mathbf{Cov}(\epsilon_y, \epsilon_{b,n})(\mathbf{I} - \mathbf{K}_n\mathbf{H})^T, \\ \mathbf{B}_{n+1} &\leftarrow \frac{(1-\alpha)\mathrm{Tr}(\mathbf{B}_n) + \alpha\mathrm{Tr}(\mathbf{A}_n)}{\mathrm{Tr}(\mathbf{A}_n)}\mathbf{A}_n \end{aligned}$$

where $\mathbf{H}$ is the linearization of the transformation operator $\mathcal{H}$ in the neighbourhood of $\mathbf{x}_{b,n}$, and $\mathbf{A}_n$ is the estimation of posterior state error covariances at iteration n. The scaling coefficient $\alpha \in (0, 1)$ is related to the confidence level of prior error amplitude estimation. According to [Cheng et al., 2019], the more confident we are in the initial guess of $\mathbf{B}$ matrix, the higher level of α should be set. Then,

$$\mathbf{K}_n = \mathbf{B}_n\,\mathbf{H}^T(\mathbf{H}\mathbf{B}_n\mathbf{H}^T + \mathbf{R})^{-1}, \tag{5.18}$$

is so called the Kalman gain matrix. The iteration value of $\mathbf{Cov}(\epsilon_{b,n}, \epsilon_y)$, also deduced via the linear formulation, depends on the current Kalman gain:

$$\begin{aligned}\mathbf{Cov}(\epsilon_{b,n}, \epsilon_y) = \mathbf{Cov}(\epsilon_y, \epsilon_{b,n})^T &= \mathbf{Cov}\Big(\Big[(\mathbf{I} - \mathbf{K}_{n-1}\mathbf{H})\epsilon_{b,n-1} + \mathbf{K}_{n-1}\epsilon_y\Big], \epsilon_y\Big) \\ &= (\mathbf{I} - \mathbf{K}_{n-1}\mathbf{H})\mathbf{Cov}(\epsilon_{b,n-1}, \epsilon_y) + \mathbf{K}_{n-1}\mathbf{R}. \end{aligned} \tag{5.19}$$

PUB method

Instead of classical variational DA framework, the PUB method is built on a general BLUE formulation where state-observation error covariance could take place and be considered in the loss function for finding the optimal state. More precisely, the background and the observation vectors are combined into a single observation vector $\mathbf{z}$, with the extended definition of transformation operator $\tilde{\mathbf{H}}$, observation error $\mathbf{w}$ and error covariance $\mathbf{C}$:

$$\mathbf{z} \equiv \begin{pmatrix} \mathbf{x}_b \\ \mathbf{y} \end{pmatrix}, \quad \tilde{\mathbf{H}} \equiv \begin{pmatrix} \mathbf{I} \\ \mathbf{H} \end{pmatrix}, \quad \mathbf{w} \equiv \begin{pmatrix} \epsilon_b \\ \epsilon_y \end{pmatrix}, \quad \mathbf{C} \equiv \begin{pmatrix} \mathbf{B} & \mathbf{C}_{o-b}^T \\ \mathbf{C}_{o-b} & \mathbf{R} \end{pmatrix} \tag{5.20}$$

where $\mathbf{C}_{o-b} = \mathbf{Cov}(\epsilon_b, \epsilon_y)$ is set to zero at the beginning of the iterative process. The new objective function of DA could thus be written as:

$$J(\mathbf{x}) = \frac{1}{2}||\mathbf{z} - \tilde{\mathbf{H}}\mathbf{x}||_{\mathbf{C}^{-1}}. \tag{5.21}$$

We remind that while the state-observation error covariance does not exist (i.e. $\mathbf{C}_{o-b} = 0_{\dim(\mathbf{x})\times\dim(\mathbf{y})}$), the loss function of Eq.5.21 is completely equivalent to the 3D-Var formulation given in Eq.5.1. The PUB iterative formulation (both state estimation and error covariance) is obtained via the general form of BLUE where only the estimation of state variables and its associated error co-

variance are updated:

$$\mathbf{x}_{b,n+1} \leftarrow \mathbf{x}_{a,n} = \underset{\mathbf{x}}{\operatorname{argmin}}\left(\frac{1}{2}||\mathbf{z}_n - \tilde{\mathbf{H}}\mathbf{x}||_{\mathbf{C}_n^{-1}}\right), \tag{5.22}$$

$$\mathbf{z}_{n+1} = \begin{pmatrix} \mathbf{x}_{b,n+1} \\ \mathbf{y} \end{pmatrix}, \tag{5.23}$$

$$\mathbf{A}_n = (\tilde{\mathbf{H}}^T \mathbf{C}_n^{-1} \tilde{\mathbf{H}})^{-1}, \tag{5.24}$$

$$\mathbf{B}_{n+1} \leftarrow \frac{(1-\alpha)\mathrm{Tr}(\mathbf{B}_n) + \alpha \mathrm{Tr}(\mathbf{A}_n)}{\mathrm{Tr}(\mathbf{A}_n)} \mathbf{A}_n, \tag{5.25}$$

$$\mathbf{C}_{o-b,n+1}^T = (\tilde{\mathbf{H}}^T \mathbf{C}_n^{-1} \tilde{\mathbf{H}})^{-1} \tilde{\mathbf{H}}^T \mathbf{C}_n^{-1} \begin{pmatrix} \mathbf{C}_{o-b,n}^T \\ \mathbf{R} \end{pmatrix}, \tag{5.26}$$

$$\mathbf{C}_{n+1} = \begin{pmatrix} \mathbf{B}_{n+1} & \mathbf{C}_{o-b,n+1}^T \\ \mathbf{C}_{o-b,n+1} & \mathbf{R} \end{pmatrix}. \tag{5.27}$$

In the framework of twin experiments, [Cheng et al., 2019] shows that the improvement in terms of both the state estimation and the posterior error covariance specification could be achieved via CUTE or PUB as long as the **R** matrix is well known.

5.4 Data assimilation schemes for MORDOR-TS

We want to improve river discharges computed by MORDOR-TS in terms of both reanalysis and forecast by correcting initial reservoirs and daily precipitations over a given time period of T days also called assimilation window. Precipitations are daily prescribed over the 28 spatial mesh cells of the basin catchment and daily measured discharges at 9 gauges positioned on the Tarn river and its affluents are described in Fig. 5.1.

At this stage, we consider that MORDOR-TS takes 8 parameters to set the initial reservoir levels at the beginning of the assimilation window and T daily precipitations over the window as inputs and calculates T daily discharges at the 9 gauges as output (see Fig. 5.4 [b]). We introduce the following variables for each mesh cell index i with $i = 1, \ldots, 28$ and each day t with $t = 1, \ldots, T$:

- $p_{i,t}$: the precipitation (in mm),
- $w_{i,t}^j$: the water level of filling for the reservoir j (in mm),
- $Q_{q,t}^s$: the simulated river discharge (in m^3/s) at the q^{th} outlet,
- $Q_{q,t}^o$: the observed river discharge (in m^3/s) at the q^{th} outlet.

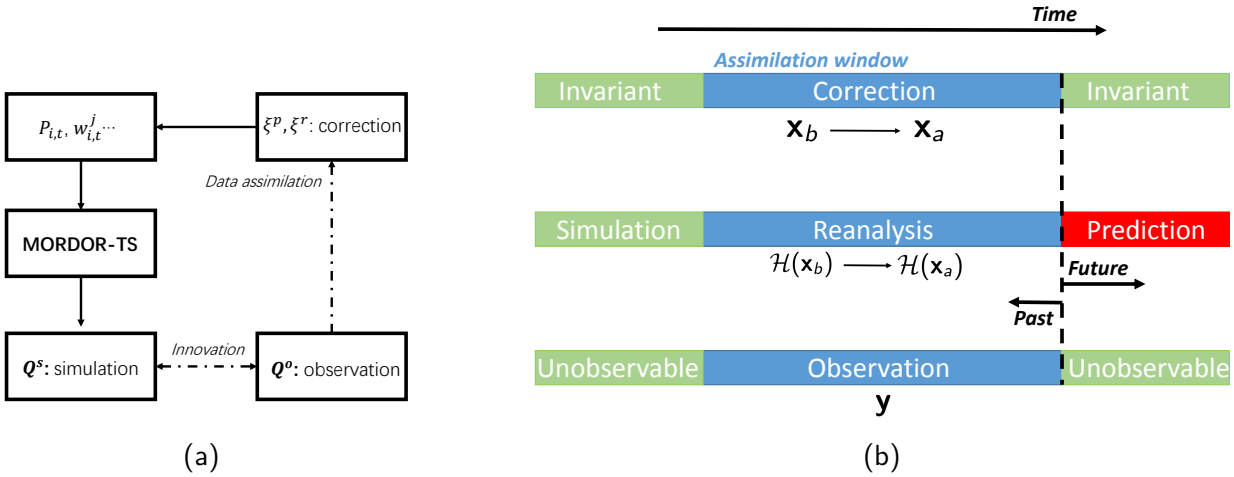


Figure 5.4: Fig. [a]: Flowchart of the MORDOR-TS hydrological model and DA articulation; Fig. [b]: DA modeling for a temporal assimilation window.

Other physical quantities, such as the observed temperatures, are considered as constant parameters in this DA modeling. Previous experiences show that performing DA correction on all input variables (i.e. $p_{i,t}$ and $w_{j,t}$) may introduce an over-parameterization and induce an "overfitting", with a high risk of deterioration of the flow forecast. For this reason, we perform an uniform spatial correction on the daily precipitations in the 28 spatial mesh cells. More precisely, the correction can be carried out with the help of two groups of increment variables:

- ξ^p_t: a spatially uniform additive correction on the precipitations at day t,
- $\xi^{r,j}$: additive corrections on the 8 parameters of the initial reservoir storage level at $t = 0$.

The scheme of DA algorithms is illustrated in Fig. 5.4 [a]. To gain a clear insight of the roles of each component (precipitation and initial reservoir level) in the control vector composed of the increments ξ^p_t and $\xi^{r,j}$ and their impact on the reconstructed discharge, three DA schemes are developed with different definitions of the control vector. More precisely, the 3D-Var-P scheme (resp. 3D-Var-R scheme) corrects solely the daily precipitation (resp. the initial storage level), while the last scheme named 3D-Var-P+R, using $\mathbf{x} = [\xi^p_t, \xi^{r,j}]$ as control vector, corrects both precipitation and initial storage level. For numerical experiments presented in this paper, T is set to 30 days. The control vector thus has a dimension from 8 to 38 depending on the modeling, and the observation vector has a dimension of 270. The main scheme characteristics are summarized in Table. 5.1.

DA scheme	Control vector : **x**	$dim(\mathbf{x})$	Observations: **y**	$dim(\mathbf{y})$	invariant parameters	constraints
3D-Var-P	ξ_t^p	T	$Q_{q,t}^o$	$9T$	$p_{i,t}, w_{j,0}, T_{i,t},$, etc	$p_{i,t} + \xi_t^p \geq 0$
3D-Var-R	$\xi^{r,j}$	8				
3D-Var-P+R	$\xi_t^p, \xi^{r,j}$	$T + 8$				

Table 5.1: Different DA schemes specifics (T stands for the size of the assimilation window).

For the DA schemes including a correction of precipitations, constraints have also been added to ensure the positiveness of precipitation in each spatial mesh cell, i.e.:

$$\mathbf{min}_{(i=1...28)}(p_{i,t}) + \xi_t^p \geq 0 \quad \text{for} \quad \forall t = 1..T. \tag{5.28}$$

These constraints ensure the physical feasibility of the analyzed state. However, this approach might not be optimal when observed precipitation quantities are highly non-homogeneous, i.e. when:

$$\exists t, \mathbf{max}_i(p_{i,t}) \gg \mathbf{min}_i(p_{i,t}).$$

In this case, the constraint of Eq.5.28 leads to:

$$\xi_t^p \geq -\mathbf{min}_i(p_{i,t}).$$

Thus only little correction can be performed while the highest quantity of precipitation $\mathbf{max}(p_{i,t})$ is overestimated. This difficulty could be overcome to some extent by correcting the initial reservoir level as in the 3D-Var-P+R scheme.

5.5 Error covariance tuning for the hydrological model

In this section, we explain how the covariance tuning strategies are implemented in the hydrological model with the 3D-Var modeling. We continuously shift by one day the assimilation period of T days in order to cover all possible assimilation windows, collecting sufficient assimilation residuals (i.e. $||\mathbf{y} - \mathcal{H}(\mathbf{x}_a)||_2$) for posterior covariances analysis, e.g. $\mathrm{DI01}$ or D05. For example, if the first assimilation window extends from January 1st, 1990 (denoted 1/1/1990) to January 30th, 1990 (denoted as 30/1/1990), then the following one will extend from 2/1/1990 to 31/1/1990. Therefore, each precipitation observation will be in fact assimilated several times. No information

about the error covariances **B** and **R**, neither the correlation kernel nor the error amplitude, is originally available for this application. As shown in section 5.4, prior assumptions are required for each of these methods. We thus decide to test these tuning approaches in a specific order, from an offline adjusting of error amplitudes to a more refined online (real-time) tuning as shown in Fig. 5.5.

5.5.1 Initial set up

In DA approaches, when sufficient error statistic is not available, the error covariance is often defined using some preselected symmetric definite matrices. If prior errors are supposed to be uncorrelated, an identity matrix is used directly as the initial covariance matrix in many cases (e.g. [Chandramouli et al., 2020]). Otherwise, correlated prior errors are often represented by a correlation kernel of Matérn-type, e.g. the exponential kernel (Matérn $1/2$), the Balgovind kernel (Matérn $3/2$) or the Gaussian kernel (Matérn $5/2$) (see [Cheng et al., 2019]). Among them, the Balgovind type, also known as second-order autoregressive (SOAR) correlation function, is widely applied in NWP and geosciences because of its smoothness and distance regularity. The latter is crucial in all Bayesian based methods, including DA. These beneficial properties of the Balgovind kernels have been noticed for a wide range of DA applications, including NWP [Stewart et al., 2013], nuclear modeling [Ponçot et al., 2013],[Gong et al., 2020a], chemical design [Singh et al., 2011], for covariance computation. This correlation kernel $\phi(.)$ can be written as:

$$\phi(r) = (1 + \frac{r}{L}) \exp(-\frac{r}{L}) \tag{5.29}$$

where r represents the spatial distance between the correlated points and L denotes the correlation scale length. In this work, we choose to represent the initial error correlation either by a diagonal matrix (if independent), or by a Balgovind kernel (if correlated).

5.5.2 Offline DI01

For this approach, the idea is to monitor the ratio between ξ_t^p (whose error standard deviations are respectively denoted as $\sigma_{b,1}$ and $\sigma_{b,2}$) and $\xi^{r,j}$ introduced in section 5.4 . To this end, we rely on the $\mathrm{DI01}$ algorithm used as a pre-stage offline step, cf. Fig. 5.5, applied in each subspace based on the modeling of 3D-Var-P or 3D-Var-R, both using the full observation vector of dimension 270 (9 outlets $\times$ 30 days). The application of $\mathrm{DI01}$ in subspaces has been studied in [Desroziers and Ivanov, 2001] for multivariate systems and in [Cheng et al., 2020b] combined with unsupervised clustering techniques. In this paper, the objective of applying $\mathrm{DI01}$ in these subspaces is to specify the sub background-observation ratios ($\sigma_{b,1}/\sigma_o$, $\sigma_{b,2}/\sigma_o$), and eventually to adjust the

inter-blocks ratio, i.e.:

$$\frac{\sigma_{b,1}}{\sigma_{b,2}} = \frac{\sigma_{b,1}}{\sigma_o} \times \frac{\sigma_o}{\sigma_{b,2}}. \tag{5.30}$$

More precisely, in order to gain a more robust inter-blocks ratio, the $\mathrm{DI01}$ algorithms for 3D-Var-P and 3D-Var-R are carried out in all assimilation windows from 2003 to 2004, leading to a total of 730 assimilations. The average of the estimated ratio $\sigma_{b,1}/\sigma_o$, $\sigma_{b,2}/\sigma_o$ is taken through a logarithmic function with base 10,

$$\mathbb{E}\Big[\log(\frac{\sigma_{b,k}}{\sigma_o})\Big] = \frac{1}{2}\left(\mathbb{E}\left[\log\left(\prod_{n=1}^{n_{\max}} \frac{s_{b,n}^k}{s_{o,n}^k}\right)\right]\right) \tag{5.31}$$

where $k = 1, 2$ and $n_{\max}$ is set to be a fixed value for each subspace. The reason of taking a logarithmic function is to balance the impact of some extreme values. These values are computed offline, leading to an initial setup for the error covariances:

$$\mathbf{B} = \begin{pmatrix} \exp\left(\mathbb{E}\Big[\log(\frac{\sigma_{b,1}}{\sigma_o})\Big]\right) \times \mathbf{B}_1 & 0_{30\times 8} \\ 0_{8\times 30} & \exp\left(\mathbb{E}\Big[\log(\frac{\sigma_{b,2}}{\sigma_o})\Big]\right) \times \mathbf{B}_2 \end{pmatrix} \quad \text{and} \quad \mathbf{R} = \mathbf{I} \tag{5.32}$$

where $\mathbf{B}_1(30 \times 30), \mathbf{B}_2(8 \times 8)$ are the initial guesses of background error correlation concerning respectively the daily precipitation ξ_t^p and the 8 parameters $\xi^{r,j}$. We impose a positive temporal correlation in $\mathbf{B}_1$ by taking a Balgovind kernel with correlation scale $L = 5$ while the estimation error of the reservoir parameters is supposed to be uncorrelated, i.e. $\mathbf{B}_2 = \mathbf{I}$. Since we are only interested in the analyzed state $\mathbf{x}_a$, the true error amplitudes of $\mathbf{B}$ and $\mathbf{R}$ are not important as long as the ratio $\frac{\sigma_{b,k}}{\sigma_o}$ is well specified. In fact, if $\mathbf{B}$ and $\mathbf{R}$ are multiplied by the same factor, no impact will appear in the optimization result in Eq.5.1, i.e. the analyzed state $\mathbf{x}_a$. The estimated value $\prod_{n=1}^{n_{\max}} \frac{s_{b,n}^k}{s_{o,n}^k}$ for each assimilation is drawn in Fig. 5.6 [a,b] where the x-axis represents the date of the beginning of the assimilation window. We also represent the averaged evolution of $s_{b,n}^1, s_{o,n}^1, s_{b,n}^2, s_{o,n}^2$ and their logarithms against $\mathrm{DI01}$ iterations in Fig. 5.6 [c,d,e,f] where the sky blue area represents the standard deviation of $s_{b,n}, s_{o,n}$. We observe a fast convergence of s_b, s_o in average for both modelings, coherent with the results in [Desroziers and Ivanov, 2001]. The maximum number of $\mathrm{DI01}$ iteration is set to be $n_{\max} = 15$ for 3D-Var-P due to the high variance of s_b^2 as shown in Fig. 5.6 [e]. According to previous experiences, the error variance of ξ_t^p (resp. $\xi^{r,j}$) is set to be 10^3 (resp. 10^2) higher than the one of observation. Finally, we obtain the

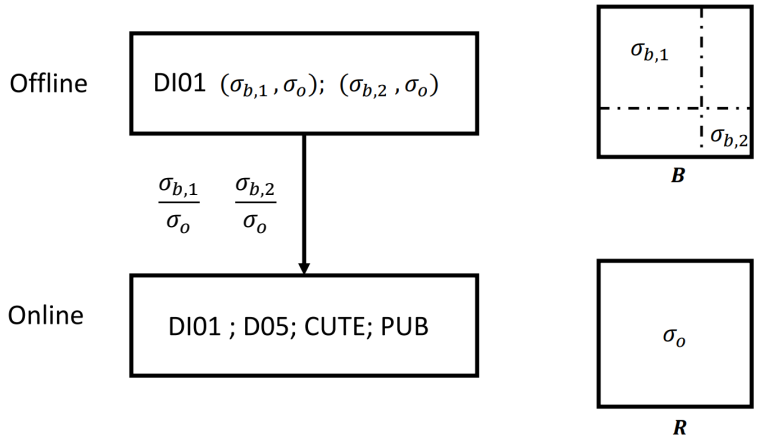


Figure 5.5: Diagram of the combination of offline and online covariance tuning methods.

averaged background-observation covariance ratio, suggested by DI01,

$$\mathbb{E}\left[\prod_{n=1}^{n_{\max}=5} \frac{s_{b,n}^1}{s_{o,n}^1}\right] = -2.24, \quad \mathbb{E}\left[\prod_{n=1}^{n_{\max}=15} \frac{s_{b,n}^2}{s_{o,n}^2}\right] = 1.56, \tag{5.33}$$

leading to an initial covariance set up,

$$\mathbb{E}\Big[\log(\frac{\sigma_{b,1}}{\sigma_o})\Big] = \log(10^3) + (-\frac{2.24}{2}) = 1.88, \quad \mathbb{E}\Big[\log(\frac{\sigma_{b,2}}{\sigma_o})\Big] = \log(10^2) + (-\frac{1.56}{2}) = 2.78 \tag{5.34}$$

5.5.3 D05 online estimation for the observation matrix

Relying on the error amplitudes adjusted offline in section 5.5.2, we apply the residual analysis of [Desroziers et al., 2005] for flow-independent observation covariance (of dimension 270×270) specification. To this end, 3D-Var-P+R is implemented for all assimilation windows from 1990 to 2000 (for a total of 3652) for O-A and O-B residual computations. The result of the **R** matrix estimation through Eq.5.12, as well as the associated error correlation is shown in Fig. 5.7 [a,b,c,d]. The observation gauges, each of 30 days, follow the order of Fig. 5.3 [a] in the covariance matrices where the auto-covariance of *Tarn at Millau* is indicated in Fig. 5.7 [a] and [b]. Due to its geographical position, the observation of *Tarn at Millau* has the highest error variance. According to numerical results, a strong spatial error correlation among different gauges exists while the temporal correlation could almost be neglected. Since the estimation is performed using real-world data, the estimated **R** matrix (Fig. 5.7 [b]) is slightly non-symmetric. We then regularize this matrix as shown in Eq.5.15 and Eq.5.16, with $\alpha = 0.1$ and $\mathbf{C} = \mathrm{Tr}(\mathbf{R}) \times \mathbf{I}$ for a second iteration of D05 on the same data. The output **R** matrix of the second iteration

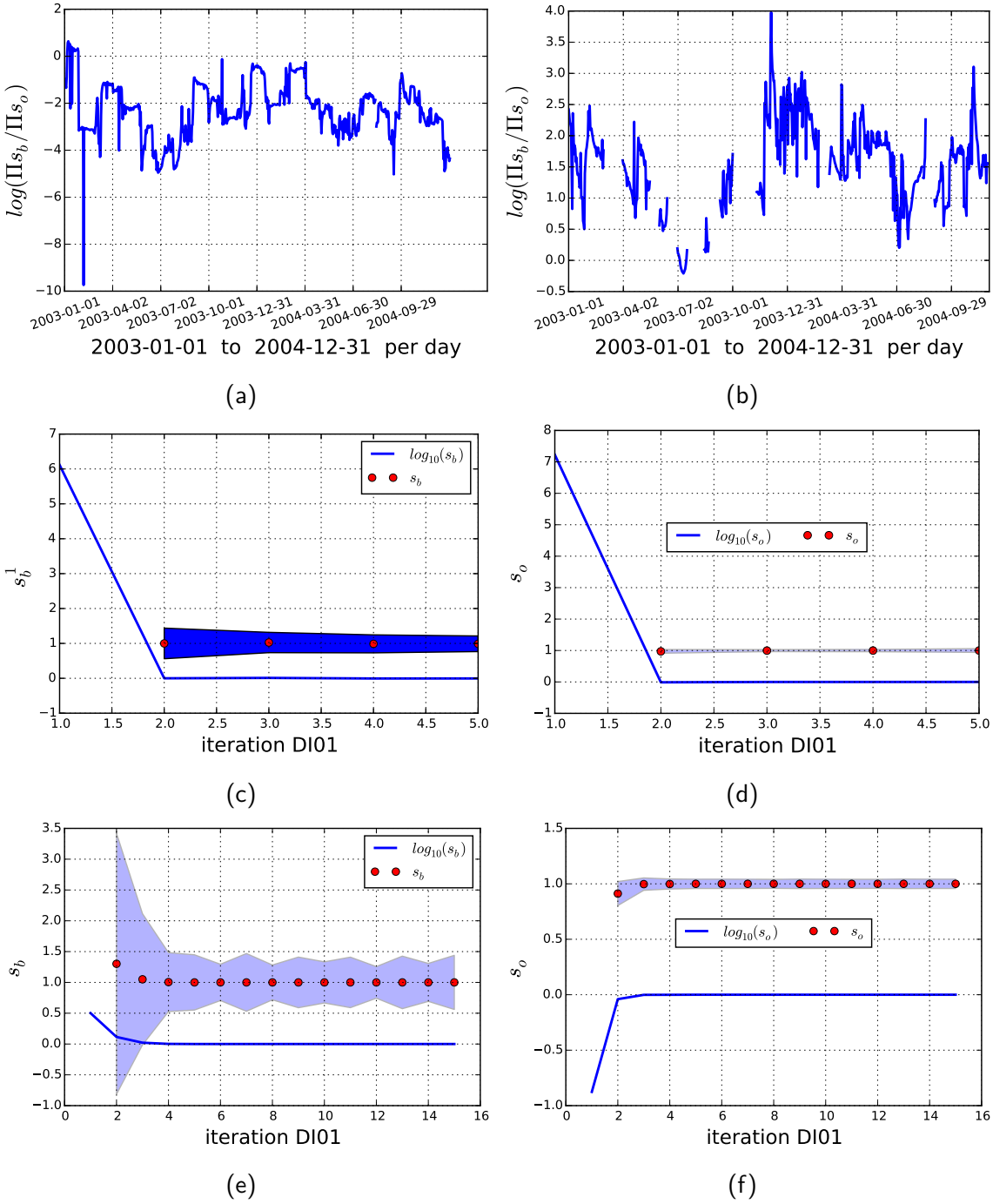


Figure 5.6: Fig. [a,b]: Evolution of $\frac{s_{b,n}^k}{s_{o,n}^k}$ ([a]: k=1, [b]: k=2); Fig. [c,d,e,f]: averaged evolution of respectively $s_{b,n}^1, s_{o,n}^1, s_{b,n}^2, s_{o,n}^2$ against $\mathrm{DI01}$ iterations n where the sky blue area represents the variable deviation.

(Fig. 5.7 [f]) is very similar to the regularized matrix after the first D05 iteration (Fig. 5.7 [e]), both in terms of error amplitude and correlation structure. The fast convergence of D05 is achieved thanks to the previous tuning of DI01. As pointed out by [Desroziers et al., 2005], an adequate initial guess accelerates the algorithm convergence. We could thus conclude that this **R** matrix is stable under the Desroziers diagnosis, allowing to consider it as the "true" observation matrix as mentioned in [Gauthier et al., 2018].

5.5.4 Online DI01

In section 5.5.2, DI01 is implemented in subspaces to provide a better initial guess of **B** and **R**. Nevertheless, this averaged error amplitude ratio might not fit every pair of $(\mathbf{x}_b, \mathbf{y})$ in different windows. In fact, the estimated $\prod_{n=1}^{n_{\max}} \frac{s_{b,n}^k}{s_{o,n}^k}$ show certain variance against time in Fig. 5.6 [a,b]. Therefore, we apply a new (online) DI01 step in order to adjust the $||\mathbf{B}||/||\mathbf{R}||$ ratio subject to each pair of $(\mathbf{x}_b, \mathbf{y})$. This means the expectation in DI01, as shown in Eq.5.11, is estimated using one scalar realisation of J_b and J_o. The result of Eq.5.34 is used as the initial $||\mathbf{B}||/||\mathbf{R}||$ ratio where the inter-blocks fraction $\sigma_{b,1}/\sigma_{b,2}$ remains invariant during iterations. The maximum number of iterations is fixed to $n_{\max} = 10$ for this online application.

5.5.5 Online CUTE or PUB

Here, we apply the CUTE or PUB methods (both with the confidence coefficient $\alpha = 0.2$) using the result obtained in section 5.5.2 as initial set up. Similar to DI01 online, CUTE or PUB provide an individual correction, according to each assimilation window with a maximum of 5 iterations. The maximum iteration number is smaller than regular cases (see [Cheng et al., 2019]) since we consider that the $||\mathbf{B}||/||\mathbf{R}||$ ratio is well adjusted via offline DI01.

This strategy of subsequent tuning algorithms, where the error covariance matrices are progressively specified as shown in Fig. 5.5, could be followed for other types of industrial applications under similar prior conditions. We are now interested in quantifying the performance of data assimilation in terms of flow reanalysis and, more importantly, flow forecast.

5.6 Flow forecast

As explained previously, the goal of DA for this problem is to improve the flow *forecast* accuracy. In this section, we show the forecast improvement issued from the online covariance tuning CUTE, PUB, DI01 and the D05 observation covariance estimation, based on the error amplitudes $(\sigma_{b,1}, \sigma_{b,2}, \sigma_o)$ adjusted via offline DI01.

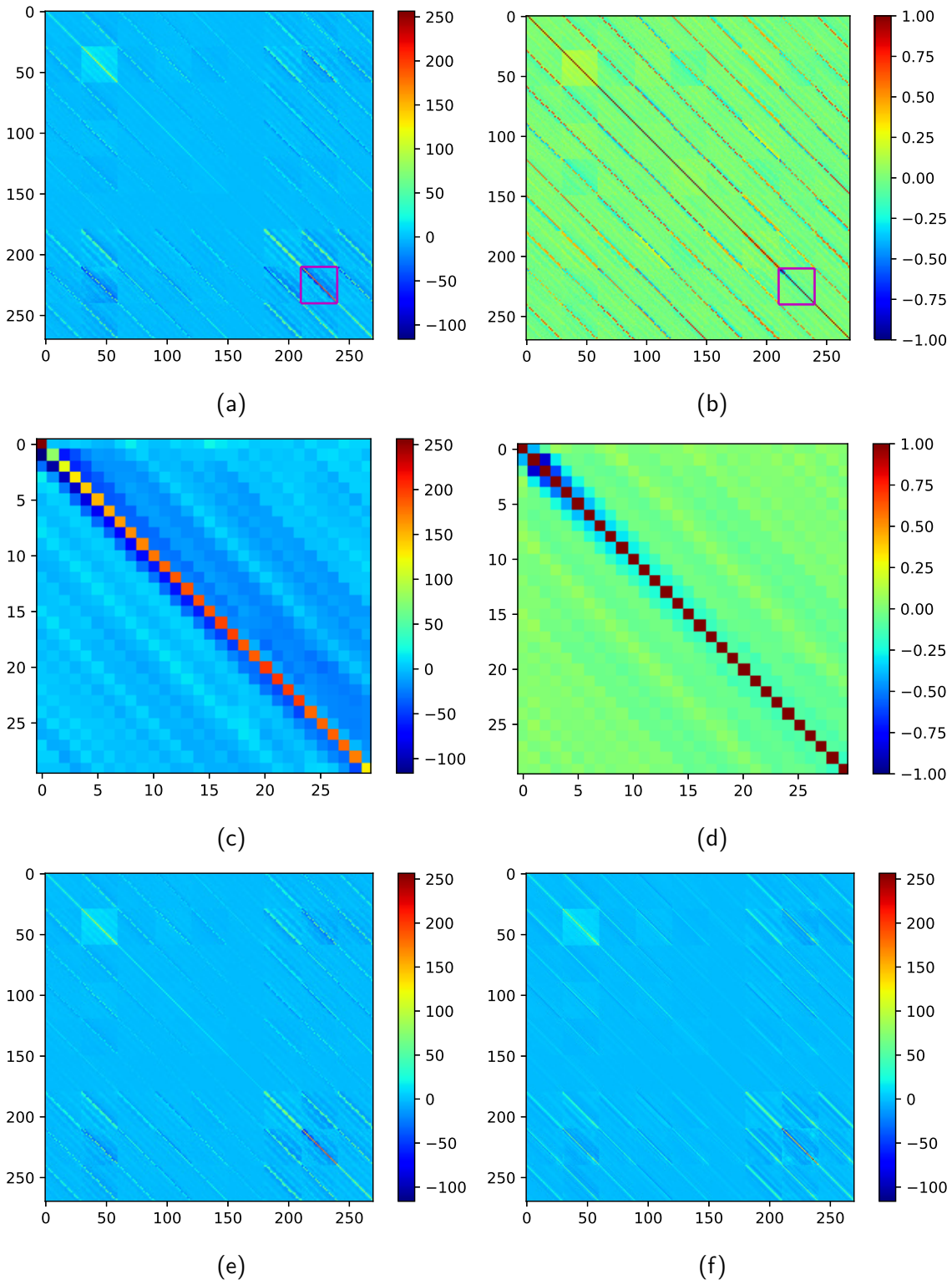


Figure 5.7: Estimated observation error covariance [a] and correlation [b] with *Tarn at Millau* [c,d] station closeups for the first D05 estimation. Fig. [e] shows the regularized matrix of 9 gauges while Fig. [f] represents the output of the second D05 iteration.

	reanalysis (30 days)		forecast (3 days)	
Online DA methods	Millau gauge	9 gauges	Millau gauge	9 gauges
DI01	60.56%	45.34%	32.44%	20.63%
D05	60.39%	44.95%	32.97%	20.59%
CUTE	68.59%	51.9%	43.28%	29.52%
PUB	64.3%	48.83%	34.61%	23.74%

Table 5.2: Averaged (over all assimilation windows in 1990) flow improvements of reanalysis and forecast (at Millau or over the nine gauges) for the different online DA methods.

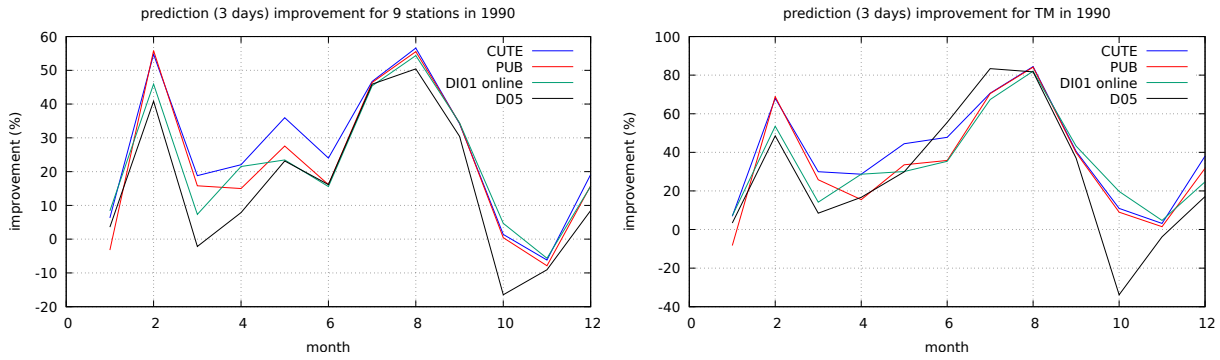
5.6.1 Forecast improvement rate

The forecast improvement rate denoted as ∇ is obtained by calculating the normalized difference of observation-(model forecast) and observation-(corrected forecast), i.e.

$$\nabla = \frac{\mathbb{E}(||\mathbf{y} - \mathbf{H}\mathbf{x}_b||_2) - \mathbb{E}(||\mathbf{y} - \mathbf{H}\mathbf{x}_a||_2)}{\mathbb{E}(||\mathbf{y} - \mathbf{H}\mathbf{x}_b||_2)}. \tag{5.35}$$

where $||.||_2$ stands for the euclidean vector norm. The $\mathbf{x}_b, \mathbf{x}_a$ and $\mathbf{y}$ in Eq.5.35 is taken in a small chosen window just after the assimilation window as shown in Fig. 5.4 [b]. We concentrate on the short-range forecast of 3 days for all 9 gauges, regarding Fig. 5.3[a]. The improvement rate is estimated by averaging all assimilation windows of the year 1990, i.e. 365 assimilations of 30 days to be precise, shown in Table. 5.2. All covariance tuning approaches manage to improve extensively the forecast and reconstruction precision in DA modeling, compared to the original prediction of MORDOR. This improvement is more significant at the Millau outlet than other catchments. In general, DI01 and D05 are over-performed by CUTE and PUB. More precisely, among all approaches, CUTE shows a consistent advantage in terms of both reanalysis and forecast.

The averaged improvement rate per month is shown in Fig. 5.8 where each x-axis tick specifies the month at which the assimilation window starts. For example, the value at $x = 1$ represents the averaged forecast improvement of all assimilation correction started in January 1990, i.e. $1^{st}, 2^{nd}$... until 31^{th} January. From Fig. 5.8, observing similar evolution of different methods, a seasonal effect is well noticed. In fact, the period of June to August usually stands for the longest drought in a year where the river flow is more consistent. This effect helps the correction of DA algorithms and explains the good performance of all approaches during this period. The bad scores observed for all methods on October in Fig. 5.8 are probably related to a very particular meteorological episode in the south of France (Cévenol episode) that can be seen in Fig. 5.2, where very heavy rainfall causes multiple high floods over a very short period of time (a few days). Such an event occurring over a few days does not fit very well with a 30-day assimilation window. In summary, all DA approaches shown in Fig. 5.8 provide a significant enhance (around 30% to 40% in average for most months) compared to the model (background) flow forecast.



(a) (b)

Figure 5.8: Averaged improvement rate per month in 1990 in all 9 stations [a] and at *Tarn at Millau* [b].

5.6.2 Forecast in all gauges

We display the observed flow (x-axis) of the first day after the assimilation window against the predicted one (y-axis) using different methods in Fig. 5.9 and Fig. 5.10. The 10% highest observed flow are presented in red while the others in blue. The number of gauges follow the same order in Fig. 5.3 [a]. We notice a consistent advantage of DA approaches with advanced covariance tuning methods compared to the background simulation, especially when flooding (in red) except for some extreme levels (outlier events). These results show the improvement of the DA relying on the CUTE, PUB and DI01 algorithms, particularly for medium-level flood events. These improvements are more significant at *Tarn at Montbrun, Tarnon at Florac, Breze at Meyrueis, Tarn at Millau*, i.e. the constraints 1, 3 in Fig. 5.9 and 5, 7 in Fig. 5.10.

5.6.3 Some examples

Several examples of 30-day assimilation windows are displayed in Fig. 5.11 [a,b,e,f] where a purple vertical line clearly separates the reanalysis (left side corresponding to past) and the forecast (right side corresponding to future) at *Tarn at Millau*. The yellow stars depict the evolution of daily observation. The sub-figures Fig. 5.11 [c,d,g,h] show the difference between reconstructed/predicted flow and the observation, associated to these four scenarios. These scenarios are chosen based on significant differences between the simulation (blue curve) and observation (yellow points). Moreover, these scenarios refer to different hydrological regimes, from - long-range drought (Fig. 5.11 [e]) to - the onset of a flood event (Fig. 5.11 [a,b]). In fact, abrupt drought–flood transitions (and vice versa) stand for difficult scenarios to deal with. A sudden rainfall, somewhat unexpected by the system, might deteriorate the actual state, either by an overreacting (e.g. around April 7^{th} in Fig. 5.11 [a]) or an underacting (e.g. around September 27^{th} in Fig. 5.11 [f]). Since the daily measured precipitation represents a sort of integration in 24 hours, some representation error (see [Janjić et al., 2018] for a clear definition) may also take

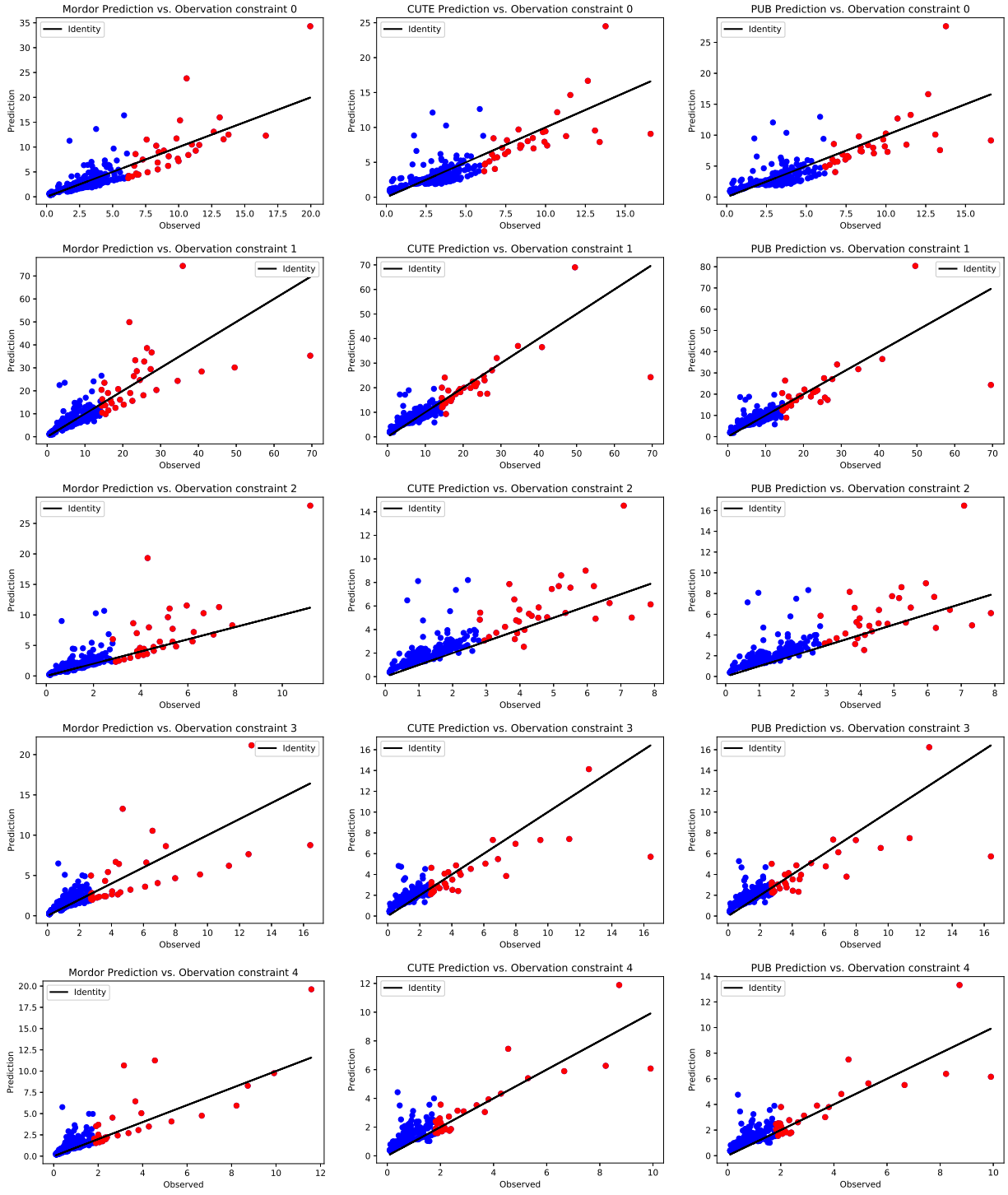


Figure 5.9: Forecast: 1st day after the assimilation window for observation stations in 1990 for constraints (gauges) 0 to 4, following the same order as Fig. 5.3. The first column represent the original MODRDOR forecast without data assimilation. The second and the third column represent the forecast with respectively CUTE and PUB DA correction.

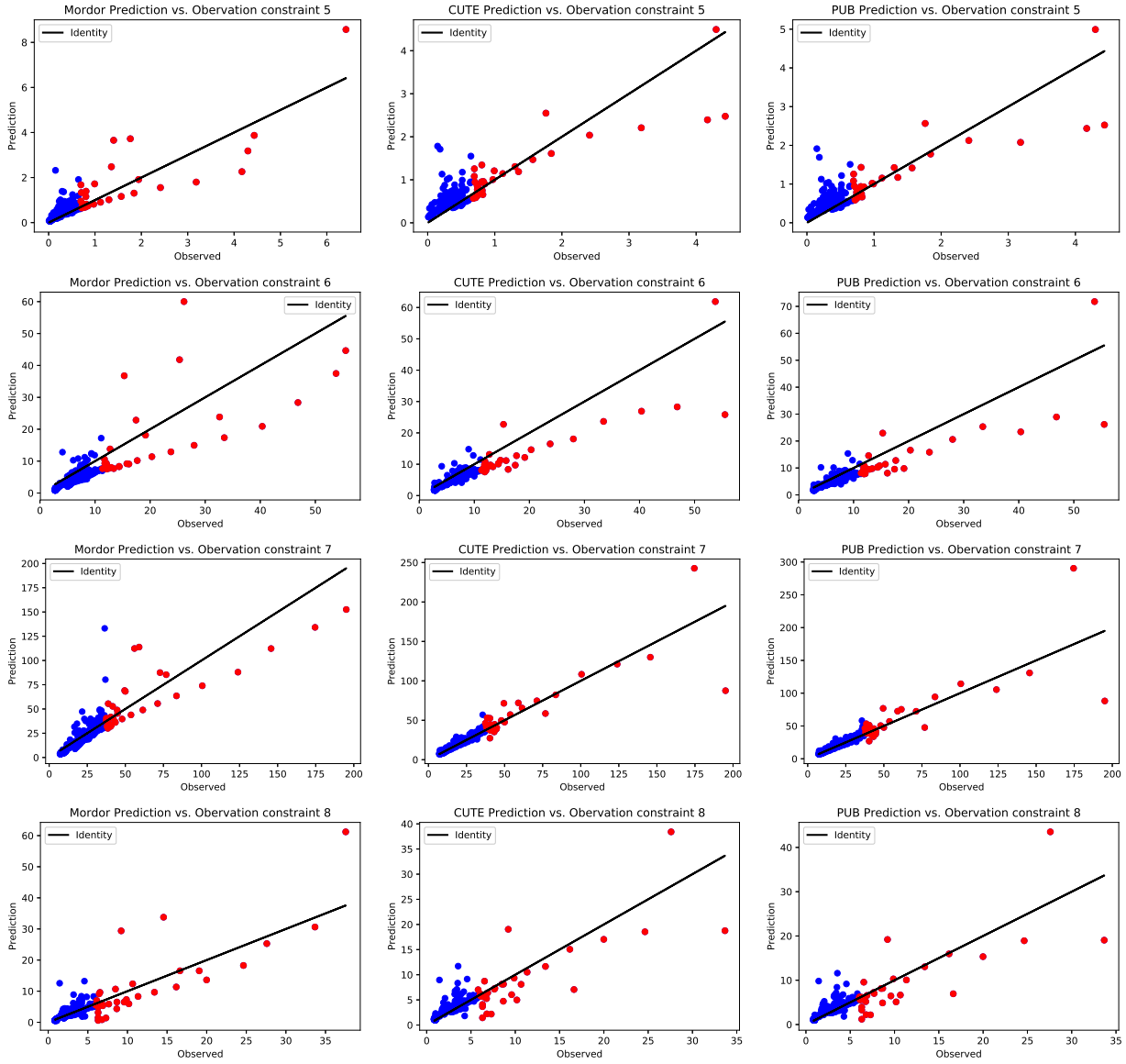


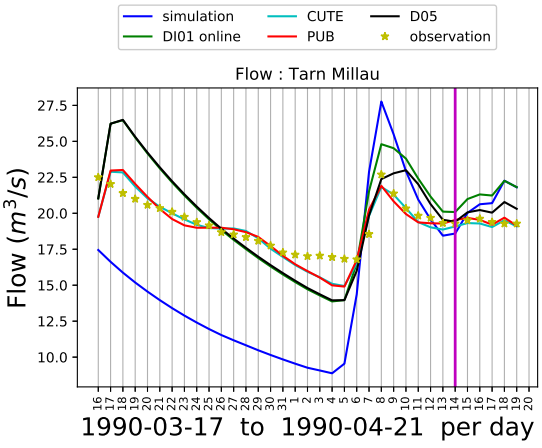
Figure 5.10: Same description as Fig. 5.9 for constraints 5 to 8.

place. For example, if the rain starts very early in the morning and lasts for a short moment, the flow increase might be delayed in MODOR-TS simulation, compared to reality.

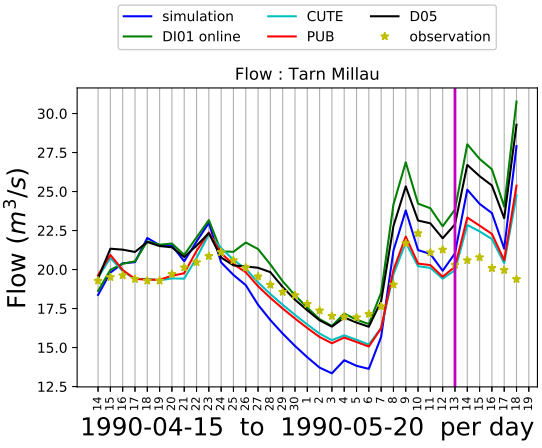
We now take a clearer look at these scenarios. Fig. 5.11 [a] and Fig. 5.11 [e] show consistent underestimation (could be overestimation elsewhere) of MORDOR-TS simulation for *Tarn at Millau* during drought periods. A similar phenomenon could also be found in Fig. 5.2. Since the O-B innovation is monotonous (either overestimating or underestimating) for a relatively long period, these scenarios could be easily corrected by DA approaches once error covariances are well-specified. This also explains that the best forecast performance is found in summer as shown in Fig. 5.8. Fig. 5.11 [b] and Fig. 5.11 [f] show two DA windows where the observed river flow are somehow more chaotic, compared to Fig. 5.11 [a], Fig. 5.11 [e]. In these cases, a more precise temporal error covariance obtained by tuning algorithms, helps the reconstructed flow to be more consistent with the observed one, leading to a better flow forecast as shown by the curves of CUTE and PUB after the assimilation window. In general, it is clear that the model simulation (blue curve) is remarkably improved by DA correction, in both reanalysis and forecast. Comparing different approaches, CUTE and PUB methods have a more accurate short-range forecast than DI01 online or D05 by providing a better history matching in the assimilation window. From this fact, the correction of CUTE, PUB might be considered closer to reality.

5.7 Discussion

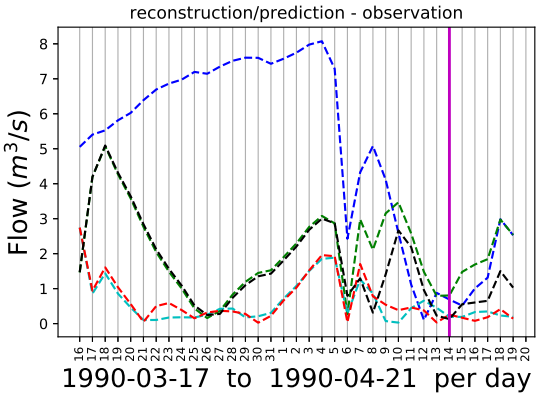
Data assimilation is commonly applied in hydrology to improve the flow simulation ([Leisenring and Moradkhani, 2011]). but the error covariance estimation is often challenging for these applications. It is very common for DA algorithms, to have to start with an arbitrarily chosen error covariance matrix, sometimes as trivial as the identity matrix. Several approaches have been proposed to improve the covariance computation or diagnosis. However, these methods often require some severe preliminary conditions, such as the knowledge of the error correlation form (DI01), the knowledge of observation error variances (CUTE, PUB) or the flow-independent assumption (D05). These conditions could be already difficult to be individually fulfilled and thus even more difficult to be jointly fulfilled. However, we find that, if these approaches are applied in a certain order, following the principle of first correcting error amplitudes and then correcting their covariance structure, former methods may provide a better initial set up to the last ones. This scheme could be split into two parts: an offline tuning for balancing different state variables and for adjusting the initial $||\mathbf{B}||/||\mathbf{R}||$ ratio, and an online tuning for the covariance structure in each assimilation window. The main contribution of this paper is about proposing a scheme of these covariance tuning methods and applying them in a real-world hydrological model. We concentrate on variational approaches in this study. As for covariance tuning algorithms, DI01, D05, CUTE and PUB are chosen because they do not necessarily require a long data assimilation chain or a large ensemble of simulated trajectories. All these methods are applied for a sufficient number of iterations until some stopping criteria is satisfied. Numerical results show that the matching of observed and reconstructed flow is progressively improved along the schema


simulation
DI01 online
CUTE
PUB
D05
observation
Flow : Tarn Millau
Flow (m^3/s)
1990-03-17 to 1990-04-21 per day

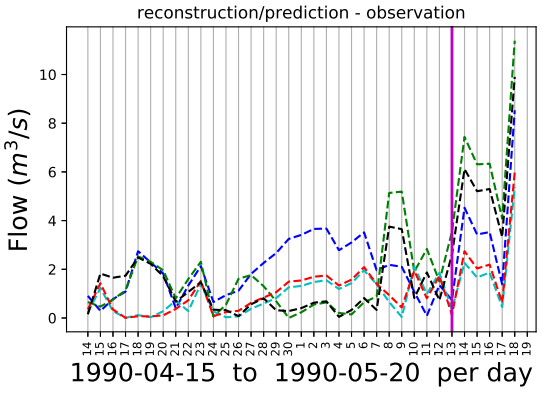
(a)


simulation
DI01 online
CUTE
PUB
D05
observation
Flow : Tarn Millau
Flow (m^3/s)
1990-04-15 to 1990-05-20 per day

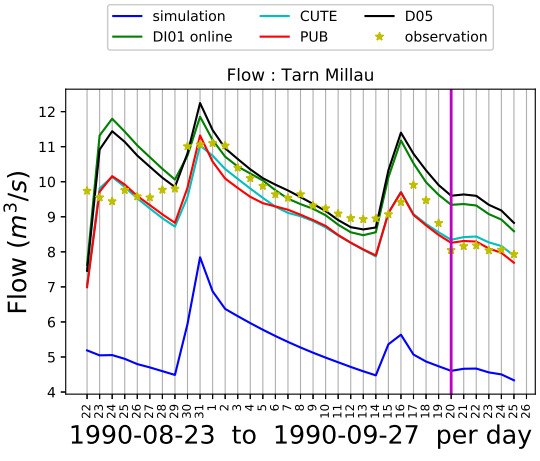
(b)


reconstruction/prediction - observation
Flow (m^3/s)
1990-03-17 to 1990-04-21 per day

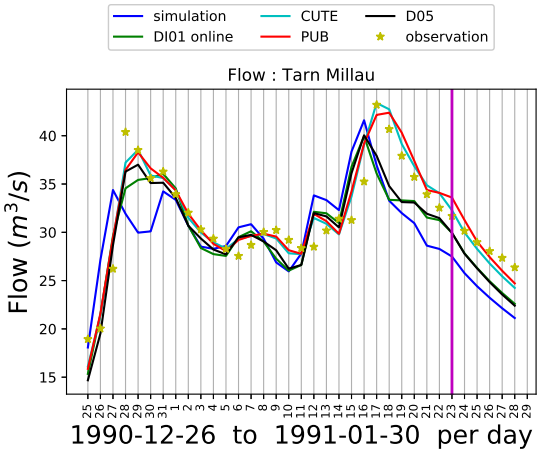
(c)


reconstruction/prediction - observation
Flow (m^3/s)
1990-04-15 to 1990-05-20 per day

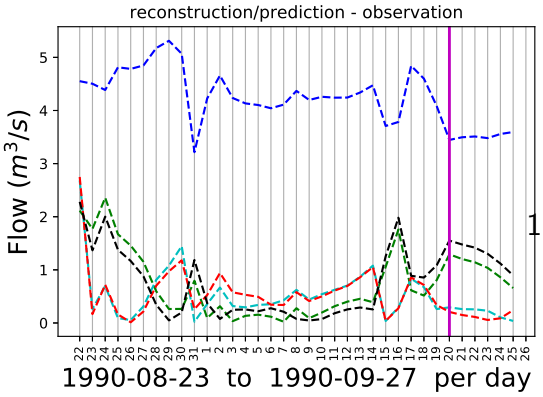
(d)

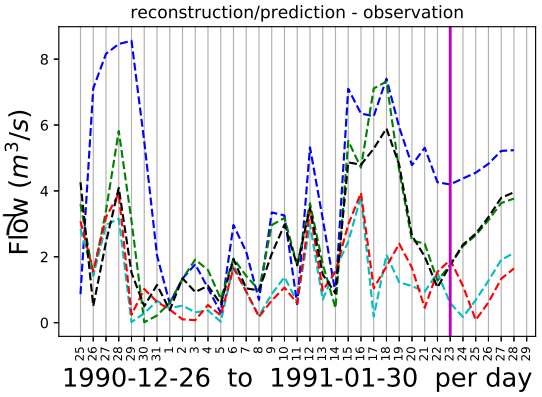

simulation
DI01 online
CUTE
PUB
D05
observation
Flow : Tarn Millau
Flow (m^3/s)
1990-08-23 to 1990-09-27 per day

(e)


simulation
DI01 online
CUTE
PUB
D05
observation
Flow : Tarn Millau
Flow (m^3/s)
1990-12-26 to 1991-01-30 per day

(f)


reconstruction/prediction - observation
Flow (m^3/s)
1990-08-23 to 1990-09-27 per day


reconstruction/prediction - observation
Flow (m^3/s)
1990-12-26 to 1991-01-30 per day

of tuning algorithms. More importantly, we gain a significantly more accurate and more robust flow forecast for the Tarn river which is crucial for industrial applications at EDF concerning the management of hydroelectrical power plants. Among all online specification methods, CUTE and PUB, iterating consistently both the analyzed state and its associated covariance, provide the most accurate short-range forecast in average. Furthermore, when the iteration number is fixed, CUTE and PUB is computationally cheaper than DI01 because the optimization cost for finding the analyzed state reduces against CUTE/PUB iterations.

A second contribution of this paper consists of the study of D05 convergence under regularization, following the recent work of [Ménard, 2016] and [Bathmann, 2018]. We prove, by giving a counter-example, that the theoretical convergence of D05 is no longer ensured under a certain type of regularization. As mentioned in [Bathmann, 2018], this regularization is often necessary in practice, due to sampling errors in covariance estimation. Moreover, although the proof of [Bathmann, 2018] is algebraically correct, the formulation of iteration might not be valid because of the non positive definiteness of intermediate matrices. This fact is also shown by an example in the appendix. These findings obviously suggest more careful attention while applying D05 approach, especially when the number of iterations is large. In this study, while applying D05 with regularization, we obtain an observation matrix which is stable under the Desroziers criteria in the hydrological model.

Future work will involve exploration of other combinations of covariance tuning algorithms. In this study, we have also tried to combine CUTE (or PUB) with D05. However, the flow forecast results were less optimal, compared to CUTE and PUB with a diagonal $\mathbf{R}$ matrix, as shown in this paper. The main reason could come from the representation error in the covariance estimation, which is hard to quantify and to be eliminated via statistical approaches. We also intend to apply the combination of covariance tuning methods in a more general framework other than hydrology, such as nuclear engineering, object tracking etc. These applications often involve more complicated dynamical systems than the conceptual hydrological model presented in this paper. Since covariance specification algorithms are computationally expensive, especially for online tuning, we have the idea to reduce the computational cost via an efficient representation of real-time observations using data compression techniques [Fowler, 2019].

5.8 Appendix: Convergence of D05 iterative method

5.8.1 Justification of convergence in the ideal case

The convergence of D05 iterative method is proved by [Bathmann, 2018] in the ideal case, i.e., when the expectation in Eq.5.14 is error-free and the current iterative matrix $\mathbf{R}_n$ stays always invertible.

Following the notation of [Bathmann, 2018], let

$$\mathbf{G} = \mathbf{H}\mathbf{B}\mathbf{H}^T \tag{5.36}$$

$$\mathbf{D} = \mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}_{\mathrm{E}}, \tag{5.37}$$

respectively denote the projection of the background matrix in the observation space and its sum with the exact observation matrix $\mathbf{R}_{\mathrm{E}}$. Therefore, we have necessarily

$$\mathbf{D} - \mathbf{G} = \mathbf{R}_{\mathrm{E}}. \tag{5.38}$$

According to [Bathmann, 2018], the updating formulation of Eq.5.14 is equivalent to:

$$\mathbf{R}_{n+1} = \mathbf{R}_n(\mathbf{G} + \mathbf{R}_n)^{-1}\mathbf{D}. \tag{5.39}$$

where n is the current iteration. It is obvious that the exact observation matrix $\mathbf{R}_{\mathrm{E}}$ is a fixed point of Eq.5.39. In fact, when $\mathbf{R}_{\mathrm{E}}$ is SPD, the iterative process of Eq.5.39 converges necessarily to the exact covariance $\mathbf{R}_{\mathrm{E}}$. The interested readers are referred to [Bathmann, 2018] and [Ménard, 2016]. We describe briefly their algebraic proof based on the two following lemmas.

Lemma 1. If $\mathbf{D}$ and $\mathbf{G}$ is SPD and $\mathbf{D} - \mathbf{G}$ is also SPD, then $\lambda_{max}(\mathbf{D}^{-1}\mathbf{G}) < 1$. Otherwise $\lambda_{max}(\mathbf{D}^{-1}\mathbf{G}) \geq 1$.

Lemma 2. For the matrix sequence $\mathbf{R}_n$ defined in Eq.5.39, let $\mathbf{M} = \mathbf{D}^{-1}\mathbf{G}$, then

$$\mathbf{R}_n^{-1} = \mathbf{R}_{\mathrm{E}}^{-1} + \mathbf{M}^n[\mathbf{R}_0^{-1} - \mathbf{R}_{\mathrm{E}}^{-1}] \tag{5.40}$$

The convergence could thus be derived as shown in ***Theorem 1***.

Theorem 1. If the fixed point $\mathbf{R}_{\mathrm{E}} = \mathbf{D} - \mathbf{G}$ is SPD then the D05 iterations converge to $\mathbf{R}_{\mathrm{E}}$. Otherwise the iterations diverge to a singular matrix.

Proof by Bathmann: If $\mathbf{R}_{\mathrm{E}}$ is SPD. By ***Lemma 2.***

$$\mathbf{R}_n^{-1} = \mathbf{R}_{\mathrm{E}}^{-1} + \mathbf{M}^n[\mathbf{R}_0^{-1} - \mathbf{R}_{\mathrm{E}}^{-1}] \tag{5.41}$$

where $\mathbf{M} = \mathbf{D}^{-1}\mathbf{G}$. As $\lambda_{max}(\mathbf{M}) < 1$, by ***Lemma 1.***, $\mathbf{M}^n \longrightarrow 0$, thus $\mathbf{R}_{A,n}^{-1} \longrightarrow \mathbf{R}_E^{-1}$ since they are both non-singular. If $\mathbf{R}_{\mathrm{E}}$ is not SPD, then $\lambda_{max}(\mathbf{M}) \geq 1$, thus $\mathbf{M}^n$ diverges. We can deduce that $||\mathbf{R}_n^{-1}|| \longrightarrow \infty$ and therefore $\{\mathbf{R}_n\}$ diverges to a singular matrix.

The case when $\mathbf{B}$ matrix is incorrectly specified is discussed in [Ménard, 2016], where it is proved that $\mathbf{R}_n$ will become rank deficient if the eigenvalues of $\mathbf{B}$ are overestimated.

5.8.2 Necessary regularization

As pointed out by [Bathmann, 2018], the Eq.5.39 could not ensure the symmetricity of the updated matrix $\mathbf{R}_{n+1}$. An operation to enforce the symmetricity is necessary which leads the Eq.5.39 to:

$$\mathbf{R}_{n+1} = \frac{1}{2}(\mathbf{R}_n + \mathbf{R}_n^T)\Big(\mathbf{G} + \frac{1}{2}(\mathbf{R}_n + \mathbf{R}_n^T)\Big)^{-1}\mathbf{D}. \tag{5.42}$$

Being discussed in [Bathmann, 2018], the study of the convergence of Eq.5.42 remains, for instance, an open question. It is mentioned in [Bathmann, 2018] and [Ménard, 2016] that an extra-regularization, e.g. via a hybrid method, is also needed to ensure that all the eigenvalues to be strictly positive.

5.8.3 Limitations of Desroziers method

Non-convergence of regularized matrix sequence

We found that unlike Eq.5.39, the regularized sequence of Eq.5.42 could have another fixed, different from the true observation covariance (i.e. $\mathbf{R}_{\mathrm{E}} = \mathbf{D} - \mathbf{G}$). The proof is given by a counter-example:

$$\mathbf{G} = \begin{bmatrix} 1.5 & 1 \\ 1 & 4 \end{bmatrix}, \mathbf{D} = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}, \mathbf{R} = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}, \tag{5.43}$$

which satisfies

$$\mathbf{R} = \frac{1}{2}(\mathbf{R} + \mathbf{R}^T)\Big(\mathbf{G} + \frac{1}{2}(\mathbf{R} + \mathbf{R}^T)\Big)^{-1}\mathbf{D}, \tag{5.44}$$

and $\mathbf{R}$ is not SPD. Therefore, the proof of [Bathmann, 2018] is no longer valid for regularized matrix sequence as the equation has other fixed points other than $\mathbf{R}_{\mathrm{E}}$.

Negative eigenvalues

The appearance of negative eigenvalues is known as an important challenge of D05 iterative methods (see [Bathmann, 2018] and [Ménard, 2016]). In this work, we found that the assumption of symmetric positiveness of $\mathbf{R}_n$ (for all n) should be added in the proof of section 5.8.1 otherwise this proof could be completely misleading. In fact, when the $\mathbf{R}$ matrix possess negative eigenvalues, the term $\frac{1}{2}(\mathbf{y} - \mathcal{H}(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{y} - \mathcal{H}(\mathbf{x}))$ which no longer represents a real norm, could have negative values, leading to a different expression of the analyzed state $\mathbf{x}_a$. As consequence, the Desroziers diagnosis formulation, which is established via a "BLUE" type resolution, is no longer valid. This effect is illustrated with a simple 2D example:

$$\mathbf{x}_b = \begin{bmatrix}0\\0\end{bmatrix}, \mathbf{B} = \mathbf{I}_{2,2}, \mathbf{H} = \begin{bmatrix}1 & 0\\1 & 1\end{bmatrix}, \mathbf{R} = \begin{bmatrix}1 & 0\\0 & -1\end{bmatrix}, \mathbf{y} = \begin{bmatrix}0\\0\end{bmatrix}. \tag{5.45}$$

$$\text{let} \quad \mathbf{x} = \begin{bmatrix}\mathbf{x}_1\\\mathbf{x}_2\end{bmatrix}, \quad \text{thus} \quad \mathcal{J}(x) = \mathbf{x}_1^2 - 2\mathbf{x}_1\mathbf{x}_2. \tag{5.46}$$

It is obvious that the objective function $\mathcal{J}$ does not process a minimum in $\mathbb{R}/\infty$. However, the Desroziers diagnosis take into account the BLUE formulation, which writes as

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}_b) = \begin{bmatrix}0\\0\end{bmatrix}. \tag{5.47}$$

Therefore, because of this incoherence emerged by the non-symmetricity of $\mathbf{R}_n$, although the mathematical proof in section 5.8.1 is without fault, the application of Desroziers iterative method may probably not lead to the true observation covariance even in the ideal case described in [Bathmann, 2018].

5.8.4 MORDOR storage modeling

This section is not included in the submitted paper. We illustrate the storage modeling and the exchange between reservoirs in Fig. 5.12.

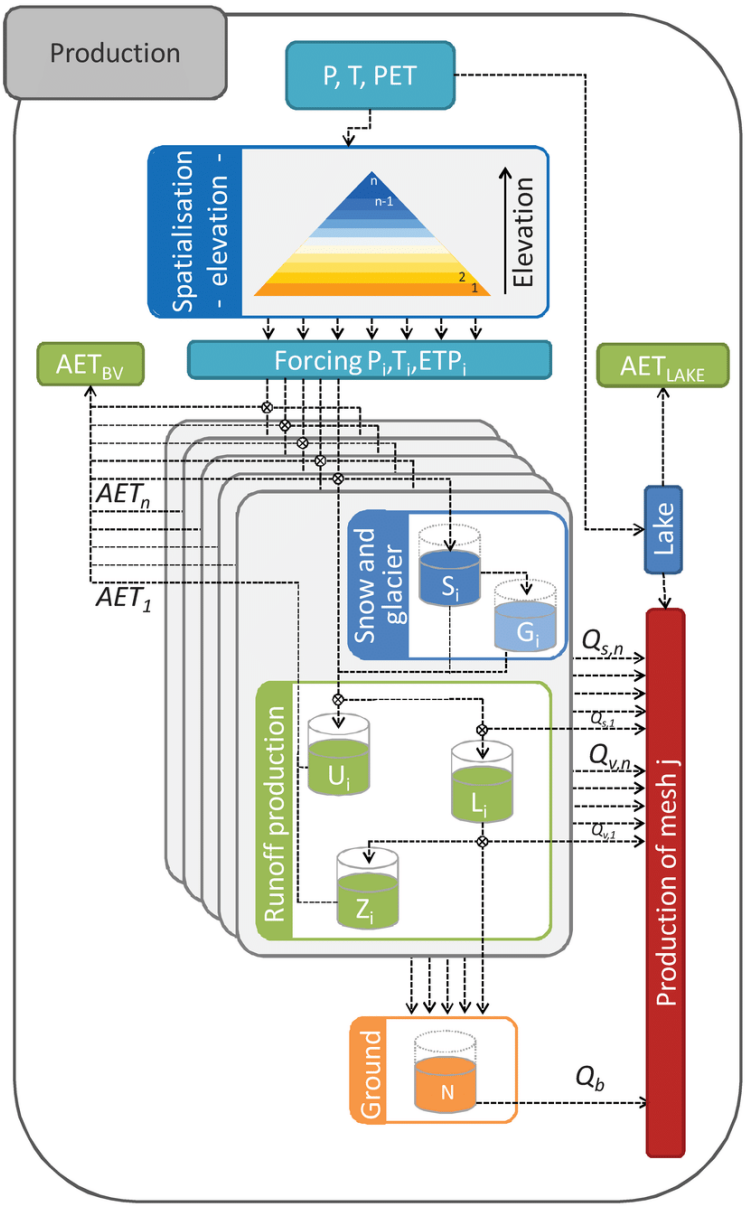


Figure 5.12: The MORDOR-TS reservoir modelling, from [Rouhier et al., 2017].

Chapter 6

Conclusion and future work

6.1 Conclusion

Data assimilation is an important tool to combine real-time or historical observations with simulation results to improve the static or dynamic estimation of system states. The assimilation precision depends essentially on the specification of error covariances. The latter is often challenging in some industrial problems because neither the knowledge of the full historical dynamic nor a large ensemble of simulated trajectories is easily available. Hence, classical calibration or ensemble-based approaches may not be appropriate to this end. Furthermore, instead of using a long DA chain, we intend to improve the short-range forecast of industrial problems. In this thesis, we review in detail some well-known covariance diagnosis and tuning methods, such DI01, D05, mostly in the variational assimilation framework. Relying on the assumption of flow-independence, these methods, based on statistics of innovation quantities from different time stamps, are used to estimate either the full covariance structure or some associated key parameters including error amplitude, correlation scale etc. However, originally developed in meteorology, these methods depend on specific prior assumptions. For example, the DI01 requires the knowledge of error correlation and the D05 iterative method assumes that the $\mathbf{B}$, $\mathbf{H}$ matrices are well-known. Some of these conditions can be difficult to fulfil in other domains of our interest, such as nuclear engineering, hydrology or civil engineering.

In this thesis, we develop two new methods CUTE and PUB, consisting of several iterations of classical assimilation formula with the same set of observations. We further take into account the error covariance between the updated state and the observations which appears due to the iterative process itself. We first test this method in a twin experiment framework with the nonlinear discretized shallow water equation. According to numerical results, both the assimilation accuracy and the output error covariance identification improve against the iteration of CUTE and PUB.

To extend on multivariate and multidimensional systems, we introduce the concept of graph-based localizations. Compared to classical domain localization approaches, this method does not rely on prior assumptions on the error covariance (e.g. correlation scale, preselected subspaces). Instead, the local spaces are automatically detected via graph-based unsupervised algorithms. The basic idea is to group the state variables impacted by the same observations to one subspace. This concept is implemented using DI01 in chapter 4. The combination with other covariance tuning algorithms is to be explored hereafter.

In chapter 5, we apply these covariance tuning methods including CUTE and PUB with an offline DI01 in subspaces as a preliminary step in an industrial hydrological model. Following the principle of "first adjusting error amplitudes then covariance structure", we propose an "optimal" combination of tuning algorithms with some specific order of activation. This strategy improves significantly short-range flow forecast as expected, which is crucial in industrial problems at EDF. Among all online tuning algorithms, CUTE and PUB show some advantages in both history matching (reanalysis) and forecast.

In summary, the main contribution of this thesis is the development of new methods to improve the precision and the efficiency of error covariance computation in data assimilation in an industrial context.

6.2 Future work

The study and implementation of these newly developed methods can be pursued, on a broader range of application fields, such as xenon dynamics forecast in nuclear engineering or deformation detection in material laws calibration. More particularly, we believe that the graph-based localization can be much advantageous when dealing with multi-grade data sets. The choice and effects of particular parameters (e.g. the confidence rate in CUTE/PUB and the extra filtering for extreme values in graph-based localization) in these algorithms is still open. We look for theoretical analysis as well as real-world experiments to ensure appropriate values of these parameters.

Furthermore, since domain localization techniques are somewhat limited to sparse $\mathbf{H}$ matrices, we also look for more general approaches to reduce the computational cost and thus accelerate real-time assimilations which will allow more refined covariance tuning algorithms to take place. For instance, we have considered observation data compression techniques which are widely applied in data assimilation problems for forecasting/reanalyzing complex dynamical systems. These techniques, including POD-type ([Collard et al., 2010]) or information-based ([Waller et al., 2017]) data compression, attempts to find a efficient representation of observation data, with minimum loss of assimilation accuracy. Our current experiences show that these compression strategies perform well in both twin experiments and the hydrological application. The next step could be performing covariance specification methods with compressed data.

Another perspective is about the combination of data assimilation with machine learning (ML) methods, more precisely, with deep learning algorithms. Very recent work of [Brajard et al., 2019] introduces a new hybrid method of this kind to improve the approximation of dynamical models via deep learning regression. The idea can be roughly summarized as using DA corrections to optimize/filter the learning targets of neural networks based on available observations. In this DA-ML iterative process, the error covariance specification is far from obvious as the updated trajectory is an output of neural networks. We believe that this specification could be improved by non-parametric approaches such as D05, CUTE or PUB.

Bibliography

[Arcucci et al., 2017] Arcucci, R., D'Amore, L., Pistoia, J., Toumi, R., and Murli, A. (2017). On the variational data assimilation problem solving and sensitivity analysis. *Journal of Computational Physics*, 335:311–326.

[Arcucci et al., 2018] Arcucci, R., Mottet, L., Pain, C., and Guo, Y.-K. (2018). Optimal reduced space for variational data assimilation. *Journal of Computational Physics*, 379.

[Argaud, 2019] Argaud, J.-P. (2019). User documentation, in the SALOME 9.3 platform, of the ADAO module for "Data Assimilation and Optimization". Technical report 6125-1106-2019-01935-EN, EDF / R&D.

[Argaud et al., 2018] Argaud, J.-P., Bouriquet, B., Caso, F., Gong, H., Maday, Y., and Mula, O. (2018). Sensor placement in nuclear reactors based on the generalized empirical interpolation method. *Journal of Computational Physics*, 363.

[Argaud et al., 2016] Argaud, J.-P., Bouriquet, B., Courtois, M., and Le Roux, J.-C. (2016). 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*, volume 7. ASME.

[Argaud et al., 2017] Argaud, J. P., Bouriquet, B., Gong, H., Maday, Y., and Mula, O. (2017). Stabilization of (g)eim in presence of measurement noise: Application to nuclear reactor physics. In Bittencourt, M. L., Dumont, N. A., and Hesthaven, J. S., editors, *Spectral and High Order Methods for Partial Differential Equations ICOSAHOM 2016*, pages 133–145, Cham. Springer International Publishing.

[Asch et al., 2016] Asch, M., Bocquet, M., and Nodet, M. (2016). *Data assimilation: methods, algorithms, and applications*. Fundamentals of Algorithms. SIAM.

[Bannister, 2008] Bannister, R. N. (2008). A review of forecast error covariance statistics in atmospheric variational data assimilation. i: Characteristics and measurements of forecast error covariances. *Quarterly Journal of the Royal Meteorological Society*, 134(637):1951–1970.

[Bathmann, 2018] Bathmann, K. (2018). Justification for estimating observation-error covariances with the Desroziers diagnostic. *Quarterly Journal of the Royal Meteorological Society*, 144(715):1965–1974.

[Bishop, 2019] Bishop, C. H. (2019). Data assimilation strategies for state dependent observation error variances. *Quarterly Journal of the Royal Meteorological Society*, 1-11.

[Blayo et al., 2012] Blayo, E., Bocquet, M., Cosme, E., and Cugliandolo, L. F. (2012). Advanced data assimilation for geosciences: Lecture notes of the les Houches school of physics: Special issue, June 2012.

[Blondel et al., 2008] Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008.

[Bocquet et al., 2010] Bocquet, M., Pires, C. A., and Wu, L. (2010). Beyond Gaussian Statistical Modeling in Geophysical Data Assimilation. *Monthly Weather Review*, 138(8):2997–3023.

[Bouttier and Courtier, 2002] Bouttier, F. and Courtier, P. (2002). Data assimilation concepts and methods. In *Meteorological Training Course Lecture Series*. ECMWF.

[Brajard et al., 2019] Brajard, J., Carrassi, A., Bocquet, M., and Bertino, L. (2019). Combining data assimilation and machine learning to emulate a dynamical model from sparse and noisy observations: a case study with the Lorenz 96 model. *Geoscientific Model Development Discussions*, 2019:1–21.

[Browet and Van Dooren, 2014] Browet, A. and Van Dooren, P. (2014). Low-rank similarity measure for role model extraction. In *21st International Symposium on Mathematical Theory of Networks and Systems, July 7-11, 2014. Groningen, The Netherlands*.

[Buehner et al., 2010] Buehner, M., Houtekamer, P. L., Charette, C., Mitchell, H. L., and He, B. (2010). Intercomparison of Variational Data Assimilation and the Ensemble Kalman Filter for Global Deterministic NWP. Part II: One-Month Experiments with Real Observations. *Monthly Weather Review*, 138(5):1567–1586.

[Byrd et al., 1995] Byrd, R. H., Lu, P., and Nocedal, J. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific and Statistical Computing*, 16(5):1190–1208.

[Carrassi et al., 2018] Carrassi, A., Bocquet, M., Bertino, L., and Evensen, G. (2018). Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *Wiley Interdisciplinary Reviews: Climate Change*, 9(5):e535.

[Castronova and Goodall, 2014] Castronova, A. M. and Goodall, J. L. (2014). A hierarchical network-based algorithm for multi-scale watershed delineation. *Computers & Geosciences*, 72:156 – 166.

[CEA/DEN et al., 2020] CEA/DEN, EDF R&D, and Open Cascade (2020). SALOME, The Open Source Integration Platform for Numerical Simulation. http://www.salome-platform.org/.

[Chabot et al., 2017] Chabot, V., Berre, L., and Desroziers, G. (2017). 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.

[Chabot et al., 2015] Chabot, V., Nodet, M., Papadakis, N., and Vidard, A. (2015). Accounting for observation errors in image data assimilation. *Tellus A: Dynamic Meteorology and Oceanography*, 67(1):23629.

[Chandramouli et al., 2020] Chandramouli, P., Memin, E., and Heitz, D. (2020). 4D large scale variational data assimilation of a turbulent flow with a dynamics error model. *Journal of Computational Physics*, 412:109446.

[Chapnik et al., 2004] Chapnik, B., Desroziers, G., Rabier, F., and Talagrand, O. (2004). Property and first application of an error-statistics tuning method in variational assimilation. *Quarterly Journal of the Royal Meteorological Society*, 130(601):2253 – 2275.

[Chapnik et al., 2006] Chapnik, B., Desroziers, G., Rabier, F., and Talagrand, O. (2006). Diagnosis and tuning of observational error in a quasi-operational data assimilation setting. *Quarterly Journal of the Royal Meteorological Society*, 132(615):543–565.

[Cheng et al., 2019] Cheng, S., Argaud, J.-P., Iooss, B., Lucor, D., and Ponçot, A. (2019). Background error covariance iterative updating with invariant observation measures for data assimilation. *Stochastic Environmental Research and Risk Assessment*, 33(11):2033–2051.

[Cheng et al., 2020a] Cheng, S., Argaud, J.-P., Iooss, B., Lucor, D., and Ponçot, A. (2020a). Error covariance tuning in variational data assimilation: application to an operating hydrological model, accepted for publication, link. *Stochastic Environmental Research and Risk Assessment*.

[Cheng et al., 2020b] Cheng, S., Argaud, J.-P., Iooss, B., Ponçot, A., and Lucor, D. (2020b). A graph clustering approach to localization for adaptive covariance tuning in data assimilation based on state-observation mapping, preprint.

[Cheng et al., 2017] Cheng, S., Laurent, A., and Dooren, P. V. (2017). Role model detection using low rank similarity matrix.

[Cherian et al., 2011] Cherian, A., Sra, S., Banerjee, A., and Papanikolopoulos, N. (2011). Efficient similarity search for covariance matrices via the jensen-bregman logdet divergence. pages 2399–2406.

[Chui and Chen, 1991] Chui, C. and Chen, G. (1991). *Kalman Filtering With Real-Time Applications*. Springer.

[Cioaca and Sandu, 2014] Cioaca, A. and Sandu, A. (2014). Low-rank approximations for computing observation impact in 4D-Var data assimilation. *Computers & Mathematics with Applications*, 67(12):2112 – 2126.

[Clayton et al., 2012] Clayton, A. M., Lorenc, A. C., and Barker, D. M. (2012). 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.

[Clifford S. et al., 2009] Clifford S., T., Tivadar M, T., and Róbert, B.-F. (2009). Graphclus, a matlab program for cluster analysis using graph theory. *Computers & Geosciences*, 35(6):1205 – 1213.

[Cobb et al., 2014] Cobb, L., Krishnamurthy, A., Mandel, J., and Beezley, J. D. (2014). Bayesian tracking of emerging epidemics using ensemble optimal statistical interpolation. *Spatial and Spatio-temporal Epidemiology*, 10:39 – 48.

[Collard et al., 2010] Collard, A. D., McNally, A. P., Hilton, F. I., Healy, S. B., and Atkinson, N. C. (2010). The use of principal component analysis for the assimilation of high-resolution infrared sounder observations for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 136(653):2038–2050.

[Coppersmith and Winograd, 1990] Coppersmith, D. and Winograd, S. (1990). Matrix multiplication via arithmetic progressions. *Journal of Symbolic Computation*, 9(3):251 – 280.

[Courtier et al., 1998] Courtier, P., Andersson, E., Heckley, W., Vasiljevic, D., Hamrud, M., Hollingsworth, A., Rabier, F., Fisher, M., and Pailleux, J. (1998). The ECMWF implementation of three-dimensional variational assimilation (3D-var). i: Formulation. *Quarterly Journal of the Royal Meteorological Society*, 124:1783 – 1807.

[Cuthill and McKee, 1969] Cuthill, E. and McKee, J. (1969). Reducing the bandwidth of sparse symmetric matrices. In *Proceedings of the 1969 24th National Conference*, ACM '69, pages 157–172, New York, NY, USA. ACM.

[Daget, 2008] Daget, N. (2008). *Estimation d'ensemble des paramètres des covariances d'erreur d'ébauche dans un système d'assimilation variationnelle de données océaniques*. PhD thesis, Université de Toulouse, France.

[Daley, 1992] Daley, R. (1992). The lagged innovation covariance: A performance diagnostic for atmospheric data assimilation. *Monthly Weather Review*, 120(1):178–196.

[Derber and Rosati, 1989] Derber, J. and Rosati, A. (1989). A Global Oceanic Data Assimilation System. *Journal of Physical Oceanography*, 19(9):1333 – 1347.

[Desroziers et al., 2005] Desroziers, G., Berre, L., Chapnik, B., and Poli, P. (2005). Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society*, 131(613):3385 – 3396.

[Desroziers and Ivanov, 2001] Desroziers, G. and Ivanov, S. (2001). Diagnosis and adaptive tuning of observation-error parameters in a variational assimilation. *Quarterly Journal of the Royal Meteorological Society*, 127(574):1433 – 1452.

[Dreano et al., 2017] Dreano, D., Tandeo, P., Pulido, M., Ait-El-Fquih, B., Chonavel, T., and Hoteit, I. (2017). 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.

[Evensen, 1994] Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5):10143–10162.

[Farchi and Bocquet, 2019] Farchi, A. and Bocquet, M. (2019). On the efficiency of covariance localisation of the ensemble Kalman filter using augmented ensembles. *Frontiers in Applied Mathematics and Statistics*, 5:3.

[Fisher, 2003] Fisher, M. (2003). Background error covariance modelling. In *Seminar on Recent developments in data assimilation for atmosphere and ocean (Shinfield Park, Reading, 8-12 September)*. ECMWF.

[Fisher et al., 2005] Fisher, M., Leutbecher, M., and Kelly, G. A. (2005). 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.

[Fitt et al., 2010] Fitt, A. D., Norbury, J., Ockendon, H., and Wilson, E. (2010). *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.

[Fortunato, 2010] Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, 486(3):75 – 174.

[Fowler, 2019] Fowler, A. (2019). Data compression in the presence of observational error correlations. *Tellus A: Dynamic Meteorology and Oceanography*, 71(1):1634937.

[Garand et al., 2007] Garand, L., Heilliette, S., and Buehner, M. (2007). Interchannel error correlation associated with airs radiance observations: Inference and impact in data assimilation. *Journal of Applied Meteorology and Climatology*, 46(6):714–725.

[Garavaglia et al., 2017] Garavaglia, F., Le Lay, M., Gottardi, F., Garçon, R., Gailhard, J., Paquet, E., and Mathevet, T. (2017). Impact of model structure on flow simulation and hydrological realism: from a lumped to a semi-distributed approach. *Hydrology and Earth System Sciences*, 21(8):3937–3952.

[Garçon, 1996] Garçon, R. (1996). Prévision opérationnelle des apports de la Durance à Serre-Ponçon à l'aide du modèle MORDOR. Bilan de l'année 1994-1995. *La Houille Blanche*, (5):71–76.

[Gaspari and Cohn, 1999] Gaspari, G. and Cohn, S. E. (1999). Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125(554):723–757.

[Gaspari and E. Cohn, 1999] Gaspari, G. and E. Cohn, S. (1999). Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125:723–757.

[Gauthier et al., 2018] Gauthier, P., Du, P., Heilliette, S., and Garand, L. (2018). Convergence Issues in the Estimation of Interchannel Correlated Observation Errors in Infrared Radiance Data. *Monthly Weather Review*, 146(10):3227–3239.

[Gerke, 2011] Gerke, M. (2011). Using horizontal and vertical building structure to constrain indirect sensor orientation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66:307–316.

[Goeury et al., 2017] Goeury, C., Ponçot, A., Argaud, J.-P., Zaoui, F., Ata, R., and Audouin, Y. (2017). 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*.

[Gong et al., 2020a] Gong, H., Yu, Y., and Li, Q. (2020a). Reactor power distribution detection and estimation via a stabilized gappy proper orthogonal decomposition method. *Nuclear Engineering and Design*, 370:110833.

[Gong et al., 2020b] Gong, H., Yu, Y., Li, Q., and Quan, C. (2020b). An inverse-distance-based fitting term for 3D-Var data assimilation in nuclear core simulation. *Annals of Nuclear Energy*, 141:107346.

[Greybush et al., 2011] Greybush, S. J., Kalnay, E., Miyoshi, T., Ide, K., and Hunt, B. R. (2011). Balance and ensemble Kalman filter localization techniques. *Monthly Weather Review*, 139(2):511–522.

[Gueuning et al., 2019] Gueuning, M., Cheng, S., Lambiotte, R., and Delvenne, J.-C. (2019). Rock–paper–scissors dynamics from random walks on temporal multiplex networks. *Journal of Complex Networks*, 8(2).

[Hamill et al., 2001] Hamill, T. M., Whitaker, J. S., and Snyder, C. (2001). Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Monthly Weather Review*, 129(11):2776–2790.

[Hastie et al., 2001] Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., New York, NY, USA.

[Hoffman et al., 2013] Hoffman, R., V. Ardizzone, J., Leidner, S., Smith, D., and Atlas, R. (2013). Error estimates for ocean surface winds: Applying desroziers diagnostics to the cross-calibrated, multiplatform analysis of wind speed. *Journal of Atmospheric and Oceanic Technology*, 30(11):2596–2603.

[Hollingsworth and Lönnberg, 1986] Hollingsworth, A. and Lönnberg, P. (1986). The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field. *Tellus A: Dynamic Meteorology and Oceanography*, 38(2):111–136.

[Hollingsworth and Lönnberg, 1989] Hollingsworth, A. and Lönnberg, P. (1989). The verification of objective analyses: Diagnostics of analysis system performance. *Meteorology and Atmospheric Physics*, 40:3–27.

[Houser et al., 2012] Houser, P., Lannoy, G., and Walker, J. (2012). *Hydrologic Data Assimilation*.

[Hunt et al., 2005] Hunt, B., Kostelich, E., and Szunyogh, I. (2005). Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter. *Physica D: Nonlinear Phenomena*, 230:112–126.

[Ishibashi, 2015] Ishibashi, T. (2015). Tensor formulation of ensemble-based background error covariance matrix factorization. *Monthly Weather Review*, 143(12):4963–4973.

[Jafarpour and Khodabakhshi, 2011] Jafarpour, B. and Khodabakhshi, M. (2011). A probability conditioning method (pcm) for nonlinear flow data integration into multipoint statistical facies simulation. *Mathematical Geoscience*, 43:133–164.

[Janjić et al., 2018] Janjić, T., Bormann, N., Bocquet, M., Carton, J. A., Cohn, S. E., Dance, S. L., Losa, S. N., Nichols, N. K., Potthast, R., Waller, J. A., and Weston, P. (2018). On the representation error in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 144(713):1257–1278.

[Jiang et al., 2020] Jiang, N., Studer, E., and Podvin, B. (2020). Physical modeling of simultaneous heat and mass transfer: species interdiffusion, soret effect and dufour effect. *International Journal of Heat and Mass Transfer*, 156:119758.

[Kalnay and Yang, 2010] Kalnay, E. and Yang, S.-C. (2010). Accelerating the spin-up of Ensemble Kalman Filtering. *Quarterly Journal of the Royal Meteorological Society*, 136(651):1644–1651.

[Ketchen and Shook, 1996] Ketchen, D. J. and Shook, C. L. (1996). The application of cluster analysis in strategic management research: an analysis and critique. *Strategic Management Journal*, 17(6):441–458.

[Kumar, 2018] Kumar, D. (2018). Ensemble-based assimilation of nonlinearly related dynamic data in reservoir models exhibiting non-gaussian characteristics. *Mathematical geosciences*, 51:75–107.

[Leisenring and Moradkhani, 2011] Leisenring, M. and Moradkhani, H. (2011). Snow water equivalent prediction using bayesian data assimilation methods. *Stochastic Environmental Research and Risk Assessment*, 25(2):253–270.

[Lerat, 2009] Lerat, J. (2009). *Quels apports hydrologiques pour les modèles hydrauliques? Vers un modèle intégré de simulation des crues*. PhD thesis, Université Pierre et Marie Curie.

[Li et al., 2016] Li, W., Sankarasubramanian, A., Ranjithan, R. S., and Sinha, T. (2016). 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.

[Liu and Xue, 2016] Liu, C. and Xue, M. (2016). Relationships among four-dimensional hybrid ensemble–variational data assimilation algorithms with full and approximate ensemble covariance localization. *Monthly Weather Review*, 144(2):591–606.

[Lucor and Le Maître, 2018] Lucor, D. and Le Maître, O. P. (2018). Cardiovascular modeling with adapted parametric inference. *ESAIM: ProcS*, 62:91–107.

[Melkumyan and Ramos, 2011] Melkumyan, A. and Ramos, F. (2011). Multi-kernel gaussian processes. pages 1408–1413.

[Mirouze, 2010] Mirouze, I. (2010). *Régularisation de problèmes inverse à l'aide de l'équation de diffusion, avec application à l'assimilation variationnelle de données océaniques*. PhD thesis, Université de Toulouse, France.

[Mirouze and Weaver, 2010] Mirouze, I. and Weaver, A. (2010). Representation of correlation functions in variational assimilation using an implicit diffusion operator. *Quarterly Journal of the Royal Meteorological Society*, 136(651):1421–1443.

[Ménard, 2016] Ménard, R. (2016). Error covariance estimation methods based on analysis residuals: theoretical foundation and convergence properties derived from simplified observation networks. *Quarterly Journal of the Royal Meteorological Society*, 142(694):257–273.

[Oliver and Webster, 2015] Oliver, M. and Webster, R. (2015). *Basic Steps in Geostatistics: The Variogram and Kriging*. Springer Briefs in Agriculture.

[Paquet, E., 2004] Paquet, E. (2004). Évolution du modèle hydrologique MORDOR : modélisation du stock nival à différentes altitudes. *La Houille Blanche*, (2):75–82.

[Parrish and Derber, 1992] Parrish, D. F. and Derber, J. C. (1992). The National Meteorological Center's spectral statistical-interpolation analysis system. *Monthly Weather Review*, 120(8):1747–1763.

[Parés et al., 2017] Parés, F., Garcia-Gasulla, D., Vilalta, A., Moreno, J., Ayguadé, E., Labarta, J., Cortés, U., and Suzumura, T. (2017). Fluid communities: A competitive, scalable and diverse community detection algorithm. In *Complex Networks & Their Applications VI - Proceedings of Complex Networks, Lyon, France, November 29 - December 1, 2017*, volume 689 of *Studies in Computational Intelligence*, pages 229–240. Springer.

[Pennec et al., 2006] Pennec, X., Fillard, P., and Ayache, N. (2006). A Riemannian framework for tensor computing. *International Journal of Computer Vision*, 66(1):41–66.

[Phillips et al., 2015] Phillips, J., Schwanghart, W., and Heckmann, T. (2015). Graph theory in the geosciences. *Earth-Science Reviews*, 143:147 – 160.

[Ponçot et al., 2013] Ponçot, A., Argaud, J.-P., Bouriquet, B., Erhard, P., Gratton, S., and Thual, O. (2013). Variational assimilation for xenon dynamical forecasts in neutronic using advanced background error covariance matrix. *Annals of Nuclear Energy*, 60:39–50.

[Qian et al., 2019] Qian, Y., Expert, P., Rieu, T., Panzarasa, P., and Barahona, M. (2019). Quantifying the alignment of graph and features in deep learning. *arXiv preprint arXiv:1905.12921*.

[R. Eyre and I. Hilton, 2013] R. Eyre, J. and I. Hilton, F. (2013). Sensitivity of analysis error covariance to the mis-specification of background error covariance. *Quarterly Journal of the Royal Meteorological Society*, 139(671):524–533.

[Rabier, 2005] Rabier, F. (2005). Overview of global data assimilation developments in numerical weather-prediction centres. *Quarterly Journal of the Royal Meteorological Society*, 131(613):3215–3233.

[Rochoux et al., 2018] Rochoux, M., Collin, A., Zhang, C., Trouvé, A., Lucor, D., and Moireau, P. (2018). Front shape similarity measure for shape-oriented sensitivity analysis and data assimilation for Eikonal equation. *ESAIM: ProcS*, 63:258–279.

[Rouhier, 2018] Rouhier, L. (2018). *Régionalisation d'un modèle hydrologique distribué pour la modélisation de bassins non jaugés. Application aux vallées de la Loire et de la Durance*. PhD thesis, Sorbonne Université.

[Rouhier et al., 2017] Rouhier, L., Le Lay, M., Garavaglia, F., Moine, N., Hendrickx, F., Monteil, C., and Ribstein, P. (2017). Impact of mesoscale spatial variability of climatic inputs and parameters on the hydrological response. *Journal of Hydrology*, 553:13 – 25.

[S. Hodges and J. Reich, 2010] S. Hodges, J. and J. Reich, B. (2010). Adding spatially-correlated errors can mess up the fixed effect you love. *The American Statistician*, 64:325–334.

[Saint-Venant, 1871] Saint-Venant, A. B. (1871). Théorie du mouvement non permanent des eaux, avec application aux crues des rivières et à l'introduction de marées dans leurs lits. *Comptes rendus de l'Académie des Sciences*, 73:147—-154 and 237—-240.

[Sandu and Chai, 2011] Sandu, A. and Chai, T. (2011). Chemical data assimilation—an overview. *Atmosphere*, 2(3):426—-463.

[Sandu and Cheng, 2015] Sandu, A. and Cheng, H. (2015). An error subspace perspective on data assimilation. *International Journal for Uncertainty Quantification*, 5:491–510.

[Schirber et al., 2013] Schirber, S., Klocke, D., Pincus, R., Quaas, J., and Anderson, J. L. (2013). Parameter estimation using data assimilation in an atmospheric general circulation model: From a perfect toward the real world. *Journal of Advances in Modeling Earth Systems*, 5(1):58–70.

[Sénégas et al., 2001] Sénégas, J., Wackernagel, H., Rosenthal, W., and Wolf, T. (2001). Error covariance modeling in sequential data assimilation. *Stochastic Environmental Research and Risk Assessment*, 15(1):65–86.

[Singh et al., 2011] Singh, K., Jardak, M., Sandu, A., Bowman, K., Lee, M., and Jones, D. (2011). Construction of non-diagonal background error covariance matrices for global chemical data assimilation. *Geoscientific Model Development*, 4(2):299–316.

[Sinsbeck and Tartakovsky, 2015] Sinsbeck, M. and Tartakovsky, D. (2015). Impact of data assimilation on cost-accuracy tradeoff in multifidelity models. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1):954–968.

[Sørensen and Madsen, 2004] Sørensen, J. V. T. and Madsen, H. (2004). Data assimilation in hydrodynamic modelling: on the treatment of non-linearity and bias. *Stochastic Environmental Research and Risk Assessment*, 18(4):228–244.

[Stein, 1999] Stein, M. L. (1999). *Interpolation of Spatial Data*. Springer.

[Stewart et al., 2013] Stewart, L. M., Dance, S. L., and Nichols, N. K. (2013). Data assimilation with correlated observation errors: experiments with a 1-D shallow water model. *Tellus A: Dynamic Meteorology and Oceanography*, 65(1):19546.

[T. Ihler et al., 2005] T. Ihler, A., Kirshner, S., Ghil, M., Robertson, A., and Smyth, P. (2005). Graphical models for statistical inference and data assimilation. *Physica D: Nonlinear Phenomena*, 230(1):72–87.

[Talagrand, 1998] Talagrand, O. (1998). A posteriori evaluation and verification of analysis and assimilation algorithms. In *Workshop on Diagnosis of Data Assimilation Systems*, pages 17–28, Shinfield Park, Reading.

[Tandeo et al., 2018] Tandeo, P., Ailliot, P., Bocquet, M., Carrassi, A., Miyoshi, T., Pulido, M., and Zhen, Y. (2018). A review of innovation-based methods to jointly estimate model and observation error covariance matrices in ensemble data assimilation. *arXiv preprint arXiv:1807.11221, accepted for submission to Monthly Weather Review*.

[Tibshirani et al., 2001] Tibshirani, R., Walther, G., and Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society Series B*, 63:411–423.

[van Leeuwen, 2019] van Leeuwen, P. J. (2019). Non-local observations and information transfer in data assimilation. *Frontiers in Applied Mathematics and Statistics*, 5:48.

[Vo and Durlofsky, 2014] Vo, H. and Durlofsky, L. (2014). A new differentiable parameterization based on principal component analysis for the low-dimensional representation of complex geological models. *Mathematical Geosciences*, 46:775–813.

[Waller et al., 2017] Waller, J. A., Dance, S. L., and Nichols, N. K. (2017). On diagnosing observation-error statistics with local ensemble data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 143(708):2677–2686.

[Waller et al., 2016] Waller, J. A., Simonin, D., Dance, S. L., Nichols, N. K., and Ballard, S. P. (2016). 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.

[Weaver and Courtier, 2001] Weaver, A. and Courtier, P. (2001). Correlation modelling on the sphere using a generalized diffusion equation. *Quarterly Journal of the Royal Meteorological Society*, 127(575):1815 – 1846.

[Weaver and Mirouze, 2013] Weaver, A. T. and Mirouze, I. (2013). 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.

[Weston et al., 2014] Weston, P. P., Bell, W., and Eyre, J. R. (2014). Accounting for correlated error in the assimilation of high-resolution sounder data. *Quarterly Journal of the Royal Meteorological Society*, 140(685):2420–2429.

[Wishart, 1928] Wishart, J. (1928). The generalised product moment distribution in samples from a normal multivariate population. *Biometrika*, 20A(1/2):32–52.

[Xu et al., 2017] Xu, M., Yuan, B., Wang, L., and Zhang, L. (2017). Data assimilation for Fukushima nuclear accident assessments. In *International Conference on Nuclear Engineering, Shanghai, China*, volume Volume 4: Nuclear Safety, Security, Non-Proliferation and Cyber Security; Risk Management, 57823.

[Zhu et al., 1997] Zhu, C., Byrd, R. H., and Nocedal, J. (1997). L-BFGS-B: Algorithm 778: L-BFGS-B, FORTRAN routines for large scale bound constrained optimization. *ACM Transactions on Mathematical Software*, 23(4):550–560.

université PARIS-SACLAY

ÉCOLE DOCTORALE
Sciences mécaniques et énergétiques, matériaux et géosciences (SMEMAG)

Titre : Spécification et localisation de covariance des erreurs en assimilation de données avec une application industrielle

Mots clés : assimilation de données, matrices de covariance d'erreurs, prévision dynamique, optimisation, incertitudes

Résumé : Les méthodes d'assimilation de données et plus particulièrement les méthodes variationnelles sont mises à profit dans le domaine industriel pour deux grands types d'applications que sont la reconstruction de champ physique et le recalage de paramètres. Une des difficultés de mise en œuvre des algorithmes d'assimilation est que la structure de matrices de covariance d'erreurs, surtout celle d'ébauche, n'est souvent pas ou mal connue. Dans cette thèse, on s'intéresse à la spécification et la localisation de matrices de covariance dans des systèmes multivariés et multidimensionels, et dans un cadre industriel. Dans un premier temps, on cherche à adapter/améliorer notre connaissance sur les covariances d'analyse à l'aide d'un processus itératif. Dans ce but nous avons développé deux nouvelles méthodes itératives pour la construction de matrices de covariance d'erreur d'ébauche. L'efficacité de ces méthodes est montrée numériquement en expériences jumelles avec des erreurs indépendantes ou relatives aux états vrais. On propose ensuite un nouveau concept de localisation pour le diagnostic et l'amélioration des covariances des erreurs. Au lieu de s'appuyer sur une distance spatiale, cette localisation est établie exclusivement à partir de liens entre les variables d'état et les observations. Finalement, on applique une combinaison de ces nouvelles approches et de méthodes plus classiques existantes, pour un modèle hydrologique multivarié développé à EDF. L'assimilation de données est mise en œuvre pour corriger la quantité de précipitation observée afin d'obtenir une meilleure prévision du débit d'une rivière en un point donné.

Title: Error covariance specification and localizationin data assimilation with industrial application

Keywords: data assimilation, error covariance matrix, dynamic forecast, optimization, uncertainty

Abstract: Data assimilation techniques are widely applied in industrial problems of field reconstruction or parameter identification. The error covariance matrices, especially the background matrix in data assimilation are often difficult to specify. In this thesis, we are interested in the specification and localization of covariance matrices in multivariate and multidimensional systems in an industrial context. We propose to improve the covariance specification by iterative processes. Hence, we developed two new iterative methods for background matrix recognition. The power of these methods is demonstrated numerically in twin experiments with independent errors or relative to true states. We then propose a new concept of localization and applied it for error covariance tuning. Instead of relying on spatial distance, this localization is established purely on links between state variables and observations. Finally, we apply these new approaches, together with other classical methods for comparison, to a multivariate hydrological model. Variational assimilation is implemented to correct the observed precipitation in order to obtain a better river flow forecast.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery