



**HAL**  
open science

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

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

► **To cite this version:**

Sibo Cheng, Jean-Philippe Argaud, Bertrand Iooss, Didier Lucor, Angélique Ponçot. Error covariance tuning in variational data assimilation: application to an operating hydrological model. Stochastic Environmental Research and Risk Assessment, 2021, 10.1007/s00477-020-01933-7 . hal-02992507

**HAL Id: hal-02992507**

**<https://hal.science/hal-02992507v1>**

Submitted on 6 Nov 2020

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

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

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

Sibo Cheng<sup>1,2</sup>, Jean-Philippe Argaud<sup>1</sup>, Bertrand Iooss<sup>1,3</sup>, Didier Lucor<sup>2</sup>, Angélique Ponçot<sup>1</sup>

<sup>1</sup> EDF R&D

<sup>2</sup> LIMSI, CNRS, Université Paris-Saclay

<sup>3</sup> Institut de Mathématiques de Toulouse, Université Paul Sabatier

**Keywords**— Data assimilation, Error covariance, Covariance modeling, Hydrology, Catchment modeling, Flow forecast, Precipitation-flow simulation

## 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.

## 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 2. We then recall some standard formulation of variational assimilation, following with the posterior covariance diagnosis in section 3. In section 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 and 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.

## 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.

### 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].

### 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. 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.

As an example, we show in Fig. 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

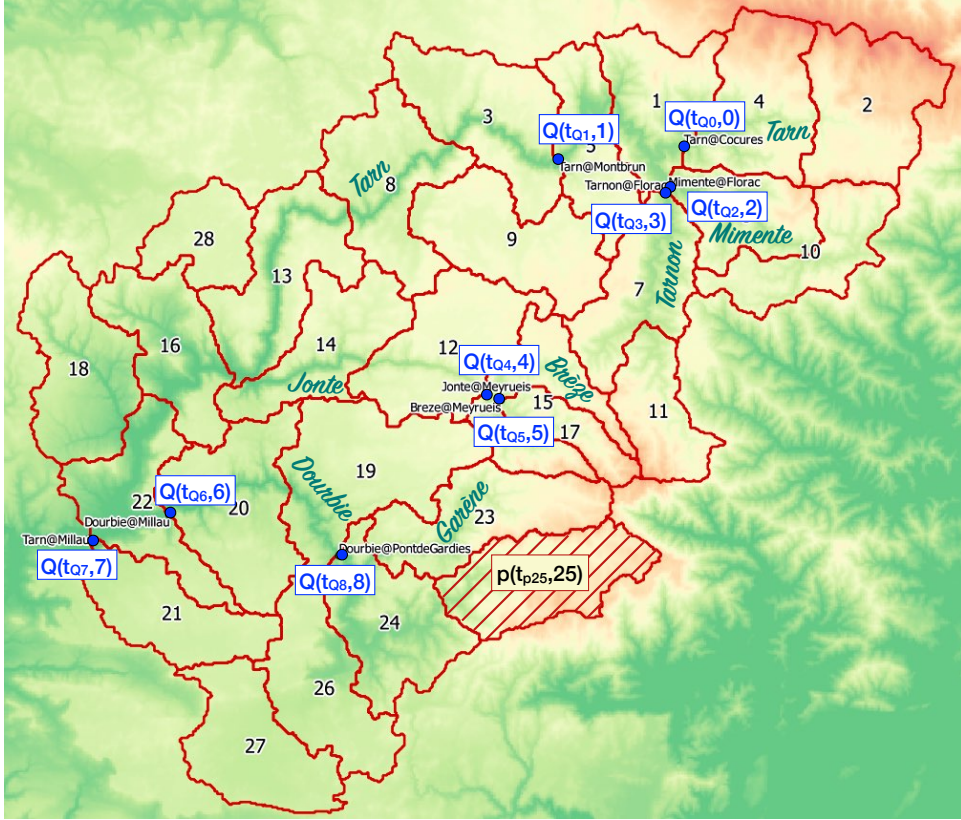


Figure 1: Spatial discretization of the Tarn basin with its 28 mesh cells for forcing data (rainfall and temperature) and the locations of the 9 observation gauges (blue dots). The area in darker green represents the elevation of the Tarn river and its affluents.

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. 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. 3), we compute the correlation between precipitation and flow observations where the latter is shifted 9 days ahead. As we observe from Fig. 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).

### 3 Data assimilation and covariance tuning

#### 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:

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

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

$$= J_b(\mathbf{x}) + J_o(\mathbf{x}), \quad (3)$$

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

$$\mathbf{B} = \text{Cov}(\epsilon_b, \epsilon_b) \quad (4)$$

$$\mathbf{R} = \text{Cov}(\epsilon_y, \epsilon_y), \quad (5)$$

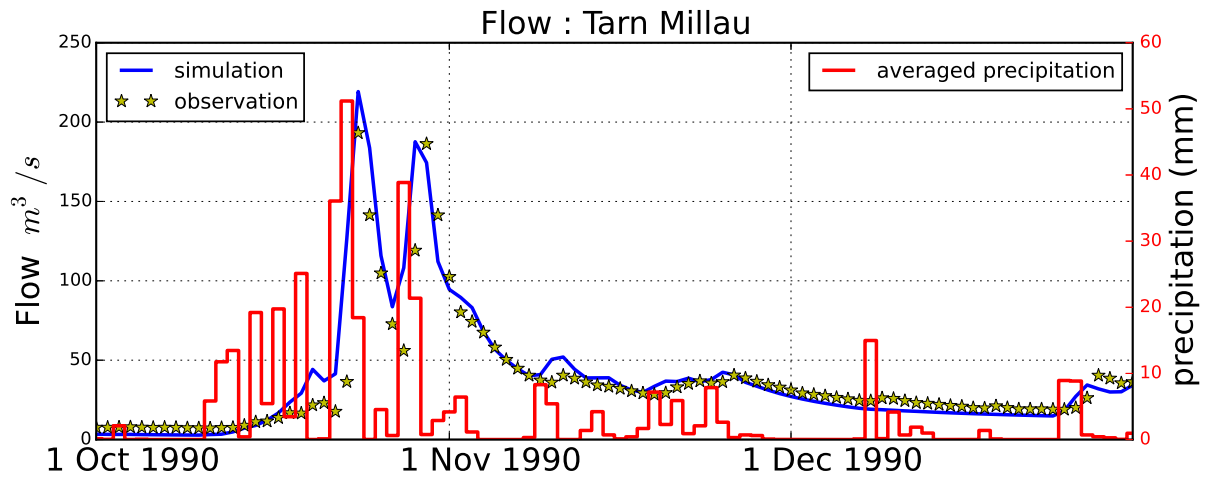


Figure 2: Example of simulation predicted by MORDOR-TS using daily precipitation, and observed Tarn discharges at Millau for three months in 1990. Simultaneous observed precipitations are in red bars (with the scale on the right vertical axis).

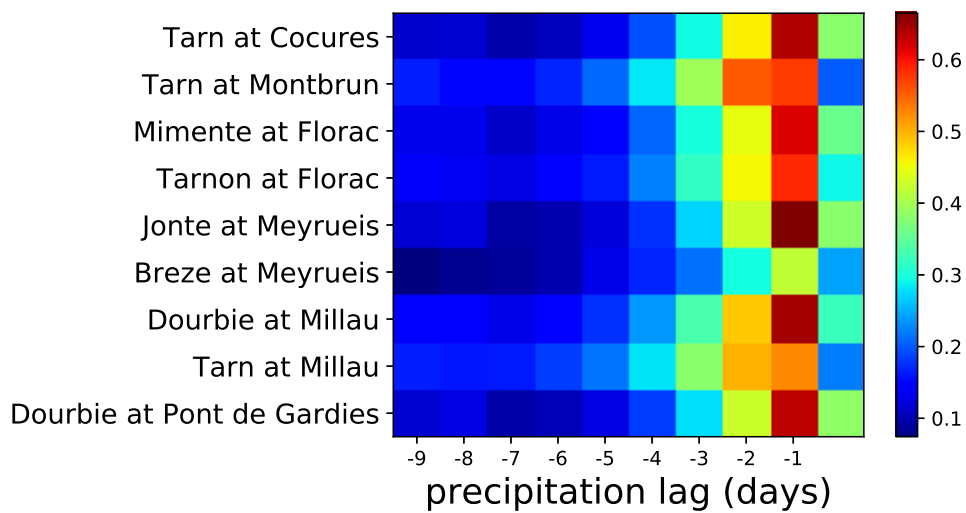


Figure 3: The lag correlation between daily precipitation and observed river flow at 9 observation gauges.

where

$$\epsilon_b = \mathbf{x}_b - \mathbf{x}_t \quad (6)$$

$$\epsilon_y = \mathcal{H}(\mathbf{x}_t) - \mathbf{y}, \quad (7)$$

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

$$\epsilon_b \sim \mathcal{N}(0, \mathbf{B}) \quad (8)$$

$$\epsilon_y \sim \mathcal{N}(0, \mathbf{R}). \quad (9)$$

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.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.

$$\mathbf{x}_a = \underset{\mathbf{x}}{\operatorname{argmin}}(J(\mathbf{x})). \quad (10)$$

The two covariance matrices  $\mathbf{B}$  and  $\mathbf{R}$  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.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]).

## 3.2 Covariances matrices diagnosis/tuning

The specification of error covariances (both  $\mathbf{B}$  and  $\mathbf{R}$ ) 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.

### 3.2.1 Desroziers & Ivanov covariance tuning (DI01)

First developed in the meteorological domain, the DI01 tuning algorithm [Desroziers and Ivanov, 2001] consists in adjusting the ratio between  $\mathbf{B}$  and  $\mathbf{R}$  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]), DI01 could be applied with a single real-time assimilation trajectory. More precisely, the iterative process could be synthesized as:

$$s_{b,n} = \frac{2\mathbb{E}[J_b(\mathbf{x}_a)]}{\operatorname{Tr}(\mathbf{K}_n \mathbf{H})}, \quad s_{o,n} = \frac{2\mathbb{E}[J_o(\mathbf{x}_a)]}{\operatorname{Tr}(\mathbf{I} - \mathbf{H}\mathbf{K}_n)},$$

$$\mathbf{B}_{n+1} = s_{b,n} \mathbf{B}_n, \quad \mathbf{R}_{n+1} = s_{o,n} \mathbf{R}_n, \quad (11)$$

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  $\operatorname{Tr}(\mathbf{B})/\operatorname{Tr}(\mathbf{R})$  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  $Tr(\mathbf{B}(s_b))$  and  $Tr(\mathbf{R}(s_o))$ . 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., 2020] has extended the application of 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  $\xi_t^p$  and  $\xi^{r,j}$ . The marginal error deviations are considered homogeneous inside each group of variables.

### 3.2.2 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:

$$\mathbb{E}\left([y - \mathcal{H}(\mathbf{x}_a)][y - \mathcal{H}(\mathbf{x}_b)]^T\right) = \mathbf{R}. \quad (12)$$

Hence, the difference between the left side and the right side of Eq.12, where  $\|\cdot\|_F$  denotes the Frobenius norm:

$$\|\mathbf{R} - \mathbb{E}([y - \mathbf{H}(\mathbf{x}_a)][y - \mathbf{H}(\mathbf{x}_b)]^T)\|_F \quad (13)$$

can be used as a validation indicator of  $\mathbf{R}$  matrix estimation. Expression in Eq.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]:

$$\mathbf{R}_{n+1} = \mathbb{E}\left([y - \mathcal{H}(\mathbf{x}_{a,n})][y - \mathcal{H}(\mathbf{x}_b)]^T\right). \quad (14)$$

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.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:

$$\mathbf{R}_n \leftarrow \frac{1}{2}(\mathbf{R}_n + \mathbf{R}_n^T). \quad (15)$$

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  $\mathbf{R}_{r,n}$ :

$$\mathbf{R}_{r,n} \leftarrow (1 - \mu)\mathbf{R}_n + \mu\mathbf{C}, \quad (16)$$

following Eq.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 8.3.1 of the present paper, we show that the matrix sequence built according to the regularization formula given in Eq.42, does not necessarily converge to the exact observation covariance matrix: a counter-example is given where a fixed point of Eq.14 other than the exact observation covariance is found.

### 3.2.3 CUTE and PUB iterative covariance tuning methods

Recently proposed iterative methods called CUTE (Covariance Updating iTerativeE) 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:

$$\mathbf{x}_{b,n+1} = \mathbf{x}_{a,n} = \underset{\mathbf{x}}{\operatorname{argmin}} \left( \frac{1}{2} \|\mathbf{x} - \mathbf{x}_{b,n}\|_{\mathbf{B}_n^{-1}}^2 + \frac{1}{2} \|\mathbf{y} - \mathcal{H}(\mathbf{x})\|_{\mathbf{R}^{-1}}^2 \right), \quad (17)$$

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:

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

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  $\alpha$  should be set. Then,

$$\mathbf{K}_n = \mathbf{B}_n \mathbf{H}^T (\mathbf{H} \mathbf{B}_n \mathbf{H}^T + \mathbf{R})^{-1}, \quad (18)$$

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:

$$\begin{aligned} \mathbf{Cov}(\epsilon_{b,n}, \epsilon_y) &= \mathbf{Cov}(\epsilon_y, \epsilon_{b,n})^T = \mathbf{Cov} \left( [(\mathbf{I} - \mathbf{K}_{n-1} \mathbf{H}) \epsilon_{b,n-1} + \mathbf{K}_{n-1} \epsilon_y], \epsilon_y \right) \\ &= (\mathbf{I} - \mathbf{K}_{n-1} \mathbf{H}) \mathbf{Cov}(\epsilon_{b,n-1}, \epsilon_y) + \mathbf{K}_{n-1} \mathbf{R}. \end{aligned} \quad (19)$$

## 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  $\mathbf{C}$ :

$$\mathbf{z} \equiv \begin{pmatrix} \mathbf{x}_b \\ \mathbf{y} \end{pmatrix}, \quad \tilde{\mathbf{H}} \equiv \begin{pmatrix} \mathbf{I} \\ \mathbf{H} \end{pmatrix}, \quad \mathbf{w} \equiv \begin{pmatrix} \epsilon_b \\ \epsilon_y \end{pmatrix}, \quad \mathbf{C} \equiv \begin{pmatrix} \mathbf{B} & \mathbf{C}_{o-b}^T \\ \mathbf{C}_{o-b} & \mathbf{R} \end{pmatrix} \quad (20)$$

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:

$$J(\mathbf{x}) = \frac{1}{2} \|\mathbf{z} - \tilde{\mathbf{H}} \mathbf{x}\|_{\mathbf{C}^{-1}}. \quad (21)$$

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.21 is completely equivalent to the 3D-Var formulation given in Eq.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 covariance are updated:

$$\mathbf{x}_{b,n+1} \leftarrow \mathbf{x}_{a,n} = \underset{\mathbf{x}}{\operatorname{argmin}} \left( \frac{1}{2} \|\mathbf{z}_n - \tilde{\mathbf{H}} \mathbf{x}\|_{\mathbf{C}_n^{-1}} \right), \quad (22)$$

$$\mathbf{z}_{n+1} = \begin{pmatrix} \mathbf{x}_{b,n+1} \\ \mathbf{y} \end{pmatrix}, \quad (23)$$

$$\mathbf{A}_n = (\tilde{\mathbf{H}}^T \mathbf{C}_n^{-1} \tilde{\mathbf{H}})^{-1}, \quad (24)$$

$$\mathbf{B}_{n+1} \leftarrow \frac{(1 - \alpha) \operatorname{Tr}(\mathbf{B}_n) + \alpha \operatorname{Tr}(\mathbf{A}_n)}{\operatorname{Tr}(\mathbf{A}_n)} \mathbf{A}_n, \quad (25)$$

$$\mathbf{C}_{o-b,n+1}^T = (\tilde{\mathbf{H}}^T \mathbf{C}_n^{-1} \tilde{\mathbf{H}})^{-1} \tilde{\mathbf{H}}^T \mathbf{C}_n^{-1} \begin{pmatrix} \mathbf{C}_{o-b,n}^T \\ \mathbf{R} \end{pmatrix}, \quad (26)$$

$$\mathbf{C}_{n+1} = \begin{pmatrix} \mathbf{B}_{n+1} & \mathbf{C}_{o-b,n+1}^T \\ \mathbf{C}_{o-b,n+1} & \mathbf{R} \end{pmatrix}. \quad (27)$$

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  $\mathbf{R}$  matrix is well known.

## 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. 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. 4 [b]). We introduce the following variables for each mesh cell index  $i$  with  $i = 1, \dots, 28$  and each day  $t$  with  $t = 1, \dots, T$ :

- $p_{i,t}$ : 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  $q^{\text{th}}$  outlet,
- $Q_{q,t}^o$ : the observed river discharge (in  $m^3/s$ ) at the  $q^{\text{th}}$  outlet.

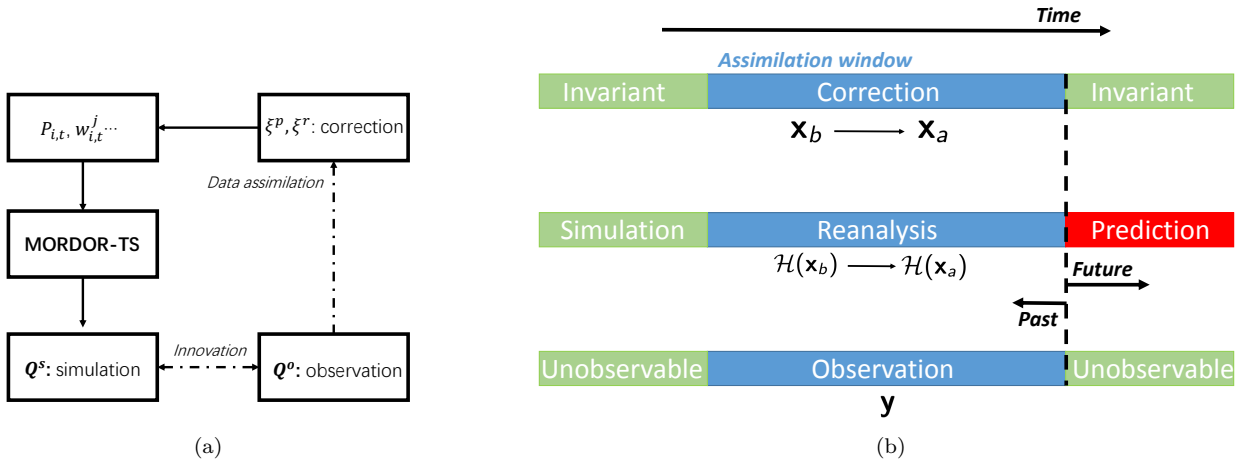


Figure 4: Fig. [a]: Flowchart of the MORDOR-TS hydrological model and DA articulation; Fig. [b]: DA modeling for a temporal assimilation window.

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:

- $\xi_t^p$ : 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. 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  $\xi_t^p$  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  $\mathbf{x} = [\xi_t^p, \xi^{r,j}]$  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. 1.

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

Table 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.:

$$\min_{(i=1\dots 28)}(p_{i,t}) + \xi_t^p \geq 0 \quad \text{for } \forall t = 1..T. \quad (28)$$

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.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 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. 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  $\mathbf{B}$  and  $\mathbf{R}$ , neither the correlation kernel nor the error amplitude, is originally available for this application. As shown in section 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.

### 5.1 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 Balgovid kernel (Matérn 3/2) or the Gaussian kernel (Matérn 5/2) (see [Cheng et al., 2019]). Among them, the Balgovid 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 Balgovid 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  $\phi(\cdot)$  can be written as:

$$\phi(r) = \left(1 + \frac{r}{L}\right) \exp\left(-\frac{r}{L}\right) \quad (29)$$

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 Balgovid kernel (if correlated).

### 5.2 Offline DI01

For this approach, the idea is to monitor the ratio between  $\xi_t^p$  (whose error standard deviations are respectively denoted as  $\sigma_{b,1}$  and  $\sigma_{b,2}$ ) and  $\xi^{r,j}$  introduced in section 4. To this end, we rely on the DI01 algorithm used as a pre-stage offline step, cf. Fig. 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 DI01 in subspaces has been studied in [Desroziers and Ivanov, 2001] for multivariate systems and in [Cheng et al., 2020] combined with unsupervised clustering techniques. In this paper, the objective of applying 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.:

$$\frac{\sigma_{b,1}}{\sigma_{b,2}} = \frac{\sigma_{b,1}}{\sigma_o} \times \frac{\sigma_o}{\sigma_{b,2}}. \quad (30)$$

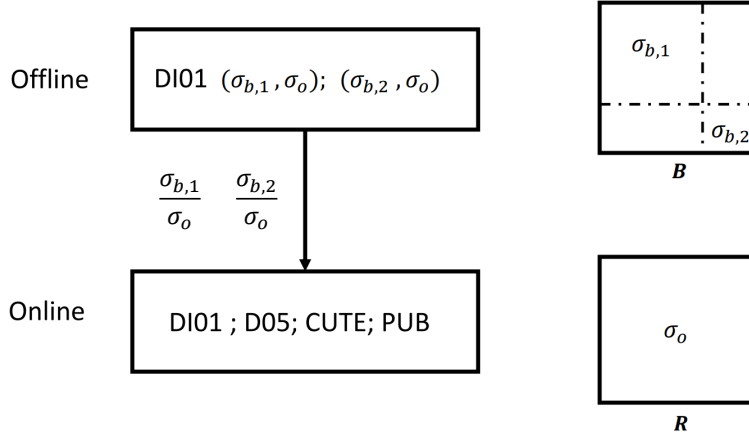


Figure 5: Diagram of the combination of offline and online covariance tuning methods.

More precisely, in order to gain a more robust inter-blocks ratio, the 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,

$$\mathbb{E}\left[\log\left(\frac{\sigma_{b,k}}{\sigma_o}\right)\right] = \frac{1}{2} \left( \mathbb{E}\left[\log\left(\prod_{n=1}^{n_{\max}} \frac{s_{b,n}^k}{s_{o,n}^k}\right)\right] \right) \quad (31)$$

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:

$$\mathbf{B} = \begin{pmatrix} \exp\left(\mathbb{E}\left[\log\left(\frac{\sigma_{b,1}}{\sigma_o}\right)\right]\right) \times \mathbf{B}_1 & 0_{30 \times 8} \\ 0_{8 \times 30} & \exp\left(\mathbb{E}\left[\log\left(\frac{\sigma_{b,2}}{\sigma_o}\right)\right]\right) \times \mathbf{B}_2 \end{pmatrix} \quad \text{and} \quad \mathbf{R} = \mathbf{I} \quad (32)$$

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  $\xi_t^p$  and the 8 parameters  $\xi^{r,j}$ . We impose a positive temporal correlation in  $\mathbf{B}_1$  by taking a Balgovid 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.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. 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 DI01 iterations in Fig. 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 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. 6 [e]. According to previous experiences, the error variance of  $\xi_t^p$  (resp.  $\xi^{r,j}$ ) is set to be  $10^3$  (resp.  $10^2$ ) higher than the one of observation. Finally, we obtain the averaged background-observation covariance ratio, suggested by DI01,

$$\mathbb{E}\left[\prod_{n=1}^{n_{\max}=5} \frac{s_{b,n}^1}{s_{o,n}^1}\right] = -2.24, \quad \mathbb{E}\left[\prod_{n=1}^{n_{\max}=15} \frac{s_{b,n}^2}{s_{o,n}^2}\right] = 1.56, \quad (33)$$

leading to an initial covariance set up,

$$\mathbb{E}\left[\log\left(\frac{\sigma_{b,1}}{\sigma_o}\right)\right] = \log(10^3) + \left(-\frac{2.24}{2}\right) = 1.88, \quad \mathbb{E}\left[\log\left(\frac{\sigma_{b,2}}{\sigma_o}\right)\right] = \log(10^2) + \left(-\frac{1.56}{2}\right) = 2.78 \quad (34)$$

### 5.3 D05 online estimation for the observation matrix

Relying on the error amplitudes adjusted offline in section 5.2, we apply the residual analysis of [Desroziers et al., 2005] for flow-independent observation covariance (of dimension  $270 \times 270$ ) 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  $\mathbf{R}$  matrix estimation through Eq.12, as well as the associated error correlation is shown in Fig. 7 [a,b,c,d]. The observation gauges, each of 30 days, follow the order of Fig. 3 [a] in the covariance matrices where the auto-covariance of *Tarn at Millau* is indicated in Fig. 7 [a] and [b]. Due to its geographical

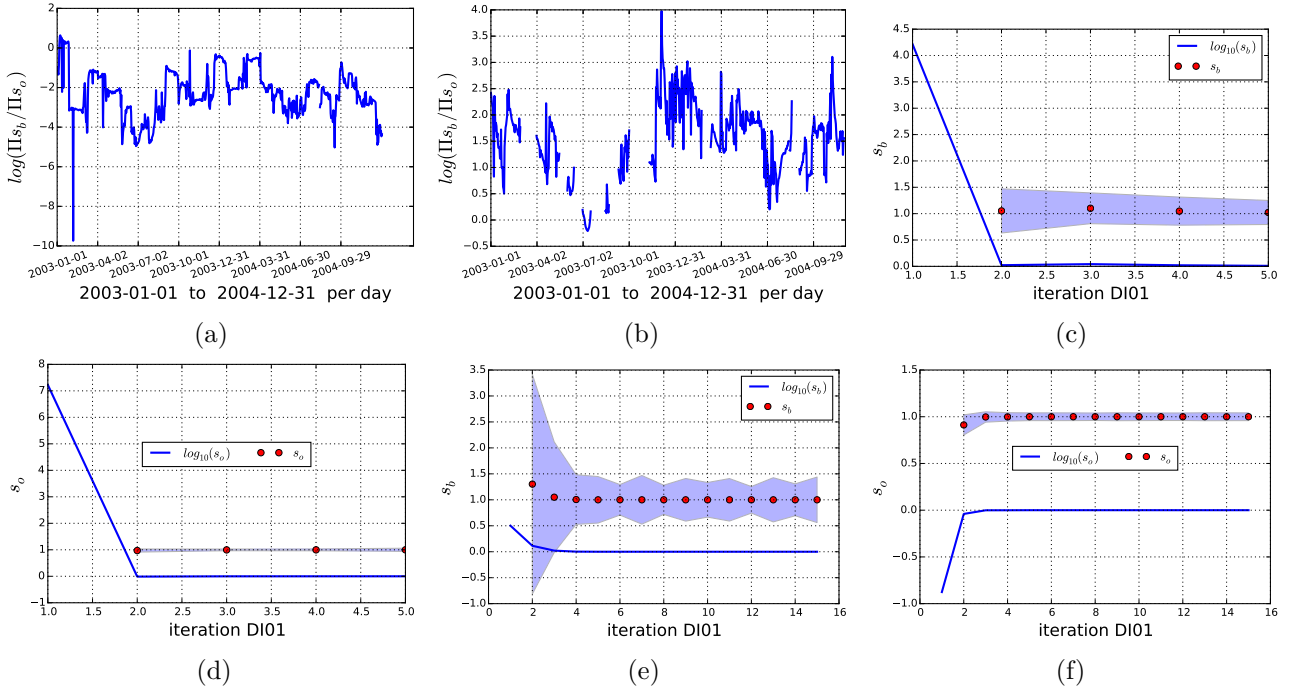


Figure 6: Fig. [a,b]: Evolution of  $\frac{s_{b,n}^k}{s_{o,n}^k}$  ([a]:  $k=1$ , [b]:  $k=2$ ); Fig. [c,d,e,f]: averaged evolution of respectively  $s_{b,n}^1, s_{o,n}^1, s_{b,n}^2, s_{o,n}^2$  against DI01 iterations  $n$  where the sky blue represents the variable deviation.

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  $\mathbf{R}$  matrix (Fig. 7 [b]) is slightly non-symmetric. We then regularize this matrix as shown in Eq.15 and Eq.16, with  $\alpha = 0.1$  and  $\mathbf{C} = \text{Tr}(\mathbf{R}) \times \mathbf{I}$  for a second iteration of D05 on the same data. The output  $\mathbf{R}$  matrix of the second iteration (Fig. 7 [f]) is very similar to the regularized matrix after the first D05 iteration (Fig. 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  $\mathbf{R}$  matrix is stable under the Desroziers diagnosis, allowing to consider it as the "true" observation matrix as mentioned in [Gauthier et al., 2018].

## 5.4 Online DI01

In section 5.2, DI01 is implemented in subspaces to provide a better initial guess of  $\mathbf{B}$  and  $\mathbf{R}$ . 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. 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.11, is estimated using one scalar realisation of  $J_b$  and  $J_o$ . The result of Eq.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 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 5.2 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, 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.

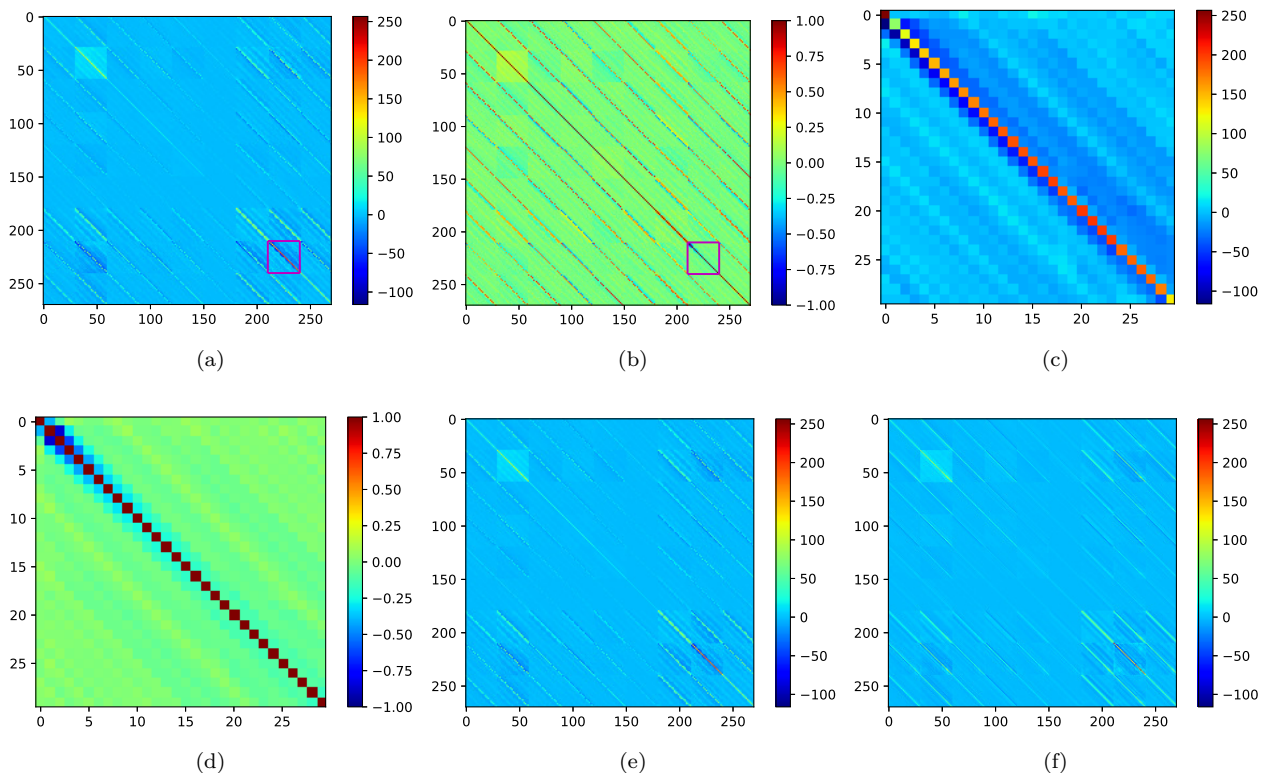


Figure 7: Estimated observation error covariance [a] and correlation [b] with *Tarn at Millau* [c,d] station closeups for the first D05 estimation. Fig. [e] shows the regularized matrix of 9 gauges while Fig. [f] represents the output of the second D05 iteration.

Online DA methods	reanalysis (30 days)		forecast (3 days)	
	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 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.

## 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.

### 6.1 Forecast improvement rate

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

$$\nabla = \frac{\mathbb{E}(\|\mathbf{y} - \mathbf{H}\mathbf{x}_b\|_2) - \mathbb{E}(\|\mathbf{y} - \mathbf{H}\mathbf{x}_a\|_2)}{\mathbb{E}(\|\mathbf{y} - \mathbf{H}\mathbf{x}_b\|_2)}. \quad (35)$$

where  $\|\cdot\|_2$  stands for the euclidean vector norm. The  $\mathbf{x}_b$ ,  $\mathbf{x}_a$  and  $\mathbf{y}$  in Eq.35 is taken in a small chosen window just after the assimilation window as shown in Fig. 4 [b]. We concentrate on the short-range forecast of 3 days for all 9 gauges, regarding Fig. 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. 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.

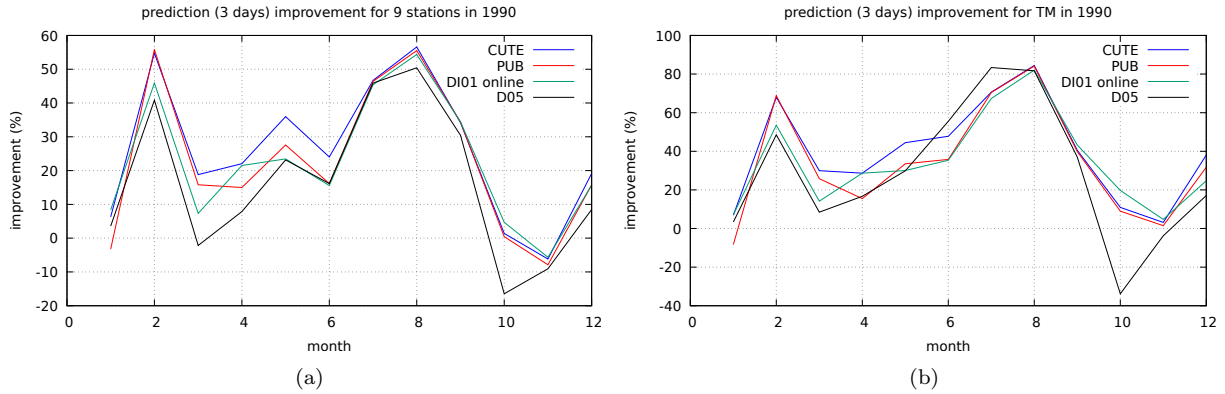


Figure 8: Averaged improvement rate per month in 1990 in all 9 stations [a] and at *Tarn at Millau* [b].

The averaged improvement rate per month is shown in Fig. 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.  $1^{st}, 2^{nd} \dots$  until  $31^{th}$  January. From Fig. 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. 8 are probably related to a very particular meteorological episode in the south of France (Cévenol episode) that can be seen in Fig. 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. 8 provide a significant enhance (around 30% to 40% in average for most months) compared to the model (background) flow forecast.

## 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. 9 and Fig. 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. 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. 9 and 5, 7 in Fig. 10.

## 6.3 Some examples

Several examples of 30-day assimilation windows are displayed in Fig. 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. 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. 11 [e]) to - the onset of a flood event (Fig. 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. 11 [a]) or an underacting (e.g. around September  $27^{th}$  in Fig. 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 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. 11 [a] and Fig. 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. 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. 8. Fig. 11 [b] and Fig. 11 [f] show two DA windows where the observed river flow are somehow more

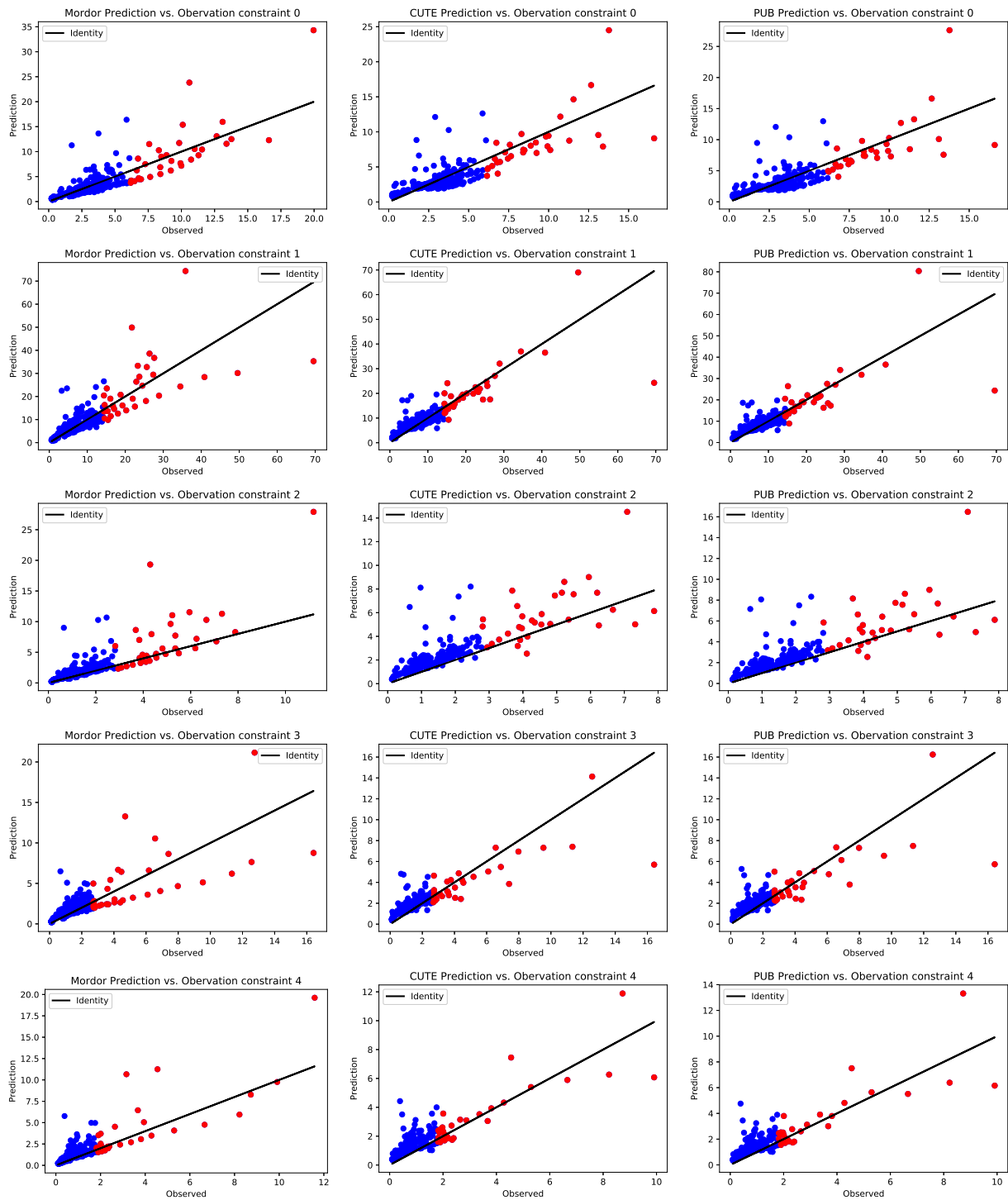


Figure 9: Forecast: 1st day after the assimilation window for observation stations in 1990 for constraints (gauges) 0 to 4, following the same order as Fig. 3.



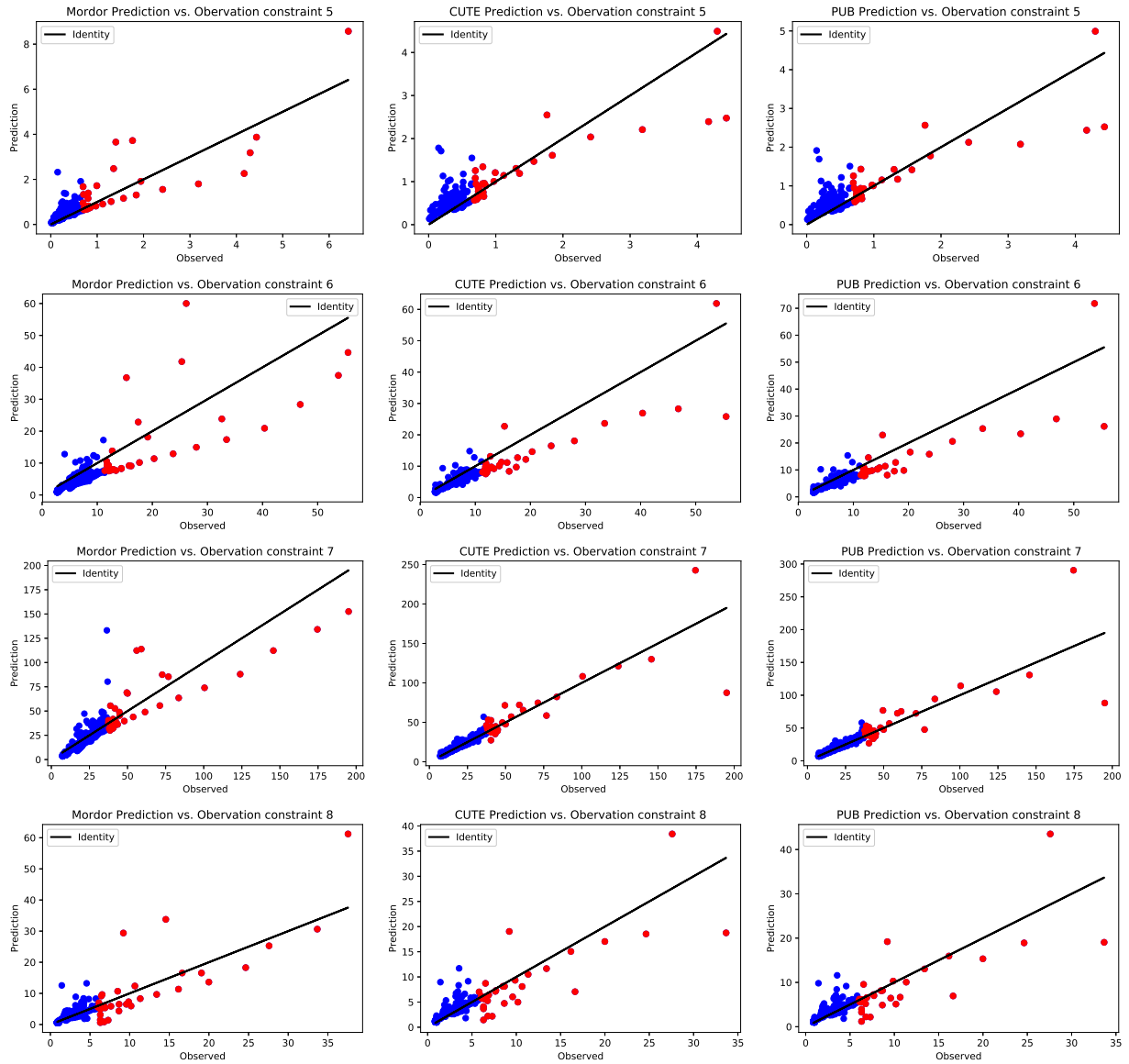


Figure 10: Same description as Fig. 9 for constraints 5 to 8.

chaotic, compared to Fig. 11 [a], Fig. 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.

## 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 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].

## 8 Appendix: Convergence of D05 iterative method

### 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.14 is error-free and the current iterative matrix  $\mathbf{R}_n$  stays always invertible.

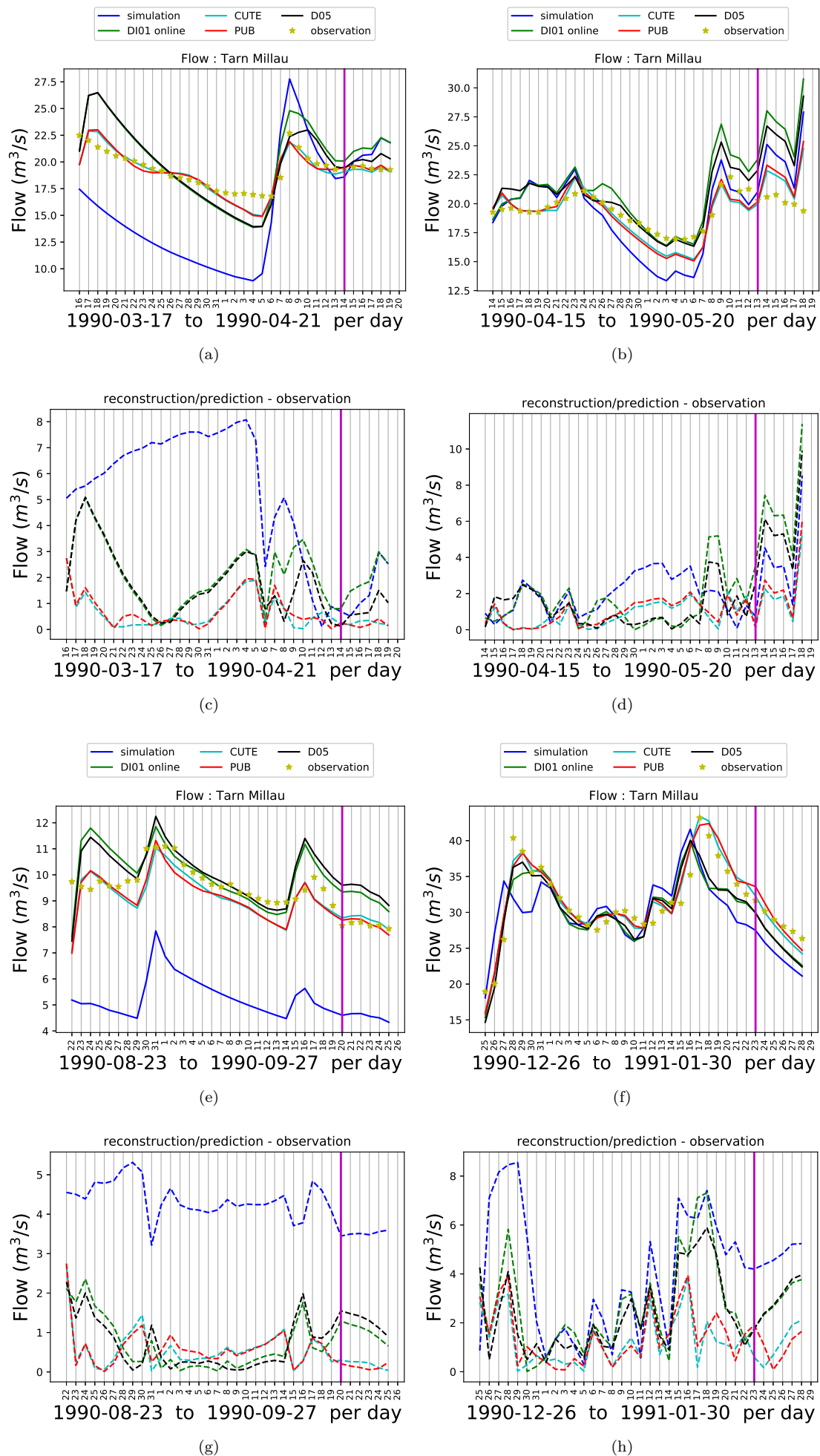


Figure 11: Examples of reanalysis and forecast at *Tarn at Millau* ([a,b,e,f]) in 1990 where the reanalysis and the forecast are separated by the vertical line. Sub-figures [c,d,g,h] represent the difference between reconstructed/predicted flow and the observation.

Following the notation of [Bathmann, 2018], let

$$\mathbf{G} = \mathbf{H}\mathbf{B}\mathbf{H}^T \quad (36)$$

$$\mathbf{D} = \mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}_E, \quad (37)$$

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

$$\mathbf{D} - \mathbf{G} = \mathbf{R}_E. \quad (38)$$

According to [Bathmann, 2018], the updating formulation of Eq.14 is equivalent to:

$$\mathbf{R}_{n+1} = \mathbf{R}_n(\mathbf{G} + \mathbf{R}_n)^{-1}\mathbf{D}. \quad (39)$$

where  $n$  is the current iteration. It is obvious that the exact observation matrix  $\mathbf{R}_E$  is a fixed point of Eq.39. In fact, when  $\mathbf{R}_E$  is SPD, the iterative process of Eq.39 converges necessarily to the exact covariance  $\mathbf{R}_E$ . 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.39, let  $\mathbf{M} = \mathbf{D}^{-1}\mathbf{G}$ , then

$$\mathbf{R}_n^{-1} = \mathbf{R}_E^{-1} + \mathbf{M}^n[\mathbf{R}_0^{-1} - \mathbf{R}_E^{-1}] \quad (40)$$

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

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

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

$$\mathbf{R}_n^{-1} = \mathbf{R}_E^{-1} + \mathbf{M}^n[\mathbf{R}_0^{-1} - \mathbf{R}_E^{-1}] \quad (41)$$

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

## 8.2 Necessary regularization

As pointed out by [Bathmann, 2018], the Eq.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.39 to:

$$\mathbf{R}_{n+1} = \frac{1}{2}(\mathbf{R}_n + \mathbf{R}_n^T) \left( \mathbf{G} + \frac{1}{2}(\mathbf{R}_n + \mathbf{R}_n^T) \right)^{-1} \mathbf{D}. \quad (42)$$

Being discussed in [Bathmann, 2018], the study of the convergence of Eq.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.

## 8.3 Limitations of Desroziers method

### 8.3.1 Non-convergence of regularized matrix sequence

We found that unlike Eq.39, the regularized sequence of Eq.42 could have another fixed, different from the true observation covariance (i.e.  $\mathbf{R}_E = \mathbf{D} - \mathbf{G}$ ). The proof is given by a counter-example:

$$\mathbf{G} = \begin{bmatrix} 1.5 & 1 \\ 1 & 4 \end{bmatrix}, \mathbf{D} = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}, \mathbf{R} = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}, \quad (43)$$

which satisfies

$$\mathbf{R} = \frac{1}{2}(\mathbf{R} + \mathbf{R}^T) \left( \mathbf{G} + \frac{1}{2}(\mathbf{R} + \mathbf{R}^T) \right)^{-1} \mathbf{D}, \quad (44)$$

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  $\mathbf{R}_E$ .

### 8.3.2 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 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:

$$\mathbf{x}_b = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{B} = \mathbf{I}_{2,2}, \mathbf{H} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \mathbf{R} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \quad (45)$$

$$\text{let } \mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}, \text{ thus } \mathcal{J}(x) = \mathbf{x}_1^2 - 2\mathbf{x}_1\mathbf{x}_2. \quad (46)$$

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

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}_b) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \quad (47)$$

Therefore, because of this incoherence emerged by the non-symmetry of  $\mathbf{R}_n$ , although the mathematical proof in section 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].

## References

- [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., 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.
- [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.
- [Bouttier and Courtier, 2002] Bouttier, F. and Courtier, P. (2002). Data assimilation concepts and methods. In *Meteorological Training Course Lecture Series*. ECMWF.
- [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.
- [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/>.
- [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.
- [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., 2020] 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, preprint.
- [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.
- [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.
- [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.
- [Fowler, 2019] Fowler, A. (2019). Data compression in the presence of observational error correlations. *Tellus A: Dynamic Meteorology and Oceanography*, 71(1):1634937.
- [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.
- [Gauthier et al., 2018] Gauthier, P., Du, P., Heillette, 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.
- [Goery et al., 2017] Goery, 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.
- [Houser et al., 2012] Houser, P., Lannoy, G., and Walker, J. (2012). *Hydrologic Data Assimilation*.
- [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.
- [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.
- [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*, 136:1421 – 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.

- [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.
- [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.
- [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.
- [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.
- [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.
- [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*.
- [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.