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

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

▶ **To cite this version:**

Sibo Cheng, Jean-Philippe Argaud, Bertrand Iooss, Angélique Ponçot, Didier Lucor. A graph clustering approach to localization for adaptive covariance tuning in data assimilation based on state-observation mapping. Mathematical Geosciences, Springer Verlag, 2021. meteo-02460851v2

HAL Id: meteo-02460851

https://hal-meteofrance.archives-ouvertes.fr/meteo-02460851v2

Submitted on 4 Jun 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.

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

Sibo Cheng[1,2], Jean-Philippe Argaud[1], Bertrand Iooss[1,3],
Angélique Ponçot[1], Didier Lucor[2]

[1] EDF R&D

[2] LIMSI, CNRS, Université Paris-Saclay

[3] Institut de Mathématiques de Toulouse, Université Paul Sabatier

June 4, 2020

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 results in more accurate both observation- and background-error parameters tuning.

Keywords— Data assimilation, Covariance matrices, Graph community detection, Unsupervised learning

1 Introduction

Data assimilation (DA) techniques, originally developed for numerical weather prediction (NWP), have been widely applied in the field of geosciences ([1], [4]) for instance for field reconstruction or parameter identification. The applications, often complex, multi-physics and nonlinear with different model resolutions and prediction time horizons, vary from reservoir modelling ([26]), to geological feature prediction ([37]), to operational oceanography. The essential idea of DA is to improve the numerical prediction by combining the information embedded in a prior estimation (also known as the background state) and real time observations. Unfortunately, the gigantic size (often $\mathcal{O}(10^{6-9})$ for multi-dimensional problems) of geosciences DA problems makes a full Bayesian approach computationally unaffordable. Instead, a variational approach weights these two information sources, thanks to background and observation error covariances, $\mathbf{B}$ and $\mathbf{R}$ respectively.

These prior covariance matrices can be estimated from the help of a correlation kernel (e.g. [33], [20]) or a diffusion operator (e.g. [27]). The computation of covariances may also be performed/improved by ensemble methods ([8]), or some iterative methods for which a part of $\mathbf{B}$ and/or $\mathbf{R}$ is supposed to be perfectly known e.g. [13], [12], [7]. These approaches quite often rely on converged state ensemble statistics, noiseless dynamical system or assumption of error amplitude ([7]). These conditions are usually difficult to be satisfied for high-dimensional geophysical systems.

An important ingredient used to make DA 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 ([17]) or problems of parameter identification ([32]), 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 [36].

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 ([18]), 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) ([15]) where the covariance localization is crucial to produce more accurate analyses.

The second class of families (domain localization) is explicit and performs DA 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. [38], 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 [38], difficulties appear with the domain localization when the region of influence is far offset from the domain of dependence. In fact, the former 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 [38]. Empirical choice of cut-off or distance thresholds may result in removal of true physical long-range correlations, thus inducing imbalance in the analysis [21].

This conclusion points to the relevance of more efficient and less arbitrary segregation operators. In practice, DA 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 [22], section 14). Fig. 1 illustrates with a very simple schematic the class of problems which could benefit from such an approach. It depicts 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 DA 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 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

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

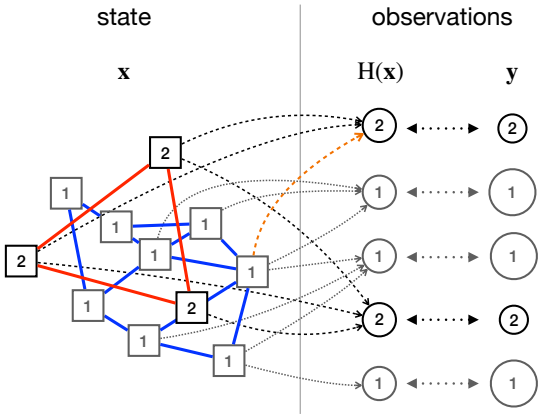


Figure 1: Simple sketch illustrating the type of relations between state variables and observations considered for data assimilation in this paper. The observation operator $\mathcal{H}$ maps some state variables $\mathbf{x}$ to the space of observations so that they can be compared with the experimental measurements $\mathbf{y}$. A graph clustering approach is put to use as a *localizer* to reveal unknown state variable/observation communities.

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 (see the overview paper of [29]).

In a more general framework, graphical models are used in DA 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 ([24]). In this circumstance, graphical models could be considered as a variable dependency based localization methods. Another advantage of graphical representations, as pointed out by [24], 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 efficient 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 effort has been made on posterior diagnosing and iterative adjustment of error covariance quantification, especially by the meteorology community. (e.g. [13], [12]). Among these tuning methods, the one of Desroziers and Ivanov (also known as **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 DA 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 [12]. The later is based on statistics of prior and posterior innovation quantities. In fact, the deployment of **DI01** in subspaces has already been introduced in [6] for block diagonal structures of $\mathbf{B}$ and $\mathbf{R}$. In this paper, we adopt a **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 2 We then explain how this Jacobian matrix, considered as a state-observation mapping, can be used to build an observation-based state network where the subspaces decomposition is carried out by applying graph-based community detection algorithms. The localized version of **DI01** is then introduced (section 4) and investigated in a twin experiments framework (section 5). We close the paper with a discussion (section 6).

2 Data assimilation framework

The goal of DA 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 while $\mathbf{y}_j$ is called an observation. The principle of DA 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:

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

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.

$$\mathbf{B} = \mathbf{cov}(\epsilon_b, \epsilon_b), \qquad (3)$$
$$\mathbf{R} = \mathbf{cov}(\epsilon_y, \epsilon_y), \qquad (4)$$

where

$$\epsilon_b = \mathbf{x}_b - \mathbf{x}_t, \qquad (5)$$
$$\epsilon_y = \mathcal{H}(\mathbf{x}_t) - \mathbf{y}. \qquad (6)$$

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 data assimilation, thus they can be perfectly characterized by the covariance matrices, i.e.

$$\epsilon_b \sim \mathcal{N}(0, \mathbf{B}), \qquad (7)$$
$$\epsilon_y \sim \mathcal{N}(0, \mathbf{R}). \qquad (8)$$

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 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 (1) can be perfectly solved by linear formulation of BLUE:

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}_b) \qquad (9)$$

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

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}. \qquad (10)$$

Several diagnosis or tuning methods, such as the ones of [12], [13], [14] 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. [38], [31]). 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.

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 [5]). 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.

3.1 State space decomposition via graph clustering algorithms

3.1.1 Principles

Here, 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 [39], 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 DA 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 $\mathbf{H}$ 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:

$$\mathcal{D}_{\text{dependence}}(\mathbf{y}_i) = \{\mathbf{x}_k : \mathbf{H}_{i,k} \neq 0\}, \tag{11}$$

is introduced to quantify the range of this state-observation connection which is purely decided by $\mathbf{H}$ instead of the spatial distance. [39] 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 [36] where the author proposes an extra step to assimilate observations outside the region of influence.

3.1.2 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 Jacobian $\mathbf{H}$. 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. 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.

$$\exists k, \quad \text{such that} \quad \frac{\partial \mathbf{y}_k}{\partial \mathbf{x}_i} \neq 0, \qquad \frac{\partial \mathbf{y}_k}{\partial \mathbf{x}_j} \neq 0. \tag{12}$$

For a linearized state-observation operator $\mathbf{H}$, it is simply equivalent to

$$\exists k, \quad \text{such that} \quad \mathbf{H}_{k,i} \neq 0, \qquad \mathbf{H}_{k,j} \neq 0. \tag{13}$$

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. In fact, $\mathbf{H}$ can be seen as some form of weighted mapping between the space of state variables and the one of observations, which allows us to define the connection strength, quantified by a function $\mathcal{S}$ as a sum of conjugated coefficients multiplication in $\mathbf{H}$, i.e,

$$\begin{aligned} \mathcal{S} \quad &: \quad \mathbb{R}^{n_{\mathbf{x}}} \times \mathbb{R}^{n_{\mathbf{x}}} \mapsto \mathbb{R}^{+} \\ &\mathcal{S}(\mathbf{x}_i, \mathbf{x}_j) \equiv \mathcal{S}_{i,j} = \sum_{k, i \neq j} |\mathbf{H}|_{k,i} |\mathbf{H}|_{k,j}, \end{aligned} \tag{14}$$

where $|\cdot|$ represents the absolute value function on the whole matrix, so the function is symmetric. 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.

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 $\mathbf{A}_{\mathcal{G}_{\mathcal{S}}} = (a^{\mathcal{G}_{\mathcal{S}}}_{\mathbf{x}_i, \mathbf{x}_j})$:

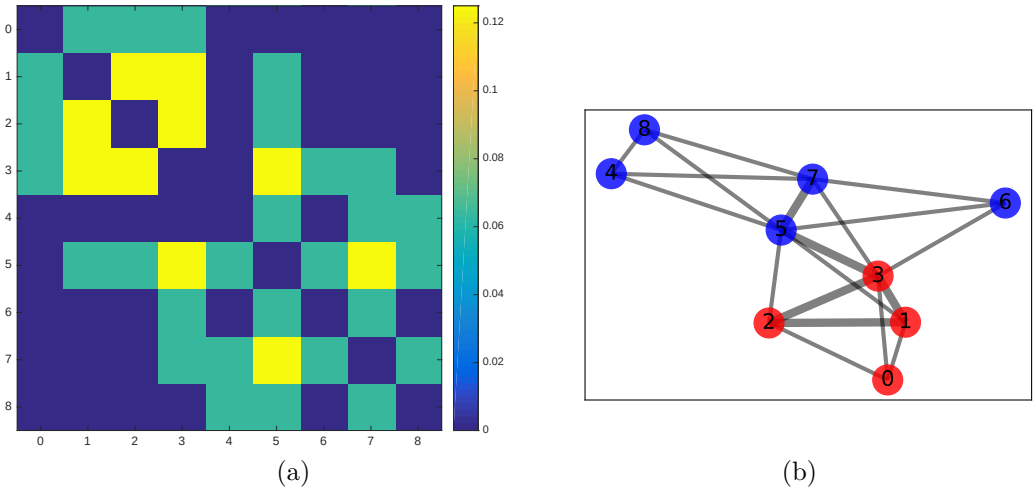
$$a^{\mathcal{G}_{\mathcal{S}}}_{\mathbf{x}_i, \mathbf{x}_j} = \begin{cases} \mathcal{S}_{i,j} \text{ if } \{\mathbf{x}_i, \mathbf{x}_j\} \in \mathbf{E}, \\ 0 \quad \text{otherwise.} \end{cases} \tag{15}$$

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 $\mathcal{G}_{\mathcal{S}}\left[\mathbf{x}^i\right] := \left(\mathbf{x}^i, \mathbf{E}(\mathbf{x}^i), \mathcal{S}_{|\mathbf{E}(\mathbf{x}^i)}\right)$, where $\mathbf{E}(\mathbf{x}^i) := \left\{\{\mathbf{x}_k, \mathbf{x}_l\} \in \mathbf{E} : \mathbf{x}_k, \mathbf{x}_l \in \mathbf{x}^i\right\}$. 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 $\mathbf{x}^j$. 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 ([16]).

3.1.3 Clustering algorithms

One of the main paradigms of clustering is to find groups/clusters intra-cluster density vs. inter-cluster sparsity. 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 ([2]) and the Fluid community algorithm ([28]). These methods are mostly based on random walks 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 ([11]) 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 [28] 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 [3].



(a) (b)

Figure 2: (a) Observation-based state adjacency matrix obtained from the transformation operator $\mathbf{H}$ in Eq. (17). (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.

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 ([25]) or the Gap statistic method ([35]). For instance the following performance metric will be used later for the experiments in section 5.2:

$$\text{performance} := \frac{m(\mathcal{C}) + \bar{m}^c(\mathcal{C})}{\frac{1}{2} n_{\mathbf{x}}(n_{\mathbf{x}} - 1)}. \tag{16}$$

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.

3.1.4 A simple example of state-observation graph clustering

For illustration purpose, inspired by the pedagogical approach of [38], we consider the following simple system with $\mathbf{x} \in \mathbb{R}^{n_{\mathbf{x}}=9}$ and $\mathbf{y} \in \mathbb{R}^{n_{\mathbf{y}}=4}$:

$$\mathbf{H} = 0.25 \times \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \end{bmatrix}, \tag{17}$$

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

$$\begin{array}{rcl} 0.25(\mathbf{x}_0 + \mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3) & = & \mathbf{y}_0 \\ 0.25(\mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3 + \mathbf{x}_5) & = & \mathbf{y}_1 \\ 0.25(\mathbf{x}_3 + \mathbf{x}_5 + \mathbf{x}_6 + \mathbf{x}_7) & = & \mathbf{y}_2 \\ 0.25(\mathbf{x}_4 + \mathbf{x}_5 + \mathbf{x}_7 + \mathbf{x}_8) & = & \mathbf{y}_3. \end{array} \tag{18}$$

The obtained observation-based adjacency matrix is represented in Fig. 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. 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 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.

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.

3.2.1 Observation reduction

Applying this strategy, the observation components $\mathbf{y}_k$ with connections to several state variable clusters must be identified and cancelled, i.e. all observations such that,

$$\frac{\partial \mathbf{y}_{k=0,\ldots,n_{\mathbf{y}}-1}}{\partial \mathbf{x}^i_{l=1,\ldots,n^i_{\mathbf{x}},i=1,\ldots,p}} \neq 0, \tag{19}$$

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 DA procedure. Back to Eq. (18), 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:

$$\begin{array}{rcl} \mathbf{x}^1 & = & \{\mathbf{x}_{k=0,\ldots,3}\}, \quad \tilde{\mathbf{y}}^1 = \{\mathbf{y}_0\}, \\ \mathbf{x}^2 & = & \{\mathbf{x}_{k=4,\ldots,8}\}, \quad \tilde{\mathbf{y}}^2 = \{\mathbf{y}_3\}. \end{array} \tag{20}$$

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

$$\tilde{\mathbf{H}} = 0.25 \times \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \end{bmatrix}. \tag{21}$$

3.2.2 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:

$$\hat{\mathbf{y}}_l = \mathbf{y}_l - \sum_{i=1,\ldots,p, i\neq j} \sum_{k \mid \mathbf{x}_k \in \mathbf{x}^i} \mathbf{H}_{l,k} \mathbb{E}_b[\mathbf{x}_k], \tag{22}$$

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 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. (18) can be adjusted to:

$$\begin{array}{l} 0.25(\mathbf{x}_0+\mathbf{x}_1+\mathbf{x}_2+\mathbf{x}_3)=\hat{\mathbf{y}}_0=\mathbf{y}_0 \\ 0.25(\mathbf{x}_1+\mathbf{x}_2+\mathbf{x}_3)=\hat{\mathbf{y}}_1=\mathbf{y}_1-0.25\mathbb{E}_b[\mathbf{x}_5] \\ 0.25(\mathbf{x}_5+\mathbf{x}_6+\mathbf{x}_7)=\hat{\mathbf{y}}_2=\mathbf{y}_2-0.25\,\mathbb{E}_b[\mathbf{x}_3] \\ 0.25(\mathbf{x}_4+\mathbf{x}_5+\mathbf{x}_7+\mathbf{x}_8)=\hat{\mathbf{y}}_3=\mathbf{y}_3. \end{array} \tag{23}$$

Thus the new operator can be written as:

$$\hat{\mathbf{H}}_{\text{adjustment}}=0.25\times\begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \end{bmatrix}. \tag{24}$$

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

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.

4.1 Desroziers & Ivanov diagnosis and tuning approach

The [13] 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. [23]) and meteorology. Consecutive works have been carried out to improve its performance and feasibility in problems of large dimension such as the study of [6]. 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 [34] and [13] 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:

$$\begin{aligned} \mathbb{E}\left[J_b(\mathbf{x}_a)\right] &= \frac{1}{2}\mathbb{E}\left[(\mathbf{x}_a-\mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x}_a-\mathbf{x}_b)\right] \\ &= \frac{1}{2}\mathrm{Tr}(\mathbf{KH}), \end{aligned} \tag{25}$$

$$\begin{aligned} \mathbb{E}\left[J_o(\mathbf{x}_a)\right] &= \frac{1}{2}\mathbb{E}\left[(\mathbf{y}-\mathbf{H}\mathbf{x}_b)^T\mathbf{R}^{-1}(\mathbf{y}-\mathbf{H}\mathbf{x}_b)\right] \\ &= \frac{1}{2}\mathrm{Tr}(\mathcal{I}-\mathbf{HK}), \end{aligned} \tag{26}$$

where $\mathbf{x}_a$ is the output of a 3D-VAR algorithm with a linear observation operator $\mathbf{H}$. In practice, this is seldomly the case and one has to deal with imperfect knowledge of the covariance matrices. 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

$$s_{b,q}=\frac{2J_b(\mathbf{x}_a)}{\mathrm{Tr}(\mathbf{K}_q\mathbf{H})}, \tag{27}$$

$$s_{o,q}=\frac{2J_o(\mathbf{x}_a)}{\mathrm{Tr}(\mathcal{I}-\mathbf{H}\mathbf{K}_q)}, \tag{28}$$

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

$$\mathbf{B}_{q+1}=s_{b,q}\mathbf{B}_q, \quad \mathbf{R}_{q+1}=s_{o,q}\mathbf{R}_q. \tag{29}$$

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 [13].

Compared to other posterior diagnosis or iterative methods, e.g. [12], [7], no estimation of full matrices is needed and only the estimation of two scalar values (J_b, J_o) is required in **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 **DI01** could be designed by choosing a minimum threshold of $\mathbf{max}(||s_{b,q} - 1||, ||s_{o,q} - 1||)$. According to [6], the convergence of s_b and s_o can be very fast, especially in the ideal case where the correlation patterns of $\mathbf{B}$ and $\mathbf{R}$ are perfectly known. Under this assumption, [6] 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.

4.2 Adaptation of DI01 algorithm to localized subspaces

The application of DA algorithms, as well as the full observation matrix diagnosis has been discussed in [38]. Following the notation of their paper, we introduce the binary selection matrix Φ_x^i, Φ_y^i of the i^{th} subvector with

$$\mathbf{x}^i = \Phi_x^i \mathbf{x}, \quad \mathbf{y}^i = \Phi_y^i \mathbf{y} \tag{30}$$

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. 2, we have

$$\Phi_x^1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{31}$$

In the case of data reduction strategy (Eq.20),

$$\Phi_{y,\text{reduction}}^1 = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}, \tag{32}$$

while for data adjustment strategy,

$$\Phi_{y,\text{adjustment}}^1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}. \tag{33}$$

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:

$$\mathbf{B}^i = \Phi_x^i \mathbf{B} \Phi_x^{i,T}, \quad \mathbf{R}^i = \Phi_y^i \mathbf{R} \Phi_y^{i,T}. \tag{34}$$

Therefore, the associated analyzed subvector $\mathbf{x}_a^i$ could be obtained by applying DA 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 DA system), we don't necessarily have

$$\mathbf{x}_a^i = \Phi_x^i \mathbf{x}_a. \tag{35}$$

For more details of decomposition formulations, the interested readers are referred to [38]. 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:

$$\begin{aligned} \mathbb{E}\left[J_b(\mathbf{x}_a^i)\right] &= \mathbb{E}\left[(\mathbf{x}_a^i - \mathbf{x}_b^i)^T (\mathbf{B}^i)^{-1} (\mathbf{x}_a^i - \mathbf{x}_b^i)\right] \\ &= \frac{1}{2}\mathrm{Tr}(\mathbf{K}^i \mathbf{H}^i), \end{aligned} \tag{36}$$

$$\begin{aligned} \mathbb{E}\left[J_o(\mathbf{x}_a^i)\right] &= \mathbb{E}\left[(\mathbf{y}^i - \mathbf{H}^i \mathbf{x}_b^i)^T (\mathbf{R}^i)^{-1} (\mathbf{y}^i - \mathbf{H}^i \mathbf{x}_b^i)\right] \\ &= \frac{1}{2}\mathrm{Tr}(\mathcal{I}^i - \mathbf{H}^i \mathbf{K}^i), \end{aligned} \tag{37}$$

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. The identity matrix $\mathcal{I}^i$ is of the same dimension as $\mathbf{B}^i$.

We can also define the local tuning algorithm:

$$s_{b,q}^i = \frac{2J_b(\mathbf{x}_a^i)}{\mathrm{Tr}(\mathbf{K}_q{}^i\mathbf{H}^i)}, \tag{38}$$

$$s_{o,q}^i = \frac{2J_o(\mathbf{x}_a^i)}{\mathrm{Tr}(\mathcal{I}^i - \mathbf{H}^i\mathbf{K}_q{}^i)}, \tag{39}$$

$$\mathbf{B}_{q+1}^i = s_{b,q}^i\mathbf{B}_q^i, \tag{40}$$

$$\mathbf{R}_{q+1}^i = s_{o,q}^i\mathbf{R}_q^i. \tag{41}$$

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 $(\mathbf{D}_\mathbf{B}^i, \mathbf{D}_\mathbf{R}^i)$, the positive definiteness of $\mathbf{B}$ and $\mathbf{R}$ is thus guaranteed as the correlation structure remains invariant, cf. Algorithm 1.

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 $\left\{s_{b,q}^i, s_{o,q}^i\right\}$ with $\mathbf{B}_q^i, \mathbf{B}_q^i$ via Eq. (38-39)
  end
  Updating of full covariance matrices from blockwise tuned covariance in current cluster:

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

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

$$(\mathbf{D}_\mathbf{B}^i)_{j,j} = \begin{cases} \prod_{q=1}^{q_{\max}} s_{b,q}^i & \text{if } \{\mathbf{x}_j\} \subset \mathbf{x}^i \\ 1 \quad \text{otherwise} \end{cases}$$

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

end

outputs: Improved error covariances

4.3 Complexity analysis

Reducing computational cost could be seen as an important vocation of localization techniques, especially for domain localization methods ([38]). 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. [10]. Therefore, the global **DI01** covariance tuning, for a given state vector of size $n_{\mathbf{x}}$, is of computational complexity:

$$\mathcal{C}_{\text{global}}(n_{\mathbf{x}}) = q_{\max} \times n_{\mathbf{x}}^{\mu}. \tag{42}$$

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:

$$\mathcal{C}_{\text{localized}}(n_{\mathbf{x}^1}, ..., n_{\mathbf{x}^p}) = \sum_{i=1}^{p} q_{\max}^{i} \times (n_{\mathbf{x}}^{i})^{\mu}. \tag{43}$$

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 clusters are of comparable size, Eq. 43 could be simplified as:

$$\mathcal{C}_{\text{localized}}(n_{\mathbf{x}}) = \Big(\sum_{i=1}^{p} q_{\max}^{i}\Big) \times \left(\mathcal{O}\left(\left(\frac{n_{\mathbf{x}}}{p}\right)^{\mu}\right)\right) = \frac{\sum_{i=1}^{p} q_{\max}^{i}}{p} \times \frac{\mathcal{O}(n_{\mathbf{x}}^{\mu})}{\mathcal{O}(p^{\mu-1})}. \tag{44}$$

Considering the number of **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. 44 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 DA algorithms are often combined with other techniques, such as adjoint modelling. In these cases, the marginal computation cost of each iteration of **DI01** (both in subspaces and the global space) could be reduced further. Nevertheless, the value of μ in Eq. 44 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.

5 Illustration with numerical experiments

5.1 Test case description

Similar to the works of [9] and [38], 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.

5.1.1 Construction of H

A sparse Jacobian matrix $\mathbf{H}$ 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 $\mathbf{H}$ 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 [19]) due to its non-homogeneity in the space. In order to be as general as possible, we consider $\left|\mathbf{x}^1\right| = \left|\mathbf{x}^2\right| = 50$ and $\left|\mathbf{y}^1\right| = \left|\mathbf{y}^2\right| = 25$. 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(a). More specifically, the state-observation mapping is constructed as follows: we use a binomial distribution with two levels of success probability:

$$Pr(\mathbf{H}_{i,j}=1) = \left\{ \begin{array}{lll} 15\% & \text{if} \quad \mathbf{x}_i \in \mathbf{x}^1 \quad \text{and} \quad y_j \in \mathbf{y}^1 \\ 15\% & \text{if} \quad \mathbf{x}_i \in \mathbf{x}^2 \quad \text{and} \quad y_j \in \mathbf{y}^2 \\ 1\% & \text{otherwise} \quad \text{(perturbations)}. \end{array} \right. \tag{45}$$

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 **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:

$$\epsilon_b = \mathbf{x}_b - \mathbf{x}_t \sim \mathcal{N}(0^{n_\mathbf{x}=100}, \mathbf{B}_\mathrm{E}), \tag{46}$$

$$\epsilon_y = \mathbf{y} - \mathbf{H}\mathbf{x}_t \sim \mathcal{N}(0^{n_\mathbf{y}=50}, \mathbf{R}_\mathrm{E}), \tag{47}$$

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 **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 $\mathbf{y} - \mathbf{H}\mathbf{x}_t$).

5.1.2 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:

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

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:

$$(\mathbf{C_B})_{i,j} = (\mathbf{C_R})_{i,j} = \left(1 + \frac{r}{L}\right) \exp^{-\frac{r}{L}}, \tag{48}$$

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 [30], [33]).

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

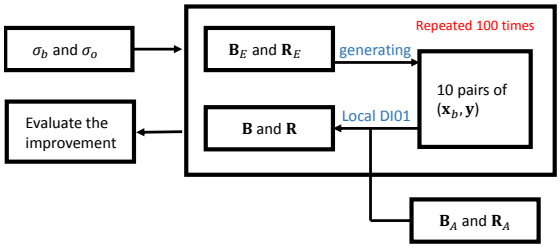


Figure 3: Flowchart of Monte-Carlo experiments for adaptive data assimilation with fixed parameters: $\sigma_b, \sigma_o, \mathbf{B}_A, \mathbf{R}_A$.

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:

$$\gamma_{\tilde{\mathbf{B}}} = \left(\Delta_{\mathbf{B}} - \Delta_{\tilde{\mathbf{B}}}\right)/\Delta_{\mathbf{B}}, \tag{49}$$

with $\Delta_{\cdot} = \mathbb{E}\left[\|\cdot - \mathbf{B}_{\mathrm{E}}\|_F\right]$ representing the expected matrices difference in the Frobenius norm, and similarly for the adjustment strategy and for the observations covariance. The larger the gain, the more advantage can be expected from the new approach compared to the standard **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. 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 $\sigma_{b,A} = \sigma_{o,A} = 0.05$.

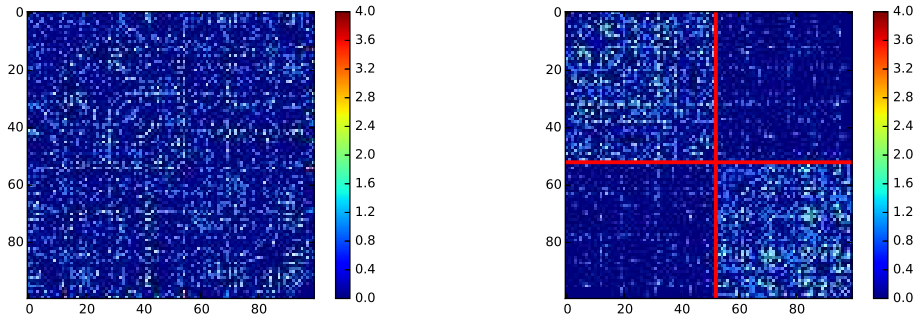
Meanwhile, the average exact prior error deviation ($\sigma_{b,E}, \sigma_{o,E} = \sqrt{\sigma_{o,1}\sigma_{o,2}}$) varies in the range ($[0.025, 0.1]$). In other words, we test a range spanning a domain with over-estimation of 100% of error deviation to an under-estimation of 100%. We remind that the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

5.2 Results

Thanks to the adjacency matrix of the observation-based state network as shown in Fig. 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 ([16]) of the obtained partition. It is used as an indicator for finding the optimal community number (as shown in Fig. 5), which is a standard approach for graph problems.

According to the result presented in Fig. 5, we chose to separate the state variables into two subsets, which is the correct number of communities when we simulate the Jacobian matrix $\mathbf{H}$. We emphasize that despite the fact that the $\mathbf{H}$ 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. The result of graph partition algorithm of two communities is summarized in Table. 1 and Fig. 4(b). From Table. 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. 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. 6 and Fig. 7 collect the results of the Monte-Carlo tests described in 5.1.2 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 two strategies for fitting the observation data. We are interested in the



(a) (b)

Figure 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.

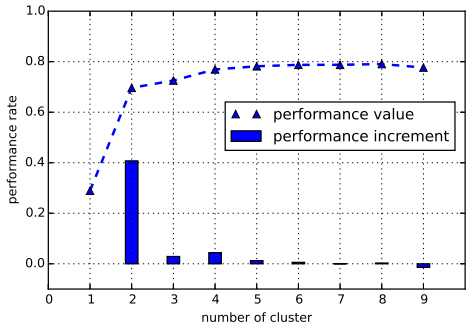


Figure 5: The evolution of performance value and its increment against the number of communities chosen.

potential advantage of the new methods compared to the standard algorithm. The normalized difference of covariance specification error is drawn as mentioned in 5.1.2 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.

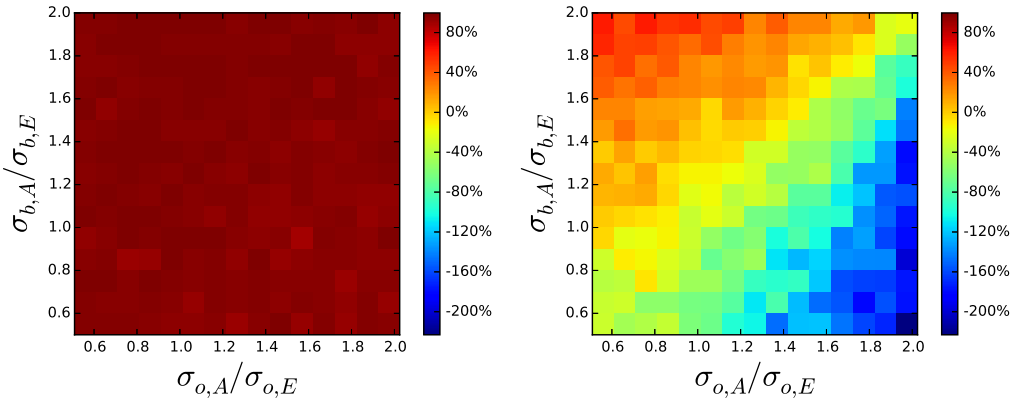
5.2.1 Measure of improvement of the localized approaches for the estimation of the background B matrix

From Fig. 6, one observes that in this test, the localized **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.

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

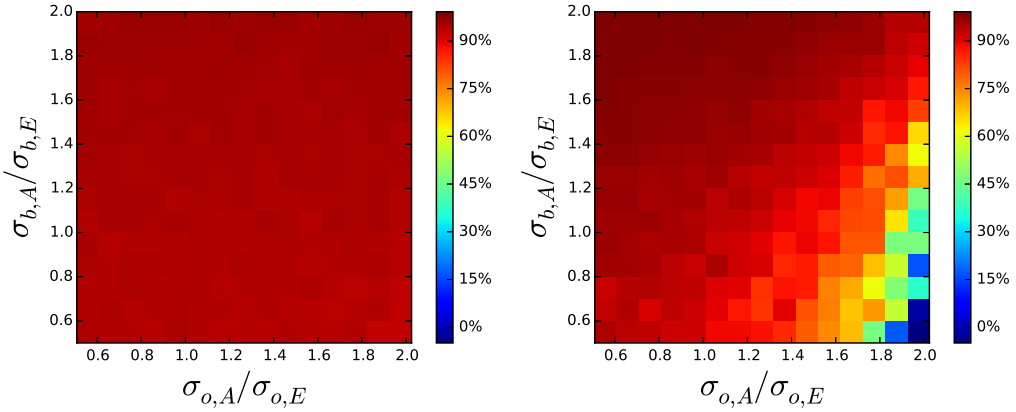
From Fig. 7, one observes significant advantages in most cases for both new adaptive approaches. In fact, according to the hypothesis of our experiments, the non-homogeneity of observation errors is completely neglected by a standard **DI01** . This non-homogeneity could be covered using the graph-based new approach. Similar to the matrix **B**, less optimal results are 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.



(a) (b)

Figure 6: Average improvement (in % according to the measures introduced in 5.1.2) of the background error covariance $\mathbf{B}$ corrected by the proposed localized approach relative to the standard global tuning, (a): with data reduction ($\tilde{\delta_{\mathbf{B}}}$); (b): with data adjustment ($\hat{\delta_{\mathbf{B}}}$); 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 $[0.025, 0.1]$.



(a) (b)

Figure 7: Same figure as in Fig. 6 for observation error covariance improvement $\tilde{\delta_{\mathbf{R}}}$ (a) and $\hat{\delta_{\mathbf{R}}}$ (b).

Strategy chosen \ Size of subsets	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

Table 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. 5.

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

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

5.2.3 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 5.1, except that the ratio of $\sigma^{i=1}_{o,E}(\mathbf{y}_u)\Big/\sigma^{i=2}_{o,E}(\mathbf{y}_v)$ 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 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 $\sigma_{b,\mathrm{A}}/\sigma_{b,\mathrm{E}}, \sigma_{o,\mathrm{A}}/\sigma_{o,\mathrm{E}}$ similarly to Fig. 6 and Fig. 7. On the other hand, while the prior estimation of $\sigma_{b,A}, \sigma_{o,A}$ is of extremely poor quality, for example,$\sigma_{o,\mathrm{A}}/\sigma_{o,\mathrm{E}} > 100$ (or $< 1/100$), we recommend to consider the standard DI01 in the first place.

6 Discussion

Localization technique is an important numerical tool which contributes to the success of solving high-dimensional DA 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 becomes increasingly challenging, for instance due to the use of satellite observations.

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 homo-

geneity 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 **DI01** is chosen because the ratio of available data to problem size is often limited for geosciences applications. Furthermore, the correction of **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 DA 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.

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.

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 DA 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.

References

[1] E. Blayo, B. Marc, and C. Emmanuel. Advanced data assimilation for geosciences: Lecture notes of the les houches school of physics: Special issue. Oxford Scholarship Online, 2012.

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

[3] A. Browet and P. Van Dooren. 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*, 2014.

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

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

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

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

[8] A. M. Clayton, A. C. Lorenc, and D. M. Barker. Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the met office. *Quarterly Journal of the Royal Meteorological Society*, 139(675):1445–1461, 2012.

[9] S. Clifford, T. M.Toth, and R. Busa-Fekete. GraphClus, a MATLAB program for cluster analysis using graph theory. *Computers & Geosciences*, 35(6):1205 – 1213, 2009.

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

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

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

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

[14] D. Dreano, P. Tandeo, M. Pulido, B. Ait-El-Fquih, T. Chonavel, and I. Hoteit. Estimating model error covariances in nonlinear state-space models using Kalman smoothing and the expectation-maximisation algorithm. *Quarterly Journal of the Royal Meteorological Society*, 143(705):1877 – 1885, 2017.

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

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

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

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

[19] M. Gerke. Using horizontal and vertical building structure to constrain indirect sensor orientation. *Journal of Photogrammetry and Remote Sensing*, 66:307–316, 05 2011.

[20] H. Gong, Y. Yu, Q. Li, and C. Quan. An inverse-distance-based fitting term for 3D-Var data assimilation in nuclear core simulation. *Annals of Nuclear Energy*, 141:107346, 02 2020.

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

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

[23] R. Hoffman, J. V. Ardizzone, S. Leidner, D. Smith, and R. Atlas. 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, 2013.

[24] A.-T. Ihler, S. Kirshner, M. Ghil, A. Robertson, and P. Smyth. Graphical models for statistical inference and data assimilation. *Physica D: Nonlinear Phenomena*, 230:72–87, 01 2005.

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

[26] D. Kumar. Ensemble-based assimilation of nonlinearly related dynamic data in reservoir models exhibiting non-gaussian characteristics. *Mathematical geosciences*, 51:75–107, 09 2018.

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

[28] F. Parés, D. G. Gasulla, A. Vilalta, J. Moreno, E. Ayguadé, J. Labarta, U. Cortés, and T. Suzumura. Fluid communities: A competitive, scalable and diverse community detection algorithm. In *Complex Networks & Their Applications VI*, pages 229–240, Cham, 2018. Springer International Publishing.

[29] J. Phillips, W. Schwanghart, and T. Heckmann. Graph theory in the geosciences. *Earth-Science Reviews*, 143:147 – 160, 02 2015.

[30] A. Ponçot, J.-P. Argaud, B. Bouriquet, P. Erhard, S. Gratton, and O. Thual. Variational assimilation for xenon dynamical forecasts in neutronic using advanced background error covariance matrix. *Annals of Nuclear Energy*, 60:39–50, 10 2013.

[31] A. Sandu and H. Cheng. An error subspace perspective on data assimilation. *International Journal for Uncertainty Quantification*, 5:491–510, 01 2015.

[32] S. Schirber, D. Klocke, R. Pincus, J. Quaas, and J. L. Anderson. 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, 2013.

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

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

[35] R. Tibshirani, G. Walther, and T. Hastie. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society Series B*, 63:411–423, 02 2001.

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

[37] H. Vo and L. Durlofsky. A new differentiable parameterization based on principal component analysis for the low-dimensional representation of complex geological models. *Mathematical Geosciences*, 46:775–813, 10 2014.

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

[39] J. A. Waller, D. Simonin, S. L. Dance, N. K. Nichols, and S. P. Ballard. Diagnosing observation error correlations for doppler radar radial winds in the met office UKV model using observation-minus-background and observation-minus-analysis statistics. *Monthly Weather Review*, 144(10):3533–3551, 2016.