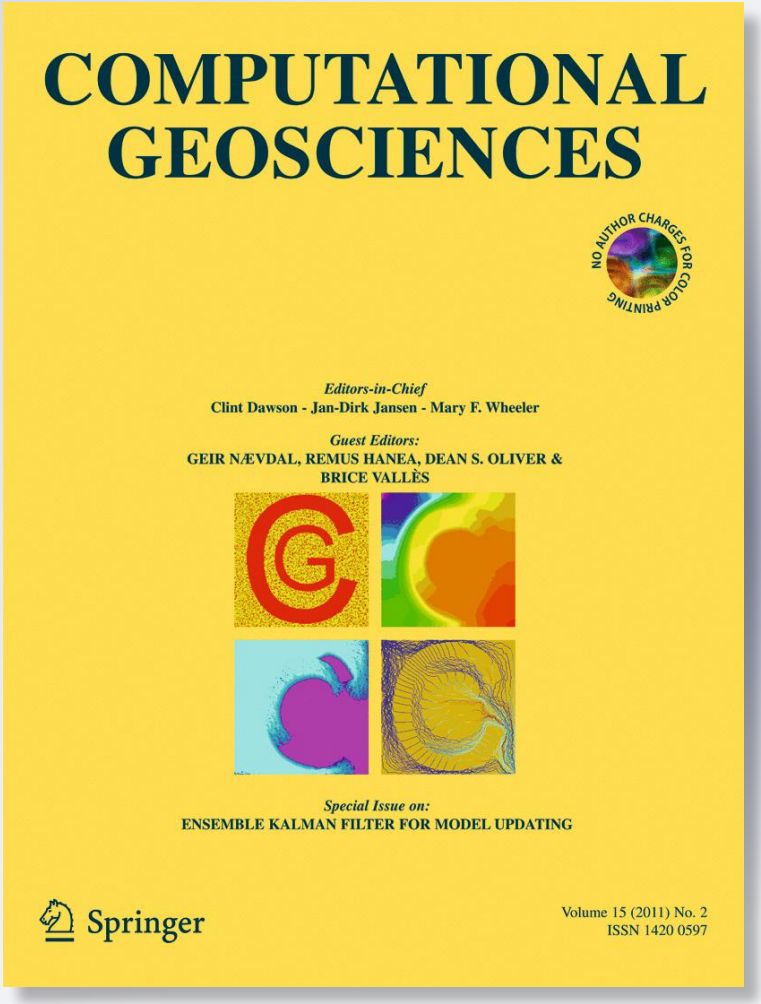
Relation between two common localisation methods for the EnKF

Computational Geosciences
Modeling, Simulation and Data Analysis

ISSN 1420-0597
Volume 15
Number 2

Comput Geosci (2011)
15:225-237
DOI 10.1007/
s10596-010-9202-6



Springer

Comput Geosci (2011) 15:225–237
DOI 10.1007/s10596-010-9202-6

ORIGINAL PAPER

Relation between two common localisation methods for the EnKF

Pavel Sakov · Laurent Bertino

Received: 17 December 2009 / Accepted: 7 July 2010 / Published online: 24 July 2010

Abstract This study investigates the relation between two common localisation methods in ensemble Kalman filter (EnKF) systems: covariance localisation and local analysis. Both methods are popular in large-scale applications with the EnKF. The case of local observations with non-correlated errors is considered. Both methods are formulated in terms of tapering of ensemble anomalies, which provides a framework for their comparison. Based on analytical considerations and experimental evidence, we conclude that in practice the two methods should yield very similar results, so that the choice between them should be based on other criteria, such as numerical effectiveness and scalability.

Keywords Data assimilation · Ensemble Kalman filter · EnKF · Localisation · Covariance localisation · Local analysis

1 Introduction

Ensemble Kalman filter (EnKF, [8]) is a state space formulation of the Kalman filter that uses an ensemble of model states to store, propagate and update the estimates of the state and state error covariance. The ensemble formulation is particularly beneficial for large-scale systems, when explicit storage and manipulation of the covariance matrix is impossible or not feasible. Apart from numerical and logistical advantages, the state space formulation permits: (1) propagation of the state and state error covariance with a nonlinear model without need for tangent linear model; (2) use of small ensembles due to localisation; and (3) assimilation of asynchronous observations. As a result, the EnKF increasingly becomes a method of choice for large-scale data assimilation systems, along with variational methods.

The biggest limitation for most EnKF-based systems is perhaps the resource limited ensemble size. This is true even for medium-size systems, with the model state vector size of order of just tens of thousands, not to mention the large-scale applications. This limitation requires use of methods known as localisation, which artificially reduce the spatial domain of influence of observations during the update. The localisation makes it possible to dramatically reduce the necessary ensemble size and create operational systems with as few as a hundred ensemble members or less.

Note that the concept of localisation formulated above only makes sense for *local* observations, that is for observations which can be attributed to a certain spatial location or, more generally, to a certain element or a "close" group of elements in the state vector. It also requires existence of some physically sensible norm to characterise the "distance" between model elements. There are generalisations of the distance based localisation, associated with the concept of adaptive localisation (e.g. [2, 4]), but those methods are beyond the focus of this study.

P. Sakov (✉) · L. Bertino
Nansen Environmental and Remote Sensing Center,
Thormøhlensgate 47, Bergen 5006, Norway
e-mail: pavel.sakov@nersc.no

L. Bertino
e-mail: laurent.bertino@nersc.no

Currently the localisation methods are used in the EnKF mainly on an ad-hoc basis. There are a number of different concepts in literature used for justification of localisation methods in the EnKF; however to the best of our knowledge there are no estimates of the degree of suboptimality introduced in the analysis. This indeed applies to other deviations from the Kalman filter assumptions (e.g. linearity, Gaussianity) and other enhancements of the Kalman filter analysis (e.g. inflation); however, unlike the above deviations and enhancements, the effect of localisation almost never can be regarded as "weak", as it typically reduces the effective dimension of the state vector in the analysis by a substantial factor and ensures that the ensemble is no longer rank deficient.

Further on, little is known about the two most common distance-based localisation methods used in practice, covariance localisation (CL) and local analysis (LA), including whether they are equivalent or not ([22], Section 3d). This study is rather specific in its goals; it aims at giving an answer to that question. It does not consider other important localisation related issues, like its effect on dynamical balance and optimality.

The outline of this work is as follows. Section 2 provides a brief background on the EnKF and the two localisation methods in hand; Sections 3 and 4 formulate the methods; Section 5 investigates the relation between them; the discussion and conclusions are presented in Sections 6 and 7.

2 Background

The Kalman filter update equations are commonly written as follows:

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}\left(\mathbf{d} - \mathbf{H}\mathbf{x}^f\right), \tag{1}$$

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^f, \tag{2}$$

where $\mathbf{x}$ is the state estimate; superscripts f and a to refer to forecast and analysis variables, correspondingly; $\mathbf{K}$ is the Kalman gain,

$$\mathbf{K} = \mathbf{P}\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\mathbf{P}\mathbf{H}^{\mathrm{T}} + \mathbf{R}\right)^{-1}; \tag{3}$$

$\mathbf{d}$ is vector of observations; $\mathbf{H}$ is the observation matrix; $\mathbf{P}$ is the state error covariance; $\mathbf{I}$ is the identity matrix; superscript "T" denotes matrix transposition; and $\mathbf{R}$ is the observation error covariance.

In the EnKF, the state estimate $\mathbf{x}$ and state error covariance $\mathbf{P}$ are carried by the ensemble of model states $\mathbf{E}$:

$$\mathbf{x} = \frac{1}{m}\mathbf{E}\mathbf{1}, \tag{4}$$

$$\mathbf{P} = \frac{1}{m-1}\mathbf{A}\mathbf{A}^{\mathrm{T}}, \tag{5}$$

where m is the ensemble size; $\mathbf{E}$ is a matrix containing each ensemble member as a column; $\mathbf{1}$ is the vector with all elements equal to 1; and $\mathbf{A}$ is the matrix of ensemble anomalies:

$$\mathbf{A} = \mathbf{E}\left(\mathbf{I} - \frac{1}{m}\mathbf{1}\mathbf{1}^{\mathrm{T}}\right). \tag{6}$$

There are two main reasons for using localisation in EnKF, commonly listed in literature: "spurious covariances" (e.g. [2, 3, 14]) and insufficient ensemble rank (e.g. [16], [10, p. 241]). The concept of insufficient ensemble rank refers, in general terms, to the inability to match the model rank[1] by the ensemble size. The importance of having an ensemble of sufficient rank is well recognised and represents one of the most basic rules in the EnKF system design.

The "spurious covariances" refer to the covariances between distant or physically not connected state vector elements that arise because of the finite size of the ensemble. They result in excessive reduction of the state error covariance in the analysis that may lead to the ensemble collapse and filter divergence.

The concept of spurious covariances is closely related with the interpretation of the representation (5) as a statistical estimator, when the ensemble of model states is considered as a statistical ensemble. We see it as more relevant for the traditional ("perturbed observations") EnKF [6], but less so for the ensemble square root filters (ESRF) [20]. Similarly, it is more relevant for the case of an imperfect model with stochastic error, than for the case of a perfect model.

Unlike the traditional EnKF, the ESRF is fully equivalent to the Kalman filter if (1) the assumptions of the Kalman filter (linear model and observations,

[1] We loosely define model rank as the number of degrees of freedom of the model. The related concepts are the model subspace dimension, the number of growing Lyapunov exponents, and the Kaplan–Yorke dimension (e.g. [7]).

Gaussian observation error etc.) are satisfied and (2) the initial state and state error covariance estimates are exactly factorised by (4) and (5); at least for the perfect model case. In this context, the relation (5) becomes a factorisation rather than a statistical estimator [19], and the concept of spurious correlations becomes irrelevant, regardless of the ensemble size. For an imperfect model, or if nonlinearity or non-Gaussianity are present, the equivalence between ESRF and Kalman filter becomes approximate, and the system can again benefit from a bigger ensemble regardless of the model rank. However, we underline that the benefits from a bigger ensemble can be characterised as incremental, while it is critical for the functioning of an EnKF system to have sufficient ensemble rank.

A few words about the notations used in this study. To present the material in a reasonable way we introduce some non-traditional notations. We use an upper accent to denote the local version of the corresponding variable; e.g. $\overset{i}{\mathbf{A}}$ means local ensemble anomalies used for the update of the ith element of the state vector or the ith row of the ensemble anomalies matrix. Similarly, $\overset{i}{\mathbf{K}}$ means the local Kalman gain, and $\overset{i}{\mathbf{H}}$ – the local observation matrix. We use upper-case bold symbols for matrices and lower-case bold symbols for vectors, with the exception of $\boldsymbol{\rho}$, which we use for the so called correlation matrix, for historical reasons. For an arbitrary matrix $\mathbf{X}$, $\mathbf{X}_i$ means the ith column; $\mathbf{X}_{i,:}$ – the ith row; and $\mathbf{X}_{ij}$ –the element at the ith row and jth column. For an arbitrary vector $\mathbf{x}$, $\mathbf{x}_i$ denotes the ith element.

Throughout this study we assume that the observations are local and that they have non-correlated errors ($\mathbf{R}$ is diagonal). These are quite common (although not universal) assumptions for many typical applications.

For illustrative purposes, we will present some graphical representations of covariances, gains, ensemble transform matrices (ETM) and other objects obtained in a particular experimental setup. This setup is described in the Appendix. We underline that it is used for illustrative purposes only; and although designed to be "sensible", it is not related to any particular physical model.

3 Covariance localisation

CL, also known as covariance filtering, modifies the update equations by replacing the state error covariance by its element-wise (Schur, Hadamard) product with some distance-based correlation matrix $\boldsymbol{\rho}$ [12, 14]:

$$\mathbf{P} \rightarrow \boldsymbol{\rho} \circ \mathbf{P}. \tag{7}$$

It is one of the two most common localisation methods suitable for large-scale applications.

The rationale behind this operation is that it increases the rank of the modified covariance matrix and masks spurious correlations between distant state vector elements. Figure 1 shows an example of its effect for a model with one-dimensional periodic geometry. To be applicable to a particular scheme, CL in form (7) requires the update equations to be formulated in terms

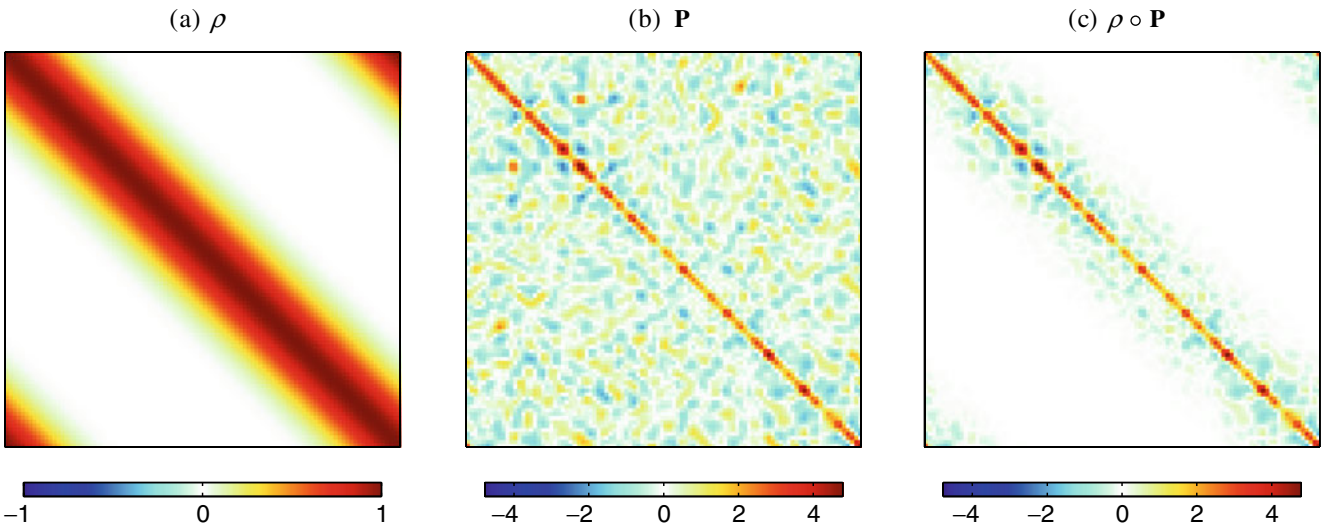


Fig. 1 Example of effect of CL on the state error covariance. See Appendix for details

of $\mathbf{P}$, which excludes some schemes, such as ETKF or, more generally, right-multiplied ESRFs [19].

For the consistency of the update operations, the localised covariance needs to remain positive (semi)definite. It is known that the element-wise product of positive (semi)definite matrices is positive (semi)definite (e.g. [13], Section 7.5.3); therefore normally a positive definite matrix is used as the localising matrix $\boldsymbol{\rho}$. Because of its positive definiteness and the range of its elements, $0 \leq \boldsymbol{\rho}_{ij} \leq 1$, it is often referred to as a correlation matrix. The distance from a given element of the state vector at which the entries in the correlation matrix fall below a certain threshold is commonly used to characterise the effective distance of influence of observations and referred to as the localisation radius. Hereafter, we define this threshold as equal to $e^{-1/2} \approx 0.61$. In atmospheric or oceanic applications, it is common to use different localisation scales horizontally and vertically, or to use a horizontal localisation only; the detailed description of these applications is beyond the scope of this study.

The localised state error covariance matrix is almost never calculated explicitly, except, perhaps, for small systems. Instead, when the observations are local enough, the terms $(\boldsymbol{\rho} \circ \mathbf{P}^f)\mathbf{H}^{\mathrm{T}}$ and $\mathbf{H}(\boldsymbol{\rho} \circ \mathbf{P}^f)\mathbf{H}^{\mathrm{T}}$ in (3) can be approximated as follows:

$$\left(\boldsymbol{\rho} \circ \mathbf{P}^f\right) \mathbf{H}^{\mathrm{T}} \approx \frac{1}{m-1} \sum_{o=1}^{p} \overset{io}{\mathbf{A}}{}^f \left(\overset{\{o\}}{\mathbf{H}} \mathbf{A}^f\right)^{\mathrm{T}}, \tag{8}$$

$$\mathbf{H}\left(\boldsymbol{\rho} \circ \mathbf{P}^f\right) \mathbf{H}^{\mathrm{T}} \approx \frac{1}{m-1} \left(\tilde{\mathbf{H}} \boldsymbol{\rho} \tilde{\mathbf{H}}^{\mathrm{T}}\right) \circ \left[\left(\mathbf{H}\mathbf{A}^f\right)\left(\mathbf{H}\mathbf{A}^f\right)^{\mathrm{T}}\right], \tag{9}$$

where $\overset{io}{\mathbf{A}}{}^f$ denotes ensemble anomalies with each member tapered about the location i_o of the oth observation, with the taper function equal to the oth column of $\boldsymbol{\rho}\mathbf{H}^{\mathrm{T}}$; $\overset{\{o\}}{\mathbf{H}}$ denotes $\mathbf{H}$ with all rows but the oth set to zero; $\tilde{\mathbf{H}}$ is the observation location matrix, such that $\tilde{\mathbf{H}}_{oi} = 1, i = i_o; 0, i \neq i_o$ (Reviewer 2, pers. comm.); and p is the number of observations. By "local enough" observations we mean the common situation when the characteristic distance between non-zero entries in each row of the observation matrix $\mathbf{H}$ is much smaller than either the localisation radius or the characteristic correlation radius of the state error. In the case of direct observation of the state vector elements, that is when each row of the observations matrix $\mathbf{H}$ contains a single non-zero value equal to 1, the approximations (8) and (9) become equalities.

Because $\overset{io}{\mathbf{A}}{}^f_{i_o,:} = \mathbf{A}^f{}_{i_o,:}$, (8) can be rewritten as follows:

$$\left(\boldsymbol{\rho} \circ \mathbf{P}^f\right) \mathbf{H}^{\mathrm{T}} \approx \frac{1}{m-1} \sum_{o=1}^{p} \overset{io}{\mathbf{A}}{}^f \left(\overset{io}{\mathbf{A}}{}^f\right)^{\mathrm{T}} \overset{\{o\}}{\mathbf{H}}{}^{\mathrm{T}}. \tag{10}$$

Here, the products $\overset{io}{\mathbf{A}}{}^f \left(\overset{io}{\mathbf{A}}{}^f\right)^{\mathrm{T}}/(m-1)$ can be interpreted as "local" state error covariances used for assimilation of the corresponding observations. Note that all update schemes that allow formulation in terms of state error covariance do in fact use it only in blocks of $\mathbf{P}^f\mathbf{H}^{\mathrm{T}}$. Therefore, the above interpretation can be applied in these cases to the whole scheme rather than only to the update of the ensemble mean (1). Such schemes include the traditional EnKF [6], the Potter scheme (EnSRF [21]), the left-multiplied ESRF [19], and the "deterministic" EnKF (DEnKF, [18]); but not the right-multiplied ESRFs such as the ETKF [5], where the update of ensemble anomalies is conducted in ensemble space.

4 Local analysis

LA is another common localisation method. It uses local approximation of the state error covariance for each updated state vector element by building a virtual local spatial window around this element; this is equivalent to setting ensemble anomalies outside local window to zero during update. Unlike CL, which can not be applied to some schemes, LA is a scheme-independent method.

The initial formulations of LA [1, 9, 17] did not modify observations and ensemble members within local windows. This, however, leads to discontinuities in the analysis when an observation "enters" or "leaves" the local window as one moves from updating one state vector element to another. This undesirable phenomena has been addressed in the LETKF [15], where the impact of observations close to the boundary of the local domain is reduced by artificially increasing its error variance. We will show below that the same effect can achieved by the ensemble tapering.

Let us consider a local update of the ith element of the state vector and use the accent "i" to denote the local variables used in the update. Then, generically, this update can be written as follows:

$$\mathbf{x}^a_i = \mathbf{x}^f_i + \overset{i}{\mathbf{K}}_{i,:}\left(\overset{i}{\mathbf{d}} - \overset{i}{\mathbf{H}}\overset{i}{\mathbf{x}}{}^f\right); \tag{11}$$

$$\mathbf{A}^a_{i,:} = \mathbf{A}^f_{i,:}\overset{i}{\mathbf{T}}. \tag{12}$$

Both the Kalman gain $\overset{i}{\mathbf{K}}$ and ETM $\overset{i}{\mathbf{T}}$ above (for certainty we assume using symmetric ETKF) can be written by using blocks

$$\overset{i}{\mathbf{S}} \equiv \overset{i}{\mathbf{R}}^{-1/2}\overset{i}{\mathbf{H}}\overset{i}{\mathbf{A}}^{f}/\sqrt{m-1}: \tag{13}$$

$$\overset{i}{\mathbf{K}}_{i,:} = \mathbf{A}^{f}_{i,:}\overset{i}{\mathbf{S}}^{\mathrm{T}}\left(\mathbf{I}+\overset{i}{\mathbf{S}}\overset{i}{\mathbf{S}}^{\mathrm{T}}\right)^{-1}\overset{i}{\mathbf{R}}^{-1/2}/\sqrt{m-1}; \tag{14}$$

$$\overset{i}{\mathbf{T}} = \left(\mathbf{I}+\overset{i}{\mathbf{S}}^{\mathrm{T}}\overset{i}{\mathbf{S}}\right)^{-1/2}. \tag{15}$$

Consequently, if $\overset{i}{\mathbf{R}}$ is diagonal, then scaling the observation error variance for the oth observation $\overset{i}{\mathbf{R}}_{oo}$ by a factor k will change $\overset{i}{\mathbf{S}}$ in the same way as multiplying the oth row of the product $\overset{i}{\mathbf{H}}\overset{i}{\mathbf{A}}$ by $k^{-1/2}$. If observations are local and the factor k depends only on the distance r between the updated element and the observation, then scaling observations by using a taper function $f(r)$ will change $\overset{i}{\mathbf{S}}$ in the same way as tapering of ensemble anomalies with $f(r)^{-1/2}$.

Apart from $\overset{i}{\mathbf{S}}$, the Kalman gain in (14) depends on the local observation error covariance $\overset{i}{\mathbf{R}}^{-1/2}$. Therefore, to make tapering of ensemble anomalies completely equivalent to tapering of observation error variance, one needs also to taper innovations $\overset{i}{\mathbf{d}} - \overset{i}{\mathbf{H}}\overset{i}{\mathbf{x}}^{f}$ in (11) in the same way as the ensemble observation anomalies $\overset{i}{\mathbf{H}}\overset{i}{\mathbf{A}}^{f}$.

The expressions for the update of ensemble anomalies depend on a particular scheme used. Hence, to demonstrate equivalence of tapering of observation error variance and tapering of ensemble anomalies for the update of ensemble anomalies one needs to consider a particular scheme. For the ETKF this equivalence follows from the form of the ETM (15), which depends only on $\overset{i}{\mathbf{S}}$. For the traditional EnKF, the proof is similar to that for the update of the mean (discussed above), provided that the perturbations of observations are tapered in the same way as the innovations.

The discussion above shows that in LA with both traditional EnKF and symmetric ETKF tapering of observations error variance with a distance-dependent taper function $f(r)$ has the same effect as tapering of ensemble anomalies and innovations with the taper function $f^{-1/2}(r)$, provided that observations are local and observation errors are non-correlated. For the rest of this study, we will use the formulation of LA in terms of tapering of ensemble anomalies, as it allows us to make some direct comparisons with CL.

Figure 2 shows an example of local state error covariance used for updating the ith element in the case with a one-dimensional geometry. $\overset{i}{\mathbf{f}}$ in the left panel denotes a vector of taper coefficients for the state vector used in LA for this update. Note that during LA ensemble anomalies and ensemble mean are updated only for the centre point of the local domain. The other (non-central) elements of the local ensemble anomalies and local ensemble mean are not updated. In this sense, the local analysis is not entirely consistent: the analysed ensemble is different from that calculated in any given local update (except for the centre element of the domain).

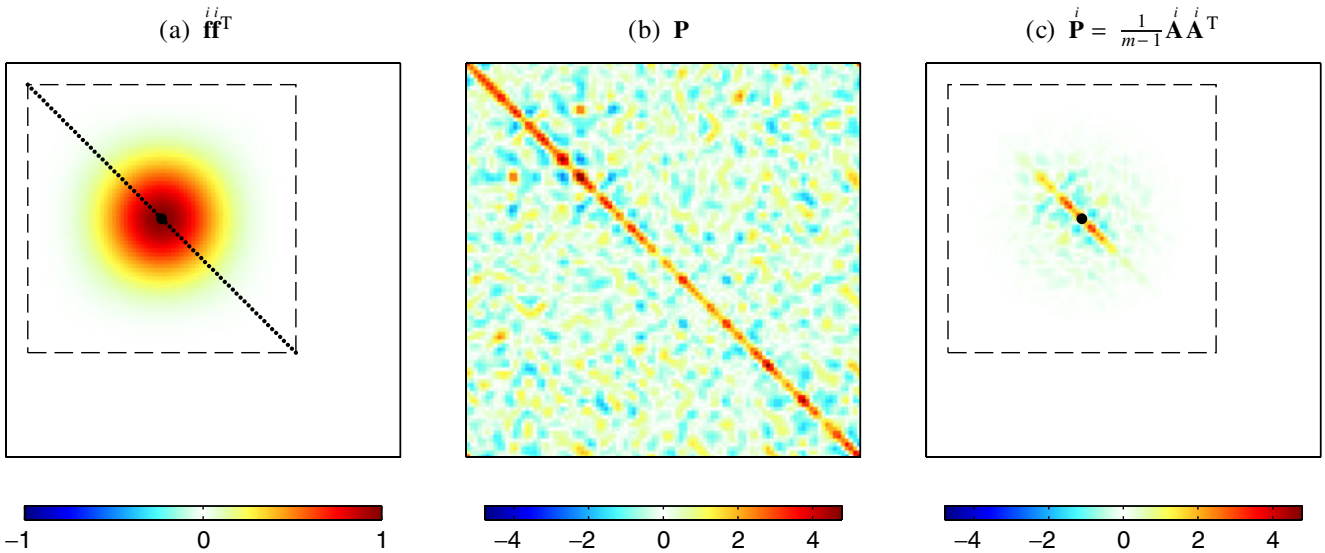


Fig. 2 Example effect of LA on the state error covariance used for updating the ith ($i = 40$) element of the state vector (see Appendix for details). The *dashed rectangle* shows the boundaries of the local domain

The conditions of local observations and non-correlated observation errors are not particularly restrictive for conducting a comparison between CL and LA, because they are essentially implied by these methods. The concept of distance between a state vector element and an observation is fundamental for the distance based localisation and implies that the observations are local. Tapering of observation error variance of individual observations in LA does only make sense if observation errors are non-correlated.[2]

5 Relation between CL and LA

In this section we investigate the relation between CL and LA, or how similar the analyses produced with these methods are.

5.1 Update of ensemble mean

The update of the ith element of ensemble mean can be written as

$$\mathbf{x}_i^a = \mathbf{x}_i^f + \overset{i}{\mathbf{K}}_{i,:}\left(\overset{(i)}{\mathbf{d}} - \overset{(i)}{\mathbf{H}}\overset{(i)}{\mathbf{x}}{}^f\right), \tag{16}$$

where brackets around the accent i mean that either tapered or non-tapered variables may be used, depending on the method. For CL it follows from (10):

$$\left[\left(\boldsymbol{\rho} \circ \mathbf{P}^f\right)\mathbf{H}^{\mathrm{T}}\right]_{i,:} \approx \frac{1}{m-1}\mathbf{A}_{i,:}^f\left(\mathbf{H}\overset{i}{\mathbf{A}}{}^f\right)^{\mathrm{T}}, \tag{17}$$

so that after straightforward transformations the local Kalman gain can be written as

$$\overset{i}{\mathbf{K}}_{i,:} \approx \mathbf{A}_{i,:}^f\overset{i}{\mathbf{S}}{}^{\mathrm{T}}\left[\mathbf{I} + \left(\tilde{\mathbf{H}}\boldsymbol{\rho}\tilde{\mathbf{H}}^{\mathrm{T}}\right) \circ \left(\mathbf{S}\mathbf{S}^{\mathrm{T}}\right)\right]^{-1}\mathbf{R}^{-1/2}/\sqrt{m-1}. \tag{18}$$

Comparing this expression with the expression (14) for LA, one can see a number of differences. Firstly, the inverted term is global in CL and local in LA. However, the difference is not important if the assimilation is "weak", that is if $\|(\mathbf{H}\boldsymbol{\rho}\mathbf{H}^{\mathrm{T}}) \circ (\mathbf{S}\mathbf{S}^{\mathrm{T}})\| \ll \|\mathbf{I}\|$. Secondly, the innovations are not tapered in CL but are tapered in LA. The importance of this depends on the rate of decay of Kalman gain with distance of observations from the updated element. If the characteristic spatial scale of variability of physical fields is much smaller than the localisation radius, then the contribution from distant observations to the increment is small, and their tapering is not essential; and vice versa. The same observation also applies to the inverted term: in the case of small spatial scale of variability of physical fields, the difference between using $[\mathbf{I} + (\tilde{\mathbf{H}}\boldsymbol{\rho}\tilde{\mathbf{H}}^{\mathrm{T}}) \circ (\mathbf{S}\mathbf{S}^{\mathrm{T}})]^{-1}$ in (18) and $(\mathbf{I} + \overset{i}{\mathbf{S}}\overset{i}{\mathbf{S}}{}^{\mathrm{T}})^{-1}$ in (14) can be small even if the assimilation is not "weak".

Figure 3 depicts an example of the gain matrices for CL and LA for two different values of the observation error variance, $\mathbf{R}_{oo} = 1$ and $\mathbf{R}_{oo} = 10^{-4}$, $o = 1, \ldots, 30$. In the case $\mathbf{R}_{oo} = 1$ we observe a considerable reduction in the ensemble spread at the observation locations, given by $\sigma \equiv \sqrt{\mathrm{tr}(\mathbf{H}\mathbf{P}\mathbf{H}^{\mathrm{T}})}$: $k_\sigma \equiv \sigma^f/\sigma^a \approx 2.1$, both for CL and LA.[3] This situation can be characterised rather as a "moderate" than "weak" assimilation, when one expects $k_\sigma - 1 \ll 1$. The reduction of the spread by a factor of about 2 is perhaps stronger than what should be typically expected in robust geophysical data assimilation systems; still, even for this case the gains for both methods in Fig. 3c are quite close. Another observation is that the gain quickly reduces with the distance to observation, so that tapering of innovation with CL would only produce a rather small effect.

The case $\mathbf{R}_{oo} = 10^{-4}$ considered in Fig. 3 is a rather extreme example of "strong" data assimilation, with $k_\sigma \approx 2 \cdot 10^2$. The two methods produce quite different results here, although the structure of the gain in (f) still looks well centred and not completely non-sensible. However, the structure of the gain matrix for LA in the whole (e) shows signs of loosing diagonal look, while it retains such look for CL (d). This difference is reflected in bigger gains for distant elements with LA (f, red line) compared with CL (f, blue line).

In Fig. 3, we presented the "local" gain for LA, used for updating a specified element of the state vector. Alternatively, one could reconstruct the "global" gain by appending the appropriate rows of the local gain matrices $\overset{i}{\mathbf{K}}_{i,:}$, $i = 1, \ldots, n$. We believe however that the overall structure of this composite gain is probably clear enough from the "local" version. The same applies to presenting the "local" version of the ETMs in the following section.

[2] From that perspective, tapering of ensemble anomalies is perhaps a more general approach, because it does not imply that restriction. From another hand, algorithmic benefits of assuming non-correlated observation errors are often too substantial, and correlation coefficients between observation errors too uncertain to account for the observation error correlation in practice.

[3] In a more general case of observations with different error variance the definition of σ could be modified as follows: $\sigma = \sqrt{\mathrm{tr}(\mathbf{H}\mathbf{P}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1})}$. This makes the value of σ invariant towards breaking / joining observations at the same location.

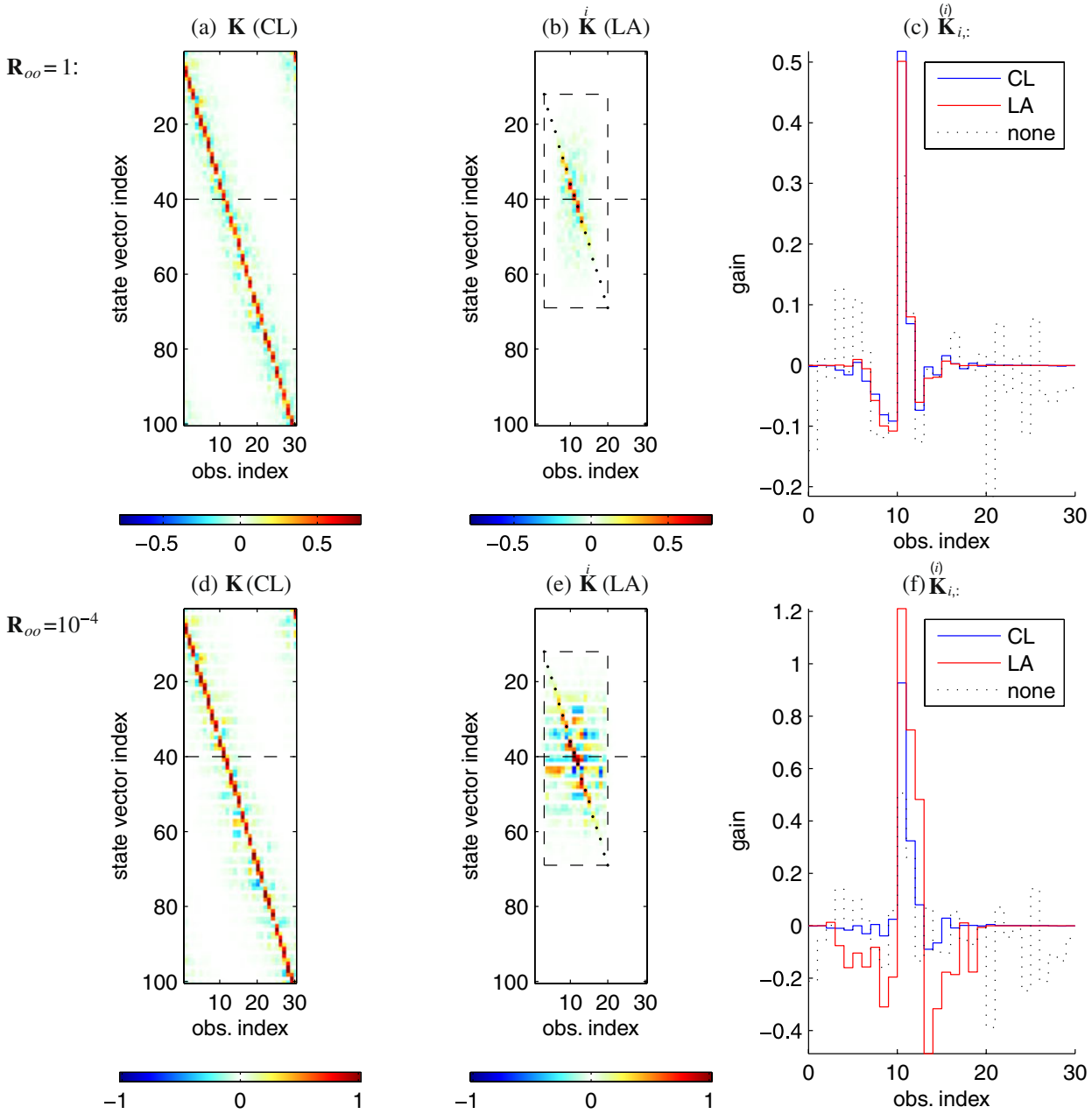


Fig. 3 Example of Kalman gain matrices for CL (**a**, **d**) and LA (**b**, **e**) methods for two different values of observation error variance, $\mathbf{R}_{oo} = 1$ (**a**–**c**) and $\mathbf{R}_{oo} = 10^{-4}$ (**d**–**f**), $o = 1, \ldots, 30$. **c**, **f** Show gain from each observation for both methods as well as for the case with no localisation, for element $i = 40$ of the state vector; the location of this element is shown by *dashed horizontal line* in panels (**a**, **b**, **d**, **e**). The *dashed rectangle* in **b** and **e** shows the boundaries of the local domain

5.2 Update of ensemble anomalies

It is usually convenient to consider the update of ensemble anomalies and the update of ensemble mean in the EnKF separately (with a possible exception of the traditional EnKF). However, in many cases the transform matrices for the update of ensemble mean can be written in terms of Kalman gain. For example, the update of ensemble anomalies in the ESRF can be written using the following left-multiplied ETM [19]:

$$\mathbf{A}^a = \mathbf{T}\mathbf{A}^f, \quad \mathbf{T} = (\mathbf{I} - \mathbf{K}\mathbf{H})^{1/2} \tag{19}$$

It is possible in such cases to make general conclusions about the update of ensemble anomalies with CL or LA based on the behaviour of Kalman gain (analysed in the previous section).

Below we conduct several numerical experiments aimed at, firstly, to confirm this conclusion; secondly, to better understand the mechanics behind it; and thirdly, to investigate whether the specifics of the behaviour of the two methods versus the assimilation strength are about the same for the update of ensemble mean and the update of ensemble anomalies. We will use the left-multiplied ESRF, a scheme that allows both CL and LA, and therefore makes it possible to compare not only the results of the ensemble update with both methods, but also the ETMs. This becomes possible because, unlike right-multiplied ETMs, which perform ensemble transformation in ensemble space, left-multiplied ETMs conduct ensemble transformation in state space. Note that while left-multiplied ETMs represent a useful analysis tool in the context of this study, because of their size ($n \times n$) their direct use in large-scale systems is not feasible, even for LA.

It is convenient to use an alternative to (19) representation for the left multiplied ETM

$$\mathbf{T} = \left(\mathbf{I} + \mathbf{P}^f \mathbf{H}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{H}\right)^{-1/2} \tag{20}$$

because of the linearity of the expression within the inverse square root function in regard to $\mathbf{P}^f$. This expression can be obtained from (19) by applying the matrix inversion lemma ([13], Section 0.7.4). For CL and LA (20) modifies as follows:

$$\text{CL:} \quad \mathbf{T} = \left[\mathbf{I} + \left(\boldsymbol{\rho} \circ \mathbf{P}^f\right) \mathbf{H}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{H}\right]^{-1/2}; \tag{21}$$

$$\text{LA:} \quad \mathbf{T}_{i,:} = \left[\left(\mathbf{I} + \overset{i}{\mathbf{P}}{}^f \mathbf{H}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{H}\right)^{-1/2}\right]_{i,:}. \tag{22}$$

Figure 4 shows the ETMs for both methods *before* calculating the inverse square root. Its right panel (c) compares the 40th row of these matrices used for the updating the corresponding row of the ensemble anomalies matrix. It can be seen that while the structures of the matrices in the figure are different, their rows corresponding to the updated element do coincide. This happens because for local observations

$$\left[\sum_{o=1}^{p} \overset{io}{\mathbf{A}}{}^f \left(\overset{io}{\mathbf{A}}{}^f\right)^{\mathrm{T}} \overset{\{o\}}{\mathbf{H}}{}^{\mathrm{T}}\right]_{i,:} = \left[\overset{i}{\mathbf{A}}{}^f \left(\overset{i}{\mathbf{A}}{}^f\right)^{\mathrm{T}} \mathbf{H}^{\mathrm{T}}\right]_{i,:},$$

so that

$$\left[\left(\boldsymbol{\rho} \circ \mathbf{P}^f\right) \mathbf{H}^{\mathrm{T}}\right]_{i,:} = \left(\overset{i}{\mathbf{P}}{}^f \mathbf{H}^{\mathrm{T}}\right)_{i,:}. \tag{23}$$

In the case of "weak" assimilation, the expressions (21) and (22) can be approximated by the first two terms of the expansion of the inverse square root into Taylor series:

$$\text{CL:} \quad \mathbf{T} = \mathbf{I} - \frac{1}{2}\left(\boldsymbol{\rho} \circ \mathbf{P}^f\right) \mathbf{H}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{H} + O\left(\left\| \left(\boldsymbol{\rho} \circ \mathbf{P}^f\right) \mathbf{H}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{H} \right\|^2\right);$$

$$\text{LA:} \quad \mathbf{T}_{i,:} = \left(\mathbf{I} - \frac{1}{2} \overset{i}{\mathbf{P}}{}^f \mathbf{H}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{H}\right)_{i,:} + O\left(\left\| \overset{i}{\mathbf{P}}{}^f \mathbf{H}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{H} \right\|^2\right).$$

Because of (23), it follows from here that the ensemble transform coefficients for both methods become equal in this case.

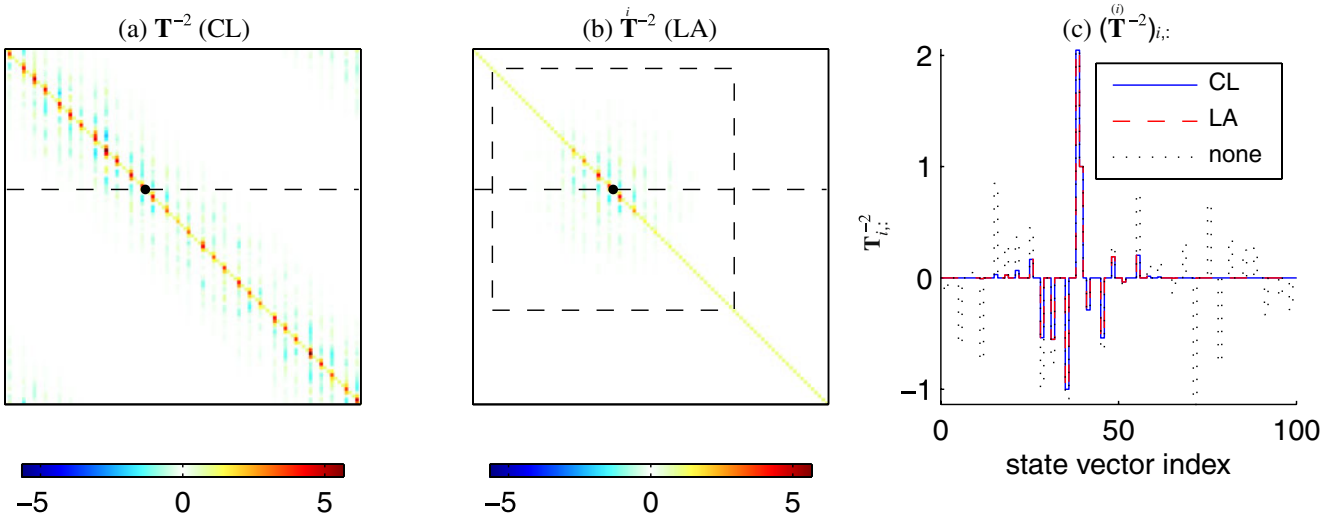


Fig. 4 Example of the ETMs for CL and LA *before* calculating the inverse square root in (21) and (22) ($\mathbf{R}_{oo} = 1$). The *black dot* shows the diagonal element corresponding to the state vector element ($i = 40$) updated with LA; the *horizontal dotted line* shows the corresponding row of the ETMs; and the *dotted rectangle* in **b** shoes the boundary of the corresponding local domain

In a general case (of assimilation of arbitrary "strength"), the inverse square root in (21) and (22) breaks the asymptotic equivalence between the ETMs for CL and LA observed for "weak" assimilation. Figure 5 compares the ETMs for CL and LA for three different values of observation error variance. The first row (a–c) with $\mathbf{R}_{oo} = 1$ corresponds to the case of "moderate" assimilation. The results obtained with both methods are still quite similar at this level of observation error variance. Both ETMs have diagonal structure, and the transform coefficients are shown in (c) to be close to each other.

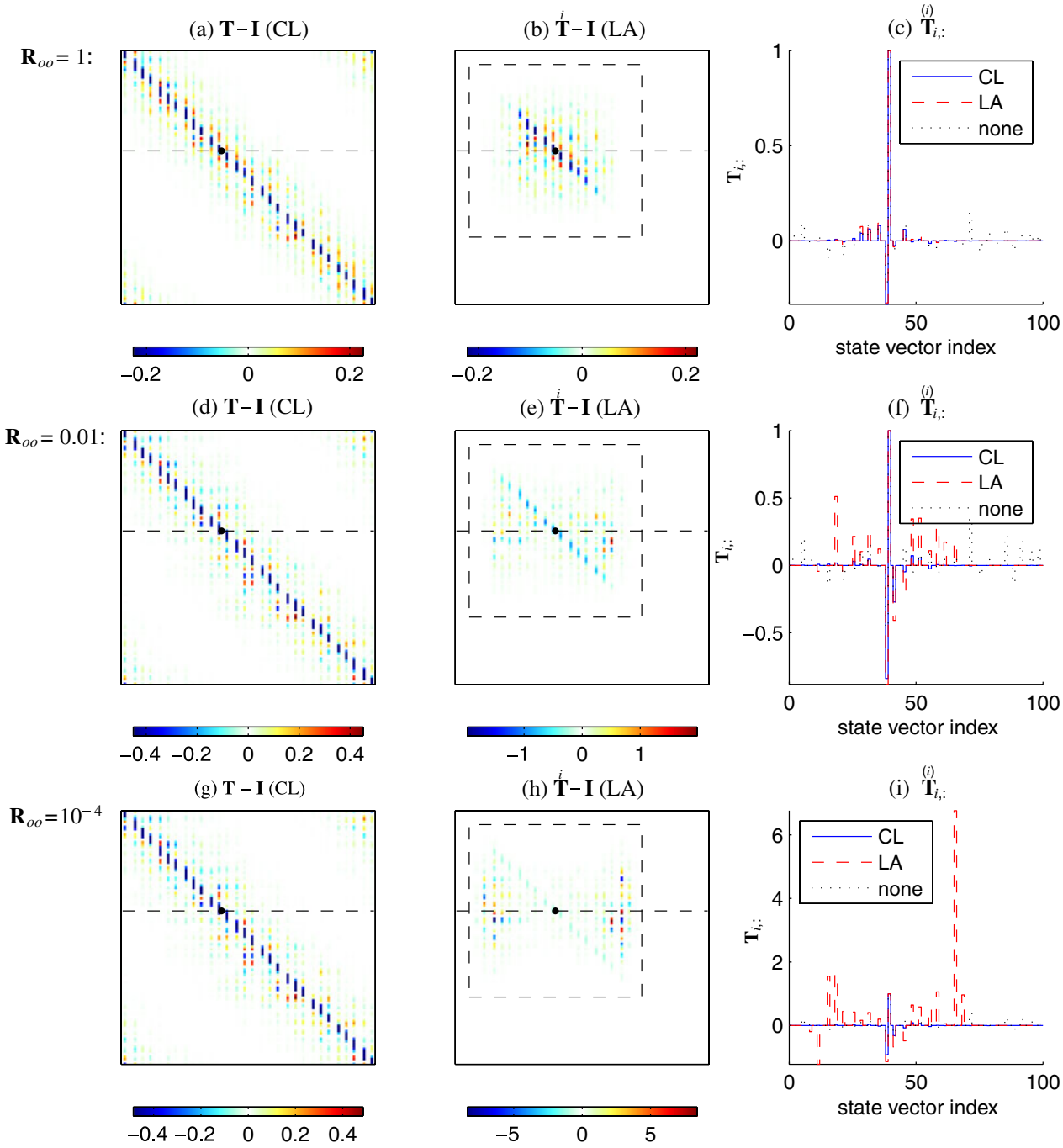


Fig. 5 Example of ETMs for CL (**a**, **d**, **g**) and LA (**b**, **e**, **h**) methods for different values of observation error variance, $\mathbf{R}_{oo} = 1, 10^{-2}, 10^{-4}$, $o = 1, \ldots, 30$. **c**, **f**, **i** Show transform coefficients for both methods as well as for the case with no localisation, for element $i = 40$ of the state vector; the location of this element is shown by *dashed horizontal line* in **a**, **b**, **d**, **e**, **g**, **h**. The *dashed rectangle* in **b**, **e**, **h** shows the boundaries of the local domain

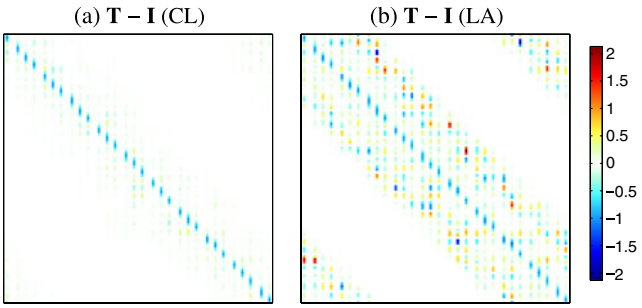


Fig. 6 Side-by-side comparison of the ETM for CL and the "global" ETM for LA for the case $\mathbf{R}_{oo} = 10^{-2}$

The second row (d–f) shows results for observation error variance set at $\mathbf{R}_{oo} = 10^{-2}$, which corresponds to a very "strong" assimilation ($k_\sigma \approx 20$). We observe that increasing the "strength" of assimilation results in some qualitative changes in the transformation of ensemble anomalies. While the structure of the ETM for CL preserves its diagonal character, the coefficients of ETM for LA start to "fill" the whole local domain and become "spiky", as can be seen from comparison of (f) and (c). Clearly, LA is falling apart at this level of assimilation.

These tendencies become even more obvious after a further increase of the assimilation strength to the extremely high level. The corresponding ETMs and coefficients are shown in the third row of the figure (g–i). They still have diagonal structure for CL, but become quite sporadic for LA.

To present the difference in structure of the ETMs for CL and LA in a more evident way, we plot side by side the ETM matrix for CL and the "global" ETM matrix for CL in Fig. 6, for the case of observation error variance $\mathbf{R}_{oo} = 10^{-2}$. The ith row of the "global" LA ETM matrix represents the ith row of the ETM matrix, used for updating the ith row of the ensemble anomalies matrix: $\mathbf{T}_{i,:} = \overset{i}{\mathbf{T}}_{i,:}$, $i = 1, \ldots, n$. While the values of the ETM coefficients around the main diagonal seem to be quite close for both methods at this (very strong) assimilation level, LA shows a lot of seemingly sporadic contributions from the elements in the whole local domain, up to its boundaries.

6 Discussion

The theoretical consideration of CL and LA in Sections 3 and 4 shows that, firstly, these methods are essentially different and, secondly, they become very close (or, with the innovation tapering, equivalent) in the case of "weak" data assimilation. The illustrating example of application of both methods considered in Section 5 shows that the results obtained with them can be quite close even in the case of "moderate" assimilation. An increase in assimilation strength causes the updates with CL and LA to diverge. When this happens, the ETMs obtained with CL remain mainly diagonal, while those obtained with LA loose their diagonal structure and become "spiky". Comparing the updates of ensemble mean and ensemble anomalies, it seems that the update of the mean shows relatively more robustness towards the increase of the assimilation strength.

The two methods only start to show some degree of difference in results at a rather strong level of assimilation. At the same time, even the weakest assimilation considered in this study (corresponding to $\mathbf{R}_{oo} = 1$ in our experimental setup) is probably higher than that expected in typical geophysical applications. "Strong" data assimilation corresponds to a strong reduction in the uncertainty of the state. This, in turn, implies a high degree of optimality of the system that assumes, in the Kalman filter framework, linearity, gaussianity, and an unbiased model—the conditions that are difficult to achieve in practice. (Recall that the covariance update is based on balancing the second order terms in regard to ensemble anomalies in the cost function—see e.g. [15], Section 2.1.)

As a consequence, the difference in CL and LA that becomes evident at a relatively extreme levels of assimilation should not be of concern in practice, and the choice between these two localisation methods needs to be done based on other criteria, such as numerical effectiveness and scalability. For example, LA allows formulation of the analysis scheme in ensemble space, while CL requires use of state space, which may be not as convenient with large-scale systems. CL may have an edge in computational effectiveness for medium-scale systems, when it is possible to store the Kalman gain in random access memory, while LA is naturally suitable for parallelisation and hence may be a good choice for large-scale systems.

Note that tapering of the forecast ensemble anomalies in (19) and innovations in (16) necessary for the asymptotic equivalence of CL to LA in the case of "weak" assimilation does essentially affect the numerical effectiveness of CL implementations, so that the use of tapering of innovations with CL in practice is probably out of question.

We did not analyse how the observed difference in updates with CL and LA is reflected in the quality of the analyses, as such experiments are beyond the scope of the paper. It was noticed though that CL seemingly produces less off-diagonal artifacts. Here, we

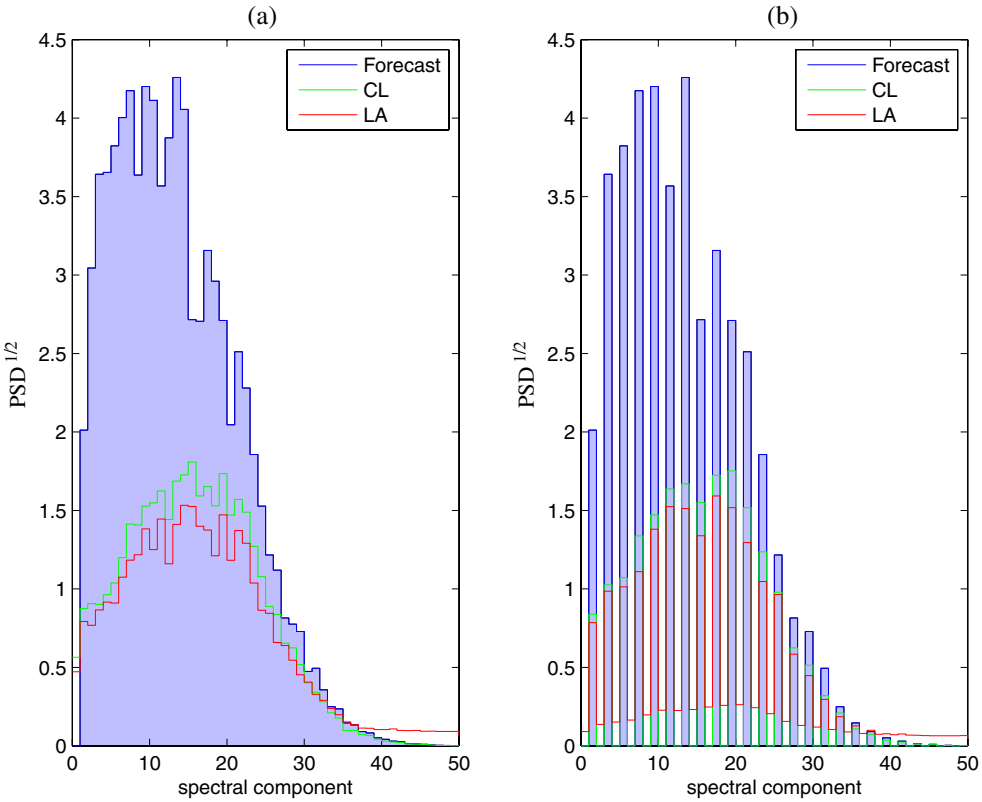


Fig. 7 Fourier spectrum of ensemble anomalies for the forecast and analyses with CL and LA. **a** For the initial ensemble used in all other experiments; **b** for the initial ensembles formed by odd harmonics only. Here, PSD is the ensemble mean power spectral density, normalised to preserve the total variance

provide just another confirmation of this impression. The left panel of Fig. 7 shows the square root of the ensemble mean Fourier power spectrum density of ensemble anomalies for the forecast and for analysis with both methods in hand, for the case of observation error variance $\mathbf{R}_{oo} = 10^{-2}$. We observe a "leak" of the variance into components with high wave numbers for LA and the absence of it for CL.

To confirm this observation, we modify this experiment by zeroing even harmonics and leaving only odd harmonics in the initial ensemble: $a^{(k)} = 0,\ k = 2, 4, 6\ldots$ in the sum in (24). The results of this experiment are shown in the right panel of Fig. 7. We observe that in the case of LA the Fourier spectrum of analysed ensemble anomalies exhibits substantial diffusion from the odd harmonics into even harmonics, while the analysis with CL is almost free from it.

Overall, it seems that at strong levels of assimilation CL produces a more robust analysis, although, as we have stressed a number of times, this difference normally should not be of a concern in practice.

Concerning the underlying reason behind the formal non-equivalence of the two methods, it is probably due to the fact that CL conducts the update in a global framework, while LA is completely local. This is particularly evident from the comparison of the ETMs, where the update coefficients for a given element of the state vector, initially the same, change in a different way during the inverse square root operation.

To compare CL and LA we used formulation of LA in terms of tapering of ensemble anomalies. Equally, after noticing that the main update schemes can be written in blocks of $\mathbf{P}^f\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}$ (in the case of local observations and non-correlated observation errors), CL can be expressed in terms of tapering of observation error variance, although we did not elaborate on this possibility.

7 Conclusions

This study investigated the relationship between two common localisation methods—covariance localisation and local analysis. It was found that in practical geophysical data assimilation both methods can be expected to produce similar results. They could be characterised as formally equivalent for the "weak" data assimilation, subject to the (normally omitted) tapering of innovations and forecast ensemble anomalies in CL; however that would downgrade the performance of the method. In many typical situations this extra tapering can be considered as unimportant because the

localisation radius is normally greater than the state error correlation radius, so that the influence of observations at distances of order of localisation radius can expected to be already rather small.

It was also shown that LA can be formulated in an equivalent way in terms of both tapering of observation error variance and tapering of ensemble anomalies. The latter makes possible the comparison between the two methods.

Acknowledgements The authors gratefully acknowledges funding from the eVITA-EnKF project by the Research Council of Norway. The authors thank two anonymous reviewers for helpful suggestions.

Appendix: Description of the experimental setup

This section describes the experimental setup used in examples. The "model state" is designed to yield perturbations of different scales, from the domain size to the grid cell size, similar to many geophysical situations. For simplicity, we employ a one-dimensional model: graphical representations of distance-based correlation matrices and local covariances from, say, a two-dimensional model can look quite complicated. Because periodic functions are used for constructing the model states, the domain is also assumed to be periodic. The state dimension, localisation radius and ensemble size are chosen in such way that use of localisation would be necessary and could produce sensible results if this ensemble occurred in practice; at least for a "weak" assimilation.

The state dimension is set to $n = 100$, and the same ensemble of $m = 21$ members is used as the forecast ensemble in all examples. Each state sample $\mathbf{x}$ is generated as a sum of $n_k = 50$ sinusoidal harmonics with random amplitude and phase:

$$x_i = \sum_{k=1}^{n_k} a^{(k)} \sin\left(\frac{2\pi k}{n} i + \phi^{(k)}\right), \quad 1 \le i \le n, \tag{24}$$

$$a^{(k)} = \text{rand}() \exp\left[-\frac{1}{2}\left(\frac{k - k_{\max}}{k_w}\right)^2\right],$$

$$\phi^{(k)} = \text{rand}() \cdot 2\pi,$$

where rand() is a function returning a uniformly distributed random number, $0 \le \text{rand}() < 1$, and $k_w = k_{\max} = 10$.

During analysis, $p = 30$ observations are assimilated. The observations are located at the state vector elements with indices closest to the locations of uniformly distributed points on interval (0.5, 100.5). The observations have non-correlated error and the same error variance. Three values of observation error variance are used: $\mathbf{R}_{oo} = 1; 10^{-2}; 10^{-4},\ o = 1, \ldots, 30$.

Localisation is performed using taper function by Gaspari and Cohn [11]:

$$f(s) = \begin{cases} 1 - \frac{5}{3}s^2 + \frac{5}{8}s^3 + \frac{1}{2}s^4 - \frac{1}{4}s^5, & 0 \le s \le 1, \\ -\frac{2}{3}s^{-1} + 4 - 5s + \frac{5}{3}s^2 & \\ \quad + \frac{5}{8}s^3 - \frac{1}{2}s^4 + \frac{1}{12}s^5, & 1 \le s \le 2, \\ 0, & 2 \le s, \end{cases} \tag{25}$$

with the localisation radius of $r_{loc} = 10$ (defined here as corresponding to $e^{1/2}$-folding distance of the taper function). Because the $e^{1/2} \approx 1.649$-folding distance of the Gaspari and Cohn's taper (25) is approximately 0.5752, $r_{loc} = 10$ corresponds to scaling $s = r \cdot 0.5752/10 = r/17.386$ between the argument s of taper function (25) and the model distance r. In our model the coordinates of state vector elements are assumed to be given by their indices, so that the maximal distance between two elements on the periodic domain is 50.

The Matlab code is available from http://enkf.nersc.no/Code/Loc_paper_Matlab_code.

References

1. Anderson, J.L.: A local least squares framework for ensemble filtering. Mon. Weather Rev. **131**, 634–642 (2003)
2. Anderson, J.L.: Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. Physica D **230**, 99–111 (2007)
3. Bishop, C.H., Hodyss, D.: Flow-adaptive moderation of spurious ensemble correlations and its use in ensemble-based data assimilation. Q. J. Royal Meteorol. Soc. **133**, 2029–2044 (2007)
4. Bishop, C.H., Hodyss, D.: Ensemble covariances adaptively localized with ECO-RAP. part 1: tests on simple error models. Tellus **61A**, 84–96 (2009)
5. Bishop, C.H., Etherton, B., Majumdar, S.J.: Adaptive sampling with the ensemble transform Kalman filter. Part I: theoretical aspects. Mon. Weather Rev. **129**, 420–436 (2001)
6. Burgers, G., van Leeuwen, P.J., Evensen, G.: Analysis scheme in the ensemble Kalman filter. Mon. Weather Rev. **126**, 1719–1724 (1998)
7. Cross, M.: Introduction to Chaos. Available at: http://www.cmp.caltech.edu/~mcc/Chaos_Course/Outline.html (2009)
8. Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics. J. Geophys. Res. **99**, 10143–10162 (1994)
9. Evensen, G.: The Ensemble Kalman filter: theoretical formulation and practical implementation. Ocean Dyn. **53**, 343–367 (2003). doi:10.1007/s110236-003-0036-9

10. Evensen, G.: Data Assimilation: the Ensemble Kalman Filter, 2nd edn. Springer, Dordrecht (2009)
11. Gaspari, G., Cohn, S.E.: Construction of correlation functions in two and three dimensions. Q. J. Royal Meteorol. Soc. **125**, 723–757 (1999)
12. Hamill, T.M., Whitaker, J.S.: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. Mon. Weather Rev. **129**, 2776–2790 (2001)
13. Horn, R.A., Johnson, C.R.: Matrix Analysis. Cambridge University Press, Cambridge, MA (1985)
14. Houtekamer, P.L., Mitchell, H.L.: A sequential ensemble Kalman filter for atmospheric data assimilation. Mon. Weather Rev. **129**, 123–137 (2001)
15. Hunt, B.R., Kostelich, E.J., Szunyogh, I.: Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter. Physica D **230**, 112–126 (2007)
16. Oke, P.R., Sakov, P., Corney, S.P.: Impacts of localisation in the EnKF and EnOI: experiments with a small model. Ocean Dyn. **57**, 32–45 (2007). doi:10.1007/sl10236-006-0088-8
17. Ott, E., Hunt, B.R., Szunyogh, I., Zimin, A.V., Kostelich, E.J., Corazza, M., Kalnay, E., Patil, D.J., Yorke, J.A.: A local ensemble Kalman filter for atmospheric data assimilation. Tellus **56A**, 415–428 (2004)
18. Sakov P, Oke P.R.: A deterministic formulation of the ensemble Kalman filter: an alternative to ensemble square root filters. Tellus **60A**, 361–371 (2008)
19. Sakov, P., Oke, P.R.: Implications of the form of the ensemble transformations in the ensemble square root filters. Mon. Weather Rev. **136**, 1042–1053 (2008)
20. Tippett, M.K., Anderson, J.L., Bishop, C.H., Hamill, T.M., Whitaker, JS.: Ensemble square root filters. Mon. Weather Rev. **131**,1485–1490 (2003)
21. Whitaker, J.S., Hamill, T.M.: Ensemble data assimilation without perturbed observations. Mon. Weather Rev. **130**, 1913–1924 (2002)
22. Whitaker, J.S., Hamill, T.M., Wei, X., Song, Y., Toth, Z.: Ensemble data assimilation with the NCEP global forecast system. Mon. Weather Rev. **136**, 463–482 (2008)