



HAL
open science

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

Axel Hutt

► **To cite this version:**

Axel Hutt. Divergence of the Ensemble Transform Kalman Filter (LETKF) by nonlocal observations. *Frontiers in Applied Mathematics and Statistics*, 2020, 10.3389/fams.2020.00042 . hal-02861799

HAL Id: hal-02861799

<https://inria.hal.science/hal-02861799>

Submitted on 9 Jun 2020

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

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

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

2 ABSTRACT

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

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

1 INTRODUCTION

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

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

40 Merging the model forecast state and observations, the ensemble Kalman filter tears the analysis, i.e.
41 the newly estimated state, towards the model forecast state and thus underestimates the forecast error
42 covariance matrix due to a limited ensemble size [14]. This is enforced by model errors [15, 16] and leads
43 to filter divergence. Moreover, if the forecast error covariances are too large, the forecasts have too less
44 weight in the assimilation step and the filter tears the analysis towards the observations. This also results to
45 filter divergence. In general terms, filter divergence occurs when an incorrect background state can not be
46 adjusted to a better estimate of the true state by assimilating observations.

47 Ensemble member inflation and localisation improves the filter performance. The present work considers a
48 perfect model and thus neglects model errors. By virtue of this study construction, all divergence effects
49 observed result from undersampling and localisation. The present work chooses a small ensemble size
50 compared to the model dimension, fixes the ensemble inflation to a flow-independent additive inflation and
51 investigates the effect of localisation.

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

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

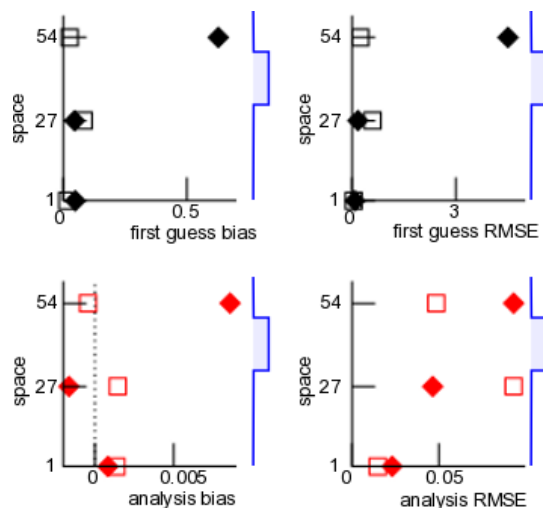


Figure 1. Example for effect of nonlocal observations on departure statistics. Verification of local observations at three spatial positions ($x = 1, 27$ and 54) for local observations only (open squares) and local observations and nonlocal observation (solid diamonds). The blue-coloured line sketch 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 parameters are given in section 2.7.

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

2 METHODS

77 2.1 The model

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

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

81 with $k = 1, \dots, N$, $x_k = x_{k+N}$ and $\alpha_k = k/N$. We choose $I = 8.0$ and $N = 80$ and the initial condition
 82 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
 83 random variable $\xi \sim \mathcal{N}(0, 0.01)$. Figure 2(A) shows the model field dependent on time.

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

89 2.2 The Local Ensemble Transform Kalman Filter (LETKF)

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

$$J(\mathbf{x}) = (\mathbf{x} - \mathbf{x}^b)^t \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}^b) + \left(\mathbf{y} - \hat{H}(\mathbf{x}^b)\right)^t \mathbf{R}^{-1} (\mathbf{y} - \hat{H}(\mathbf{x}^b)) \quad (2)$$

93 with $\mathbf{x} \in \mathbb{R}^N$, the background error covariance $\mathbf{B} \in \mathbb{R}^{N \times N}$ and the observation error covariance
 94 $\mathbf{R} \in \mathbb{R}^{S \times S}$. The observation operator $\hat{H} : \mathbb{R}^N \rightarrow \mathbb{R}^S$ is linear in the present work and projects a model
 95 state into the observation space and thus links model and observations.

96 The LETKF estimates the background error covariance \mathbf{B} by background-ensemble perturbations of
 97 number L

$$\mathbf{B} \approx \frac{1}{L-1} \mathbf{X}^b (\mathbf{X}^b)^t \quad (3)$$

98 with $\mathbf{X}^b \in \mathbb{R}^{N \times L}$. The columns of \mathbf{X}^b are the background ensemble member perturbations $\{\mathbf{x}^{b,l} -$
 99 $\bar{\mathbf{x}}^b\} \forall l = 1, \dots, L$, $\{\mathbf{x}^{b,l}\}$ is the set of background ensemble members and $\bar{\mathbf{x}}^b$ is the mean over the
 100 ensemble.

101 Then the coordinate transformation from physical space to ensemble space

$$\mathbf{x} = \bar{\mathbf{x}}^b + \mathbf{X}^b \mathbf{w} \quad (4)$$

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

$$J(\mathbf{w}) = (L-1) \mathbf{w} \mathbf{w}^t + \left(\mathbf{y} - \bar{\mathbf{y}}^b - \mathbf{Y} \mathbf{w}\right)^t \mathbf{R}^{-1} (\mathbf{y} - \bar{\mathbf{y}}^b - \mathbf{Y} \mathbf{w}) \quad (5)$$

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

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

105 which is valid for linear observation operators.

106 The minimization of the cost function (5) yields

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

107 with

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

108 Equation (4) provides the analysis ensemble mean

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^b + \mathbf{X}^b \bar{\mathbf{w}}^a . \quad (8)$$

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

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

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

113 Finally the analysis ensemble members in physical space read

$$\mathbf{x}^{a,l} = \bar{\mathbf{x}}^b + \mathbf{X}\bar{\mathbf{w}}^a + \mathbf{X}\mathbf{W}^{a,l} , \quad l = 1, \dots, L , \quad (9)$$

114 see [7, 8] for more details.

115 Specifically, we have chosen $L = 10$ ensemble members and number of observations $S = 1$ or $S = 2$.

116 2.3 Observation data

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

122 The present work considers observations

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

123 where $\boldsymbol{\eta} \in \mathbb{S}$ is Gaussian white noise with the true variance \mathbf{R}_t and $\hat{H}(\mathbf{x})$ is a linear observation operator
 124 $\hat{H}(\mathbf{x}) = \mathbf{H}\mathbf{x}$, $\mathbf{H} \in \mathbb{R}^{S \times N}$. In the following, the linear operator \mathbf{H} is called sensitivity function and
 125 we adopt this name from meteorological data assimilation of nonlocal satellite data. The present work
 126 considers either nonlocal observations only ($S = 1$)

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

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

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

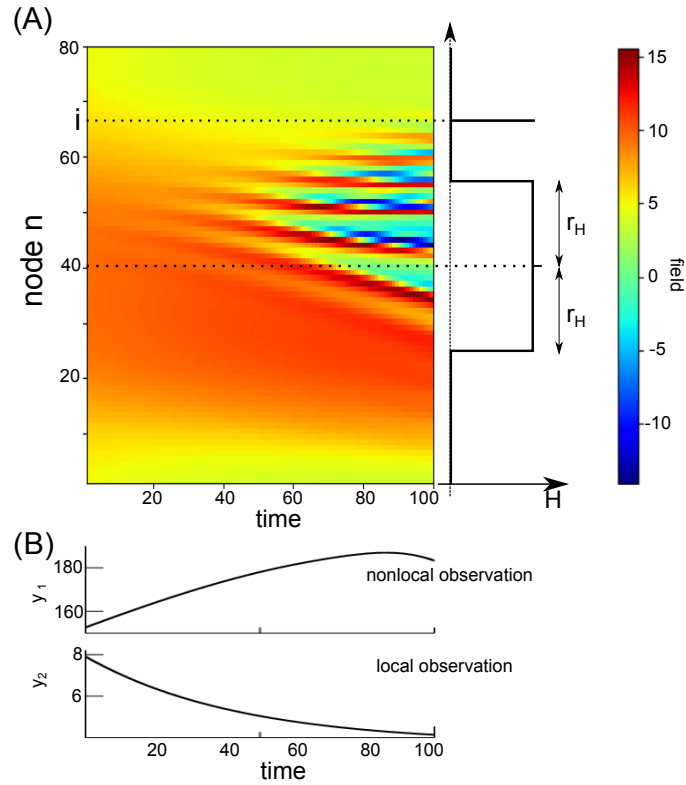


Figure 2. The model field \mathbf{V} from Eq. (1) and an illustration of the observation operator \mathbf{H} from Eq. (12) with the sensitivity function width r_H . (A) Exemplary space-time distribution of the model solution (left hand side) with the parameters $I = 8.0$ and $N = 80$, the sketched position 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.

128 where the local observation is captured at spatial location i , cf. Fig. 2 for illustration. In the subsequent
 129 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
 130 local observation. Moreover, in the following a grid point whose activity contributes to an observation is
 131 called an *observed grid point* and all others are called *unobserved grid points*. Mathematically, observed
 132 (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 $2r_H$

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

$$H_{1n} = 0 \quad \text{otherwise .}$$

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

138 This approach has been taken in most cases in the work. Since, however, this implicit filter error may already
 139 contribute to a filter instability or even may induce it, a short partial study has assumed perfect knowledge of
 140 the observation error. To this end, in this short partial study we have assumed $(\mathbf{R}_t)_{jj} = 0.1$, $j = 1, \dots, S$
 141 and perfect knowledge of this error, i.e. $\mathbf{R} = \mathbf{R}_t$.

142 Although techniques have been developed to estimate \mathbf{R} adaptively [25], we do not employ such a scheme
 143 for simplicity.

144

145 2.4 Localisation

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

152 The observation error matrix \mathbf{R} is diagonal, i.e. observation errors between single observations are
 153 uncorrelated. Then at each grid point i the localisation scheme considers observations y_n at location j
 154 only if the distance between location i and j does not exceed the localisation radius r_l . Then the error of
 155 observation n is $R_{nn} = R_{nn}^0 / \rho_{ij}$, where $\rho_{ij} = GC(d_{ij}, r) + \varepsilon$ for $d_{ij} \leq r_l$ is the weighting function with
 156 the Gaspari-Cohn function $GC(d, r)$ [29], $\varepsilon > 0$ is a small constant ensuring a finite observation error and
 157 d_{ij} is the spatial distance between i and j . The Gaspari-Cohn function approximates a Gaussian function
 158 with standard deviation $r\sqrt{3/10}$ by a polynomial with finite support. The parameter $2r = r_l$ is the radius
 159 of the localisation function with $0 \leq GC(z, r) \leq 1$, $0 \leq z \leq r_l$. Consequently the observation error takes
 160 its minimum R_{nn}^0 at distance $d_{ij} = 0$ and increases monotonously with distance to its maximum R_{nn}^0 / ε at
 161 $d_{ij} = r_l$. In the present implementation, we use $\varepsilon = 10^{-7}$ and observation errors $R_{11}^0 = 0.1$ for a single
 162 nonlocal observation $S = 1$ and $R_{nn}^0 = 0.1\delta_{nm}$, $n = 1, 2$ for local and nonlocal observation with $S = 2$.

163 The observation error close to the border of the localisation area about a grid point i is large by definition
 164 $R_{nn} = R_{nn}^0 / (GC(d \rightarrow r_l/2, r_l/2) + \varepsilon)$. In numerical practice, the assimilation effect of large values
 165 $R_{nn} > R_{nn}^0 / GC_{\text{low}}$ is equivalent for some distances from the grid point i in a reasonable approximation if
 166 GC_{low} is low enough. By virtue of the monotonic decrease of $GC(d, r_l/2)$ with respect to distance $d \geq 0$,
 167 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
 168 corrected localisation radius r_c , the observation errors R_{nn} are that large that observations at such distances
 169 do poorly contribute to the analysis. For instance, if $GC_{\text{low}} = 0.01$, then $r_l = 5 \rightarrow r_c = 3$, $r_l = 10 \rightarrow$
 170 $r_c = 7$ and $r_l = 15 \rightarrow r_c = 11$. It is important to note that this corrected localisation radius depends on the
 171 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
 172 following study cases results are given for original localisation radii r_l , while the usage of the corrected
 173 localisation radius is stated explicitly. [The existence 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.](#)

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

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

186 2.5 LETKF for a single observation

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

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

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

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

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

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

195 Now utilising the Woodbury matrix identity [38]

$$\begin{aligned} & (\mathbf{B} + \mathbf{UCV})^{-1} = \\ & \mathbf{B}^{-1} - \mathbf{B}^{-1} \mathbf{U} (\mathbf{C}^{-1} + \mathbf{V} \mathbf{B}^{-1} \mathbf{U})^{-1} \mathbf{V} \mathbf{B}^{-1} \end{aligned}$$

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

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

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

$$\bar{x}_i^a = \bar{x}_i^b + \mathbf{X}_i \mathbf{Y}^t \alpha_i \quad (16)$$

198 with

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

199 Since $R_{N/2} = R_{11}^0$ and $R_{N/2 \pm r_i} = R_{nn}^0 / \varepsilon = 10^7 R_{nn}^0$, α_i takes its maximum at the observation location
 200 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
 201 of the localisation area.

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

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

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

$$\sqrt{\mathbf{Q}_i} = \mathbf{U} \sqrt{\mathbf{D}} \mathbf{U}^t, \quad (18)$$

206 where $\sqrt{\mathbf{D}} \in \mathbb{R}^{L \times L}$ is diagonal and its matrix elements are the eigenvalues of \mathbf{Q} . The columns of matrix
207 $\mathbf{U} \in \mathbb{R}^{L \times L}$ are the normalised eigenvectors of \mathbf{Q} . Then Eq. (7) yields

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

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

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

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

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

213 This leads to

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

214 2.6 Additive covariance inflation

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

$$\mathbf{X}_{\text{add}}^b = \mathbf{X}^b + \mathbf{\Gamma} .$$

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

220 2.7 Numerical experiments

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

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

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

236 with $r_H = 10$ and $n_1 = 1$, $n_2 = 27$, $n_3 = 54$. The localisation radius is identical to the sensitivity function
 237 $r_l = r_H$ and data assimilation is performed during 250 filter cycles with an initial phase of 50 forecast steps.
 238 For stabilisation reasons, we have increased the model integration time step to $\Delta t = 10^{-2}/12$ but reduced
 239 the number of model integrations to 10 steps, cf. [19], thus essentially retaining the time interval between
 240 observations. Other parameters are identical to the standard setting described in the previous sections.

241 [As mentioned above, typically the measurement process is not known in all details. For instance, the](#)
 242 [observation error is assumed to be \$R = 0.1\$ for the nonlocal observations, whereas the true model exhibits](#)
 243 [noise-free observations with \$R_t = 0\$. This is the valid setting for all simulations but few set of trials shown](#)
 244 [in Fig. 5. In a set of experiments \(Fig. 5\(solid, dashed and dashed-dotted line\)\), observations are noisy with](#)
 245 [noise perturbation variance 0.1 and hence \$R_t = R = 0.1\$. Moreover, the additive inflation factor is chosen](#)
 246 [to \$f_{\text{add}} = 0.1\$ but in two single sets of experiments \(cf. Fig. 5\(dashed and dashed-dotted line\)\), where](#)
 247 [the additive inflation factor is \$f_{\text{add}} = 0.05\$. In addition, the weighting function of nonlocal observations is a boxcar window function with](#)

248 sharp borders but in a single set of experiments, where the weighting function is a smooth Gaspari-Cohn
 249 function, cf. Fig. 5(dashed-dotted line).

250 The verification measures bias and RMSE are computed for the local observations only according to
 251 Eqs. (21),(22).

252 2.8 Divergence criteria and verification

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

260 The present work focuses primarily on a non-catastrophic filter divergence type showing a strong increase
 261 of the innovation magnitude to values much larger than the observation equivalent of the attractor. This
 262 divergence may be temporally intermittent with finite duration. Since this intermittent innovation divergence
 263 results in increased first guess departures and hence worsens forecasts, it is important to detect these
 264 divergences and control them. By definition the innovation process diverges if $\max_{l,k} |[\mathbf{y}_n - \mathbf{H}\mathbf{x}^{b,l}]_k| > \sigma_{th}$
 265 for any observation n with $\sigma_{th} = 1000\sqrt{R_{nn}^0}$. Then the numerical simulation is stopped. The time of filter
 266 divergence is called T_b in the following. This criterion for innovation divergence is hard: if the innovation
 267 reaches the threshold σ_{th} , then innovation divergence occurs. The corresponding divergence rate γ is the
 268 ratio between the number of **divergent and non-divergent** trials. For instance, for $\gamma = 1$ all numerical trials
 269 diverge whereas $\gamma = 0$ reflect stability in all numerical trials.

270 Moreover, it is possible that $|[\mathbf{y}_n - \mathbf{H}\mathbf{x}^{b,l}]_k|$ grows intermittently but does not reach the divergence
 271 threshold. The first guess departure bias

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

272 and the corresponding root mean-square error

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

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

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

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

276 and

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

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

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

284

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

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

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

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

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

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

291 $\forall l \neq k, l, k = 1, \dots, L$. The alignment (anti-alignment) condition $\cos \beta_{lk} > 0.5$ ($\cos \beta_{lk} < -0.5$) implies
 292 $-60^\circ \leq \beta_{lk} \leq 60^\circ$ ($120^\circ \leq \beta_{lk} \leq 240^\circ$). Please note that $0 \leq p_{a,u} \leq 1$ and the larger p_a (p_u) the more
 293 ensemble members are aligned (anti-aligned) to each other.

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

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

297 The term $\mathbf{x}^{a,l} - \mathbf{x}^{b,l}$ is the analysis ensemble member perturbation from the background members and
 298 $\mathbf{x}^{b,l} - \bar{\mathbf{x}}^b$ is the direction of the background member perturbation. If $\cos \alpha_l \rightarrow 1$ ($\cos \alpha_l \rightarrow -1$) the analysis

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

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

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.

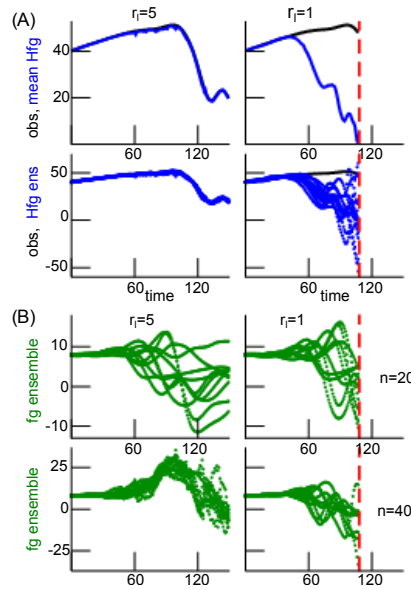


Figure 3. Temporal solutions of the filter process with $r_H = 5$ with two different localisation radii r_l . (A) Comparison of observations (black line) and model equivalents of the ensemble mean \mathbf{y}^b (top row, solid blue line) and the ensemble members $\mathbf{y}^{(b,l)}$ (bottom row, dotted 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

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

307 3.1 Nonlocal observations

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

310 Numerical results

311 In order to find out how the choice of localisation radius r_l affects the stability of the LETKF, a large
 312 number of numerical experiments assist to investigate statistically under which condition the filter diverges.
 313 Figure 3 shows the temporal evolution of the background \mathbf{x}^b and the model equivalents in observation
 314 space \mathbf{y}^b for two different localisation radii. In Fig. 3(A) observations (black line) are very close to model
 315 equivalents (blue lines) for identical localisation and sensitivity function width, i.e. $r_l = r_H$. Conversely,
 316 observations and model-equivalents diverge after some time for $r_l \neq r_H$. This is visible in the ensemble
 317 mean (Fig. 3(A), top row) and the single ensembles (Fig. 3(A), bottom row). The different filter behavior
 318 can be observed in model space as well, but there it is less obvious, cf. Fig. 3(B). The ensemble members at
 319 spatial location $n = 40$ are located in the center of the observation area. They exhibit a rather small spread
 320 around the ensemble mean for $r_l = r_H$, whereas the ensemble spread is larger for $r_l \neq r_H$. The ensemble
 321 at $n = 20$ is outside the observation area and thus is not assimilated. There, the ensemble in $r_l = r_H$ and
 322 $r_l \neq r_H$ are close to each other.

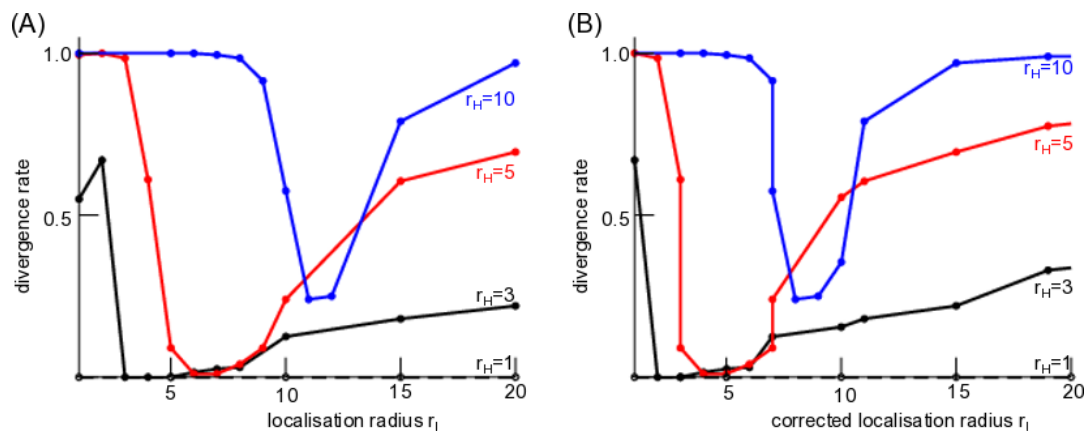


Figure 4. Stability of the LETKF of nonlocal observations dependent on the sensitivity function width r_H and the localisation radius r_l . The divergence rate γ is defined in section 2.8. (A) with original localisation radius r_l . (B) with corrected localisation radius r_c and $GC_{\text{low}} = 0.01$.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

323 This result can be generalised to a larger number of localisation and sensitivity function widths, cf. Fig. 4.
 324 For the smallest sensitivity function width and thus the smallest observation area with $r_H = 1$, no filter
 325 process diverges for a large range of localisation radii r_l , i.e. the LETKF is stable (dashed black line in
 326 Fig. 4). This case $r_H = 1$ corresponds to local observations. Now increasing the observation area with
 327 $r_H > 1$, the filter may diverge and its divergence rate γ depends on the localisation radius. We observe that
 328 the filter diverges least when the localisation radius is close to the sensitivity function width. These findings
 329 hold true for both the original localisation radius and the corrected radius r_c , cf. Methods section 2.4 and
 330 Fig. 4(A) and (B). Moreover, the filter does not exhibit catastrophic divergence before the background
 331 reaches its divergence threshold.

332 These results hold also true if observations are subjected to additive noise and the observation error
 333 is chosen to the true value, cf. Fig. 5(solid line) and if additive inflation is chosen to a lower value
 334 (Fig. 5(dashed line)). Similarly to Fig. 4, the divergence rate is minimum if the sensitivity function width is
 335 close to the original (Fig. 5(A)) or corrected (Fig. 5(B)) localisation radius r_l . The situation is different if
 336 the sensitivity function is not a non-smooth boxcar function as in the majority of the studies but a smooth
 337 Gaspari-Cohn function. Then the divergence rate is still minimum but the corresponding localisation radius
 338 of this minimum is much smaller than r_h , cf. dotted-dashed line in Fig. 5.

339 All these results consider the realistic case of a small number of ensemble members $L \ll N$. Nevertheless,
 340 it is interesting to raise the question how these results depend on the ensemble size. Figure 5(bold dotted-
 341 dashed line) indicates that a full ensemble with $L = 80$ removes the minimum with maximum divergence
 342 rate for $r_l < r_H$ and full stability for $r_l > r_H$.

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

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

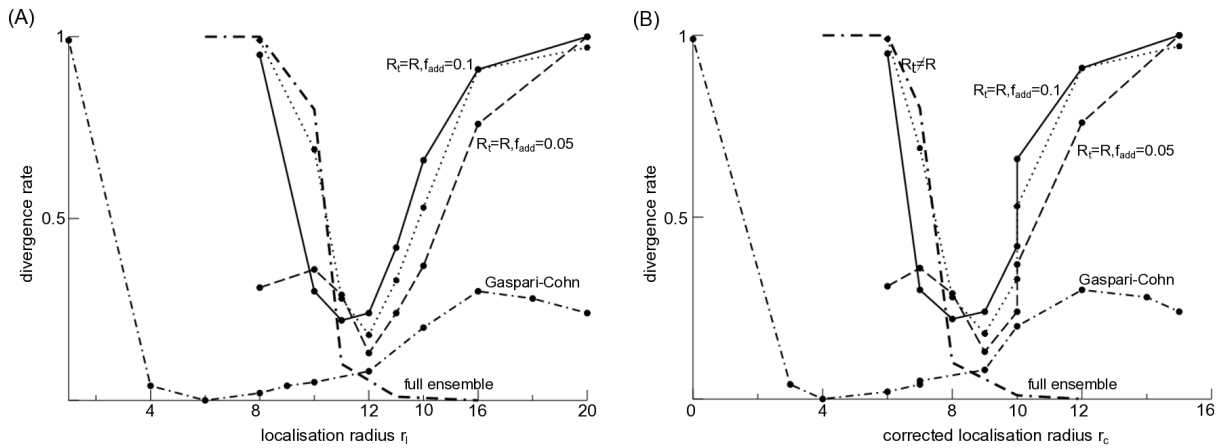


Figure 5. LETKF stability for different parameters and $r_H = 10$. The solid line denotes the divergence rate γ if the true observation error $R_t = 0.1$ is known, i.e. $R = R_t$, and the inflation rate is $f_{add} = 0.1$; the dashed line denotes the divergence rate for lower inflation rate $f_{add} = 0.05$, 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 ($R_t = 0, R = 0.1$) 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 $GC_{low} = 0.01$.

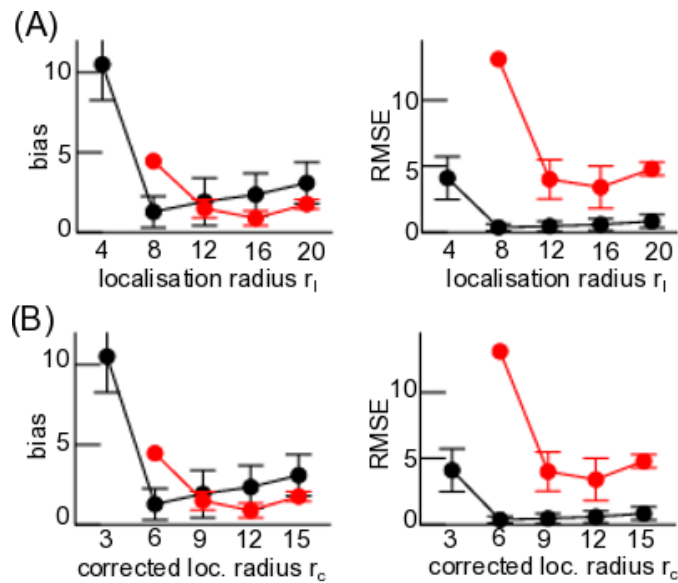


Figure 6. First guess departure statistics of trials that do not reach the divergence threshold. Here $r_H = 5$ (black) and $r_H = 10$ (red). (A) original localisation radius r_l . (B) corrected localisation radius r_c with $GC_{low} = 0.01$. All statistical measures are based on 100 trials.

351 various statistical quantities for three exemplary trials. These quantities have been proposed to reflect
 352 or explain divergence. The innovation-based measure Θ_n diverges (Fig. 7(B)) when the filter diverges
 353 (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
 354 certain number of ensemble members align and anti-align intermittently but do not align in the instance of
 355 divergence (Fig. 7(C)). In the case of similar localisation radius and sensitivity function width, a similar
 356 number of ensemble members align and anti-align but the filter does not diverge. Conversely, for $r_l \gg r_H$
 357 ensemble members both align and anti-align while the filter diverges. These results already indicate a
 358 different divergence mechanism for $r_l \leq r_H$ and $r_l > r_H$. Accordingly, for $r_l < r_H$ and $r_l \approx r_H$ background

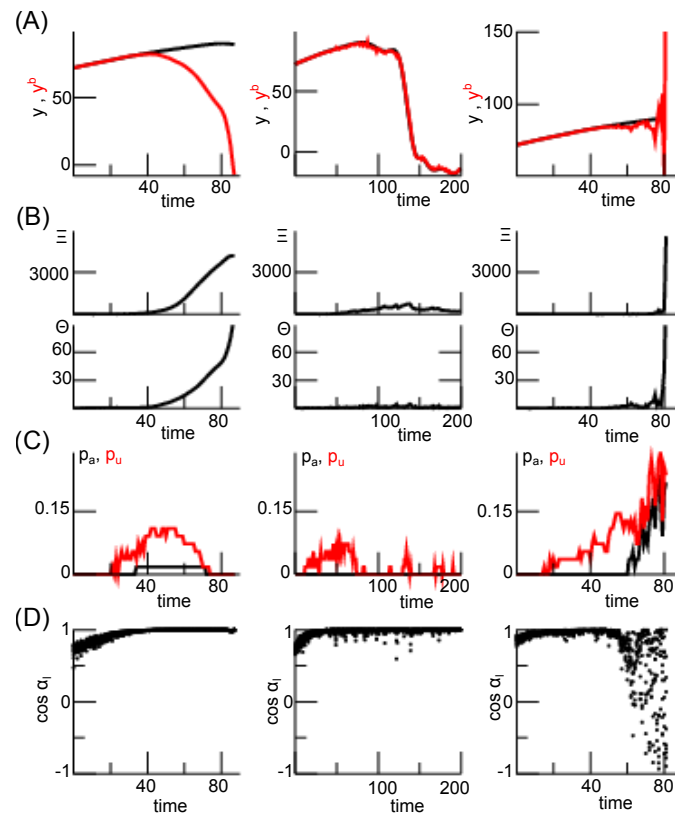


Figure 7. Various measures reflecting stability of the LETKF dependent on the localisation radius r_l in single trials. (A) observation y_o (black) and model equivalent $\mathbf{H}\bar{\mathbf{x}}^b$ (red). (B) Statistical quantities Ξ_n (top) and Θ_n (bottom), for definition see section 2.8. (C) The 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 sensitivity function width $r_H = 5$.

359 member perturbations align with the analysis member perturbations with $\cos \alpha_l \rightarrow 1$ (Fig. 7(D)), whereas
 360 $\cos \alpha_l$ fluctuates between 1 and -1 for $r_l \gg r_H$ while diverging.

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

369 Analytical description

370 According to Fig. 9, there are different possible configurations of the sensitivity function with respect
 371 to the localisation area. The localisation radius r_l may be smaller (cases 1) or larger (cases 2) than the
 372 sensitivity width r_H or both may be equal (cases 3). In addition, it is insightful to distinguish observed and
 373 unobserved grid points as already proposed in [18].

374 Now let us take a closer look at each case, cf. Fig. 9:

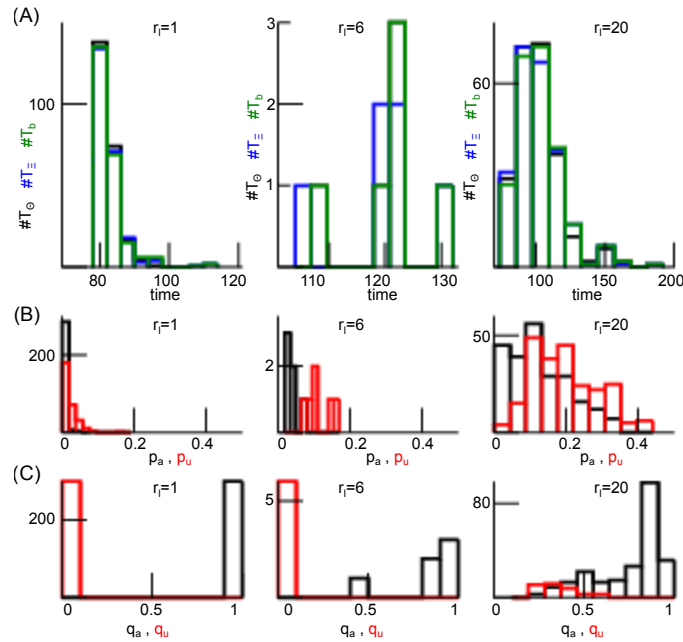


Figure 8. Divergence times and ensemble member alignment dependent on the localisation radius r_l . (A) Histogram of time of maximum Θ_n (T_Θ , black), time of maximum Ξ_n (T_Ξ , blue) and the divergence time T_b (green), see the Methods section 2.8 for definitions. (B) Histograms of alignment ratio 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.

- 375 • case **1.1**, $r_l \leq r_H$, $|i - \frac{N}{2}| \leq r_H$ and
 376 $|i - \frac{N}{2}| \leq r_l$: the localisation radius is smaller than the sensitivity function width and the observation
 377 at spatial location $N/2$ is located within the localisation radius about grid point i . Then, the analysis

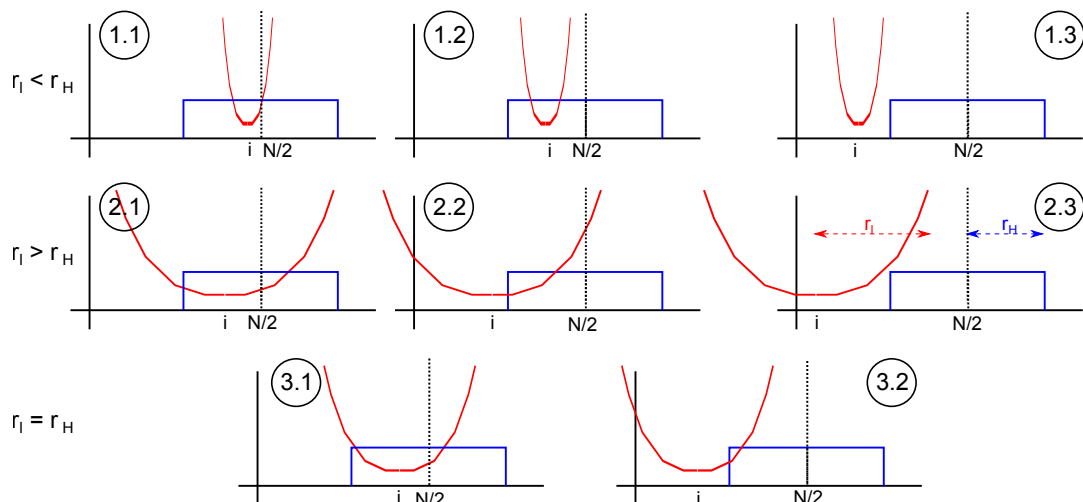


Figure 9. Sketch of different configurations of sensitivity function and localisation area. The circles denote the different cases (n.m) The sensitivity function (blue) has its center at the center of the spatial domain and the localisation function (red) is located about model grid element i .

378 ensemble (20) and its mean (16) read

$$x_{o,i}^{a,l} = \bar{x}_{o,i}^b + \mathbf{X}_{o,i} \mathbf{Y}^t \alpha_i + \mathbf{X}_{o,i} \mathbf{Y}^t \frac{\sqrt{\lambda_i}}{y^2} Y_l + \sum_{n=1}^{L-1} \mathbf{X}_{o,i} \mathbf{v}_n (\mathbf{v}_n)_l \quad (27)$$

379 and

$$\bar{x}_{o,i}^a = \bar{x}_{o,i}^b + \mathbf{X}_{o,i} \mathbf{Y}^t \alpha_i \quad (28)$$

380 with the corresponding ensemble means at observed grid points $\bar{x}_{o,i}^b$ and $\bar{x}_{o,i}^a$, the first guess
381 perturbations $\mathbf{X}_{o,i}$ and the analysis ensemble members $x_{o,i}^{a,l}$.

382 • case **1.2**, $r_l \leq r_H$, $|i - \frac{N}{2}| \leq r_H$ and

383 $|i - \frac{N}{2}| > r_l$: compared to case **1.1**, the grid point i is observed as well but the observation is outside
384 the localisation area; hence the analysis is identical to the first guess

$$\begin{aligned} \bar{x}_{o,i}^a &= \bar{x}_{o,i}^b \\ x_{o,i}^{a,l} &= \bar{x}_{o,i}^b + (\mathbf{X})_{o,i} . \end{aligned}$$

385 • case **1.3**, $r_l \leq r_H$, $|i - N/2| > r_H$ and

386 $|i - N/2| > r_l$: the grid point i is not observed and the observation is outside the localisation area
387 leading to

$$\begin{aligned} \bar{x}_{u,i}^a &= \bar{x}_{u,i}^b \\ x_{u,i}^{a,l} &= \bar{x}_{u,i}^b + (\mathbf{X})_{u,i} \end{aligned} \quad (29)$$

388 with the corresponding unobserved ensemble means $\bar{x}_{u,i}^b$ and $\bar{x}_{u,i}^a$, the unobserved ensemble
389 perturbations $\mathbf{X}_{u,i}$ and the analysis ensemble member $x_{u,i}^{a,l}$.

390 • case **2.1**, $r_l > r_H$, $|i - \frac{N}{2}| \leq r_H$ and

391 $|i - \frac{N}{2}| \leq r_l$: the localisation radius is larger than the sensitivity function width, the observation is
392 located within the localisation radius about the grid point i and all grid points are observed. This case
393 is equivalent to case **1.1** and the expressions for the analysis ensemble and mean hold as well.

394 • case **2.2**, $r_l > r_H$, $|i - \frac{N}{2}| > r_H$ and

395 $|i - \frac{N}{2}| \leq r_l$: compared to case **2.1**, the observation is located within the localisation radius but grid
396 points are unobserved. Then

$$\bar{x}_{u,i}^a = \bar{x}_{u,i}^b + \mathbf{X}_{u,i} \mathbf{Y}^t \alpha_i \quad (30)$$

397 and

$$\begin{aligned}
 x_{u,i}^{a,l} &= \bar{x}_{u,i}^b + \mathbf{X}_{u,i} \mathbf{Y}^t \alpha_i + \mathbf{X}_{u,i} \mathbf{Y}^t \frac{\sqrt{\lambda_i}}{y^2} Y_l \\
 &\quad + \sum_{n=1}^{L-1} \mathbf{X}_{u,i} \mathbf{v}_n (\mathbf{v}_n)_l.
 \end{aligned}
 \tag{31}$$

- 398 • case **2.3**, $r_l > r_H$, $|i - \frac{N}{2}| > r_H$ and
 399 $|i - \frac{N}{2}| > r_l$: in this case, the grid points are unobserved and the observation is outside the localisation
 400 area. Then the analysis is identical to the first guess and the case is equivalent to case **1.3**.
- 401 • case **3.1**, $r_l = r_H$, $|i - \frac{N}{2}| \leq r_H$: the observation is located within the localisation radius about the
 402 grid point i , the grid point is observed and the expressions in case **1.1** hold.
- 403 • case **3.2**, $r_l = r_H$, $|i - \frac{N}{2}| > r_H$: the observation is not located within the localisation radius of grid
 404 point i , then grid point is not observed and the expressions in case **1.3** hold.

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

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

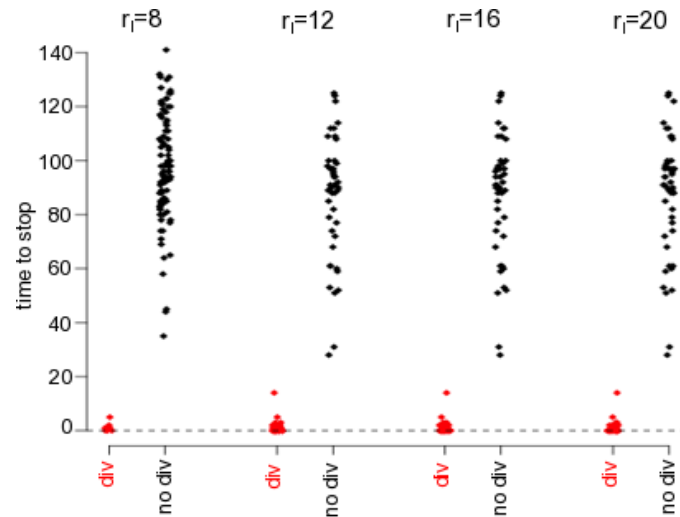


Figure 10. The divergence correlates with the weighted model-observation covariances at observed grid points A_o . The plots show the times of maxima T_o (cf. Eq. (35)) to stop, i.e. $T_{\text{stop}} - T_o$. T_o is the time when the mean model-observation covariance A_o is maximum, for divergent (red-colored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

435 nonlocal observations to a single location. The larger the localisation radius, the more distant are grid
 436 points to the observation location and the less representative is the localised observation to distant grid
 437 points.
 438 Hence these two latter cases may cause detrimental effects. Consequently, cases 1 and 2, i.e. $r_l \neq r_H$,
 439 yields bad estimates of analysis updates that make the Kalman filter diverge. Conversely, case 3, i.e.
 440 $r_l = r_H$, involves consistent updates only and detrimental effects as described for the other cases are not
 441 present. These effects may explain enhanced filter divergence for $r_l \neq r_H$ and minimum filter divergence
 442 for $r_l = r_H$ seen in Fig. 3, and the minimum divergence rate at $r_l \approx r_H$ shown in Fig. 4.

443 The important terms in case 2.2, i.e. Eqs. (30) and (31), are $\mathbf{X}_{u,i}\mathbf{Y}^t$, α_i , $\sqrt{\lambda_i}Y_l/y^2$, and
 444 $\sum_n \mathbf{X}_{u,i}\mathbf{v}_n(\mathbf{v}_n)_l$. Equivalently, the missing terms in case 1.2 are $\mathbf{X}_{o,i}\mathbf{Y}^t$, α_i , $\sqrt{\lambda_i}Y_l/y^2$ and
 445 $\sum_n \mathbf{X}_{o,i}\mathbf{v}_n(\mathbf{v}_n)_l$. For instance,

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

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

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

449 in case 2.1 and

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

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

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

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

459 The corresponding quantities

$$T_o = \arg \max_n \frac{1}{M_o} \sum_{i \in \mathcal{M}_o} (\mathbf{c}_o)_i(t_n) \alpha_i(t_n) \quad (35)$$

$$T_u = \arg \max_n \frac{1}{M_u} \sum_{i \in \mathcal{M}_u} (\mathbf{c}_u)_i(t_n) \alpha_i(t_n)$$

$$\Delta T = T_o - T_u \quad (36)$$

460 define the time instances when these maxima are reached and ΔT is their difference. For instance, if
 461 $\Delta T > 0$, then the weighted model-observation covariances at observed grid points reach their maximum
 462 before weighted model-observation covariances at unobserved grid points.

463 To illustrate the importance of A_o and its corresponding occurrence time T_o , Fig. 10 shows T_o relative to the
 464 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
 465 and for non-divergent trials $T_{\text{stop}} = 200$ is the maximum time. Figure 10 reveals that T_o is very close to
 466 the divergence time, whereas T_o is widely distributed about $T_o = 110$ ($T_{\text{stop}} - T_o = 90$) in non-divergent
 467 trials. This indicates that A_o is strongly correlated with the underlying divergence mechanism.

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

478 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

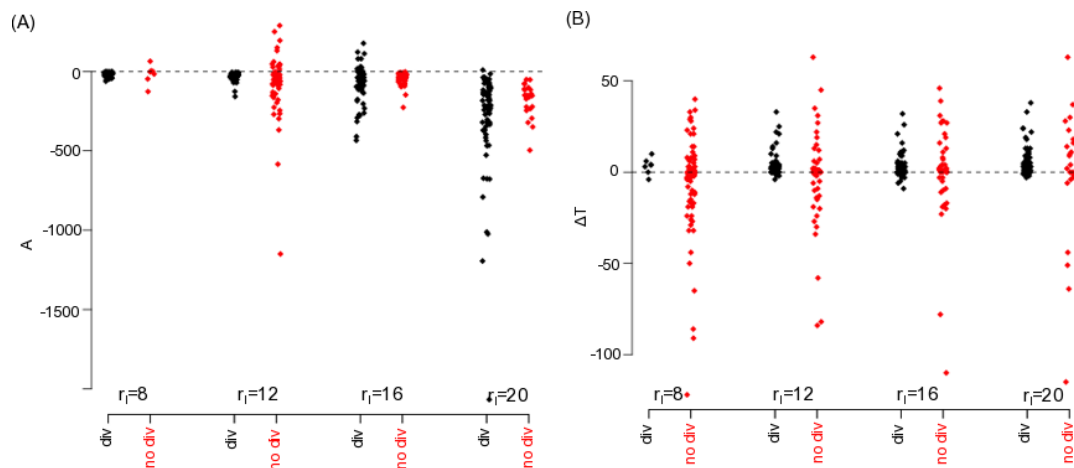


Figure 11. Comparison of weighted model-observation covariances in observed and non-observed grid points. (A) $A = A_o - A_u$ is the difference between **maximum weighted model-observation covariances** in observed and unobserved grid points. (B) $\Delta T = T_o - T_u$ is the difference of times when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

479 larger maximum faster than observed grid points. This indicates that the model-observation covariance c_u
 480 reflects the instability of the filter.

481 3.2 Local and nonlocal observations

482 Several international weather services apply ensemble Kalman filters and assimilate both nonlocal
 483 and local observations. Performing assimilation experiments similar to the experiments for nonlocal
 484 observations but now with a single additional local observation at grid point $i = N/2$ (cf. Methods section
 485 2.7), the filter divergence rate γ indicates the filter stability. Figure 12 illustrates how local observations
 486 affect the filter stability in addition to nonlocal observations. For $r_H = 1$, the filter diverges rarely due
 487 to large innovations (with fewest trials at $r_l \approx 10$) but at a larger number than in the absence of local
 488 observations, cf. Fig. 4. Moreover, increasing the localisation radius yields a higher number of trials with
 489 catastrophic filter divergence with a maximum catastrophic divergence rate at $r_l \approx 10$. In sum, the least
 490 number of **divergent** trials occur at $r_l = r_H = 1$ (blue curve in Fig. 12). A similar stability behavior occurs
 491 for $r_H = 5$ with a minimum innovation divergence rate at $r_l \approx r_H$ and a maximum catastrophic divergence
 492 rate at $r_l \approx 10$. Again, the least number of trials diverge at $r_l = r_H$.

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

4 DISCUSSION

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

509 Nonlocal observations facilitates filter divergence

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

519 Optimal localisation radius

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

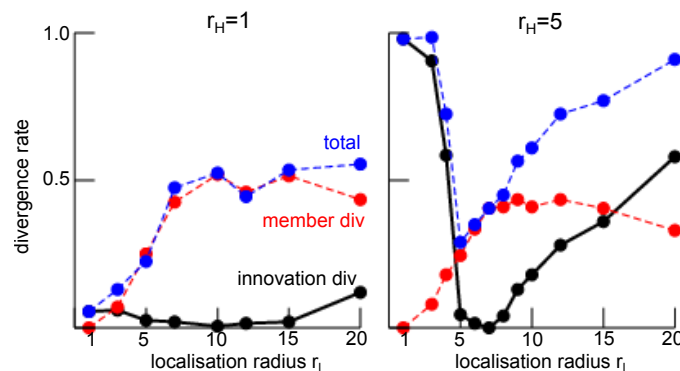


Figure 12. Rate of filter divergence γ (innovation divergence, black line) and catastrophic filter divergence (member divergence, red line) in the presence of a single local and a single nonlocal observation. The total number of **divergent** trials is the sum of innovation and member divergence-trials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

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

534 It is important to point out that, under certain conditions, it may be beneficial to further enlarge the
535 localisation area compared to the sensitivity function. If the system's activity synchronizes on a larger
536 spatial scale, then information is shared between observed and unobserved grid points and a larger
537 localisation radius would be beneficial. Examples for such synchronisation phenomena are deep clouds
538 or large-scale winds in meteorology or locally self-organised spots in physical complex systems. In other
539 words, to decide how to choose the localisation radius one should take a closer look at the system's
540 dynamics: if larger spatially synchronised phenomena are expected, then $r_l \gg r_H$ is preferable, otherwise
541 $r_l \approx r_H$.

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

548 Origin of divergence

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

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

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

582 Moreover, there is some evidence that ensemble member alignment may cause catastrophic filter
583 divergence [21, 19, 20]. Figure 8 shows such indirect markers indicating weak member anti-alignment for
584 $r_l \leq r_H$ but enhanced alignment and anti-alignment for $r_l > r_H$. The authors in [19] argue that finite
585 ensemble sizes cause the ensemble to align in case of divergence and Ng et al. [53] show that the ensemble
586 members may align with the most unstable phase space direction. However, our results reveal that member
587 alignment does not represent the major mechanism for innovation divergence. Conversely, Fig. 8 provides
588 evidence for alignment of analysis increments and background perturbations when the filter diverges. This
589 alignment indicates that the analysis members point into the same direction as the background members.
590 For instance, if background member perturbations point to less stable locations in phase space, then the
591 LETKF does not correct this direction and the new analysis state is less stable, cf. the model example
592 in [20]. This shows accordance to the reasoning in Ng et al. [53].

593 In addition to the alignment mechanism, Eq. (32) represents the covariation of ensemble perturbations in
594 spatial and observation space at observed and unobserved spatial locations. For observed spatial locations,
595 it is maximum just before the innovation divergence time. Moreover, it reaches its maximum at unobserved
596 locations almost always before the maximum at observed locations are reached (Fig. 11). It seems this new
597 feature represent an important contribution to the innovation divergence and future work will analyse this
598 covariation in more detail.

599 **Limits and outlook**

600 The present work considers the specific case of finite low ensemble size and application of the localisation
601 scheme. To understand better the origin of the filter divergence, it is insightful to study in detail the limiting
602 case of large ensemble sizes, i.e. close to the model dimension, and a neglect of localisation. Although this
603 limit is far from practice in seismology and meteorology, where the model systems are too large to study
604 this limit, nevertheless this limit study is of theoretical interest and future work will consider it in some
605 detail.

606 There is some evidence that the optimal localisation radius is flow-dependent [54, 55], whereas we assume
607 a constant radius. In addition, the constrained choice of parameters and missing standard techniques to
608 prevent divergence, such as adaptive inflation and adaptive observation error, limits the present work
609 in generality and interpretation and thus makes it hard to derive decisive conclusions. Future work will
610 implement adaptive schemes [56, 57] in a more realistic model framework.

611 In the majority of studies, the present work considers a non-smooth boxcar sensitivity function in order
612 to distinguish observed and unobserved grid points. Although this simplification allows to gain deeper
613 understanding of possible contributions to the filter divergence, the sensitivity function is unrealistic. A
614 more realistic sensitivity function is smooth and unimodal or bimodal. Figure 5 shows that such a sensitivity
615 function yields a minimum divergence rate but the localisation radius at the minimum rate is much smaller
616 than the sensitivity function width. Consequently, the line of argumentation about Fig. 9 does not apply

617 here since there is no clear distinction of observed and unobserved grid points anymore. Future work will
618 attempt to consider smooth unimodal sensitivity functions.
619 Moreover, the localisation scheme of nonlocal observations applied in the present work is very basic due to
620 its choice of the maximum sensitivity as the observations location. Future work will investigate cut-off
621 criteria as such in [12] that chooses the location of nonlocal observations in the range of the sensitivity
622 function. Fertig et al. [12] also have shown that such a cut-off criterion improves first guess departure
623 statistics and well reduces the divergence for localisation radii that are smaller than the sensitivity weighting
624 function.
625 Nevertheless the present work introduces the problem of intermittent innovation divergence, extends lines
626 of reason on the origin of filter divergence to nonlocal observation and proposes new markers of innovation
627 divergence.
628

5 ACKNOWLEDGEMENT

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

REFERENCES

- 632 [1] Bengtsson L, Ghil M, Källén E, editors. *Dynamic Meteorology: Data Assimilation Methods, Applied*
633 *Mathematical Sciences*, vol. 36 (Springer) (1981).
- 634 [2] Luo X, Bhakta T, Jakobsen M, Navdal G. Efficient big data assimilation through sparse representation:
635 A 3d benchmark case study in petroleum engineering. *PLoS One* **13** (2018) e0198586. doi:10.1371/
636 journal.pone.0198586.
- 637 [3] Hutt A, Stannat W, Potthast R, editors. *Data Assimilation and Control: Theory and Applications in*
638 *Life Sciences* (Frontiers Media) (2019). doi:10.3389/978-2-88945-985-8.
- 639 [4] Schiff SJ. *Neural Control Engineering* (Cambridge, MA: MIT Press) (2011).
- 640 [5] Nakamura G, Potthast R. *Inverse Modeling*. 2053-2563 (IOP Publishing) (2015). doi:10.1088/
641 978-0-7503-1218-9.
- 642 [6] Asch M, Bocquet M, Nodet M. *Data Assimilation: Methods, Algorithms, and Applications*
643 (Philadelphia: SIAM) (2016).
- 644 [7] Hunt B, Kostelich E, Szunyogh I. Efficient data assimilation for spatiotemporal chaos: A local
645 ensemble transform Kalman filter. *Physica D* **230** (2007) 112–126.
- 646 [8] Schraff C, Reich H, Rhodin A, Schomburg A, Stephan K, Perianez A, et al. Kilometre-scale ensemble
647 data assimilation for the cosmo model (kenda). *Q. J. R. Meteorol. Soc.* **142** (2016) 1453–1472.
648 doi:10.1002/qj.2748.
- 649 [9] Schomburg A, Schraff C, Potthast R. A concept for the assimilation of satellite cloud information
650 in an ensemble Kalman filter: single-observation experiments. *Q. J. R. Meteorol. Soc.* **141** (2015)
651 893–908. doi:10.1002/qj.2748.
- 652 [10] Miyoshi T, Sato Y. Assimilating satellite radiances with a local ensemble transform Kalman filter
653 (letkf) applied to the jma global model (gsm). *SOLA* **3** (2007) 37–40. doi:10.2151/sola.2007-010.
- 654 [11] Kurzrock F, Cros S, Ming F, Otkin J, Hutt A, Linguet L, et al. A review of the use of geostationary
655 satellite observations in regional-scale models for short-term cloud forecasting. *Meteorologische*
656 *Zeitschrift* **27** (2018) 277–298. doi:10.1127/metz/2018/0904.

- 657 [12] Fertig EJ, Hunt BR, Ott E, Szunyogh I. Assimilating non-local observations with a local ensemble
658 Kalman filter. *Tellus A* **59** (2007) 719–730. doi:10.1111/j.1600-0870.2007.00260.x.
- 659 [13] Hutt A, Schraff C, Anlauf H, Bach L, Baldauf M, Bauernschubert E, et al. Assimilation of SEVIRI
660 water vapour channels with an ensemble Kalman filter on the convective scale. *Front. Earth Sci.* **8**
661 (2019) 70. doi:10.3389/feart.2020.00070.
- 662 [14] Furrer R, Bengtsson T. Estimation of high-dimensional prior and posterior covariance matrices in
663 Kalman filter variants. *J. Multivar. Ana.* **98** (2007) 227–255.
- 664 [15] Anderson JL. An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.* **129**
665 (2001) 2884–2903.
- 666 [16] Hamill TM, Whitaker JS, Snyder C. Distance-dependent filtering of background error covariance
667 estimates in an ensemble Kalman filter. *Mon. Wea. Rev.* **129** (2001) 2776–2790. doi:10.1175/
668 1520-0493.
- 669 [17] Tong XT, Majda A, Kelly D. Nonlinear stability and ergodicity of ensemble based Kalman filters.
670 *Nonlinearity* **29** (2016) 657–691. doi:10.1088/0951-7715/29/2/657.
- 671 [18] Tong XT, Majda A, Kelly D. Nonlinear stability of ensemble Kalman filters with adaptive covariance
672 inflation. *Commun. Math. Sci.* **14** (2016) 1283–1313. doi:10.4310/cms.2016.v14.n5.a5.
- 673 [19] Gottwald G, Majda AJ. A mechanism for catastrophic filter divergence in data assimilation for sparse
674 observation networks. *Nonlin. Processes Geophys.* **20** (2013) 705–712. doi:10.5194/npg-20-705-2013.
- 675 [20] Kelly D, Majda A, Tong XT. Concrete ensemble Kalman filters with rigorous catastrophic filter
676 divergence. *Proc. Natl. Acad. Sci. USA.* **112** (2015) 10589–10594. doi:10.1073/pnas.1511063112.
- 677 [21] Majda A, Harlim J. Catastrophic filter divergence in filtering nonlinear dissipative systems. *Comm.*
678 *Math. Sci.* **8** (2008) 27–43.
- 679 [22] Migliorini S, Candy B. All-sky satellite data assimilation of microwave temperature sounding channels
680 at the met office. *Quart. J. Roy. Meteor. Soc.* **145** (2019) 867–883. doi:10.1002/qj.3470.
- 681 [23] Lorenz EN, Emanuel KA. Optimal sites for supplementary weather observations: Simulations
682 with a small model. *J. Atmos. Sci.* **555** (1998) 399–414. doi:10.1175/1520-0469(1998)055<0399:
683 SFSWO>2.0.CO;2.
- 684 [24] Bishop CH, Whitaker JS, Lei L. Gain form of the Ensemble Transform Kalman Filter and its relevance
685 to satellite data assimilation with model space ensemble covariance localization. *Mon. Wea. Rev.* **145**
686 (2017) 4575–4592. doi:10.1175/MWR-D-17-0102.1.
- 687 [25] Waller JA, Dance SL, Lawless AS, Nichols NK. Estimating correlated observation error statistics
688 using an ensemble transform Kalman filter. *Tellus A* **66** (2014) 23294. doi:10.3402/tellusa.v66.23294.
- 689 [26] Houtekamer PL, Zhang F. Review of the ensemble Kalman filter for atmospheric data assimilation.
690 *Mon. Wea. Rev.* **144** (2016) 4489–4532. doi:10.1175/MWR-D-15-0440.1.
- 691 [27] Perianez A, Reich H, Potthast R. Optimal localization for ensemble Kalman filter systems. *J. Met.*
692 *Soc. Japan* **92** (2014) 585–597. doi:10.2151/jmsj.2014-605.
- 693 [28] Greybush SJ, Kalnay E, Miyoshi T, Ide K, Hunt BR. Balance and ensemble Kalman filter localization
694 techniques. *Mon. Wea. Rev.* **139** (2011) 511–522. doi:10.1175/2010MWR3328.1.
- 695 [29] Gaspari G, Cohn S. Construction of correlation functions in two and three dimensions. *Q. J. R. Meteo.*
696 *Soc.* **125** (1999) 723–757.
- 697 [30] Nadeem A, Potthast R. Transformed and generalized localization for ensemble methods in data
698 assimilation. *Math. Meth. Appl. Sci.* **39** (2016) 619–634. doi:10.1002/mma.3496.
- 699 [31] Bishop CH, Hodyss D. Ensemble covariances adaptively localized with eco-rap. part 2: a strategy for
700 the atmosphere. *Tellus* **61A** (2009) 97–111. doi:10.1111/j.1600-0870.2008.00372.

- 701 [32] Leng H, Song J, Lu F, Cao X. A new data assimilation scheme: the space-expanded ensemble
702 localization Kalman filter. *Adv. Meteorol.* **2013** (2013) 410812. doi:10.1155/2013/410812.
- 703 [33] Miyoshi T, Yamane S. Local Ensemble Transform Kalman Filtering with an AGCM at a T159/L48
704 resolution. *Mon. Wea. Rev.* **135** (2007) 3841–3861. doi:10.1175/2007MWR1873.1.
- 705 [34] Farchi A, Boquet M. On the efficiency of covariance localisation of the ensemble Kalman filter using
706 augmented ensembles. *Front. Appl. Math. Stat.* **5** (2019) 3. doi:10.3389/fams.2019.00003.
- 707 [35] Lei L, Whitaker JS. Model space localization is not always better than observation space
708 localization for assimilation of satellite radiances. *Mon. Wea. Rev.* **143** (2015) 3948–3955.
709 doi:10.1175/MWR-D-14-00413.1.
- 710 [36] Campbell WF, Bishop CH, Hodyss D. Vertical covariance localization for satellite radiances in
711 ensemble Kalman filters. *Mon. Wea. Rev.* **138** (2010) 282–290. doi:10.1175/MWR3017.1.
- 712 [37] Houtekamer PL, Mitchell H, Pellerin G, Buehner M, Charron M, Spacek L, et al. Atmospheric data
713 assimilation with an ensemble Kalman filter: Results with real observations. *Mon. Wea. Rev.* **133**
714 (2005) 604–620. doi:10.1175/MWR-2864.1.
- 715 [38] Higham NJ. *Accuracy and stability of numerical algorithms* (SIAM), 2nd edn. (2002).
- 716 [39] Anderson JL, Anderson SL. A monte carlo implementation of the nonlinear filtering problem to
717 produce ensemble assimilations and forecasts. *Mon. Wea. Rev.* **127** (1999) 2741–2758. doi:10.1175/
718 1520-0493.
- 719 [40] Luo X, Hoteit I. Covariance inflation in the ensemble Kalman filter: A residual nudging perspective
720 and some implications. *Mon. Weath. Rev.* **141** (2013) 3360–3368. doi:10.1175/MWR-D-13-00067.1.
- 721 [41] Hamill TM, Whitaker JS. What constrains spread growth in forecasts initialized from ensemble
722 Kalman filters ? *Mon. Wea. Rev.* **139** (2011) 117–131. doi:10.1175/2010MWR3246.1.
- 723 [42] Mitchell HL, Houtekamer PL. An adaptive ensemble Kalman filter. *Mon. Weath. Rev.* **128** (2000)
724 416–433. doi:10.1175/1520-0493.
- 725 [43] Grewal MS, Andrews AP. *Kalman filtering: Theory and practice using MATLAB* (John Wiley &
726 Sons), 2nd edn. (2001).
- 727 [44] Marx B, Potthast R. On instabilities in data assimilation algorithms. *Mathematics* **8** (2012) 27–43.
728 doi:10.1007/s13137-012-0034-5.
- 729 [45] Lahoz WA, Schneider P. Data assimilation: making sense of earth observation. *Front. Environ. Sci.* **2**
730 (2014) 16. doi:10.3389/fenvs.2014.00016.
- 731 [46] Tong XT. Performance analysis of local ensemble Kalman filter. *J. Nonlinear. Sci.* **28** (2018)
732 1397–1442. doi:10.1007/s00332-018-9453-2.
- 733 [47] Ying Y, Zhang F, Anderson J. On the selection of localization radius in ensemble filtering for multiscale
734 quasigeostrophic dynamics. *Mon. Wea. Rev.* **146** (2018) 543–560. doi:10.1175/MWR-D-17-0336.1.
- 735 [48] Miyoshi T, Kondo K. A multi-scale localization approach to an ensemble Kalman filter. *SOLA* **9**
736 (2013) 170–173.
- 737 [49] Migliorini S. Information-based data selection for ensemble data assimilation. *Quart. J. Roy. Meteor.*
738 *Soc.* **139** (2013) 2033–2054.
- 739 [50] Kirchgessner P, Nerger L, Bunse-Gerstner A. On the choice of an optimal localization radius
740 in Ensemble Kalman Filter methods. *Mon. Wea. Rev.* **142** (2014) 2165–2175. doi:10.1175/
741 MWR-D-13-00246.1.
- 742 [51] Bishop CH, Whitaker JS, Lei L. Commentary: On the efficiency of covariance localisation of the
743 ensemble Kalman filter using augmented ensembles by alban farchi and marc bocquet. *Front. Appl.*
744 *Math. Stat.* **accepted** (2020).

- 745 [52] Lei L, Whitaker JS, Bishop C. Improving assimilation of radiance observations by implementing model
746 space localization in an ensemble Kalman filter. *J. Adv. Model. Earth Syst.* **10** (2018) 3221–3232.
747 doi:10.1029/2018MS001468.
- 748 [53] Ng GHC, McLaughlin D, Entekhabi D, Ahanin A. The role of model dynamics in ensemble Kalman
749 filter performance for chaotic systems. *Tellus A* **63** (2011) 958–977.
- 750 [54] Zhen Y, Zhang F. A probabilistic approach to adaptive covariance localization for serial ensemble
751 square root filters. *Mon. Wea. Rev.* **142** (2014) 4499–4518.
- 752 [55] Flowerdew J. Towards a theory of optimal localisation. *Tellus A* **67** (2015) 25257.
- 753 [56] Lee Y, Majda AJ, Qi D. Preventing catastrophic filter divergence using adaptive additive inflation for
754 baroclinic turbulence. *Mon. Wea. Rev.* **145** (2017) 669–682. doi:10.1175/MWR-D-16-0121.1.
- 755 [57] Miyoshi T. The gaussian approach to adaptive covariance inflation and its implementation with
756 the local ensemble transform Kalman filter. *Mon. Wea. Rev.* **139** (2011) 1519–1535. doi:10.1175/
757 2010MWR3570.1.