A Graph Clustering Approach to Localization for Adaptive Covariance Tuning in Data Assimilation Based on State-Observation Mapping

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

Received: 2 June 2020 / Accepted: 10 May 2021 / Published online: 21 June 2021
© International Association for Mathematical Geosciences 2021

Abstract An original graph clustering approach for the 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 cutoff radii or homogeneity assumptions. Localization is performed via graph theory, a branch of mathematics emerging as a powerful approach to capturing 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, and it is more amenable to scalable data assimilation. The application of this method does not require underlying block-diagonal structures of prior covariance matrices. To address intercluster connectivity, two alternative data adaptations are proposed. Once the localization is completed, a covariance diagnosis and tuning are performed within each cluster, whose contribution is sequentially integrated into the entire covariance matrix. Numerical twin experiment tests show the reduced cost and added flexibility of this approach compared to global covariance tuning, and more accurate results yielded for both observation- and background-error parameter tuning.

✉ Didier Lucor
didier.lucor@lisn.upsaclay.fr

¹ EDF R&D, Ile-de-France, France

² CNRS, Laboratoire Interdisciplinaire des Sciences du Numérique, Université Paris-Saclay, Orsay, France

³ Institut de Mathématiques de Toulouse, Université Paul Sabatier, Toulouse, France

Keywords Data assimilation · Covariance matrices · Graph community detection · Unsupervised learning

1 Introduction

Data assimilation techniques, originally developed for numerical weather prediction (NWP), have been widely applied in the field of geosciences (Blayo et al. 2012; Carrassi et al. 2018), for instance for the reconstruction of physical fields (e.g., temperature or velocity fields) or parameter identification. The applications, often complex, related to multiphysics and nonlinear modeling with different model resolutions and prediction time horizons, vary from reservoir modeling (Kumar 2018), to geological feature prediction (Vo and Durlowsky 2014), to environmental modeling (Casas et al. 2020) and wildfire front-tracking problems (Rochoux et al. 2018). Data assimilation methods have also been used in biomedical applications (Lucor and Le Maître 2018) and more recently to analyze the COVID-19 pandemic, including predicting disease diffusion (Wang et al. 2020) and determining optimal vaccination strategies (Cheng et al. 2021). The goal of data assimilation is to reduce the uncertainty in prediction that arises due to uncertainties in input variables such as parameters and state variables by combining the information embedded in a prior estimation (also known as the background state) and real-time observations or measures. Unfortunately, the large size of the multidimensional problems of geoscience data assimilation problems ($\mathcal{O}(10^{6-9})$) makes a full Bayesian approach computationally unaffordable. Instead, a variational approach weighs these two information sources, owing to the background state $\mathbf{x}_b$ and the observation $\mathbf{y}$ with their associated error covariances, represented by the matrices B and R , respectively.

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

When the state ensemble size is too small compared to the problem dimension, sampling errors may induce spurious long-distance error correlations resulting in poor conditioning of B and R . An important concept used to make data assimilation more efficient and robust follows the idea of localization. Localization 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, such as those involving interchannel radiance observation (Garand et al. 2007) or problems of parameter identification (Schirber et al. 2013), the correlation between different ranges/scales of the state or observation variables may not be directly interpreted in terms of spatial distances, and the assumption of weak long-distance

correlations might be less relevant. This study is instead focused on the more generic “long-range correlation” expression. Additionally, there might be situations in which a prior covariance structure has a limited spatial extent that is smaller than the support of the observation operator that maps the state to observation spaces. In this case, nonlocal observations that cannot be truly allocated to one specific spatial location, because they may result from spatial averages of linear or nonlinear functions of the system variables, can have a large influence on the assimilation van Leeuwen 2019.

Existing localization methods are mainly of two kinds: covariance localization and domain localization. The first family of localization methods is implicit and works on a regularization of the covariance matrix that is operated using a Schur matrix product with certain short-range predefined correlation matrices (Gaspari and Cohn 1999), which ensures the (semi)definitiveness of the new matrix and therefore avoids the introduction of spurious long-range correlation. These methods have been widely improved, such as ensemble-based Kalman filters (EnKFs) (Farchi and Bocquet 2019), where the covariance localization is crucial to produce more accurate analyses. The second class of families (domain localization) is explicit and performs data assimilation for each state variable by using only a local subset of available observations, typically within a fixed range of this point. In fact, much effort (e.g., Arcucci et al. 2018; Gong et al. 2020a) has been devoted to reducing the computational cost of high-dimensional data assimilation problems. For domain localization, a relevant localization length must be carefully chosen. This is the main disadvantage of the approach: if this length is too small, some important short- to medium-range correlations will be incorrectly neglected.

Recent works have shown that a local diagnosis/correction of error covariance computation could be helpful for improving the forecast quality of the global system (e.g., Waller et al. 2017), as well as reducing the computational cost. An observation introduces the concepts of the domain of dependence, 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, the set of analysis states that are updated in the assimilation using this observation. According to Waller et al. (2017), difficulties with the domain localization appear when the region of influence is far offset from the domain of dependence. In fact, the former, which represents the set of analysis states that are updated in the assimilation using the same observations, may be imposed based on prior assumptions, while the latter is obtained from the linearized transformation operator, which depicts how the state variables are “connected” via the observations. Nevertheless, relying purely on an imposed cutoff radius for localization may deteriorate this connection, resulting in a less optimal posterior estimation, especially when long-range error covariance is present, as illustrated in the numerical experiments of Waller et al. (2017). Empirically chosen cutoff or distance thresholds may result in the removal of true physical long-range correlations, thus inducing imbalance in the analysis (Greybush et al. 2011). This conclusion points to the relevance of more efficient and less arbitrary segregation operators. The spatial dependence between state variables and observations is an essential problem in the inversion of nonlinear problems, such as subsurface flows. The probability conditioning method (PCM) (Jafarpour and Khodabakhshi 2011) is another class of data assimilation using probability maps of state variables from an ensemble of updated models, and assimilating the probability with

Fig. 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

multipoint statistical techniques for generating geological patterns. This allows for the representation of realistic natural formations with non-Gaussian statistics (Kumar and Srinivasan 2020). Performing domain localization based on state-observation mapping may improve the quality of these probability maps, contributing to the overall algorithmic efficiency and training process of these approaches.

In practice, data assimilation often deals with nonuniform error fields containing underlying structures due to the heterogeneity of the data, which calls for *unsupervised* localization schemes. One of the main objectives of unsupervised learning is to split a data set into two or more classes based on a similarity measure over the data, without resorting to any *a priori* information on how it should be done (see Hastie et al. 2001, Sect. 14). Figure 1 illustrates with a very simple schematic the class of problems that could benefit from this 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. The type 2, observations, depend on a certain group of variables, while type 1 observations 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 element of our data assimilation approach will be to automatically and correctly localize these variable/observation clusters (otherwise known as subspaces or communities) which are considered in this paper to be able to reveal Intra-cluster networks.

In this study, the choice is made 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.

¹ In Fig. 1, only a single observation is contributed from both groups of state variables (cf., orange arrow).

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 of state variables and corresponding observations. This clustering of variables will allow more local assimilation, likely to be more flexible and efficient than a standard global assimilation technique. In recent years, graph theory has been introduced in geosciences for a large range of applications, such as quantifying complex network properties including similarity, centrality and clustering or identifying special graph structures such as 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 Phillips et al. 2015).

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

Once the graph clustering approach has been efficiently applied for localizing several state communities, the next step is to take advantage of it in order to improve the prior state/observation errors covariance. In our proposed approach, we perform a fine tuning of the entire matrices by sequentially updating the covariances, thanks to the correction contribution coming from each cluster, in particular the objective 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 correctly assigning a subset of observations to each community of state variables. How different modeling and computational approaches are possible along those lines will be shown in this study.

As mentioned previously, remarkable efforts have been made on posterior diagnosis and iterative adjustment of error covariance quantification, especially by the meteorology community (e.g., Desroziers and Ivanov 2001; Desroziers et al. 2005). Among these tuning methods, that of Desroziers and Ivanov (herein referred to as DI01), which consists in finding a fixed point for the assimilated state by adjusting the ratio between background and observation covariance matrix amplitudes without modifying their correlation structures, is well received in NWP. This approach presents the flexibility to be implemented either in a static or at any step of a dynamical data

assimilation process for both variational methods and Kalman-type filtering, even with limited background/observation data. A different approach with full covariance estimation/diagnosis based on large ensembles is for instance proposed in Desroziers et al. (2005). The latter is based on statistics of prior and posterior innovation quantities. In fact, the deployment of DI01 in subspaces has already been introduced in Chapnik et al. (2004) for block-diagonal structures of B and R . In this paper, the a DI01 approach is extended to a more general framework, where the block-diagonal structure of the covariance 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 first introduced, as well as its resolution in the case of a linearized Jacobian matrix, in Sect. 2. How to use this Jacobian matrix, viewed as a state-observation mapping, to build an observation-based state network is explained in Sect. 3. The subspace decomposition is carried out by applying graph-based community detection algorithms. The localized version of DI01 is then introduced (Sect. 4) and investigated in a twin experiments framework (Sect. 5). The paper then concludes with a discussion (Sect. 6).

2 Data Assimilation Framework

The goal of data assimilation algorithms is to correct the state $\mathbf{x}$ of a dynamical system with the help of a prior state $\mathbf{x}_b$ and an observation vector $\mathbf{y}$, the former being often obtained from a expert 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 x_i is called a state variable and y_j is called an observation, where i, j represent the vector indices. The principle of data assimilation algorithms is to find $\mathbf{x}_a$, an optimally weighted combination of $\mathbf{x}_b$ and $\mathbf{y}$, by optimizing a cost function J defined as

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

$$= J_b(\mathbf{x}) + J_o(\mathbf{x}), \quad (2)$$

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

$$B = \mathbf{cov}(\epsilon_b, \epsilon_b), \quad (3)$$

$$R = \mathbf{cov}(\epsilon_y, \epsilon_y), \quad (4)$$

where

$$\epsilon_b = \mathbf{x}_b - \mathbf{x}_t, \quad (5)$$

$$\epsilon_y = \mathcal{H}(\mathbf{x}_t) - \mathbf{y}. \quad (6)$$

The inverse matrices, B^{-1} and 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 that they can be perfectly characterized by the covariance matrices

$$\epsilon_b \sim \mathcal{N}(0, B), \quad (7)$$

$$\epsilon_y \sim \mathcal{N}(0, R). \quad (8)$$

The two covariance matrices B and R , which are difficult to know perfectly a priori, play an essential role 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, denoted 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 $\mathcal{H} \equiv H$ is linear and the covariances matrices B and R are well known, the optimization problem (1) can be exactly solved by linear formulation of the best linear unbiased estimator (BLUE)

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

which is also equivalent to a maximum a posteriori estimator. The Kalman gain matrix K is defined as

$$K = BH^T (HBH^T + R)^{-1}. \quad (10)$$

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

To achieve implementation, the representation of state variable/observation covariances by block-diagonal matrices is often carried out in data assimilation (for example, see Chabot et al. 2015). In this case, only uncorrelated state variables can be separated. In this work, the objective consists in applying covariance diagnosis methods in subspaces identified from the state-observation mapping with no assumption of block-diagonal structures of the covariances.

The state subspaces are detected via an unsupervised graph clustering learning technique. Here, the graph is 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 variable-observation mapping) connecting pairs of vertices. Graph clustering automatically groups 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

The idea is to perform a localization by segregating the state vector $\mathbf{x} \in \mathbb{R}^{n_x}$ (the background or analyzed subscript is dropped for ease of notation) into a partition $\mathcal{C}$ of subvectors $\mathcal{C} = \{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^p\}$, each $\mathbf{x}^i$ being non-empty. $\mathcal{C}$ is later called 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 $\{\mathbf{y}^1, \mathbf{y}^2, \dots, \mathbf{y}^p\}$.

In the work of Waller et al. (2016), a threshold of spatial distance $\tilde{r}$ is arbitrarily imposed a priori to define local subsets of state variables influenced by each observation during the data assimilation updating. In other words, each observation component 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 $\mathcal{R}_{\text{influence}}(y_i) = \{x_k : \phi(y_i, x_k) \leq \tilde{r}\}$, where ϕ measures some spatial distance, is called the region of influence of y_i .

However, that method faces significant difficulty when the Jacobian matrix H is dense or nonlocal, which means the updating of state variables depends on observations out of the region of influence. In fact, the nonlocality of matrix 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

$$\mathcal{D}_{\text{dependence}}(y_i) = \{x_k : H_{i,k} \neq 0\}, \quad (11)$$

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

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 transformation function $\mathcal{H}$ or its linearized version H). The main idea is to separate the ensemble of state variables into several subsets regarding their occurrence in the domains of dependence of different observations. In order to do so, the notion of observation-based connection between two state variables x_i and x_j is introduced when they appear in the domain of dependence of the same observation y_k

$$\exists k, \quad \text{such that} \quad \frac{\partial[\mathcal{H}(\mathbf{x})]_k}{\partial x_i} \neq 0, \quad \frac{\partial[\mathcal{H}(\mathbf{x})]_k}{\partial x_j} \neq 0, \quad (12)$$

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

$$\exists k, \quad \text{such that} \quad H_{k,i} \neq 0, \quad H_{k,j} \neq 0. \quad (13)$$

Our goal is to determine whether the state variables which are strongly connected based on the observations can be grouped, regardless of their spatial distance. In order to do so, the strengths of this connection for each pair of state variables are defined as

$$\begin{aligned} \mathcal{S} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} &\mapsto \mathbb{R}^+ \\ \mathcal{S}(x_i, x_j) \equiv \mathcal{S}_{i,j} &= \sum_{k, i \neq j} \left| \frac{\partial[\mathcal{H}(\mathbf{x})]_k}{\partial x_i} \right| \left| \frac{\partial[\mathcal{H}(\mathbf{x})]_k}{\partial x_j} \right| \\ &= \sum_{k, i \neq j} |H|_{k,i} |H|_{k,j} \quad \text{if } \mathcal{H} \equiv H \text{ is linear,} \end{aligned} \quad (14)$$

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

Moreover, it is assumed that the strength function is null when measuring the connection strength of one state variable with itself. In the case where $|H|$ exhibits extremely large values, extra smoothing (e.g., of sigmoid type) could be applied on $|H|_{k,i} |H|_{k,j}$ in order to appropriately balance the graph weight. Finally, in the case of data assimilation of multi-type variable problems, care must be taken with regard to an inhomogeneous H matrix, which would result in perturbations for graph clustering.

The idea is to either deal with each variable type individually (i.e., performing graph-clustering localization for each type of state variables) or introduce some kind of normalization to balance the structure of H . For example, the sum of each column in $|H|$ could be fixed to a constant value.

An undirected graph $\mathcal{G}$ that is a pair of sets $\mathcal{G} = (\mathbf{x}, E)$ is now considered, 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 E contains the edges of the graph (the edge cardinality, i.e., $|E| = m$, represents the size of the graph). Each edge is an unordered pair of endpoints $\{x_k, x_l\}$. The measure $\mathcal{S}$ is used as a weight function to define the weighted version of the graph $\mathcal{G}_{\mathcal{S}} = (\mathbf{x}, E, \mathcal{S})$. This translates into the weighted adjacency matrix $A_{\mathcal{G}_{\mathcal{S}}}$ of the graph, that is a $n_{\mathbf{x}} \times n_{\mathbf{x}}$ matrix $A_{\mathcal{G}_{\mathcal{S}}} = (a_{x_i, x_j}^{\mathcal{G}_{\mathcal{S}}})$ with

$$a_{x_i, x_j}^{\mathcal{G}_{\mathcal{S}}} = \begin{cases} \mathcal{S}_{i,j} & \text{if } \{x_i, x_j\} \in E, \\ 0 & \text{otherwise.} \end{cases} \quad (15)$$

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. The partition $\mathcal{C} = \{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^p\}$ of $\mathbf{x}$ is made. A cluster $\mathbf{x}^i$ with a node-induced subgraph of $\mathcal{G}_{\mathcal{S}}$ is also identified, forming the subgraph $\mathcal{G}_{\mathcal{S}}[\mathbf{x}^i] := (\mathbf{x}^i, E(\mathbf{x}^i), \mathcal{S}_{|E(\mathbf{x}^i)})$, where $E(\mathbf{x}^i) := \{\{x_k, x_l\} \in E : x_k, x_l \in \mathbf{x}^i\}$. So $E(\mathcal{C}) := \bigcup_{i=1}^p E(\mathbf{x}^i)$ is the set of intracluster edges, and $E \setminus E(\mathcal{C})$ is the set of intercluster edges of cluster $\mathbf{x}^i$, with $|E(\mathcal{C})| = m(\mathcal{C})$ and $|E \setminus E(\mathcal{C})| = \bar{m}(\mathcal{C})$, while $E(\mathbf{x}^i, \mathbf{x}^j)$ denotes the set of edges connecting nodes in $\mathbf{x}^i$ to nodes in $\mathbf{x}^j$. Let $\bar{m}^c(\mathcal{C}) = p(p-1) - \bar{m}(\mathcal{C})$, representing the number of non-connecting intercluster pairs of vertices. It is important to stress that the identification of structural clusters is made easier if graphs are sparse, i.e., if the number of edges m is of the order of the number of nodes $n_{\mathbf{x}}$ of the graph (Fortunato 2010).

3.1.3 Clustering Algorithms

One of the main paradigms of clustering is finding groups/clusters that ensure both intracluster density and intercluster 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 (Blondel et al. 2008) and the fluid community algorithm (Parés et al. 2018). These methods are mostly based on random walks (Gueuning et al. 2019) or centrality measures in a network with the advantage of low computational cost. The use of graph theory in numerical simulation problems such as the Cuthill–McKee algorithm (Cuthill and McKee 1969) already occurs, for example, for sorting multidimensional grid points in a more efficient way (in terms of reducing the matrix band). In this paper, a different approach is introduced, with the objective of identifying observation-based state variable communities, which are later considered state subsets in covariance tuning. Community detection is performed on the observation-based state network, regardless

of the algorithms chosen. Considering the computational cost, the fluid community detection algorithm (Parés et al. 2018) could be an appropriate choice for sparse transformation matrices because its complexity is linear relative to the number of edges in the network (i.e., $\mathcal{O}(|E|)$). When the state dimension is very large, the computation of $\mathcal{G}$ may be numerically infeasible. However, studies in graph theory have shown that if the Jacobian matrix is sparse, community detection algorithms could be used without computation of the full adjacency matrix (i.e., $|H||H|^T$), for example, via a k -means method applied directly to $|H|$, as shown in Browet and Van Dooren (2014).

In real applications of graph theory, the number of optimal cluster p is often not known in advance. Finding an appropriate cluster number remains a popular research topic. Several methods have been developed in order to propose some objective functions with the notion of optimal coverage, performance or intercluster conductance, such as the elbow method (Ketchen and Shook 1996) or the gap statistic method (Tibshirani et al. 2001). For instance, the following performance metric will be used later for the experiments in Sect. 5.2

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

It represents the fraction of node pairs that are clustered correctly with regard to 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 purposes, inspired by the pedagogical approach of Waller et al. (2017), the following simple system with $\mathbf{x} \in \mathbb{R}^{n_{\mathbf{x}}=9}$ and $\mathbf{y} \in \mathbb{R}^{n_{\mathbf{y}}=4}$ is considered

$$H = 0.25 \times \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \end{bmatrix}, \quad (17)$$

where the magnitude of nonzero H entries is assumed constant for simplicity, which leads to the associated state-observation transformation function

$$\begin{aligned} 0.25(x_0 + x_1 + x_2 + x_3) &= y_0 \\ 0.25(x_1 + x_2 + x_3 + x_5) &= y_1 \\ 0.25(x_3 + x_5 + x_6 + x_7) &= y_2 \\ 0.25(x_4 + x_5 + x_7 + x_8) &= y_3. \end{aligned} \quad (18)$$

The obtained observation-based adjacency matrix is represented in Fig. 2a and is quite sparse with only $m = 19$ edges.

The clustering result obtained by the fluid community detection algorithm is illustrated in Fig. 2b. Two clusters (red and blue) of state variables can be identified where

Fig. 2 **a** Observation-based state adjacency matrix (i.e., $H^T H$ with zero values along the diagonal) obtained from the transformation operator H in Eq. 17. The edges of maximum weights are shown in yellow, while blue indicates no edges between two states. **b** Corresponding network identified by the community detection algorithm. The graph edge weights (measure of the strength of observation-based state connections) are scaled by their widths

points in each community are tightly connected. In particular, some intracluster nodes with strong connections (e.g., $\{x_1, x_2\}$ or $\{x_5, x_7\}$) 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 $\{x_3, x_5\}$. These inter-cluster connections are still managed by the algorithm. In fact, partitioning with ideal (noise-free) subsets can hardly be obtained in real application problems.

After the identification of the state clusters, each of them needs to be associated with an ensemble of observations. As discussed previously, the difficulty appears for observations with domains of dependence spanning multiple clusters. For instance, the assignment of $\{y_0\}$ and $\{y_3\}$ respectively to the first (red) and second (blue) state community is without ambiguity, while both $\{y_1\}$ and $\{y_2\}$ are overlapped by the two communities. In this case, some data preprocessing is necessary. Dealing with this type of overlapping in the observation partition is therefore crucial for the covariance tuning.

3.2 Dealing with the Intercluster Observation Region of Dependence for Assimilation

Assuming that a p -cluster structure $\mathcal{C} = \{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^p\}$ is provided by the community detection algorithm, for each cluster $\mathbf{x}^i$, an associated observation subset $\mathbf{y}^i$ should be assigned, in order to perform local covariance tuning later on. As shown in the following, while the partition $\mathcal{C} = \{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^p\}$ of $\mathbf{x}$ will remain the same, the partition of the observations $\{\mathbf{y}^1, \mathbf{y}^2, \dots, \mathbf{y}^p\}$ 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, two alternative methods, “observation reduction” and “observation adjustment,” are proposed, providing appropriate observation subsets associated with each state cluster.

3.2.1 Observation Reduction

Applying this strategy, the observation components y_k [thus $[\mathcal{H}(\mathbf{x})]_k$] with connections to several state variable clusters must be identified and canceled

$$\frac{\partial[\mathcal{H}(\mathbf{x})]_{k=0,\dots,n_y-1}}{\partial \mathbf{x}_{l=1,\dots,n_x,i=1,\dots,p}^i} \neq 0, \quad (19)$$

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

Nevertheless, these observation data can still be used later on for evaluating the posterior estimation $\mathbf{x}_a$ in the data assimilation procedure. Back to Eq. 18, the observations $\{y_1\}$ and $\{y_2\}$ are voluntarily excluded to perform the covariance correction. The tuning will be performed with only two clusters of subvectors

$$\begin{aligned} \mathbf{x}^1 &= \{x_{k=0,\dots,3}\}, & \tilde{\mathbf{y}}^1 &= \{y_0\}, \\ \mathbf{x}^2 &= \{x_{k=4,\dots,8}\}, & \tilde{\mathbf{y}}^2 &= \{y_3\}. \end{aligned} \quad (20)$$

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

$$\tilde{H} = 0.25 \times \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \end{bmatrix}. \quad (21)$$

3.2.2 Observation Adjustment

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

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

$$\hat{y}_l = y_l - \sum_{i=1,\dots,p, i \neq j} \sum_{k \mid x_k \in \mathbf{x}^i} H_{l,k} \mathbb{E}_b[x_k], \quad (22)$$

where $\mathbb{E}_b[\cdot]$ denotes the empirical expected value based on the prior background ensemble at hand. This approach leads to an adjusted Jacobian matrix $\hat{H}$ that induces an 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 background measure and another one due to the sampling error. Examples are shown in Sect. 5.2

Applied respectively to our example, $\{y_1\}$ and $\{y_2\}$ can be adjusted to belong to the first and the second clusters. With the help of background state $\mathbf{x}_b$, system of Eq. 18 can be modified as

$$\begin{aligned} 0.25(\mathbf{x}_0 + \mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3) &= \hat{y}_0 = y_0 \\ 0.25(\mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3) &= \hat{y}_1 = y_1 - 0.25\mathbb{E}_b[\mathbf{x}_5] \\ 0.25(\mathbf{x}_5 + \mathbf{x}_6 + \mathbf{x}_7) &= \hat{y}_2 = y_2 - 0.25\mathbb{E}_b[\mathbf{x}_3] \\ 0.25(\mathbf{x}_4 + \mathbf{x}_5 + \mathbf{x}_7 + \mathbf{x}_8) &= \hat{y}_3 = y_3. \end{aligned} \quad (23)$$

Thus the new operator can be written as

$$\hat{H} = 0.25 \times \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \end{bmatrix}. \quad (24)$$

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

4 Localized Error Covariance Tuning

Now that the system has been localized based on the state-observation linearized measure, and thanks to graph clustering methods, it is shown in the next section how to use this localization to improve the error covariance tuning.

4.1 Desroziers and Ivanov Diagnosis and Tuning Approach

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

It was proven in Talagrand (1998) and Desroziers and Ivanov (2001) that under the assumption of perfect knowledge of the covariance matrices B and R , the following equalities are perfectly satisfied in a 3D-VAR assimilation system

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

$$\mathbb{E}[J_o(\mathbf{x}_a)] = \frac{1}{2} \mathbb{E}[(\mathbf{y} - H\mathbf{x}_b)^T R^{-1} (\mathbf{y} - H\mathbf{x}_b)] \quad (26)$$

$$= \frac{1}{2} \text{Tr}(\mathcal{I} - HK), \quad (27)$$

where $\mathbf{x}_a$ is the output of a 3D-VAR algorithm with a linear observation operator H . Equations 25 and 27 are seldom satisfied in practice, in the sense that the accurate knowledge of prior error covariances is often out of reach for real data assimilation applications. Nonetheless, if the correlation structures of these matrices are assumed to be well known, then it is possible to iteratively correct their magnitudes. Using the two indicators

$$s_{b,q} = \frac{2J_b(\mathbf{x}_a)}{\text{Tr}(K_q H)}, \quad (28)$$

$$s_{o,q} = \frac{2J_o(\mathbf{x}_a)}{\text{Tr}(\mathcal{I} - HK_q)}, \quad (29)$$

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

$$B_{q+1} = s_{b,q} B_q, \quad R_{q+1} = s_{o,q} R_q. \quad (30)$$

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

In contrast to other posterior diagnosis or iterative methods such as Desroziers et al. (2005) or Cheng et al. (2019), no estimation of full matrices is needed in DI01, and only the estimation of two scalar values (J_b , J_o) is required. 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 the trace of the matrices, which can be evaluated in efficient ways.

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

4.2 Adaptation of the DI01 Algorithm to Localized Subspaces

The application of data assimilation algorithms, as well as the full observation matrix diagnosis, has been discussed in Waller et al. (2017). Following the notation of their paper, the binary selection matrix Φ_x^i , Φ_y^i of the i th subvector is defined as

$$\mathbf{x}^i = \Phi_x^i \mathbf{x}, \quad \mathbf{y}^i = \Phi_y^i \mathbf{y}, \quad (31)$$

where i is the cluster index of the subspace. Both the data assimilation in the subspace and the localized covariance tuning can 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,

$$\Phi_x^1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}. \quad (32)$$

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

$$\Phi_{y,\text{reduction}}^1 = [1 \ 0 \ 0 \ 0], \quad (33)$$

while for the data adjustment strategy,

$$\Phi_{y,\text{adjustment}}^1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}. \quad (34)$$

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

$$B^i = \Phi_x^i B \Phi_x^{i,T}, \quad R^i = \Phi_y^i R \Phi_y^{i,T}. \quad (35)$$

Therefore, the associated analyzed subvector $\mathbf{x}_a^i$ can be obtained by applying the data assimilation procedure using $(\mathbf{x}_b^i, \mathbf{y}^i, B^i, R^i)$. We recall that, due to the cross-community noises (i.e., the updating of $\mathbf{x}_b^i$ may not only depend on $\mathbf{y}_b^i$ in the global data assimilation system), it is not necessarily true to have

$$\mathbf{x}_a^i = \Phi_x^i \mathbf{x}_a. \quad (36)$$

For more details about decomposition formulations, interested readers are referred to Waller et al. (2017). Our objective for implementing the localized covariance tuning algorithms is to gain a finer diagnosis and correction on the covariance computation.

The local DI01 diagnosis in $(\mathbf{x}_b^i, \mathbf{y}^i)$ can be expressed as

$$\begin{aligned}\mathbb{E}[J_b(\mathbf{x}_a^i)] &= \mathbb{E}[(\mathbf{x}_a^i - \mathbf{x}_b^i)^T (B^i)^{-1} (\mathbf{x}_a^i - \mathbf{x}_b^i)] \\ &= \frac{1}{2} \text{Tr}(K^i H^i),\end{aligned}\quad (37)$$

$$\begin{aligned}\mathbb{E}[J_o(\mathbf{x}_a^i)] &= \mathbb{E}[(\mathbf{y}^i - H^i \mathbf{x}_b^i)^T (R^i)^{-1} (\mathbf{y}^i - H^i \mathbf{x}_b^i)] \\ &= \frac{1}{2} \text{Tr}(\mathcal{I}^i - H^i K^i),\end{aligned}\quad (38)$$

where the optimization functions J_b and J_o , as well as the localized gain matrix K^i , have also been adjusted in these subspaces. In our approach, the localized state-observation mapping H^i is obtained thanks to the graph-based localization. For simplicity, the notation H^i is used in the following. The identity matrix $\mathcal{I}^i$ is of the same dimension as B^i . The local tuning algorithm is then defined as

$$s_{b,q}^i = \frac{2J_b(\mathbf{x}_a^i)}{\text{Tr}(K_q^i H^i)}, \quad (39)$$

$$s_{o,q}^i = \frac{2J_o(\mathbf{x}_a^i)}{\text{Tr}(\mathcal{I}^i - H^i K_q^i)}, \quad (40)$$

$$B_{q+1}^i = s_{b,q}^i B_q^i, \quad (41)$$

$$R_{q+1}^i = s_{o,q}^i R_q^i. \quad (42)$$

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 DI01 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 B^i, R^i), then the adjusted B and R are not guaranteed to be positive-definite, and the prior knowledge of the covariance structure might be deteriorated. To circumvent this problem, the correlation structure of $(C_B$ and $C_R)$ is kept fixed. We recall that a covariance matrix $\mathbf{Cov}(\mathbf{x})$ (of random vector $\mathbf{x}$), which by its nature is positive semi-definite, can be decomposed into its variance and correlation structures as

$$\mathbf{Cov}(\mathbf{x}) = D^{1/2} \cdot C \cdot D^{1/2},$$

where D is a diagonal matrix of the state error variances, and C is the correlation matrix. By correcting the variance in each subspace only through the diagonal matrices (D_B^i, D_R^i) , the positive definiteness of B and R is thus guaranteed, as the correlation structure remains invariant, as shown in Algorithm 1.

Algorithm 1: Localization and updating of B and R with cluster-based implementation of DI01 algorithm.

Inputs:

Background state: $\mathbf{x}_b$ Observation data: $\mathbf{y}$ Initially guessed matrix: B, R Jacobian matrix: H Algorithm: Community detection using 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 B^i, R^i .**for** q from 1 to $q_{\max}$ **do**| calculation and storing of $\{s_{b,q}^i, s_{o,q}^i\}$ with B_q^i, B_q^i via Eq. 39-40**end**

Updating of full covariance matrices from blockwise tuned covariance in current cluster

$$B \leftarrow (D_B^i)^{1/2} B (D_B^i)^{1/2}$$

$$R \leftarrow (D_R^i)^{1/2} R (D_R^i)^{1/2},$$

where D_B^i and D_R^i are diagonal matrices defined as

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

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

end

outputs: Improved error covariances

4.3 Complexity Analysis

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

$$\mathcal{C}_{\text{global}}(n_{\mathbf{x}}) = q_{\max} \times n_{\mathbf{x}}^\mu. \quad (43)$$

On the other hand, applying Algorithm 1 for p clusters of dimension $n_{\mathbf{x}^1}, \dots, n_{\mathbf{x}^p}$ with $\sum_{i=1}^p n_{\mathbf{x}^i} = n_{\mathbf{x}}$, the complexity of localized covariance tuning $\mathcal{C}_{\text{localized}}$ is written as

$$\mathcal{C}_{\text{localized}}(n_{\mathbf{x}^1}, \dots, n_{\mathbf{x}^p}) = \sum_{i=1}^p q_{\max}^i \times (n_{\mathbf{x}^i})^\mu. \quad (44)$$

Since the graph computation could be carried out offline as long as the operator H remains invariant, the cost of graph clustering is not considered here. Under the hypothesis that the clusters are of comparable size, Eq. 44 could be simplified as

$$C_{\text{localized}}(n_{\mathbf{x}}) = \left(\sum_{i=1}^p q_{\max}^i \right) \times \left(\mathcal{O} \left(\left(\frac{n_{\mathbf{x}}}{p} \right)^{\mu} \right) \right) = \frac{\sum_{i=1}^p q_{\max}^i}{p} \times \frac{\mathcal{O}(n_{\mathbf{x}}^{\mu})}{\mathcal{O}(p^{\mu-1})}. \quad (45)$$

Considering that the number of DI01 iterations per cluster may be represented by a random integer centered around some mean value $\mathbb{E}[q_{\max}^i]$, the first term of Eq. 45 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 $\overline{q_{\max}^i} \leq q_{\max}$, and one can easily deduce that $p^{\mu-1} \times C_{\text{local}} \leq C_{\text{global}}$. Therefore, the graph-based method is at least $\mathcal{O}(p^{\mu-1})$ times as fast as the standard approach. This derivation also holds for most posterior covariance tuning methods other than DI01. Note that data assimilation algorithms are often combined with other techniques, such as adjoint modeling. In these cases, the marginal computational cost of each iteration of DI01 (both in subspaces and in the global space) could be further reduced. Nevertheless, the value of μ in Eq. 45 will always remain strictly superior to 1, regardless of the computational 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, reducing the computational time even further.

5 Illustration with Numerical Experiments

5.1 Test Case Description

Similar to the works of Clifford et al. (2009) and Waller et al. (2017), our methodology is illustrated 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 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 H reflecting the clustering of the state-observation mapping is generated, whose components 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 $\mathbf{y} \in \mathbb{R}^{n_{\mathbf{y}}=50}$. A case for which the state-observation mapping H reflects clustering structures is considered. For this reason, two (this choice is arbitrary) subsets of observations are constructed a priori, each relating mainly to only one subset of state variables. In fact, the clustering structure of Jacobian matrices can often be found in real-world applications (see an example

of building structure data in Fig. 3 of Gerke (2011)) due to its non-homogeneity in the space. In order to be as general as possible, the size of these subsets is set to be equal (i.e., $|\mathbf{x}^1| = |\mathbf{x}^2| = 50$ and $|\mathbf{y}^1| = |\mathbf{y}^2| = 25$). For the sake of simplicity, the observation operator H (of dimension $[50 \times 100]$) is randomly filled with binary elements, forming a dominant blockwise structure with some extra-block nonzero terms. The latter is done in order to mimic realistic problems. In other words, some perturbations are introduced in the form of intercluster perturbations, and 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. 4a. More specifically, the state-observation mapping is constructed as follows: a binomial distribution with two levels of success probability is used,

$$\begin{aligned} Pr(H_{i,j} = 1) \\ = \begin{cases} 15\% & \text{if } x_i \in \mathbf{x}^1 \text{ and } y_j \in \mathbf{y}^1 \\ 15\% & \text{if } x_i \in \mathbf{x}^2 \text{ and } y_j \in \mathbf{y}^2 \\ 1\% & \text{otherwise (perturbations).} \end{cases} \end{aligned} \quad (46)$$

In the following tests, exact and assumed covariance magnitudes will be changed, but the choice of Jacobian H is kept invariant. The community detection method, also remaining invariant for all Monte Carlo tests, is provided by the fluid community detection algorithm mentioned in Sect. 3.1.3.

As explained previously, there is particular interest in applying DI01 to 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

$$\epsilon_b = \mathbf{x}_b - \mathbf{x}_t \sim \mathcal{N}(0^{n_x=100}, B_E), \quad (47)$$

$$\epsilon_y = \mathbf{y} - H\mathbf{x}_t \sim \mathcal{N}(0^{n_y=50}, R_E), \quad (48)$$

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

5.1.2 Twin Experiments Setup

In order to reflect the construction of H , the *exact* error deviation, hidden from the tuning algorithm (respectively denoted by $\sigma_{b,E}^i, \sigma_{o,E}^i$), is supposed constant in each cluster; thus

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

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

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

so that the background error is homogeneous while the observation error is different in the two communities, with a fixed ratio (in the following, this ratio value is chosen fixed as $\text{ratio} = 10$). However, the correlation structures of the covariance matrices are supposed to be known *a priori* and are assumed to follow a Balgovind structure

$$(C_B)_{i,j} = (C_R)_{i,j} = \left(1 + \frac{r}{L}\right) \exp^{-\frac{r}{L}}, \quad (49)$$

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

We recall that the output of all DI01-based approaches depends on the available background and observation data set. Three different methods described previously in this paper are compared, differentiated by the notation used for their output covariances:

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

The performance of the covariance tuning with localization is evaluated with 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

$$\gamma_{\tilde{B}} = (\Delta_B - \Delta_{\tilde{B}}) / \Delta_B, \quad (50)$$

with $\Delta_{\cdot} = \mathbb{E}[\|\cdot - B_E\|_F]$ representing the expected difference in the Frobenius norm between matrices, and similarly for the adjustment strategy and for the observations covariance. The larger the gain, the greater the advantage that can be expected from the new approach compared to the standard DI01 approach.

In the numerical results presented later, the empirical expectation of these indicators will be calculated by repeating the tests 100 time, in a Monte Carlo fashion, for each case of standard deviation parameters. In each Monte Carlo simulation, 10 pairs of background state and observation vectors are generated to evaluate the coefficients s_b^i and s_o^i necessary for diagnosing and improving B and R as shown in Fig. 3.

Fig. 3 Flowchart of Monte Carlo experiments for adaptive data assimilation with fixed parameters: σ_b , σ_o , B_A , R_A

In order to examine the performance of the proposed approach, the assumed prior covariances are always quantified as

$$B_A = \sigma_{b,A}^2 \times C_B,$$

$$R_A = \sigma_{o,A}^2 \times C_R,$$

with $\sigma_{b,A}^2 = \sigma_{o,A}^2 = 0.05$.

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, a range spanning a domain with overestimation of 100% of error deviation to an underestimation of 100% is tested. 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. 4a, the fluid community detection method of Sect. 3.1.3 is first applied in the observation-based state network to determine subspaces (communities) in the state space. For real applications, the number of communities is unknown. Here, the community detection algorithms with different assumed community numbers are implemented several times, and the performance rate (Fortunato 2010) of the obtained partition is evaluated. 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, the state variables are separated into two subsets, which is the correct number of communities when simulating the Jacobian matrix H . We emphasize that despite the fact that the H matrix was generated using two clusters, it was not trivial to rediscover them from the observation-based state network, once the information was shuffled and noisy (cf., from (a) to (b) in Fig. 4). The result of the graph partition algorithm of two communities is summarized in Table 1 and Fig. 4b. From Table 1, it is observed that two state variables from the second subset $\mathbf{x}^2$

Fig. 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. In both figures, the edge weights are presented by colors, from red (strong) to blue (none)

Fig. 5 The evolution of the performance value of the partition and its increment against the number of clusters chosen

Table 1 Quantification of the community detection algorithm results on the observation-based state network followed by the data reduction and data adjustment strategies

Strategy chosen	Size of subsets					
	Detected subsets					
	$ x^1 $	$ x^2 $	$ x $	$ y^1 $	$ y^2 $	$ y $
Data reduction	52	48	100	12	13	25
Data adjustment	52	48	100	25	25	50

The number of clusters (i.e., $k = 2$) is set according to the result in Fig. 5

are mistakenly assigned to the first one, x^1 . The last column of Table 1 shows the total number of observations ($|y| = |y^1| + |y^2|$) used in the covariance tuning. Only half of the observations are considered while applying the strategy of data adjustment.

Figures 6 and 7 show 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 covari-

Fig. 6 Average improvement (in % according to the measures introduced in 5.1.2) of the background error covariance B corrected by the proposed localized approach relative to the standard global tuning, **a** with data reduction ($\tilde{\delta}_B$), **b** with data adjustment ($\hat{\delta}_B$); A stands for assumed and E for exact values, with $(\sigma_{b,E}, \sigma_{o,E} = \sqrt{\sigma_{o,1}\sigma_{o,2}})$ both varying in $[0.025, 0.1]$

ance matrix 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 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 for localized methods. All tuning methods are applied for $q_{\max} = 10$ iterations, and the sequences $s_{b,q}$, $s_{o,q}$, $s_{b,q}^i$, $s_{b,q}^i$ have been well converged to 1.

5.2.1 Measure of Improvement of the Localized Approaches for the Estimation of the Background 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 parameter combinations, but it becomes less optimal when σ_b increases and σ_o decreases. Thus, careful attention should be given to the error level of the background state when applying the data adjustment strategy. In fact, when the background error level is high, adjustment of the observation data with a background state of large variance will most likely pollute the observations in terms of both observation accuracy and the knowledge of error covariances.

5.2.2 Measure of Improvement in the Localized Approaches for the Estimation of the Observation Matrix

From Fig. 7, one observes significant advantages in most scenarios 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 new graph-based approach. Similar to

Fig. 7 Average improvement (in % according to the measures introduced in 5.1.2) in the observation error covariance R corrected by the proposed localized approach relative to the standard global tuning, **a** with data reduction (δ_R), **b** with data adjustment (δ_R); A stands for assumed and E for exact values, with $(\sigma_{b,E}, \sigma_{o,E} = \sqrt{\sigma_{o,1}\sigma_{o,2}})$ both varying in $[0.025, 0.1]$

the matrix B , less optimal results are found when the background error is considerably higher than the observation error.

In these twin experiments, one may conclude that despite the fact that half of the observations are ignored for the covariance tuning, the data reduction strategy in general possesses an advantage over the data adjustment strategy. 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.

5.2.3 Test Case with a Larger Difference in 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 the experiments is the same as in Sect. 5.1, except that the ratio of $\sigma_{o,E}^{i=1}(y_u) / \sigma_{o,E}^{i=2}(y_v)$ is now set to be 100 instead of 10. The same number of experiments are carried out as in the previous case. The test results are summarized in Table 2, according to the cases of under- or overestimation of prior error amplitude. As expected, due to the larger difference between the two observation groups, and thanks to the assumed homogeneous observation matrix R_A , the results of the new approaches are even more impressive over a standard DI01 while keeping the same trends against the variations of $\sigma_{b,A}/\sigma_{b,E}$, $\sigma_{o,A}/\sigma_{o,E}$, similarly to Figs. 6 and 7. On the other hand, while the prior estimation of $\sigma_{b,A}$, $\sigma_{o,A}$ is of extremely poor quality, for example, $\sigma_{o,A}/\sigma_{o,E} > 100$ (or $< 1/100$), it is recommended to consider the standard DI01 in the first place.

Table 2 Averaged gain improvement in error covariances ((B, R) in %) with $\sigma_{o,E}^{i=1}(y_u) = 100\sigma_{o,E}^{i=2}(y_v)$ thanks to two graph clustering localization strategies (observation reduction and observation adjustment). Both $\sigma_{b,E}$ and $\sigma_{o,E}$ vary in $[0.025, 0.1]$

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

6 Discussion

The localization technique is an important numerical tool that contributes to the success of solving high-dimensional data assimilation problems for which ensemble estimates are unreliable. It is based on the assumption that correlations between dynamical system variables eventually decay with 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 nonlocal observations and/or observations from different sources and at different scales (e.g., satellite observations) becomes increasingly challenging.

Considering the application of data assimilation to hydrological modeling (Cheng et al. 2020) as a motivating example, it is known that hydrological changes induced by precipitation (in the form of rain or snow) in the various watersheds of a region affect hydraulic conditions and the accompanying flood levels, sediment transport rates and habitat conditions within various distributed streams. While any particular location along a stream channel may depend on several close or distant watersheds, there are often geographically closed watersheds that do not contribute to the same stream due to the position of the ridgeline separating their neighboring drainage basins (Cheng 2020). Therefore, the assimilation of discrete streamflow measurements to correct water levels in drainage basins and reservoirs remains challenging due to the complex network structure of the hydrological domain (Castronova and Goodall 2014; Joo et al. 2020). For this particular application, a domain localization technique based solely on the spatial distance seems to be a poor approach. Similar arguments may apply to the modeling of the subsurface flow and transport properties involved in groundwater flow and contaminant transport, energy recovery from geothermal and hydrocarbon reservoirs, and the geologic storage of CO_2 in deep underground formations.

In this work, a generalized concept of domain localization relying on graph clustering state decomposition techniques is proposed. 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 to use this decomposition to perform efficient adaptive error covariance tuning. Compared to classic domain localization, the novel method is more effective when long-distance observations and error correlation exist, either in B or in 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 cutoff radii or homogeneity assumptions. In this paper, the fluid method is chosen for applications because of its computational simplicity, especially for sparse graphs. In terms of covariance diagnosis, DI01 is chosen because the ratio of available data to the problem size is often limited for geoscience applications. Furthermore, the correction of DI01 in subspaces allows more flexible tuning of error covariances without deteriorating prior knowledge of the error correlation. Finally, it is shown that our approach reduces the computational complexity and provides some speedup. It is best suited for problems of an intermediate size, such as those involving transformed data sets, as mentioned earlier.

In this paper, our methodology is applied to a simple twin experiment data assimilation problem for which the Jacobian matrix of the observation operator is chosen to reflect a dual clustering of the state-observation mapping, whose components are then randomly mixed to hide any particular structural pattern. Simply speaking, there exist two *hidden* communities of state variables, each of which is preferably connected to its own observation community. The problem is far from trivial, as the segregation resulting in clustering is not related to spatial separations, and the exact background error magnitude is supposed to be homogeneous in our tests, but the clusters have different exact observation errors. Moreover, there exists some interconnectivity between the clusters. Considering the latter, two simple numerical approaches are proposed: a data *reduction* or a data *adjustment* strategy. The problem is investigated for a wide range of assumed prior covariances. The graph clustering approach with adaptive covariance tuning is much more efficient than a global adaptive covariance tuning approach, especially in the case of DI01 tuning where B and R are jointly corrected. Alternative methods may also be considered where the emphasis would be placed on a correction of either B or R in relation to the choice of information (e.g., related to state variables or observations) used to define the similarity measure needed by the graph clustering algorithm.

Here, our 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 intercluster connectivity, which reveals that real application problems will never be fully separable. Here, the choice is made to circumvent this 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 have to be considered, much likely involving a search for the optimal ordering of the subspace covariance tuning.

Finally, our localization approach will perform better if the assimilation problem, represented by a graph, is well separable under our cluster analysis in the sense that the

data assimilation problem is decomposed into a certain number of subset problems minimizing the overlap between subsets. This will depend somewhat 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, the measure of similarity is based on the linearized observation operator. As a complement to this approach, it may be interesting to create a measure combining prior knowledge of error covariances with the state-observation mapping (i.e., $|H| B |H|^T$ instead of $|H| |H|^T$). This might provide a way to scalably optimize the 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.

Acknowledgements The authors are grateful to an anonymous associate editor and the two reviewers for their useful remarks on the manuscript.

References

- Arcucci R, Mottet L, Pain C, Guo Y-K (2018) Optimal reduced space for variational data assimilation. *J Comput Phys* 379:51–69
- Blayo E, Marc B, Emmanuel C (2012) Advanced data assimilation for geosciences: lecture notes of the Les Houches School of physics: Special issue. Oxford Scholarship Online
- Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *J Stat Mech Theory Exp* 2008(10):P10008
- Browet A, Van Dooren P (2014) Low-rank similarity measure for role model extraction. In: 21st international symposium on mathematical theory of networks and systems, July 7–11, 2014. Groningen, The Netherlands
- Carrassi A, Bocquet M, Bertino L, Evensen G (2018) Data assimilation in the geosciences: an overview of methods, issues, and perspectives. *Wiley Interdiscip Rev Clim Change* 9(5):e535
- Casas CQ, Arcucci R, Wu P, Pain C, Guo Y-K (2020) A reduced order deep data assimilation model. *Physica D* 412:132615
- Castronova AM, Goodall JL (2014) A hierarchical network-based algorithm for multi-scale watershed delineation. *Comput Geosci* 72:156–166
- Chabot V, Nodet M, Papadakis N, Vidard A (2015) Accounting for observation errors in image data assimilation. *Tellus A Dyn Meteorol Oceanogr* 67(1):23629
- Chapnik B, Desroziers G, Rabier F, Talagrand O (2004) Property and first application of an error-statistics tuning method in variational assimilation. *Q J R Meteorol Soc* 130:2253–2275
- Cheng S (2020) Error covariance specification and localization in data assimilation with industrial application. Ph.D. Thesis, Paris-Saclay University, France
- Cheng S, Argaud J-P, Iooss B, Lucor D, Ponçot A (2019) Background error covariance iterative updating with invariant observation measures for data assimilation. *Stoch Environ Res Risk Assess* 33(11):2033–2051
- Cheng S, Argaud J-P, Iooss B, Lucor D, Ponçot A (2021) Error covariance tuning in variational data assimilation: application to an operating hydrological model. *Stoch Environ Res Risk Assess* 35:1019–1038
- Cheng S, Arcucci R, Pain CC, Guo Y-K (2021) Optimal vaccination strategies for COVID-19 based on dynamical social networks with real-time updating. [arXiv:2103.00485](https://arxiv.org/abs/2103.00485)
- Clifford S, Toth MT, Busa-Fekete R (2009) GraphClus, a MATLAB program for cluster analysis using graph theory. *Comput Geosci* 35(6):1205–1213
- Coppersmith D, Winograd S (1990) Matrix multiplication via arithmetic progressions. *J Symb Comput* 9(3):251–280
- Cuthill E, McKee J (1969) Reducing the bandwidth of sparse symmetric matrices. In *Proceedings of the 1969 24th national conference*, vol 69, pp 157–172, New York, USA

- Desroziers G, Ivanov S (2001) Diagnosis and adaptive tuning of observation-error parameters in a variational assimilation. *Q J R Meteorol Soc* 127:1433–1452
- Desroziers G, Berre L, Chapnik B, Poli P (2005) Diagnosis of observation, background and analysis-error statistics in observation space. *Q J R Meteorol Soc* 131:3385–3396
- Dreano D, Tandeo P, Pulido M, Ait-El-Fquih B, Chonavel T, Hoteit I (2017) Estimating model error covariances in nonlinear state-space models using Kalman smoothing and the expectation-maximisation algorithm. *Q J R Meteorol Soc* 143(705):1877–1885
- Evensen G (1994) Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *J Geophys Res Oceans* 99(C5):10143–10162
- Farchi A, Bocquet M (2019) On the efficiency of covariance localisation of the ensemble Kalman filter using augmented ensembles. *Front Appl Math Stat* 5:3
- Fortunato S (2010) Community detection in graphs. *Phys Rep* 486(3):75–174
- Garand L, Heilliette S, Buehner M (2007) Interchannel error correlation associated with airs radiance observations: inference and impact in data assimilation. *J Appl Meteorol Climatol* 46(6):714–725
- Gaspari G, Cohn ES (1999) Construction of correlation functions in two and three dimensions. *Q J R Meteorol Soc* 125:723–757
- Gerke M (2011) Using horizontal and vertical building structure to constrain indirect sensor orientation. *J Photogramm Remote Sens* 66:307–316
- Gong H, Yu Y, Li Q (2020a) Reactor power distribution detection and estimation via a stabilized Gappy proper orthogonal decomposition method. *Nucl Eng Des* 370:110833
- Gong H, Yu Y, Li Q, Quan C (2020b) An inverse-distance-based fitting term for 3D-Var data assimilation in nuclear core simulation. *Ann Nucl Energy* 141:107346
- Greybush SJ, Kalnay E, Miyoshi T, Ide K, Hunt BR (2011) Balance and ensemble Kalman filter localization techniques. *Mon Weather Rev* 139(2):511–522
- Gueuning M, Cheng S, Lambiotte R, Delvenne J-C (2019) Rock-paper-scissors dynamics from random walks on temporal multiplex networks. *J Complex Netw* 8(2):cnz027
- Hastie T, Tibshirani R, Friedman J (2001) The elements of statistical learning, volume 1 of springer series in statistics. Springer, New York
- Hoffman R, Ardizzone V, J, Leidner S, Smith D, Atlas R, (2013) Error estimates for ocean surface winds: applying desroziers diagnostics to the cross-calibrated, multiplatform analysis of wind speed. *J Atmos Ocean Technol* 30(11):2596–2603
- Ihler A-T, Kirshner S, Ghil M, Robertson A, Smyth P (2005) Graphical models for statistical inference and data assimilation. *Physica D* 230:72–87
- Jafarpour B, Khodabakhshi M (2011) A probability conditioning method (PCM) for nonlinear flow data integration into multipoint statistical facies simulation. *Math Geosci* 43:133–164
- Joo H, Lee M, Kim J, Jung J, Kwak J, Kim HS (2021) Stream gauge network grouping analysis using community detection. *Stoch Environ Res Risk Assess* 35(4):781–795
- Ketchen DJ, Shook CL (1996) The application of cluster analysis in strategic management research: an analysis and critique. *Strateg Manag J* 17(6):441–458
- Kumar D (2018) Ensemble-based assimilation of nonlinearly related dynamic data in reservoir models exhibiting non-gaussian characteristics. *Math Geosci* 51:75–107
- Kumar D, Srinivasan S (2020) Indicator-based data assimilation with multiple-point statistics for updating an ensemble of models with non-gaussian parameter distributions. *Adv Water Resour* 141:103611
- Lucor D, Le Maître OP (2018) Cardiovascular modeling with adapted parametric inference. *ESAIM Proc Surv* 62:91–107
- Parés F, Gasulla DG, Vilalta A, Moreno J, Ayguadé E, Labarta J, Cortés U, Suzumura T (2018) Fluid communities: a competitive, scalable and diverse community detection algorithm. In: *Complex networks & their applications VI*. Springer, Cham, pp 229–240
- Phillips J, Schwanghart W, Heckmann T (2015) Graph theory in the geosciences. *Earth Sci Rev* 143:147–160
- Ponçot A, Arnaud J-P, Bouriquet B, Erhard P, Gratton S, Thual O (2013) Variational assimilation for xenon dynamical forecasts in neutronic using advanced background error covariance matrix. *Ann Nucl Energy* 60:39–50
- Rochoux M, Collin A, Zhang C, Trouvé A, Lucor D, Moireau P (2018) Front shape similarity measure for shape-oriented sensitivity analysis and data assimilation for Eikonal equation. *ESAIM ProcS* 63:258–279
- Sandu A, Cheng H (2015) An error subspace perspective on data assimilation. *Int J Uncertain Quant* 5(6):491–510

- Schirber S, Klocke D, Pincus R, Quaas J, Anderson JL (2013) Parameter estimation using data assimilation in an atmospheric general circulation model: from a perfect toward the real world. *J Adv Model Earth Syst* 5(1):58–70
- Stewart LM, Dance SL, Nichols NK (2013) Data assimilation with correlated observation errors: experiments with a 1-D shallow water model. *Tellus A Dyn Meteorol Oceanogr* 65(1):19546
- Talagrand O (1998) A posteriori evaluation and verification of analysis and assimilation algorithms. In: *Workshop on diagnosis of data assimilation systems*. Shinfield Park, Reading, pp 17–28
- Tibshirani R, Walther G, Hastie T (2001) Estimating the number of clusters in a data set via the gap statistic. *J R Stat Soc B* 63:411–423
- van Leeuwen PJ (2019) Non-local observations and information transfer in data assimilation. *Front Appl Math Stat* 5:48
- Vo H, Durlowsky L (2014) A new differentiable parameterization based on principal component analysis for the low-dimensional representation of complex geological models. *Math Geosci* 46:775–813
- Waller JA, Simonin D, Dance SL, Nichols NK, Ballard SP (2016) Diagnosing observation error correlations for doppler radar radial winds in the Met office UKV model using observation-minus-background and observation-minus-analysis statistics. *Mon Weather Rev* 144(10):3533–3551
- Waller JA, Dance SL, Nichols NK (2017) On diagnosing observation-error statistics with local ensemble data assimilation. *Q J R Meteorol Soc* 143(708):2677–2686
- Wang S, Yang X, Li L, Nadler P, Arcucci R, Huang Y, Teng Z, Guo Y (2020) A Bayesian updating scheme for pandemics: estimating the infection dynamics of covid-19. *IEEE Comput Intell Mag* 15(4):23–33
- Weaver A, Courtier P (2001) Correlation modelling on the sphere using a generalized diffusion equation. *Q J R Meteorol Soc* 127(575):1815–1846