Divergence of the Ensemble Transform Kalman Filter (LETKF) by nonlocal observations

Axel Hutt

frontiers

Divergence of the Ensemble Transform Kalman Filter (LETKF) by nonlocal observations

A. Hutt

Department for Data Assimilation, Deutscher Wetterdienst, Offenbach am Main, Germany
Team MIMESIS, INRIA Nancy - Grand Est, Strasbourg, France

Correspondence*:
Axel Hutt, Team MIMESIS, INRIA Nancy - Grand Est, 1 Place de l'Hopital, 67000 Strasbourg, France
axel.hutt@inria.fr

ABSTRACT

Ensemble Kalman filters are powerful tools to merge model dynamics and observation data. For large system models, they are known to diverge due to subsampling errors at small ensemble size and thus possible spurious correlations in forecast error covariances. The Local Ensemble Transform Kalman filter (LETKF) remedies these disadvantages by localisation in observation space. However, its application to nonlocal observations is still under debate since it is still not clear how to optimally localize nonlocal observations. The present work studies intermittent divergence of filter innovations and shows that it increases forecast errors. Nonlocal observations enhance such innovation divergence under certain conditions, whereas similar localisation radius and sensitivity function width of nonlocal observations minimizes the divergence rate. The analysis of the LETKF reveals inconsistencies in the assimilation of observed and unobserved model grid points which may yield detrimental effects. These inconsistencies *inter alia* indicate that the localisation radius should be larger than the sensitivity function width if spatially synchronised system activity is expected. Moreover, the shift of observation power from observed to unobserved grid points hypothesised in the context of catastrophic filter divergence is supported for intermittent innovation divergence. Further possible mechanisms yielding such innovation divergence are ensemble member alignment and a novel covariation between background perturbations in location and observation space.

Keywords: ensemble Kalman filter, localisation, nonlocal observations, divergence, local observations

1 INTRODUCTION

Data assimilation (DA) merges models and observations to gain optimal model state estimates. It is well-established in meteorology [1], geophysics [2] and attracts attention in life sciences [3]. Typical applications of DA serve to estimate model parameters [4] or provide initial conditions for forecasts [5]. A prominent technique is the ensemble Kalman filter [6], which allows to assimilate observations in nonlinear models. When underlying models are high-dimensional, such as in geophysics or meteorology, spurious correlations in forecast errors are detrimental to state estimates. A prominent approach to avoid this effect is localisation of error covariances. The Local Ensemble Transform Kalman Filter (LETKF) [7] utilises a localisation scheme in observation space that is computationally effective and applicable to high-dimensional model

systems. The LETKF applies to local observations [8] measured in the physical system under study, e.g. by radiosondes, and nonlocal observations measured over a large area of the system by, e.g., weather radar or satellites [9, 10, 11]. Since nonlocal observations represent spatial integrals of activity, and the localisation scheme of the LETKF requests a single spatial location of each observation, it is conceptually difficult to apply the LETKF to nonlocal observations. In fact, present localisation definitions [12, 10] of nonlocal observations attempt to estimate the best single spatial location neglecting the spatial distribution of possible activity sources. A recent study [13] on satellite data assimilation proposes to choose the localisation radius equal to the spatial distribution width of radiation sources. This spatial source distribution is the sensitivity function of the nonlocal observation and is part of the model system. The present work considers the hypothesis that the relation between localisation radius and sensitivity function width plays an important role in the filter performance.

Merging the model forecast state and observations, the ensemble Kalman filter tears the analysis, i.e. the newly estimated state, towards the model forecast state and thus underestimates the forecast error covariance matrix due to a limited ensemble size [14]. This is enforced by model errors [15, 16] and leads to filter divergence. Moreover, if the forecast error covariances are too large, the forecasts have too less weight in the assimilation step and the filter tears the analysis towards the observations. This also results to filter divergence. In general terms, filter divergence occurs when an incorrect background state can not be adjusted to a better estimate of the true state by assimilating observations.
Ensemble member inflation and localisation improves the filter performance. The present work considers a perfect model and thus neglects model errors. By virtue of this study construction, all divergence effects observed result from undersampling and localisation. The present work chooses a small ensemble size compared to the model dimension, fixes the ensemble inflation to a flow-independent additive inflation and investigates the effect of localisation.

In addition to the filter divergence described above ensemble Kalman filter may exhibit catastrophic filter divergence which enhances the filter forecasts to numerical machine infinity [17, 18, 19, 20, 21]. This divergence is supposed to result from alignment of ensemble members and from unconserved observable energy dissipation [20]. This last criterion states that the filter diverges in a catastrophic manner if the observable energy of the system dissipates in unobserved directions, i.e. that energy moves from observed to unobserved locations. The present work raises the question whether such features of catastrophic divergence play a role in non-catastrophic filter divergence as well. Subsequent sections indicate that this is the case in the assimilation of nonlocal observations.

The underlying motivation of this work is the experience from meteorological data assimilation, that satellite data are detrimental to forecasts if assimilation procedure is not well-tuned [12, 13, 22]. This effect is supposed to result from deficits in the underlying model. The present work assumes a perfect model and investigates the question, whether assimilating nonlocal observations is still detrimental. Figure 1 shows forecast and analysis errors in numerical data assimilation experiments with this perfect model with three local observations only and with additional nonlocal observation. Nonlocal observations have positive and negative impact on the forecast error of the local observations dependent on the spatial location of the local observations with respect to the nonlocal observation. This preliminary result ,that additional observations increase the first guess error, is counter-intuitive at a first glance but consistent with practical experience in weather forecasting. This finding indicates that nonlocal observations renders the LETKF unstable and it diverges dependent on properties of the observations sensitivity function. What is the role of localisation in this context ? Is there a fundamental optimal relation between localisation and sensitivity function as found in [13] ? The present work addresses these questions in the following sections.

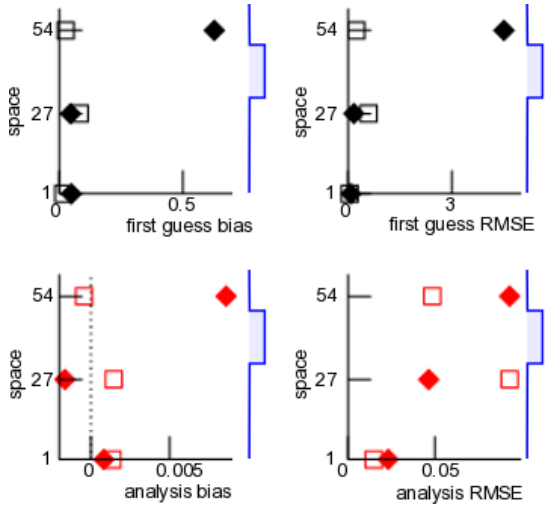

Figure 1. Example for effect of nonlocal observations on departure statistics. Verification of local observations at three spatial positions ($x = 1, 27$ and 54) for local observations only (open squares) and local observations and nonlocal observation (solid diamonds). The blue-coloured line sketch on the right hand side reflects the sensitivity function of the nonlocal observation with center at $x = 40$ and the width $r_H = 10$; the localisation radius is $r_l = 10$. Further details on the model, observations and assimilation paramaters are given in section 2.7.

The *Methods* section introduces the essential elements of the LETKF and re-calls its analytical description for a single observation in section 2.5. Section 2.8 provides conventional and new markers of filter divergence that help to elucidate possible underlying divergence mechanisms. The *Results* section presents briefly the findings, that are put into context in the *Discussion* section.

2 METHODS

2.1 The model

The storm-like Lorenz96 - model [23] is a well-established meteorological model and the present work considers an extension by a space-dependent linear damping [24]. It is a circle network with nodes of number N, whose node activity $x_k(t)$ at node k and time t obeys

$$\begin{aligned} \frac{dx_k}{dt} &= (x_{k+1} - x_{k-2})x_{k-1} \\ &\quad -(1/2 + 2\cos^4(\alpha_k \pi))x_k + I \ , \end{aligned} \tag{1}$$

with $k = 1, \ldots, N$, $x_k = x_{k+N}$ and $\alpha_k = k/N$. We choose $I = 8.0$ and $N = 80$ and the initial condition is random with $x_k(0) = 8.0 + \xi_k,\ k \neq N/2$ and $x_{N/2}(0) = 8.01 + \xi_{N/2}$ with the normal distributed random variable $\boldsymbol{\xi} \sim \mathcal{N}(0, 0.01)$. Figure 2(A) shows the model field dependent on time.

Typically, data assimilation techniques are applied to merge observations and solutions of imperfect models and the true dynamics of the underlying system is not known. To illustrate the impact of nonlocal observations, we assume (what is unrealistic in practice) that the model under consideration (1) is perfect and hence emerging differences between observations and model equivalents do not originate in the model error.

2.2 The Local Ensemble Transform Kalman Filter (LETKF)

The aim of data assimilation is to estimate a state that describes optimally both a model (or background) state $\boldsymbol{x}^b \in \mathbb{R}^N$ and corresponding observations $\boldsymbol{y} \in \mathbb{R}^S$ of number S. This *analysis* $\boldsymbol{x}^a \in \mathbb{R}^N$ minimizes the cost function

$$\begin{aligned} J(\boldsymbol{x}) = &\left(\boldsymbol{x}-\boldsymbol{x}^b\right)^t \boldsymbol{B}^{-1} \left(\boldsymbol{x}-\boldsymbol{x}^b\right) + \\ &+ \left(\boldsymbol{y}-\hat{H}(\boldsymbol{x}^b)\right)^t \boldsymbol{R}^{-1} \left(\boldsymbol{y}-\hat{H}(\boldsymbol{x}^b\right) \end{aligned} \tag{2}$$

with $\boldsymbol{x} \in \mathbb{R}^N$, the background error covariance $\boldsymbol{B} \in \mathbb{R}^{N\times N}$ and the observation error covariance $\boldsymbol{R} \in \mathbb{R}^{S\times S}$. The observation operator $\hat{H} : \mathbb{R}^N \to \mathbb{R}^S$ is linear in the present work and projects a model state into the observation space and thus links model and observations.

The LETKF estimates the background error covariance $\boldsymbol{B}$ by background-ensemble perturbations of number L

$$\boldsymbol{B} \approx \frac{1}{L-1} \boldsymbol{X}^b \left(\boldsymbol{X}^b\right)^t \tag{3}$$

with $\boldsymbol{X}^b \in \mathbb{R}^{N\times L}$. The columns of $\boldsymbol{X}^b$ are the background ensemble member perturbations $\{\boldsymbol{x}^{b,l} - \bar{\boldsymbol{x}}^b\}\ \forall l = 1,\ldots,L$, $\{\boldsymbol{x}^{b,l}\}$ is the set of background ensemble members and $\bar{\boldsymbol{x}}^b$ is the mean over the ensemble.

Then the coordinate transformation from physical space to ensemble space

$$\boldsymbol{x} = \bar{\boldsymbol{x}}^b + \boldsymbol{X}^b \boldsymbol{w} \tag{4}$$

describes a state $\boldsymbol{x}$ in the ensemble space with new coordinates $\boldsymbol{w}$ [7]. Inserting Eq. (4) into (2) yields

$$\begin{aligned} J(\boldsymbol{w}) = &(L-1)\boldsymbol{w}\boldsymbol{w}^t \\ &+ \left(\boldsymbol{y}-\bar{\boldsymbol{y}}^b-\boldsymbol{Y}\boldsymbol{w}\right)^t \boldsymbol{R}^{-1} \left(\boldsymbol{y}-\bar{\boldsymbol{y}}^b-\boldsymbol{Y}\boldsymbol{w}\right) \end{aligned} \tag{5}$$

in the new coordinate $\boldsymbol{w}$. Here $\bar{\boldsymbol{y}}^b = \hat{H}(\bar{\boldsymbol{x}}^b) \in \mathbb{R}^S$ is the model equivalent of the background ensemble mean in observation space and $\boldsymbol{Y}^b = \hat{H}(\boldsymbol{X}^b)$ is the corresponding model equivalent of $\boldsymbol{X}^b$. This implies [7]

$$\hat{H}\left(\bar{\boldsymbol{x}} + \boldsymbol{X}^b\boldsymbol{w}\right) \approx \bar{\boldsymbol{y}}^b + \boldsymbol{Y}^b\boldsymbol{w}\ ,$$

which is valid for linear observation operators.

The minimization of the cost function (5) yields

$$\bar{\boldsymbol{w}}^a \quad = \quad \boldsymbol{A}(\boldsymbol{Y}^b)^t \boldsymbol{R}^{-1} \left(\boldsymbol{y}-\bar{\boldsymbol{y}}^b\right) \tag{6}$$

with

$$\boldsymbol{A} \quad = \quad \left[(L-1)\boldsymbol{I} + \left(\boldsymbol{Y}^b\right)^t \boldsymbol{R}^{-1}\boldsymbol{Y}^b\right]^{-1} . \tag{7}$$

Equation (4) provides the analysis ensemble mean

$$\bar{\boldsymbol{x}}^a = \bar{\boldsymbol{x}}^b + \boldsymbol{X}^b \bar{\boldsymbol{w}}^a \,. \tag{8}$$

Then the square root filter-ansatz [7] yields the analysis ensemble members

$$\boldsymbol{w}^{a,l} = \bar{\boldsymbol{w}}^a + \boldsymbol{W}^{a,l} \,,$$

where $\boldsymbol{W}^{a,l}$ is the l-th column of the matrix $\boldsymbol{W}^a = [(L-1)\boldsymbol{A}]^{1/2}$. The square root of $\boldsymbol{A}$ may be computed by using the singular value decomposition $\boldsymbol{A} = \mathbf{U}\mathbf{D}\mathbf{V}^t$ with the diagonal matrix $\boldsymbol{D}$ and the eigenvector matrices $\boldsymbol{U}, \boldsymbol{V}$. This yields $\boldsymbol{A}^{1/2} = \boldsymbol{U}\boldsymbol{D}^{1/2}\boldsymbol{V}^t$.

Finally the analysis ensemble members in physical space read

$$\boldsymbol{x}^{a,l} = \bar{\boldsymbol{x}}^b + \boldsymbol{X}\bar{\boldsymbol{w}}^a + \boldsymbol{X}\boldsymbol{W}^{a,l},\ l = 1, \ldots, L \,, \tag{9}$$

see [7, 8] for more details.
Specifically, we have chosen $L = 10$ ensemble members and number of observations $S = 1$ or $S = 2$.

2.3 Observation data

In principle there are two types of observations. Local observations are measured at a single spatial location in the system, whereas nonlocal observations are integrals over a set of spatial locations. Examples for local observations are radiosondes measuring humidity and temperature in the atmosphere at a certain vertical altitude and horizontal position. Typical nonlocal observations are satellite measurements capturing the radiation in a vertical atmospheric column.

The present work considers observations

$$\boldsymbol{y} = \hat{H}(\boldsymbol{x}) + \boldsymbol{\eta} \,, \tag{10}$$

where $\boldsymbol{\eta} \in \mathbb{S}$ is Gaussian white noise with the true variance $\boldsymbol{R}_t$ and $\hat{H}(\boldsymbol{x})$ is a linear observation operator $\hat{H}(\boldsymbol{x}) = \boldsymbol{H}\boldsymbol{x},\ \boldsymbol{H} \in \mathbb{R}^{S\times N}$. In the following, the linear operator $\boldsymbol{H}$ is called sensitivity function and we adopt this name from meteorological data assimilation of nonlocal satellite data. The present work considers either nonlocal observations only ($S = 1$)

$$\begin{aligned} &H_{1n} = 1 \quad \forall\, n \in [N/2 - r_H; N/2[\ , n \in]N/2; N/2 + r_H] \\ &H_{1n} = 1 + 10^{-5} \quad , \quad n = N/2 \\ &H_{1n} = 0 \quad \text{otherwise} \end{aligned} \tag{11}$$

with sensitivity function width r_H or both observation types ($S = 2$)

$$\begin{aligned} &H_{1n} = 1 \quad \forall\, n \in [N/2 - r_H; N/2[\ , n \in]N/2; N/2 + r_H] \\ &H_{1n} = 1 + 10^{-5} \quad , \quad n = N/2 \\ &H_{2n} = 1 \quad \text{for } n = i \\ &H_{km} = 0 \quad \text{otherwise,} \end{aligned} \tag{12}$$

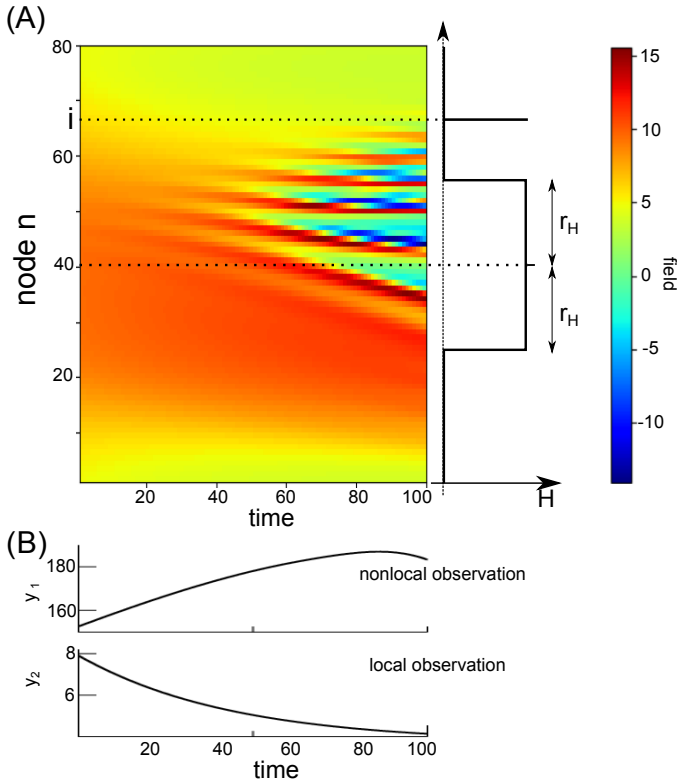

Figure 2. The model field $\mathbf{V}$ from Eq. (1) and an illustration of the observation operator $\boldsymbol{H}$ from Eq. (12) with the sensitivity function width r_H. (A) Exemplary space-time distribution of the model solution (left hand side) with the parameters $I = 8.0$ and $N = 80$, the sketched position i of a local observation and a sketched sensitivity function of nonlocal observation with the center in the middle of the spatial field with radius r_H. (B) Example observations illustrating that local and nonlocal observations are scalars and may evolve differently over time. Please also note the different order of values of the two observation types.

where the local observation is captured at spatial location i, cf. Fig. 2 for illustration. In the subsequent sections, $i = N/2$ and r_H varies in the range $1 \leq r_H \leq 10$. Please note that $r_H = 1$ approximates a local observation. Moreover, in the following a grid point whose activity contributes to an observation is called an *observed grid point* and all others are called *unobserved grid points*. Mathematically, observed (unobserved) grid points exhibit $H_{nk} \neq 0$ ($H_{nk} = 0$).

In this work, a single partial study considers a smooth sensitivity function instead of the boxcar function described above. Then the sensitivity function is the Gaspari-Cohn function $GC(n, r_H/2)$ [29] in the interval $-r_H \leq n \leq r_H$, which approximates the Gaussian function by a smooth function with finite support $2r_H$

$$H_{1n} = GC(n, r_H) \quad \forall\, n \in [-r_H; r_H]$$
$$H_{1n} = 0 \quad \text{otherwise}\,.$$

The observations $\boldsymbol{y}(t_n),\ 1, \ldots, T$ at T time instances (cf. Eq.(10)) obey the model (1) and Eq. (10) with the observation operator (11) or (12). In a large part of the work, we have assumed zero observation error $\boldsymbol{R}_t = 0$, i.e. observations are perfect in the sense that they reflect the underlying perfect model, cf. section 2.1. We take the point of view that we do not know that the model and observations are perfect and hence we guess $\boldsymbol{R}$ as it is done in cases where models and observations are not perfect.

This approach has been taken in most cases in the work. Since, however, this implicit filter error may already contribute to a filter instability or even may induce it, a short partial study has assumed perfect knowledge of the observation error. To this end, in this short partial study we have assumed $(\boldsymbol{R}_t)_{jj} = 0.1,\ j = 1, \ldots, S$ and perfect knowledge of this error, i.e. $\boldsymbol{R} = \boldsymbol{R}_t$.
Although techniques have been developed to estimate $\boldsymbol{R}$ adaptively [25], we do not employ such a scheme for simplicity.

2.4 Localisation

In the LETKF, the background covariance matrix $\boldsymbol{B}$ is expressed by L ensemble members, cf. Eq. (3), and it is rank-deficit for $L \ll N$. This leads to spurious correlations in $\boldsymbol{B}$. Spatial localisation in ensemble Kalman filters has been found to be beneficial [16, 26, 27, 28] in this context. The LETKF as defined by Hunt et al. [7] performs the localization in observation space. In detail, Hunt et al. [7] proposed to localize by increasing the observation error in matrix $\boldsymbol{R}$ dependent on the distance between the analysis grid point and observations. The present implementation follows this approach.

The observation error matrix $\boldsymbol{R}$ is diagonal, i.e. observation errors between single observations are uncorrelated. Then at each grid point i the localisation scheme considers observations y_n at location j only if the distance between location i and j does not exceed the localisation radius r_l. Then the error of observation n is $R_{nn} = R^0_{nn}/\rho_{ij}$, where $\rho_{ij} = GC(d_{ij}, r) + \varepsilon$ for $d_{ij} \leq r_l$ is the weighting function with the Gaspari-Cohn function $GC(d, r)$ [29], $\varepsilon > 0$ is a small constant ensuring a finite observation error and d_{ij} is the spatial distance between i and j. The Gaspari-Cohn function approximates a Gaussian function with standard deviation $r\sqrt{3/10}$ by a polynomial with finite support. The parameter $2r = r_l$ is the radius of the localisation function with $0 \leq GC(z, r) \leq 1,\ 0 \leq z \leq r_l$. Consequently the observation error takes its minium R^0_{nn} at distance $d_{ij} = 0$ and increases monotonously with distance to its maximum R^0_{nn}/ε at $d_{ij} = r_l$. In the present implementation, we use $\varepsilon = 10^{-7}$ and observation errors $R^0_{11} = 0.1$ for a single nonlocal observation $S = 1$ and $R^0_{nm} = 0.1\delta_{nm}, n = 1, 2$ for local and nonlocal observation with $S = 2$.

The observation error close to the border of the localisation area about a grid point i is large by definition $R_{nn} = R^0_{nn}/(GC(d \to r_l/2, r_l/2) + \varepsilon)$. In numerical practice, the assimilation effect of large values $R_{nn} > R^0_{nn}/GC_{\text{low}}$ is equivalent for some distances from the grid point i in a reasonable approximation if GC_{low} is low enough. By virtue of the monotonic decrease of $GC(d, r_l/2)$ with respect to distance $d \geq 0$, this yields the condition $GC(r_l \geq d \geq r_c, r_l/2) < GC_{\text{low}}$. In other words, for distances d larger than a corrected localisation radius r_c, the observation errors R_{nn} are that large that observations at such distances do poorly contribute to the analysis. For instance, if $GC_{\text{low}} = 0.01$, then $r_l = 5 \to r_c = 3,\ r_l = 10 \to r_c = 7$ and $r_l = 15 \to r_c = 11$. It is important to note that this corrected localisation radius depends on the width of the Gaspari-Cohn function and thus on the original localisation radius r_l, i.e. $r_c = r_c(r_l)$. In most following study cases results are given for original localisation radii r_l, while the usage of the corrected localisation radius is stated explicitly. The existencec of a corrected localisation radius r_c illustrates the insight, that there is not a single optimal localisation radius for smooth localisation functions but a certain range of equivalent localisation radii. For non-smooth localisation functions with sharp edges, e.g. a boxcar function, this variability would not exist.

The present work considers primarily nonlocal observations. Since these are not located at a single spatial site, it is non-trivial to include them in the LETKF that assumes a single observation location. To this end, several previous studies have suggested corresponding approaches [30, 31, 24, 32, 33, 34, 35, 36, 13]. A reasonable approximation for the spatial location of a nonlocal observation is the location of the maximum

sensitivity [37, 10], i.e. $\max_n H_{kn}$ of nonlocal observation k. Although this approximation has been shown to yield good results, it introduces a considerable error for broad sensitivity functions, i.e. r_H is large. In fact, this localisation scheme introduces an additional contribution to the observation error. The present implementation considers this definition. This results in the localisation of the nonlocal observation at grid point $i = N/2$.

2.5 LETKF for a single observation

In a large part of this work, we consider a single observation with $S = 1$. The subsequent paragraphs show an analytical derivation of the ensemble analysis mean and the analysis members, whose terms are interpreted in the Results section.

Considering the localisation scheme described above, at the model grid point i the analysis ensemble mean (8) reads

$$\bar{x}_i^a = \bar{x}_i^b + \left(\boldsymbol{X}_i \boldsymbol{A}_i \boldsymbol{Y}^t\right)(y_0 - \bar{y}^b)/R_i \tag{13}$$

where $\boldsymbol{Y} \in \mathbb{R}^L$ is a row vector with $Y_k = Y_{1k}^b$, with the row vector $\boldsymbol{X}_i \in \mathbb{R}^L$, $(\boldsymbol{X}_i)_k = X_{ik}$ and

$$\boldsymbol{A}_i \quad = \quad \left[(L-1)\boldsymbol{I} + \boldsymbol{Y}^t\boldsymbol{Y}/R_i\right]^{-1} \ . \tag{14}$$

The term $R_i = R_{11}^0/\rho_{i(N/2)}$ denotes the weighted observation error used at grid point i, when the observation is located at $j = N/2$, and R_{11}^0 is the error of observation y_1.

Now utilising the Woodbury matrix identity [38]

$$(\boldsymbol{B} + \boldsymbol{UCV})^{-1} =$$
$$\boldsymbol{B}^{-1} - \boldsymbol{B}^{-1}\boldsymbol{U}(\boldsymbol{C}^{-1} + \boldsymbol{V}\boldsymbol{B}^{-1}\boldsymbol{U})^{-1}\mathbf{V}\mathbf{B}^{-1}$$

for real matrices $\boldsymbol{B} \in \mathbb{R}^{n\times n}$, $\boldsymbol{U} \in \mathbb{R}^{n\times k}$, $\boldsymbol{C} \in \mathbb{R}^{k\times k}$ and $\boldsymbol{V} \in \mathbb{R}^{k\times n}$ with $n, k \in \mathbb{N}$, Eq. (14) reads

$$\begin{aligned} \boldsymbol{A}_i \quad &= \quad \frac{1}{L-1}\boldsymbol{Q}_i \\ \boldsymbol{Q}_i \quad &= \quad \boldsymbol{I} - \frac{1}{(L-1)R_i + y^2}\boldsymbol{Y}^t\boldsymbol{Y} \ , \end{aligned} \tag{15}$$

where $y = \sqrt{\boldsymbol{Y}\boldsymbol{Y}^t} \in \mathbb{R}$ is a scalar. Inserting (15) into (13), the analysis ensemble mean is

$$\bar{x}_i^a = \bar{x}_i^b + \boldsymbol{X}_i\boldsymbol{Y}^t\alpha_i \tag{16}$$

with

$$\alpha_i = \frac{y_0 - \bar{y}^b}{(L-1)R_i + y^2} \ .$$

Since $R_{N/2} = R_{11}^0$ and $R_{N/2\pm r_l} = R_{nn}^0/\varepsilon = 10^7 R_{nn}^0$, α_i takes its maximum at the observation location and is very small when the observation is at the localisation border. This means that $\bar{x}_i^a \approx \bar{x}_i^b$ at the border of the localisation area.

Now let us focus on the ensemble members. Equations (16) and (9) give the analysis ensemble members at grid point i

$$x_i^{a,l} = \bar{x}_i^b + \boldsymbol{X}_i \boldsymbol{Y}^t \alpha_i + \sqrt{L-1}\boldsymbol{X}_i \left(\sqrt{\boldsymbol{A}_i}\right)_l \ , \tag{17}$$

where $(\sqrt{\boldsymbol{A}_i})_l$ is the l-th column of matrix $\sqrt{\boldsymbol{A}_i} = \sqrt{\boldsymbol{Q}_i}/\sqrt{L-1}$.
The singular value decomposition serves as a tool to compute

$$\sqrt{\boldsymbol{Q}_i} = \boldsymbol{U}\sqrt{\boldsymbol{D}}\boldsymbol{U}^{\boldsymbol{t}} \ , \tag{18}$$

where $\sqrt{\boldsymbol{D}} \in \mathbb{R}^{L\times L}$ is diagonal and its matrix elements are the eigenvalues of $\boldsymbol{Q}$. The columns of matrix $\boldsymbol{U} \in \mathbb{R}^{L\times L}$ are the normalised eigenvectors of $\boldsymbol{Q}$. Then Eq. (7) yields

$$\begin{aligned}
\boldsymbol{Q}_i \boldsymbol{Y}^t &= \left(\boldsymbol{I} - \frac{1}{(L-1)R_i + y^2}\boldsymbol{Y}^t\boldsymbol{Y}\right)\boldsymbol{Y}^t \\
&= \boldsymbol{Y}^t - \frac{y^2}{(L-1)R_i + y^2}\boldsymbol{Y}^t \\
&= \frac{(L-1)R_i}{(L-1)R_i + y^2}\boldsymbol{Y}^t \\
&= \lambda_i \boldsymbol{Y}^t,
\end{aligned}$$

i.e. $\boldsymbol{Y}^t$ is an eigenvector of $\boldsymbol{Q}_i$ with eigenvalue $0 < \lambda_i < 1$. By virtue of the properties of R_i, λ_i takes its minimum at the observation location at $i = N/2$ and it is maximum at the localisation border.
The remaining eigenvectors of number $L-1$ are $\boldsymbol{v}_n \perp \boldsymbol{Y},\ n = 1, \ldots, L-1$ with unity eigenvalue since

$$\begin{aligned}
\boldsymbol{Q}_i \boldsymbol{v}_n &= \left(\boldsymbol{I} - \frac{1}{(L-1)R_i + y^2}\boldsymbol{Y}^t\boldsymbol{Y}\right)\boldsymbol{v}_n \\
&= \boldsymbol{v}_n - \frac{1}{(L-1)R_i + y^2}\boldsymbol{Y}^t \underbrace{\boldsymbol{Y}\boldsymbol{v}_n}_{=0} \\
&= \boldsymbol{v}_n \ .
\end{aligned}$$

Hence $\boldsymbol{U} = (\boldsymbol{Y}^t/||\boldsymbol{Y}||, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_{L-1})$ and $\sqrt{D} = \mathrm{diag}(\sqrt{\lambda_i}, 1, \ldots)$ and, after inserting into Eq. (18) and lengthy calculations

$$(\sqrt{\boldsymbol{Q}_i})_{kl} = \sqrt{\lambda_i}\frac{Y_k Y_l}{y^2} + \sum_{n=1}^{L-1}(\boldsymbol{v}_n)_k(\boldsymbol{v}_n)_l \ . \tag{19}$$

This leads to

$$x_i^{a,l} = \bar{x}_i^b + \boldsymbol{X}_i\boldsymbol{Y}^t\alpha_i + \boldsymbol{X}_i\boldsymbol{Y}^t\frac{\sqrt{\lambda_i}}{y^2}Y_l + \sum_{n=1}^{L-1}\boldsymbol{X}_i\boldsymbol{v}_n(\boldsymbol{v}_n)_l \ . \tag{20}$$

2.6 Additive covariance inflation

The ensemble Kalman filter underestimates the forecast error covariance matrix due to the limited ensemble size [39]. This problem is often adressed by covariance inflation [40, 41, 26]. The present work implements additive covariance inflation [42]. The ensemble perturbations $\boldsymbol{X}^b$ in (3) are modified by white Gaussian additive noise $\boldsymbol{\Gamma} \in \mathbb{R}^{N\times L}$

$$\boldsymbol{X}^b_{\text{add}} = \boldsymbol{X}^b + \boldsymbol{\Gamma} \, .$$

with matrix elements $\Gamma_{ij} \sim \mathcal{N}(0, f^2_{add})$ and the inflation factor $f_{add} = 0.1$.

2.7 Numerical experiments

The present study investigates solutions $\mathbf{x}(t)$ of model (1) and Eq. (10) provides the observations $\mathbf{y}(t)$. This is called the nature run. In the filter cycle, the initial analysis values are identical to the initial values of the nature run and the underlying filter model is the true model (1). In the forecast step, the model is advanced with time step $\Delta t = 10^{-3}/12$ for 100 time steps applying a 4th-order Runge-Kutta integration scheme. According to [23], the duration of one forecast step corresponds to 1 hour which is also the time between two successive observations. The analysis update is instantaneous. In an initial phase, the model evolves freely for 50 forecast steps to avoid possible initial transients. Then, the LETKF estimates the analysis ensemble according to section 2.2 during 200 cycles if not stated otherwise. One of such a numerical simulation is called a trial in the following. Each trial assumes identical initial ensemble members and the only difference in trials results from the additive noise in additive covariance inflation, cf. section 2.6.

By virtue of the primarily numerical nature of the present work, it is mandatory to vary certain parameters, such as perturbations to the observations or the factor of additive inflation. For instance, the data assimilation results in Figure 1 are based on model (1), 3 local and 1 nonlocal observation. This corresponds to the observation operator $\hat{H}$ with the sensitivity function

$$\begin{aligned} H_{1n} &= \delta_{nn_1} \\ H_{2n} &= \delta_{nn_2} \\ H_{3n} &= \delta_{nn_3} \\ H_{4n} &= 1 \quad \forall\, n \in [N/2 - r_H; N/2 + r_H] \end{aligned}$$

with $r_H = 10$ and $n_1 = 1,\ n_2 = 27,\ n_3 = 54$. The localisation radius is identical to the sensitivity function $r_l = r_H$ and data assimilation is performed during 250 filter cycles with an initial phase of 50 forecast steps. For stabilisation reasons, we have increased the model integration time step to $\Delta t = 10^{-2}/12$ but reduced the number of model integrations to 10 steps, cf. [19], thus essentially retaining the time interval between observations. Other parameters are identical to the standard setting described in the previous sections.
As mentioned above, typically the measurement process is not known in all details. For instance, the observation error is assumed to be $R = 0.1$ for the nonlocal observations, whereas the true model exhibits noise-free observations with $R_t = 0$. This is the valid setting for all simulations but few set of trials shown in Fig. 5. In a set of experiments (Fig. 5(solid, dashed and dashed-dotted line)), observations are noisy with noise perturbation variance 0.1 and hence $R_t = R = 0.1$. Moreover, the additive inflation factor is chosen to $f_{\text{add}} = 0.1$ but in two single sets of experiments (cf. Fig. 5(dashed and dashed-dotted line)), where $f_{\text{add}} = 0.05$. In addition, the weighting function of nonlocal observations is a boxcar window function with

sharp borders but in a single set of experiments, where the weighting function is a smooth Gaspari-Cohn function, cf. Fig. 5(dashed-dotted line).
The verification measures bias and RMSE are computed for the local observations only according to Eqs. (21),(22).

2.8 Divergence criteria and verification

The Kalman filter may diverge for several reasons [26, 6, 43], such as model error, insufficient sampling of error covariance or high condition number of observation operators [17, 44]. Especially the latter has been shown to be able to trigger catastrophic filter divergence of the ensemble Kalman filter exhibiting a diverging forecasts in model state space [19, 21, 20]. This divergence type exhibits a magnitude increase of model variables to machine infinity in finite time. The present implementation detects catastrophic filter divergence and stop the numerical simulation when the maximum absolute value of any single ensemble member exceeds a certain threshold $|x_k^{b,l}| > 10^{10},\ k \in [1;N],\ l \in [1;L]$.

The present work focuses primarily on a non-catastrophic filter divergence type showing a strong increase of the innovation magnitude to values much larger than the observation equivalent of the attractor. This divergence may be temporally intermittent with finite duration. Since this intermittent innovation divergence results in increased first guess departures and hence worsens forecasts, it is important to detect these divergences and control them. By definition the innovation process diverges if $\max_{l,k} |[\boldsymbol{y}_n - \boldsymbol{H}\boldsymbol{x}^{b,l}]_k| > \sigma_{th}$ for any observation n with $\sigma_{th} = 1000\sqrt{R^0_{nn}}$. Then the numerical simulation is stopped. The time of filter divergence is called T_b in the following. This criterion for innovation divergence is hard: if the innovation reaches the threshold σ_{th}, then innovation divergence occurs. The corresponding divergence rate γ is the ratio between the number of divergent and non-divergent trials. For instance, for $\gamma = 1$ all numerical trials diverge whereas $\gamma = 0$ reflect stability in all numerical trials.
Moreover, it is possible that $|[\boldsymbol{y}_n - \boldsymbol{H}\boldsymbol{x}^{b,l}]_k|$ grows intermittently but does not reach the divergence threshold. The first guess departure bias

$$\text{bias} = \frac{1}{TS}\sum_{k=1}^{T}\sum_{n=1}^{S}[\boldsymbol{y}(t_k)]_n - [\boldsymbol{H}\bar{\mathbf{x}}^b(t_k)]_n \tag{21}$$

and the corresponding root mean-square error

$$\text{RMSE} = \frac{1}{TS}\sum_{k=1}^{T}\sum_{n=1}^{S}(\boldsymbol{y}_n(t_k) - [\boldsymbol{H}\bar{\mathbf{x}}^b(t_k)]_n)^2 \tag{22}$$

quantify the forecast error in such trials. For a single observation, $\boldsymbol{y} \to y_o$. Larger values of bias RMSE indicate larger innovation values.

To quantify filter divergence, Tong et al. [18] have proposed the statistical measure

$$\Theta_n = \sqrt{\frac{1}{L}\sum_{l=1}^{L}(\boldsymbol{y}(t_n) - \boldsymbol{H}\boldsymbol{x}^{b,l}(t_n))^t(\boldsymbol{y}(t_n) - \boldsymbol{H}\boldsymbol{x}^{b,l}(t_n))}$$

and

$$\Xi_n = ||\frac{1}{L-1}\sum_{l=1}^{L}\mathbf{X}_o(t_n)\otimes\mathbf{X}_u(t_n)||$$

at time t_n, where the norm is defined by $||\mathbf{Z}|| = \sum_{n,m}|Z_{nm}|^2$ for any matrix $\boldsymbol{Z}$ and Z_{nm} are the corresponding matrix elements. The quantity Θ_n represents the ensemble spread in observation space and Ξ_n is the covariation of observed and unobserved ensemble perturbations assuming local observations. Large values of Ξ indicates catastrophic filter divergence as pointed out in [18, 20]. This definition may also apply to nonlocal observations, cf. section 2.5, although its original motivation assumes local observations. An interesting feature to estimate the degree of divergence is the time of maximum ensemble spread T_Θ and the time of maximum covariation of observed and unobserved ensemble perturbations T_Ξ:

$$\begin{aligned} T_\Theta &= \arg\max_n \Theta_n \\ T_\Xi &= \arg\max_n \Xi_n \ . \end{aligned} \tag{23}$$

Moreover, previous studies have pointed out that catastrophic filter divergence in ensemble Kalman filter implies alignment of ensemble members. This may also represent an important mechanism in non-catastrophic filter divergence. The new quantity

$$p_{a,u} = \frac{n_{a,u}}{L(L-1)/2} \tag{24}$$

is the probability of alignment and unalignment, where n_a is the number of aligned ensemble member perturbation pairs $(\mathbf{x}^{b,l}-\bar{\mathbf{x}}^b),\ (\mathbf{x}^{b,k}-\bar{\mathbf{x}}^b)$ for which

$$\cos\beta_{lk} = \frac{(\mathbf{x}^{b,l}-\bar{\mathbf{x}}^b)^t(\mathbf{x}^{b,k}-\bar{\mathbf{x}}^b)}{||\mathbf{x}^{b,l}-\bar{\mathbf{x}}^b||||\mathbf{x}^{b,k}-\bar{\mathbf{x}}^b||} \geq 0.5$$

and n_u is the number of ant-aligned member pairs with

$$\cos\beta_{lk} = \frac{(\mathbf{x}^{b,l}-\bar{\mathbf{x}}^b)^t(\mathbf{x}^{b,k}-\bar{\mathbf{x}}^b)}{||\mathbf{x}^{b,l}-\bar{\mathbf{x}}^b||||\mathbf{x}^{b,k}-\bar{\mathbf{x}}^b||} \leq -0.5$$

$\forall\ l \neq k,\ l,k = 1,\ldots,L$. The alignment (anti-alignment) condition $\cos\beta_{lk} > 0.5$ ($\cos\beta_{lk} < -0.5$) implies $-60° \leq \beta_{lk} \leq 60°$ ($120° \leq \beta_{lk}| \leq 240°$). Please note that $0 \leq p_{a,u} \leq 1$ and the larger p_a (p_u) the more ensemble members are aligned (anti-aligned) to each other.

Considering the importance of member alignment to each other for catastrophic divergence, it may be interesting to estimate the alignment degree of background member perturbation with the analysis increments $\mathbf{x}^{a,l}-\mathbf{x}^{b,l}$ by

$$\cos\alpha_l = \frac{(\mathbf{x}^{b,l}-\bar{\mathbf{x}}^b)^t(\mathbf{x}^{a,l}-\mathbf{x}^{b,l})}{||\mathbf{x}^{b,l}-\bar{\mathbf{x}}^b||||\mathbf{x}^{a,l}-\mathbf{x}^{b,l}||}\ ,\ l = 1,\ldots,L\ . \tag{25}$$

The term $\mathbf{x}^{a,l}-\mathbf{x}^{b,l}$ is the analysis ensemble member perturbation from the background members and $\mathbf{x}^{b,l}-\bar{\mathbf{x}}^b$ is the direction of the background member perturbation. If $\cos\alpha_l \to 1$ ($\cos\alpha_l \to -1$) the analysis

299 ensemble members point into the same (opposite) direction as the background ensemble members. In
300 addition,

$$q_a = \frac{n_a}{L} \quad , \quad q_u = \frac{n_u}{L} \tag{26}$$

301 are the percentages of aligned and anti-aligned ensemble members for which $\cos\alpha_l > 0.5$ (of number n_a)
302 and $\cos\alpha_l < -0.5$ (of number n_u), respectively.

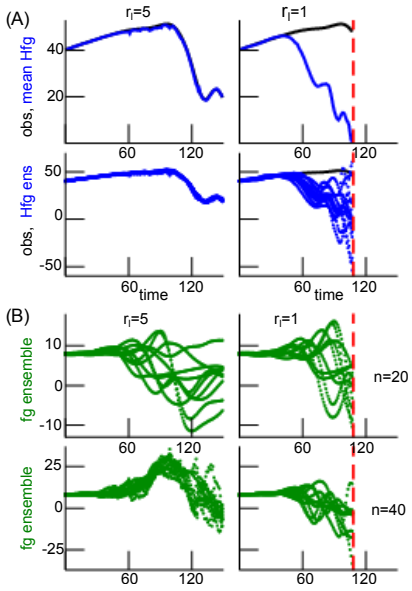

Figure 3. Temporal solutions of the filter process with $r_H = 5$ with two different localisation radii r_l. (A) Comparison of observations (black line) and model equivalents of the ensemble mean $\mathbf{y}^b$ (top row, solid blue line) and the ensemble members $\mathbf{y}^{(b,l)}$ (bottom row, dotted blue line). The time represents the number of analysis steps. (B) Ensemble members in model space at the single spatial location $n = 20$ (shown in top panel), i.e. outside the observation area with $H_{1n} = 0$, and at the single spatial location $n = 40$, i.e. in the center of the observation area (shown in bottom panel).

3 RESULTS

The stability of the ensemble Kalman filter depends heavily on the model and the nature of observations. To gain some insight into the effect of nonlocal observations, the present work considers primarily nonlocal observations only (section 3.1). Then the last section (3.2) shows briefly the divergence rates in the presence of both local and nonlocal observations.

3.1 Nonlocal observations

The subsequent sections consider nonlocal observations only and show how they affect the filter stability. To this end, the first studies are purely numerical and are complemented by an additional analytical study.

Numerical results

In order to find out how the choice of localisation radius r_l affects the stability of the LETKF, a large number of numerical experiments assist to investigate statistically under which condition the filter diverges. Figure 3 shows the temporal evolution of the background $\mathbf{x}^b$ and the model equivalents in observation space $\mathbf{y}^b$ for two different localisation radii. In Fig. 3(A) observations (black line) are very close to model equivalents (blue lines) for identical localisation and sensitivity function width, i.e. $r_l = r_H$. Conversely, observations and model-equivalents diverge after some time for $r_l \neq r_H$. This is visible in the ensemble mean (Fig. 3(A), top row) and the single ensembles (Fig. 3(A), bottom row). The different filter behavior can be observed in model space as well, but there it is less obvious, cf. Fig. 3(B). The ensemble members at spatial location $n = 40$ are located in the center of the observation area. They exhibit a rather small spread around the ensemble mean for $r_l = r_H$, whereas the ensemble spread is larger for $r_l \neq r_H$. The ensemble at $n = 20$ is outside the observation area and thus is not assimilated. There, the ensemble in $r_l = r_H$ and $r_l \neq r_H$ are close to each other.

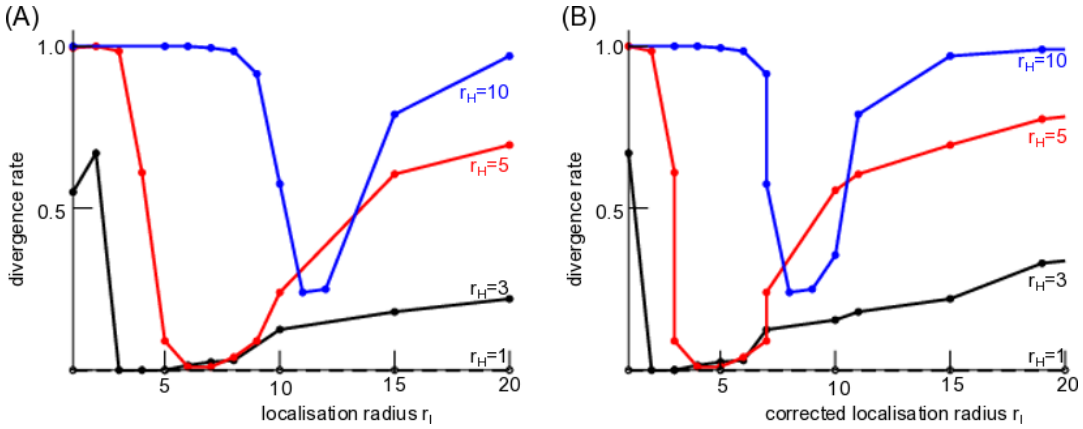

Figure 4. Stability of the LETKF of nonlocal observations dependent on the sensitivity function width r_H and the localisation radius r_l. The divergence rate γ is defined in section 2.8. (A) with original localisation radius r_l. (B) with corrected localisation radius r_c and $GC_{\text{low}} = 0.01$. Here, the observations are noise-free with $\boldsymbol{R}_t = 0$ but the chosen observation error is assumed to $\boldsymbol{R} = 0.1 \neq \boldsymbol{R}_t$ due to lack of knowledge of this true value.

This result can be generalised to a larger number of localisation and sensitivity function widths, cf. Fig. 4. For the smallest sensitivity function width and thus the smallest observation area with $r_H = 1$, no filter process diverges for a large range of localisation radii r_l, i.e. the LETKF is stable (dashed black line in Fig. 4). This case $r_H = 1$ corresponds to local observations. Now increasing the observation area with $r_H > 1$, the filter may diverge and its divergence rate γ depends on the localisation radius. We observe that the filter diverges least when the localisation radius is close to the sensitivity function width. These findings hold true for both the original localisation radius and the corrected radius r_c, cf. Methods section 2.4 and Fig. 4(A) and (B). Moreover, the filter does not exhibit catastrophic divergence before the background reaches its divergence threshold.

These results hold also true if observations are subjected to additive noise and the observation error is chosen to the true value, cf. Fig. 5(solid line) and if additive inflation is chosen to a lower value (Fig. 5(dashed line)). Similarly to Fig. 4, the divergence rate is minimum if the sensitivity function width is close to the original (Fig. 5(A)) or corrected (Fig. 5(B)) localisation radius r_l. The situation is different if the sensitivity function is not a non-smooth boxcar function as in the majority of the studies but a smooth Gaspari-Cohn function. Then the divergence rate is still minimum but the corresponding localisation radius of this minimum is much smaller than r_h, cf. dotted-dashed line in Fig. 5.
All these results consider the realistic case of a small number of ensemble members $L \ll N$. Nevertheless, it is interesting to raise the question how these results depend on the ensemble size. Figure 5(bold dotted-dashed line) indicates that a full ensemble with $L = 80$ removes the minimum with maximum divergence rate for $r_l < r_H$ and full stability for $r_l > r_H$.

The divergence criterion is conservative with a hard threshold and trials with large but sub-threshold innovations , i.e. with innovations that do not exceed the threshold, are not detected as being divergent. Nevertheless to quantify intermittent large innovations in the filter, Figure 6 shows the bias and RMSE of trials whose innovation process do not reach the divergence threshold. We observe minimum bias and RMSE for original localisation radii r_l that are similar to the sensitivity function width r_H (Fig. 6(A)). For corrected localisation radii r_c and r_H agree well at minimum bias and RMSE, cf. Fig. 6(B).

Now understanding that localisation radii $r_l \neq r_H$ may destabilize the filter, the question arises where this comes from and which mechanisms may be responsible for the innovation divergence. Figure 7 illustrates

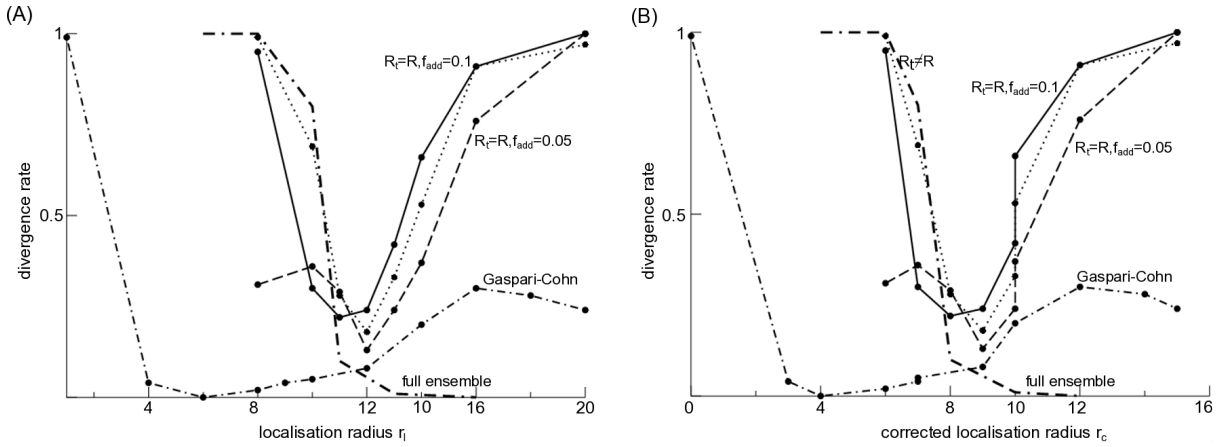

Figure 5. LETKF stability for different parameters and $r_H = 10$. The solid line denotes the divergence rate γ if the true observation error $\boldsymbol{R}_t = 0.1$ is known, i.e. $\boldsymbol{R} = \boldsymbol{R}_t$, and the inflation rate is $f_{add} = 0.1$; the dashed line denotes the divergence rate for lower inflation rate $f_{add} = 0.05$, otherwise identical to the solid line case; the dotted-dashed line marks results identical to the dashed line case but with a smooth Gaspari-Cohn sensitivity function. The dotted line is taken from Fig. 4 for comparison ($\boldsymbol{R}_t = 0$, $\boldsymbol{R} = 0.1$) and the bold dotted-dashed line represents the results with a full ensemble $L = 80$, otherwise identical to the dotted line-case. (A) original localisation radius r_l. (B) corrected localisation radius r_c with $GC_{\text{low}} = 0.01$.

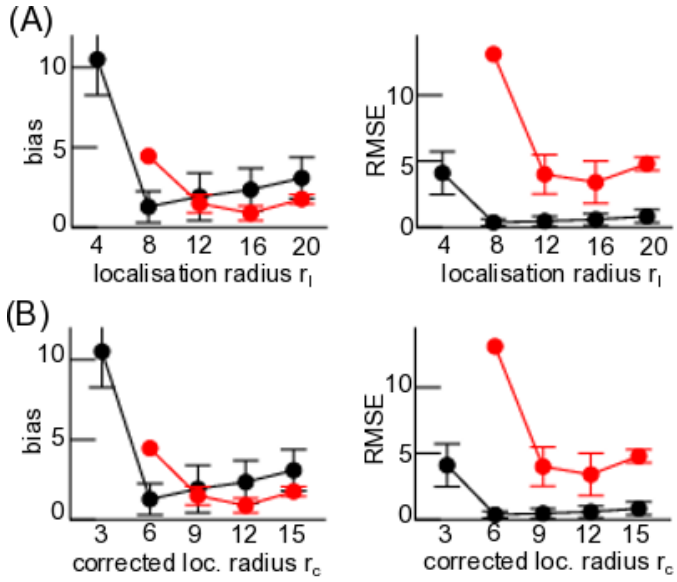

Figure 6. First guess departure statistics of trials that do not reach the divergence threshold. Here $r_H = 5$ (black) and $r_H = 10$ (red). (A) original localisation radius r_l. (B) corrected localisation radius r_c with $GC_{\text{low}} = 0.01$. All statistical measures are based on 100 trials.

various statistical quantities for three exemplary trials. These quantities have been proposed to reflect or explain divergence. The innovation-based measure Θ_n diverges (Fig. 7(B)) when the filter diverges (Fig. 7(A)) for $r_l < r_H$ and $r_l \gg r_h$, whereas Θ_n remains finite for $r_l \approx r_H$. Interestingly, for $r_l < r_H$ a certain number of ensemble members align and anti-align intermittently but do not align in the instance of divergence (Fig. 7(C)). In the case of similar localisation radius and sensitivity function width, a similar number of ensemble members align and anti-align but the filter does not diverge. Conversely, for $r_l \gg r_H$ ensemble members both align and anti-align while the filter diverges. These results already indicate a different divergence mechanism for $r_l \leq r_H$ and $r_> r_H$. Accordingly, for $r_l < r_H$ and $r_l \approx r_H$ background

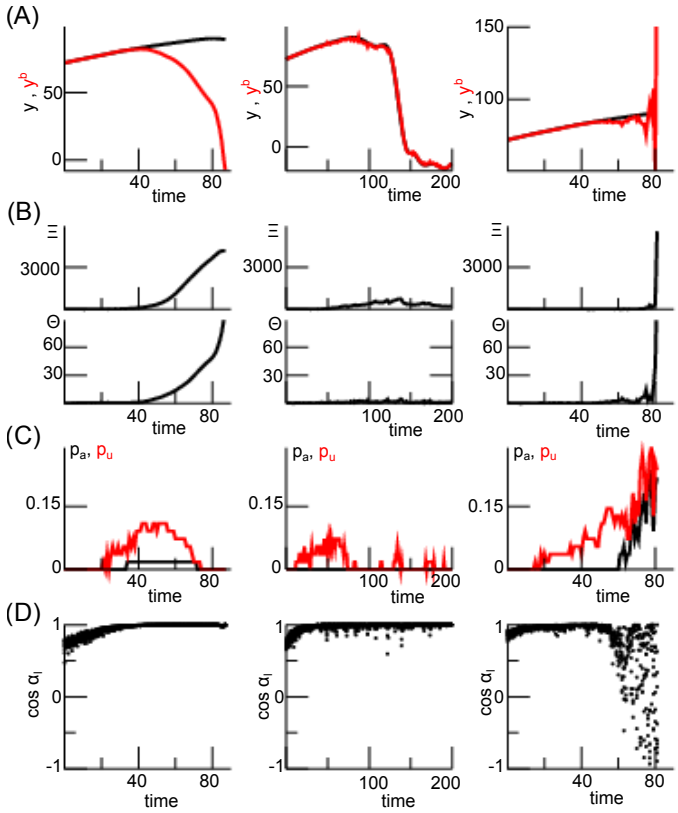

Figure 7. Various measures reflecting stability of the LETKF dependent on the localisation radius r_l in single trials. (A) observation y_o (black) and model equivalent $\mathbf{H}\bar{\mathbf{x}}^b$ (red). (B) Statistical quantities Ξ_n (top) and Θ_n (bottom), for definition see section 2.8. (C) The probability of ensemble member alignment according to Eq. (24) for aligned (black) and anti-aligned (red) members. (D) Statistical estimate of alignment between ensemble members and $\mathbf{x}^a - \mathbf{x}^b$ according to Eq. (25). The different localisation radii are $r_l = 1$ (left panel), $r_l = 6$ (center panel) and $r_l = 20$ (right panel) with the senstivity function width $r_H = 5$.

member perturbations align with the analysis member perturbations with $\cos\alpha_l \to 1$ (Fig. 7(D)), whereas $\cos\alpha_l$ fluctuates between 1 and -1 for $r_l \gg r_H$ while diverging.

Figure 8(A) shows the distribution of time instances T_Θ and T_Ξ when the respective quantities Θ_n and Ξ_n are maximum. These time instances agree well with the divergence times T_b. This confirms the single trial finding in Fig. 7(A,B) that Θ_n and Ξ_n are good markers for filter innovation divergence. Moreover only few background members align and anti-align for $r_l \leq r_H$ (small values of $p_{a,u}$), whereas many more background members align and anti-align for $r_l \gg r_H$ (Fig.8(B)). Conversely, each analysis member aligns with its corresponding background member for $r_l \leq r_H$ ($q_a = 1,\ q_u = 0$) and most analysis members still align with their background members for $r_l \gg r_H$ (Fig. 8(C)). This means that nonlocal observations do poorly affect the direction of ensemble members in these cases.

Analytical description

According to Fig. 9, there are different possible configurations of the sensitivity function with respect to the localisation area. The localisation radius r_l may be smaller (cases **1**) or larger (cases **2**) than the sensitivity width r_H or both may be equal (cases **3**). In addition, it is insightful to distinguish observed and unobserved grid points as already proposed in [18].
Now let us take a closer look at each case, cf. Fig. 9:

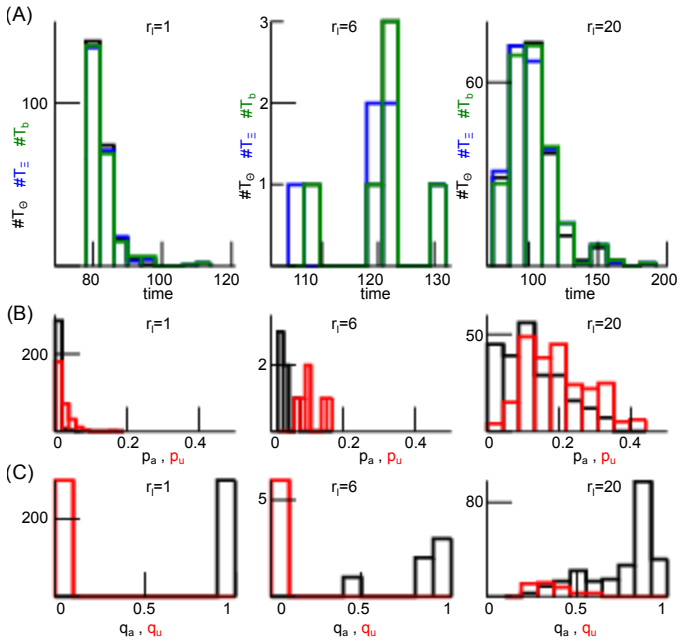

Figure 8. Divergence times and ensemble member alignment dependent on the localisation radius r_l. (A) Histogram of time of maximum Θ_n (T_Θ, black), time of maximum Ξ_n (T_Ξ, blue) and the divergence time T_b (green) , see the Methods section 2.8 for definitions. (B) Histograms of alignment ratio p_a (black) and anti-alignment ratio p_u (red) defined in Eq. (24). (C) Histograms of alignment ratio q_a (black) and anti-alignment ratio q_u (red) defined in Eq. (26). In addition $r_H = 5$ and results are based on the 200 numerical trials from Fig. 4.

- case **1.1**, $r_l \leq r_H$, $|i - \frac{N}{2}| \leq r_H$ and $|i - \frac{N}{2}| \leq r_l$: the localisation radius is smaller than the sensitivity function width and the observation at spatial location $N/2$ is located within the localisation radius about grid point i. Then, the analysis

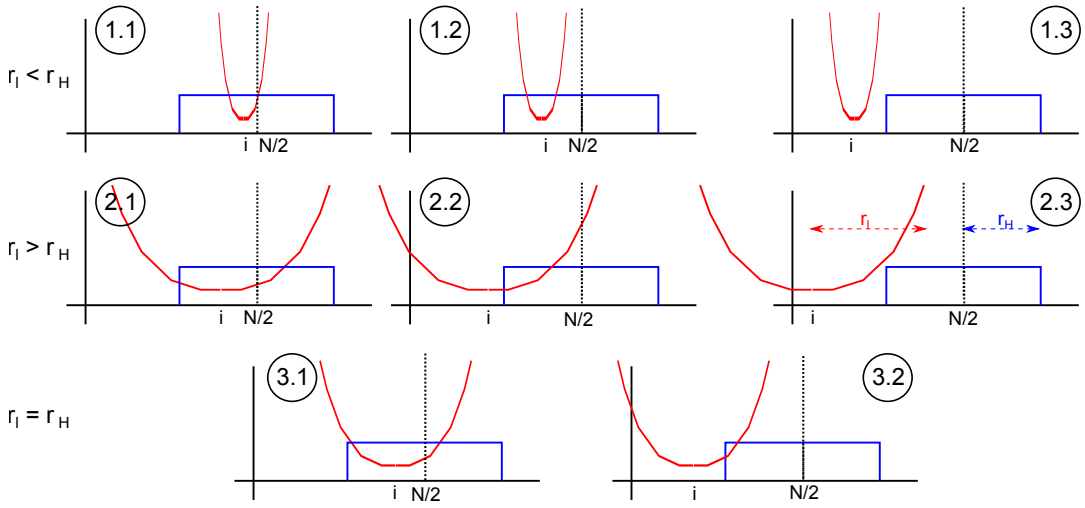

Figure 9. Sketch of different configurations of sensitivity function and localisation area. The circles denote the different cases (n.m) The sensitivity function (blue) has its center at the center of the spatial domain and the localisation function (red) is located about model grid element i.

ensemble (20) and its mean (16) read

$$\begin{aligned} x^{a,l}_{o,i} &= \bar{x}^b_{o,i} + \boldsymbol{X}_{o,i}\boldsymbol{Y}^t\alpha_i + \boldsymbol{X}_{o,i}\boldsymbol{Y}^t\frac{\sqrt{\lambda_i}}{y^2}Y_l \\ &\quad + \sum_{n=1}^{L-1}\boldsymbol{X}_{o,i}\boldsymbol{v}_n(\boldsymbol{v}_n)_l \end{aligned} \tag{27}$$

and

$$\bar{x}^a_{o,i} = \bar{x}^b_{o,i} + \boldsymbol{X}_{o,i}\boldsymbol{Y}^t\alpha_i \tag{28}$$

with the corresponding ensemble means at observed grid points $\bar{x}^b_{o,i}$ and $\bar{x}^a_{o,i}$, the first guess perturbations $\boldsymbol{X}_{o,i}$ and the analysis ensemble members $x^{a,l}_{o,i}$.

- case **1.2**, $r_l \leq r_H$, $|i - \frac{N}{2}| \leq r_H$ and $|i - \frac{N}{2}| > r_l$: compared to case **1.1**, the grid point i is observed as well but the observation is outside the localisation area; hence the analysis is identical to the first guess

$$\begin{aligned} \bar{x}^a_{o,i} &= \bar{x}^b_{o,i} \\ x^{a,l}_{o,i} &= \bar{x}^b_{o,i} + (\boldsymbol{X})_{o,i}\ . \end{aligned}$$

- case **1.3**, $r_l \leq r_H$, $|i - N/2| > r_H$ and $|i - N/2| > r_l$: the grid point i is not observed and the observation is outside the localisation area leading to

$$\begin{aligned} \bar{x}^a_{u,i} &= \bar{x}^b_{u,i} \\ x^{a,l}_{u,i} &= \bar{x}^b_{u,i} + (\boldsymbol{X})_{u,i} \end{aligned} \tag{29}$$

with the corresponding unobserved ensemble means $\bar{x}^b_{u,i}$ and $\bar{x}^a_{u,i}$, the unobserved ensemble perturbations $\boldsymbol{X}_{u,i}$ and the analysis ensemble member $x^{a,l}_{u,i}$.

- case **2.1**, $r_l > r_H$, $|i - \frac{N}{2}| \leq r_H$ and $|i - \frac{N}{2}| \leq r_l$: the localisation radius is larger than the sensitivity function width, the observation is located within the localisation radius about the grid point i and all grid points are observed. This case is equivalent to case **1.1** and the expressions for the analysis ensemble and mean hold as well.
- case **2.2**, $r_l > r_H$, $|i - \frac{N}{2}| > r_H$ and $|i - \frac{N}{2}| \leq r_l$: compared to case **2.1**, the observation is located within the localisation radius but grid points are unobserved. Then

$$\bar{x}^a_{u,i} = \bar{x}^b_{u,i} + \boldsymbol{X}_{u,i}\boldsymbol{Y}^t\alpha_i \tag{30}$$

and

$$x^{a,l}_{u,i} = \bar{x}^b_{u,i} + \boldsymbol{X}_{u,i}\boldsymbol{Y}^t\alpha_i + \boldsymbol{X}_{u,i}\boldsymbol{Y}^t\frac{\sqrt{\lambda_i}}{y^2}Y_l + \sum_{n=1}^{L-1}\boldsymbol{X}_{u,i}\boldsymbol{v}_n(\boldsymbol{v}_n)_l \,. \tag{31}$$

- case **2.3**, $r_l > r_H$, $|i - \frac{N}{2}| > r_H$ and $|i - \frac{N}{2}| > r_l$: in this case, the grid points are unobserved and the observation is outside the localisation area. Then the analysis is identical to the first guess and the case is equivalent to case **1.3**.
- case **3.1**, $r_l = r_H$, $|i - \frac{N}{2}| \leq r_H$: the observation is located within the localisation radius about the grid point i, the grid point is observed and the expressions in case **1.1** hold.
- case **3.2**, $r_l = r_H$, $|i - \frac{N}{2}| > r_H$: the observation is not located within the localisation radius of grid point i, then grid point is not observed and the expressions in case **1.3** hold.

Firstly, let us consider the limiting case of local observations with $r_H = 1$. Then case **1** does not exist. This means that case **1** emerges for nonlocal observations only and Fig. 4 demonstrates that the filter does not diverge for $1 \leq r_l \leq 20$. Moreover, the sensitivity function of the observation is non-zero at the observation location only and hence the localisation of the observation to the position of the sensitivity maximum (cf. Methods section 2.4) is trivial. In case **2**, this implies that updates at grid points far from the observation location $i \neq N/2$ consider the local observation with weighted observation error R_i. This situation changes in case of nonlocal observations with $r_H > 1$. Then case **1** exists and analysis updates in case **2** consider an errornous estimate of the nonlocal observation at the single spatial location $N/2$. The broader the sensitivity function and thus the larger r_H, the larger is the error induced by this localisation approximation. Consequently, updates at grid points far from the observation location still consider the observation with weighted observation error R_i, however the observation includes a much larger error than R_i introducing an analysis update error.

From a mathematical perspective, in cases **1.1**, **2.1** and **3.1** the LETKF updates observed grid points whereas in the cases **1.3**, **2.3** and **3.2** no update is applied. These cases appear to be consistent since grid points that contribute to the observation are updated by the observation and grid points that do not contribute to the observation are not updated. Conversely, observed grid points in case **1.2** do not consider the observation and are not updated although they contribute to the first guess in observation space. This missing update contributes to the filter error and the filter divergence as stated in previous work [12]. Moreover, the unobserved grid points in case **2.2** do consider the observation and are updated by the Kalman filter. At a very first glance, this inconsistency may be detrimental similar to case **2.1**. However, it may be arguable whether this inconsistency may contribute to the filter error. On the one hand, the background error covariance propagates information between observed to unobserved grid points in each cycle step and thus balances the missing contribution of the unobserved grid point to the observation. This may hold true for system phenomena with a large characteristic spatial scale, such as wind advection or long-range moisture transport in meteorology or, more generally, emerging long-range spatial synchronisation events. However, on the other hand, if the background error covariance represents a bad estimate, e.g. due to sampling errors or short-range synchronisation, the false (or inconsistent) update may enhance errornous propagated information and hence contributes to the filter divergence. This agrees with the vanishing divergence in case of a full ensemble (cf. Fig. 5(bold dotted-dashed line). Moreover, updates at unobserved grid points may be errornous due to model errors or the approximation error made by the reduction of

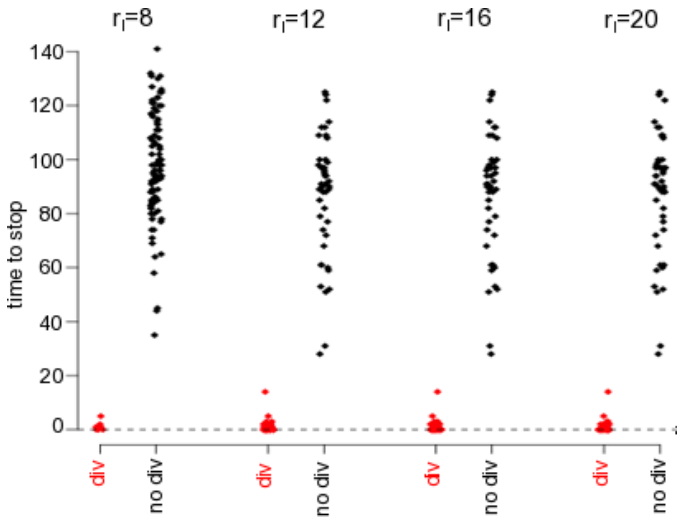

Figure 10. The divergence correlates with the weighted model-observation covariances at observed grid points A_o. The plots show the times of maxima T_o (cf. Eq. (35)) to stop, i.e. $T_{\text{stop}} - T_o$. T_o is the time when the mean model-observation covariance A_o is maximum, for divergent (red-colored with break time T_{stop}) and non-divergent (black colored with maximum time $T_{\text{stop}} = 200$) trials. Here it is $r_H = 5$.

nonlocal observations to a single location. The larger the localisation radius, the more distant are grid points to the observation location and the less representative is the localised observation to distant grid points.

Hence these two latter cases may cause detrimental effects. Consequently, cases **1** and **2**, i.e. $r_l \neq r_H$, yields bad estimates of analysis updates that make the Kalman filter diverge. Conversely, case **3**, i.e. $r_l = r_H$, involves consistent updates only and detrimental effects as described for the other cases are not present. These effects may explain enhanced filter divergence for $r_l \neq r_H$ and minimum filter divergence for $r_l = r_H$ seen in Fig. 3, and the minimum divergence rate at $r_l \approx r_H$ shown in Fig. 4.

The important terms in case **2.2**, i.e. Eqs. (30) and (31), are $\boldsymbol{X}_{u,i}\boldsymbol{Y}^t$, α_i, $\sqrt{\lambda_i}Y_l/y^2$, and $\sum_n \boldsymbol{X}_{u,i}\boldsymbol{v}_n(\boldsymbol{v}_n)_l$. Equivalently, the missing terms in case **1.2** are $\boldsymbol{X}_{o,i}\boldsymbol{Y}^t$, α_i, $\sqrt{\lambda_i}Y_l/y^2$ and $\sum_n \boldsymbol{X}_{o,i}\boldsymbol{v}_n(\boldsymbol{v}_n)_l$. For instance,

$$\boldsymbol{c}_{o,u} = \boldsymbol{X}_{o,u}\boldsymbol{Y}^t \quad = \quad \sum_{l=1}^{L}(\boldsymbol{x}_{o,u}^{b,l} - \bar{\boldsymbol{x}}_{o,u}^{b})(y^{b,l} - \bar{y}^b) \tag{32}$$

and α_i appear in both cases **2.1** and **2.2**. The terms $\boldsymbol{c}_{o,u}$ represent the covariances between model and observation perturbations over ensemble members and they may contribute differently to the intermittent divergence with increasing $|r_l - r_H|$. For a closer investigation of these terms, let us consider

$$(\boldsymbol{c}_o)_i\alpha_i \quad = \quad \boldsymbol{X}_{o,i}\boldsymbol{Y}^t\alpha_i \tag{33}$$

in case **2.1** and

$$(\boldsymbol{c}_u)_i\alpha_i = \boldsymbol{X}_{u,i}\boldsymbol{Y}^t\alpha_i \ . \tag{34}$$

in case **2.2**. These terms represent the weighted ensemble covariances between model and observation perturbations. To quantify their difference,

$$\begin{aligned} A_o &= \max_n \frac{1}{M_o} \sum_{i \in \mathcal{M}_o} (\boldsymbol{c}_o)_i(t_n)\alpha_i(t_n) \\ A_u &= \max_n \frac{1}{M_u} \sum_{i \in \mathcal{M}_u} (\boldsymbol{c}_u)_i(t_n)\alpha_i(t_n) \\ A &= A_o - A_u \end{aligned}$$

may be helpful. The term A_o (A_u) is the maximum over time of the mean of $(\boldsymbol{c}_o)_i\alpha_i$ ($(\boldsymbol{c}_u)_i\alpha_i$). This mean is computed over the set of observed (unobserved) grid points $\mathcal{M}_o$ ($\mathcal{M}_u$) with size M_o (M_u) . Consequently, A quantifies the difference of observed and unobserved weighted model-observation ensemble covariances, while the unobserved covariances are down-weighted by α_i compared to the observed covariances. This down-weighting results from the fact that unobserved grid points are more distant from the observation which yields smaller values of α_i. By definitions (33) and (34), thus $A < 0$ reflects larger weighted model-observation covariances in unobserved than observed grid points.
The corresponding quantities

$$\begin{aligned} T_o &= \arg\max_n \frac{1}{M_o} \sum_{i \in \mathcal{M}_o} (\boldsymbol{c}_o)_i(t_n)\alpha_i(t_n) \qquad (35) \\ T_u &= \arg\max_n \frac{1}{M_u} \sum_{i \in \mathcal{M}_u} (\boldsymbol{c}_u)_i(t_n)\alpha_i(t_n) \\ \Delta T &= T_o - T_u \qquad (36) \end{aligned}$$

define the time instances when these maxima are reached and ΔT is their difference. For instance, if $\Delta T > 0$, then the weighted model-observation covariances at observed grid points reach their maximum before weighted model-observation covariances at unobserved grid points.
To illustrate the importance of A_o and its corresponding occurrence time T_o, Fig. 10 shows T_o relative to the stop time T_{stop} of filter iteration, i.e. $T_{\text{stop}} - T_o$. For divergent trials, $T_{\text{stop}} = T_b$ is the time of divergence and for non-divergent trials $T_{\text{stop}} = 200$ is the maximum time. Figure 10 reveals that T_o is very close to the divergence time, whereas T_o is widely distributed about $T_o = 110$ ($T_{\text{stop}} - T_o = 90$) in non-divergent trials. This indicates that A_o is strongly correlated with the underlying divergence mechanism.

Now that that A_o is strongly correlated with the filter innovation divergence, the question arises whether the difference between weighted observed and unobserved model-observation covariances is related to the innovation divergence. Figure 11 shows the distribution of $A = A_o - A_u$ and $\Delta T = T_o - T_u$ for divergent and non-divergent experimental trials. Most trials exhibit stronger model-observation covariances in unobserved grid points than in observed grid points ($A < 0$), cf. Fig. 11(A), and the distribution variances of divergent and non-divergent trials are significantly different (Fligner-Killeen test, $p < 0.001$). Moreover, the distribution of ΔT in divergent trials is asymmetric since $\Delta T > 0$ for almost all divergent trials (see Fig. 11(B)). Hence weighted model-observation covariances in unobserved grid points reach their maximum significantly earlier than in observed grid points. Conversely the distribution of non-divergent trials is more or less symmetric about $\Delta T = 0$ (Fligner-Killeen test, $p < 0.0001$).
In this context, re-call that $A_u > A_o$ but $T_u < T_o$ in divergent trials, i.e. unobserved grid points reach their

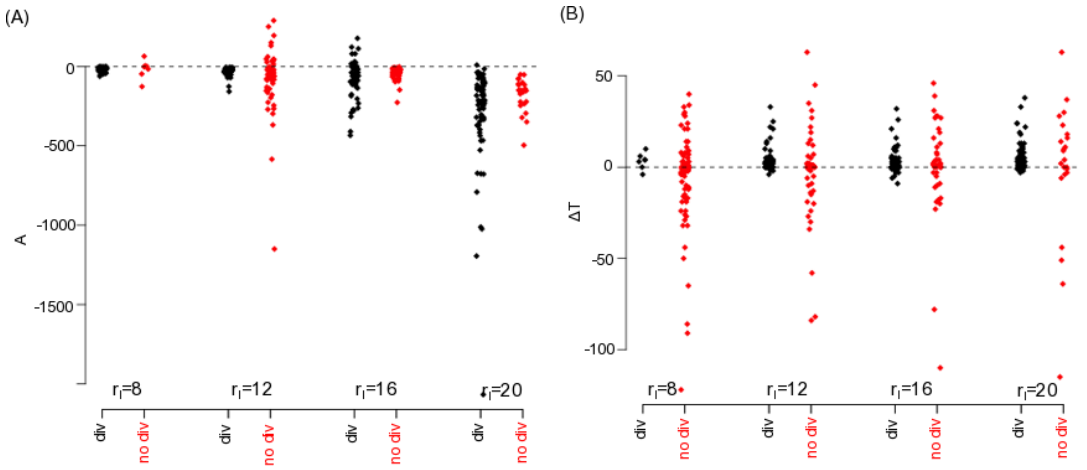

Figure 11. Comparison of weighted model-observation covariances in observed and non-observed grid points. (A) $A = A_o - A_u$ is the difference between maximum weighted model-observation covariances in observed and unobserved grid points. (B) $\Delta T = T_o - T_u$ is the difference of times when the weighted model-observation covariances reach their maximum, cf. Eq. (36). It is $r_H = 5$.

larger maximum faster than observed grid points. This indicates that the model-observation covariance $\boldsymbol{c}_u$ reflects the instability of the filter.

3.2 Local and nonlocal observations

Several international weather services apply ensemble Kalman filters and assimilate both nonlocal and local observations. Performing assimilation experiments similar to the experiments for nonlocal observations but now with a single additional local observation at grid point $i = N/2$ (cf. Methods section 2.7), the filter divergence rate γ indicates the filter stability. Figure 12 illustrates how local observations affect the filter stability in addition to nonlocal observations. For $r_H = 1$, the filter diverges rarely due to large innovations (with fewest trials at $r_l \approx 10$) but at a larger number than in the absence of local observations, cf. Fig. 4. Moreover, increasing the localisation radius yields a higher number of trials with catastrophic filter divergence with a maximum catastrophic divergence rate at $r_l \approx 10$. In sum, the least number of divergent trials occur at $r_l = r_H = 1$ (blue curve in Fig. 12). A similar stability behavior occurs for $r_H = 5$ with a minimum innovation divergence rate at $r_l \approx r_H$ and a maximum catastrophic divergence rate at $r_l \approx 10$. Again, the least number of trials diverge at $r_l = r_H$.

Figure 1 motivates the present work demonstrating that nonlocal observations yield larger first guess departures than for local observations only. Here, it is interesting to note that the numerical trial in Fig. 1 with nonlocal observations exceeds the innovation divergence threshold, cf. section 2.8, but has run over all filter cycles for illustration reasons. Moreover, several trials with the same parameters exhibit catastrophic filter divergence and the shown trial is a rare case. This divergence could have been avoided by implementing stabilising features, such as ensemble enlargement [19], adaptive localisation [28], adaptive inflation [18] or first guess check [13, 45]. However, these methods would have introduced additional assimilation effects and the gained results would not have been comparable to findings and insights in the remaining work.

4 DISCUSSION

Ensemble Kalman filtering of nonlocal observations may increase the innovation in the filter process leading to larger observation-background departure bias and RMSE, cf. Fig. 1. It is demanding to detect this innovation divergence since it is finite and transient, i.e. of finite duration. At a first glance, this negative impact is surprising since observations are thought to introduce additional knowledge to the system and thus should improve forecasts or at least retain them. To understand better why nonlocal observations may be detrimental, the present work performs numerical studies to identify markers of innovation divergence and understand their origin.

Nonlocal observations facilitates filter divergence

The majority of previous stability studies of Kalman filtering involving nonlocal observations consider catastrophic filter divergence. Kelly et al. [20] show analytically for a specific simple but non-trivial model how catastrophic filter divergence of a ensemble Kalman filter is affected by nonlocal observations. The work of Marx and Potthast [44] is an analytical discussion of the linear Kalman filter and the authors derive corresponding stability conditions. Conversely, the present work considers intermittent innovation divergence and, to our best knowledge, is one of the first to demonstrate this important effect numerically. Intermittent innovation divergence is detrimental to forecasts and are visible, e.g., in first guess departure statistics(Fig. 1). It occurs for a nonlocal observation only (Fig. 4) or for nonlocal and additional local observation (Fig. 12). This holds true for almost all localisation radii.

Optimal localisation radius

Figures 4, 5, 6 and 12 show that innovation divergence depends on the relation between sensitivity function width r_H and localisation radius r_l. The LETKF diverges least when $r_l \approx r_H$ and hence this choice of localisation radius is called optimal, i.e. the filter is least divergent. This insight agrees with the finding in an operational weather prediction framework involving the LETKF [13]. The authors consider an adaptive localisation for (nonlocal) satellite observations and choose the corresponding radius to the sensitivity function width. In two different weather situations, this tight relation improves short- and middle-range weather forecasts compared to the case of independent sensitivity width and localisation radius. Figure 9 illustrates the possible reason for the detrimental effect of different sensitivity function width and localisation radius: the LETKF is inconsistent if it updates the state at unobserved spatial

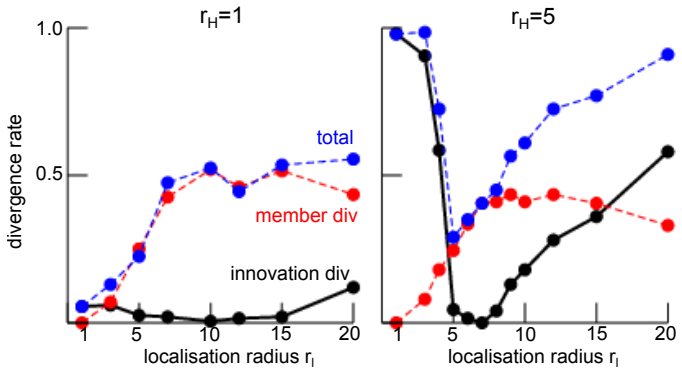

Figure 12. Rate of filter divergence γ (innovation divergence, black line) and catastrophic filter divergence (member divergence, red line) in the presence of a single local and a single nonlocal observation. The total number of divergent trials is the sum of innovation and member divergence-trials (blue line). Results are based on 200 numerical trials.

locations or does not update the state at observed spatial locations. Only if the sensitivity function and the localisation width are similar, then this detrimental effect is small. Such an inconsistency is in line with other inconsistencies in ensemble Kalman filters caused by localisation, cf. [46]. For instance, a full ensemble reduces inconsistencies for localisation radii larger than the sensitivity function width and yields filter stability (Fig. 5).

It is important to point out that, under certain conditions, it may be beneficial to further enlarge the localisation area compared to the sensitivity function. If the system's activity synchronizes on a larger spatial scale, then information is shared between observed and unobserved grid points and a larger localisation radius would be beneficial. Examples for such synchronisation phenomena are deep clouds or large-scale winds in meteorology or locally self-organised spots in physical complex systems. In other words, to decide how to choose the localisation radius one should take a closer look at the system's dynamics: if larger spatially synchronised phenomena are expected, then $r_l \gg r_H$ is preferable, otherwise $r_l \approx r_H$.

Several previous studies have derived optimal localisation radii for local observations in ensemble Kalman filter [47, 48, 49] and the specific LETKF [27, 50]. A variant of the LETKF localizes not in observation space as in the present work but in the spatial domain[34, 51, 31, 24], where the localisation of nonlocal observations has been studied as well [52]. There is the general agreement for local and non-local observations that the optimal localisation radius depends on the ensemble size and the observation error but seems to be independent on the model [50].

Origin of divergence

It is important to understand why some numerical trials diverge and some do not. Direct and indirect markers indicate which dynamical features play an important role in divergence. The most obvious direct markers are the absolute values of the innovation and the ensemble member perturbation spread Θ_n and both increase sharply during filter innovation divergence, cf. Fig. 4, 6, 7(B), 8 and 12. Similarly, the covariation of observed and unobserved background errors Ξ_n also increases during divergence. Interestingly, Θ_n and Ξ_n remain finite and take their maxima just before the instance of divergence, cf. Fig. 8. The covariation Ξ_n increases if both observed and unobserved errors increases. Kelly et al. [20] and Tong et al. [18] argue that this indicates a shift of power from observed to unobserved errors and that this shift is responsible for catastrophic divergence. The present findings support this line of argumentation and extends it to intermittent innovation divergence. This can be seen in Figure 11(A). It shows larger mean weighted model-observation error covariances (i.e. ensemble error covariances) in unobserved grid points than in observed grid points ($A < 0$) and these weighted model-observation covariances increase faster in unobserved grid points than in observed grid points. In addition, the larger the localisation radius $r_l > r_H$, the larger the ensemble error in unobserved grid points compared to observed grid points. Hence the model-observation covariance reflects a degree of instability (and thus of divergence) in the filter andn this is stronger in unobserved grid points than in observed grid points.

Figures 4, 5, 6 and 12 provide further evidence on possible error sources that yield filter divergence. The asymmetry of the divergence rates with respect to $r_l \approx r_H$ hints different underlying filter divergence contributions. If $r_l < r_H$, too few grid points are updated by the nonlocal observation (Fig. 9) although they are observed. Consequently observations include contributions from non-updated grid points which might yield large observation contributions from large model magnitudes and hence this error mechanism is rather strong. Fertig et al. [12] have identified this case as a possible source of divergence and proposed to adapt the localisation radius to the sensitivity function width. In fact, this removes case **1** in Fig. 9 and stabilises the filter for $r_l < r_H$.

For $r_l \gg r_H$, a large number of grid points are updated which, however, consider an observation with a large intrinsic error resulting from, e.g., a too low number of ensemble members. The corresponding assimilation error is more subtle than for $r_l < r_H$ and increases for larger localisation radii only. The localised nonlocal observation comprises a representation error due to the reduction of the broad sensitivity function to a single location. For small ensembles, this implicit observation error contributes to the analysis update error and, finally, to filter divergence. In sum, the two inconsistencies illustrated in Fig. 9 and derived in section 3.1 represent two possible contributions to the filter divergence for a low number of ensemble members. Conversely, for a full ensemble, intrinsic error contributions are well reduced rendering the filter more stable (Fig. 5).

Moreover, there is some evidence that ensemble member alignment may cause catastrophic filter divergence [21, 19, 20]. Figure 8 shows such indirect markers indicating weak member anti-alignment for $r_l \leq r_H$ but enhanced alignment and anti-alignment for $r_l > r_H$. The authors in [19] argue that finite ensemble sizes cause the ensemble to align in case of divergence and Ng et al. [53] show that the ensemble members may align with the most unstable phase space direction. However, our results reveal that member alignment does not represent the major mechanism for innovation divergence. Conversely, Fig. 8 provides evidence for alignment of analysis increments and background perturbations when the filter diverges. This alignment indicates that the analysis members point into the same direction as the background members. For instance, if background member perturbations point to less stable locations in phase space, then the LETKF does not correct this direction and the new analysis state is less stable, cf. the model example in [20]. This shows accordance to the reasoning in Ng et al. [53].
In addition to the alignment mechanism, Eq. (32) represents the covariation of ensemble perturbations in spatial and observation space at observed and unobserved spatial locations. For observed spatial locations, it is maximum just before the innovation divergence time. Moreover, it reaches its maximum at unobserved locations almost always before the maximum at observed locations are reached (Fig. 11). It seems this new feature represent an important contribution to the innovation divergence and future work will analyse this covariation in more detail.

Limits and outlook

The present work considers the specific case of finite low ensemble size and application of the localisation scheme. To understand better the origin of the filter divergence, it is insightful to study in detail the limiting case of large ensemble sizes, i.e. close to the model dimension, and a neglect of localisation. Although this limit is far from practice in seismology and meteorology, where the model systems are too large to study this limit, nevertheless this limit study is of theoretical interest and future work will consider it in some detail.
There is some evidence that the optimal localisation radius is flow-dependent [54, 55], whereas we assume a constant radius. In addition, the constrained choice of parameters and missing standard techniques to prevent divergence, such as adaptive inflation and adaptive observation error, limits the present work in generality and interpretation and thus makes it hard to derive decisive conclusions. Future work will implement adaptive schemes [56, 57] in a more realistic model framework.
In the majority of studies, the present work considers a non-smooth boxcar sensitivity function in order to distinguish observed and unobserved grid points. Although this simplification allows to gain deeper understanding of possible contributions to the filter divergence, the sensitivity function is unrealistic. A more realistic sensitivity function is smooth and unimodal or bimodal. Figure 5 shows that such a sensitivity function yields a minimum divergence rate but the localisation radius at the minimum rate is much smaller than the sensitivity function width. Consequently, the line of argumentation about Fig. 9 does not apply

here since there is no clear distinction of observed and unobserved grid points anymore. Future work will attempt to consider smooth unimodal sensitivity functions.
Moreover, the localisation scheme of nonlocal observations applied in the present work is very basic due to its choice of the maximum sensitivity as the observations location. Future work will investigate cut-off criteria as such in [12] that chooses the location of nonlocal observations in the range of the sensitivity function. Fertig et al. [12] also have shown that such a cut-off criterion improves first guess departure statistics and well reduces the divergence for localisation radii that are smaller than the sensitivity weighting function.
Nevertheless the present work introduces the problem of intermittent innovation divergence, extends lines of reason on the origin of filter divergence to nonlocal observation and proposes new markers of innovation divergence.

5 ACKNOWLEDGEMENT

A. Hutt would like to thank Roland Potthast for insightful hints on the stability of Kalman filters. Moreover, the author very much appreciates the valuable comments of the two (then) anonymous reviewers, whoses various insightful comments helped very much to improve the manuscript.

REFERENCES

[1] Bengtsson L, Ghil M, Källén E, editors. *Dynamic Meteorology: Data Assimilation Methods*, *Applied Mathematical Sciences*, vol. 36 (Springer) (1981).

[2] Luo X, Bhakta T, Jakobsen M, Navdal G. Efficient big data assimilation through sparse representation: A 3d benchmark case study in petroleum engineering. *PLoS One* **13** (2018) e0198586. doi:10.1371/journal.pone.0198586.

[3] Hutt A, Stannat W, Potthast R, editors. *Data Assimilation and Control: Theory and Applications in Life Sciences* (Frontiers Media) (2019). doi:10.3389/978-2-88945-985-8.

[4] Schiff SJ. *Neural Control Engineering* (Cambridge, MA: MIT Press) (2011).

[5] Nakamura G, Potthast R. *Inverse Modeling*. 2053-2563 (IOP Publishing) (2015). doi:10.1088/978-0-7503-1218-9.

[6] Asch M, Bocquet M, Nodet M. *Data Assimilation: Methods, Algorithms, and Applications* (Philadelphia: SIAM) (2016).

[7] Hunt B, Kostelich E, Szunyoghc I. Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D* **230** (2007) 112–126.

[8] Schraff C, Reich H, Rhodin A, Schomburg A, Stephan K, Perianez A, et al. Kilometre-scale ensemble data assimilation for the cosmo model (kenda). *Q. J. R. Meteorol. Soc.* **142** (2016) 1453–1472. doi:10.1002/qj.2748.

[9] Schomburg A, Schraff C, Potthast R. A concept for the assimilation of satellite cloud information in an ensemble Kalman filter: single-observation experiments. *Q. J. R. Meteorol. Soc.* **141** (2015) 893–908. doi:10.1002/qj.2748.

[10] Miyoshi T, Sato Y. Assimilating satellite radiances with a local ensemble transform Kalman filter (letkf) applied to the jma global model (gsm). *SOLA* **3** (2007) 37–40. doi:10.2151/sola.2007-010.

[11] Kurzrock F, Cros S, Ming F, Otkin J, Hutt A, Linguet L, et al. A review of the use of geostationary satellite observations in regional-scale models for short-term cloud forecasting. *Meteorologische Zeitschrift* **27** (2018) 277–298. doi:10.1127/metz/2018/0904.

[12] Fertig EJ, Hunt BR, Ott E, Szunyogh I. Assimilating non-local observations with a local ensemble Kalman filter. *Tellus A* **59** (2007) 719–730. doi:10.1111/j.1600-0870.2007.00260.x.

[13] Hutt A, Schraff C, Anlauf H, Bach L, Baldauf M, Bauernschubert E, et al. Assimilation of SEVIRI water vapour channels with an ensemble Kalman filter on the convective scale. *Front. Earth Sci.***8** (2019) 70. doi:10.3389/feart.2020.00070.

[14] Furrer R, Bengtsson T. Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. *J. Multivar. Ana.* **98** (2007) 227–255.

[15] Anderson JL. An ensemble adjustment Kalman filter for data assimiliation. *Mon. Wea. Rev.* **129** (2001) 2884–2903.

[16] Hamill TM, Whitaker JS, Snyder C. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.* **129** (2001) 2776–2790. doi:10.1175/1520-0493.

[17] Tong XT, Majda A, Kelly D. Nonlinear stability and ergodicity of ensemble based Kalman filters. *Nonlinearity* **29** (2016) 657–691. doi:10.1088/0951-7715/29/2/657.

[18] Tong XT, Majda A, Kelly D. Nonlinear stability of ensemble Kalman filters with adaptive covariance inflation. *Commun. Math. Sci.* **14** (2016) 1283–1313. doi:10.4310/cms.2016.v14.n5.a5.

[19] Gottwald G, Majda AJ. A mechanism for catastrophic filter divergence in data assimilation for sparse observation networks. *Nonlin. Processes Geophys.* **20** (2013) 705–712. doi:10.5194/npg-20-705-2013.

[20] Kelly D, Majda A, Tong XT. Concrete ensemble Kalman filters with rigorous catastrophic filter divergence. *Proc. Natl. Acad. Sci. USA.* **112** (2015) 10589–10594. doi:10.1073/pnas.1511063112.

[21] Majda A, Harlim J. Catastrophic filter divergence in filtering nonlinear dissipative systems. *Comm. Math. Sci.* **8** (2008) 27–43.

[22] Migliorini S, Candy B. All-sky satellite data assimilation of microwave temperature sounding channels at the met office. *Quart. J. Roy. Meteor. Soc.* **145** (2019) 867–883. doi:10.1002/qj.3470.

[23] Lorenz EN, Emanuel KA. Optimal sites for supplementary weather observations: Simulations with a small model. *J. Atmos. Sci.* **555** (1998) 399–414. doi:10.1175/1520-0469(1998)055⟨0399:SFSWO⟩2.0.CO;2.

[24] Bishop CH, Whitaker JS, Lei L. Gain form of the Ensemble Transform Kalman Filter and its relevance to satellite data assimilation with model space ensemble covariance localization. *Mon. Wea. Rev.* **145** (2017) 4575–4592. doi:10.1175/MWR-D-17-0102.1.

[25] Waller JA, Dance SL, Lawless AS, Nichols NK. Estimating correlated observation error statistics using an ensemble transform Kalman filter. *Tellus A* **66** (2014) 23294. doi:10.3402/tellusa.v66.23294.

[26] Houtekamer PL, Zhang F. Review of the ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.* **144** (2016) 4489–4532. doi:10.1175/MWR-D-15-0440.1.

[27] Perianez A, Reich H, Potthast R. Optimal localization for ensemble Kalman filter systems. *J. Met. Soc. Japan* **92** (2014) 585–597. doi:10.2151/jmsj.2014-605.

[28] Greybush SJ, Kalnay E, Miyoshi T, Ide K, Hunt BR. Balance and ensemble Kalman filter localization techniques. *Mon. Wea. Rev.* **139** (2011) 511–522. doi:10.1175/2010MWR3328.1.

[29] Gaspari G, Cohn S. Construction of correlation functions in two and three dimensions. *Q. J. R. Meteo. Soc.* **125** (1999) 723–757.

[30] Nadeem A, Potthast R. Transformed and generalized localization for ensemble methods in data assimilation. *Math. Meth. Appl. Sci.* **39** (2016) 619–634. doi:10.1002/mma.3496.

[31] Bishop CH, Hodyss D. Ensemble covariances adaptively localized with eco-rap. part 2: a strategy for the atmosphere. *Tellus* **61A** (2009) 97–111. doi:10.1111/j.1600-0870.2008.00372.

[32] Leng H, Song J, Lu F, Cao X. A new data assimilation scheme: the space-expanded ensemble localization Kalman filter. *Adv. Meteorol.* **2013** (2013) 410812. doi:10.1155/2013/410812.

[33] Miyoshi T, Yamane S. Local Ensemble Transform Kalman Filtering with an AGCM at a T159/L48 resolution. *Mon. Wea. Rev.* **135** (2007) 3841–3861. doi:10.1175/2007MWR1873.1.

[34] Farchi A, Boquet M. On the efficiency of covariance localisation of the ensemble Kalman filter using augmented ensembles. *Front. Appl. Math. Stat.* **5** (2019) 3. doi:10.3389/fams.2019.00003.

[35] Lei L, Whitaker JS. Model space localization is not always better than observation space localization for assimilation of satellite radiances. *Mon. Wea. Rev.* **143** (2015) 3948–3955. doi:10.1175/MWR-D-14-00413.1.

[36] Campbell WF, Bishop CH, Hodyss D. Vertical covariance localization for satellite radiances in ensemble Kalman filters. *Mon. Wea. Rev.* **138** (2010) 282–290. doi:10.1175/MWR3017.1.

[37] Houtekamer PL, Mitchell H, Pellerin G, Buehner M, Charron M, Spacek L, et al. Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Mon. Wea. Rev.* **133** (2005) 604–620. doi:10.1175/MWR-2864.1.

[38] Higham NJ. *Accuracy and stability of numerical algorithms* (SIAM), 2nd edn. (2002).

[39] Anderson JL, Anderson SL. A monte carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.* **127** (1999) 2741–2758. doi:10.1175/1520-0493.

[40] Luo X, Hoteit I. Covariance inflation in the ensemble Kalman filter: A residual nudging perspective and some implications. *Mon. Weath. Rev.* **141** (2013) 3360–3368. doi:10.1175/MWR-D-13-00067.1.

[41] Hamill TM, Whitaker JS. What constrains spread growth in forecasts initialized from ensemble Kalman filters ? *Mon. Wea. Rev.* **139** (2011) 117–131. doi:10.1175/2010MWR3246.1.

[42] Mitchell HL, Houtekamer PL. An adaptive ensemble Kalman filter. *Mon. Weath. Rev.* **128** (2000) 416–433. doi:10.1175/1520-0493.

[43] Grewal MS, Andrews AP. *Kalman filtering: Theory and practice useing MATLAB* (John Wiley & Sons), 2nd edn. (2001).

[44] Marx B, Potthast R. On instabilities in data assimilation algorithms. *Mathematics* **8** (2012) 27–43. doi:10.1007/s13137-012-0034-5.

[45] Lahoz WA, Schneider P. Data assimilation: making sense of earth observation. *Front. Environ. Sci.* **2** (2014) 16. doi:10.3389/fenvs.2014.00016.

[46] Tong XT. Performance analysis of local ensemble Kalman filter. *J. Nonlinear. Sci.* **28** (2018) 1397–1442. doi:10.1007/s00332-018-9453-2.

[47] Ying Y, Zhang F, Anderson J. On the selection of localization radius in ensemble filtering for multiscale quasigeostrophic dynamics. *Mon. Wea. Rev.* **146** (2018) 543–560. doi:10.1175/MWR-D-17-0336.1.

[48] Miyoshi T, Kondo K. A multi-scale localization approach to an ensemble Kalman filter. *SOLA* **9** (2013) 170–173.

[49] Migliorini S. Information-based data selection for ensemble data assimilation. *Quart. J. Roy. Meteor. Soc.* **139** (2013) 2033–2054.

[50] Kirchgessner P, Nerger L, Bunse-Gerstner A. On the choice of an optimal localization radius in Ensemble Kalman Filter methods. *Mon. Wea. Rev.* **142** (2014) 2165–2175. doi:10.1175/MWR-D-13-00246.1.

[51] Bishop CH, Whitaker JS, Lei L. Commentary: On the efficiency of covariance localisation of the ensemble Kalman filter using augmented ensembles by alban farchi and marc bocquet. *Front. Appl. Math. Stat.* **accepted** (2020).

[52] Lei L, Whitaker JS, Bishop C. Improving assimilation of radiance observations by implementing model space localization in an ensemble Kalman filter. *J. Adv. Model. Earth Syst.* **10** (2018) 3221–3232. doi:10.1029/2018MS001468.

[53] Ng GHC, McLaughlin D, Entekhabi D, Ahanin A. The role of model dynamics in ensemble Kalman filter performance for chaotic systems. *Tellus A* **63** (2011) 958–977.

[54] Zhen Y, Zhang F. A probabilistic approach to adaptive covariance localization for serial ensemble square root filters. *Mon. Wea. Rev.* **142** (2014) 4499–4518.

[55] Flowerdew J. Towards a theory of optimal localisation. *Tellus A* **67** (2015) 25257.

[56] Lee Y, Majda AJ, Qi D. Preventing catastrophic filter divergence using adaptive additive inflation for baroclinic turbulence. *Mon. Wea. Rev.* **145** (2017) 669–682. doi:10.1175/MWR-D-16-0121.1.

[57] Miyoshi T. The gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform Kalman filter. *Mon. Wea. Rev.* **139** (2011) 1519–1535. doi:10.1175/2010MWR3570.1.