Optimal Localization for Ensemble Kalman Filter Systems

África PERIÁÑEZ
Deutscher Wetterdienst, Data Assimilation Unit, Offenbach, Germany
Department of Mathematics, University of Reading, Berkshire, UK

Hendrik REICH
Deutscher Wetterdienst, Data Assimilation Unit, Offenbach, Germany

and

Roland POTTHAST
Deutscher Wetterdienst, Data Assimilation Unit, Offenbach, Germany
Department of Mathematics, University of Reading, Berkshire, UK

Manuscript received 17 November 2013, in final form 26 July 2014

Abstract
In ensemble Kalman filter methods, localization is applied for both avoiding the spurious correlations of distant observations and increasing the effective size of the ensemble space. The procedure is essential in order to provide quality assimilation in large systems; however a severe localization can cause imbalances that impact negatively on the accuracy of the analysis.

We want to understand the fundamental properties of localized ensemble methods and to investigate an optimal localization expression which minimizes the analysis error. The novel analytical expression derived in this work depends on the observation error, the density of measurements, and the approximation error, i.e., the error that comes from working in the ensemble space. The mathematical results are tested with two numerical simulations using a toy model. We demonstrate that observations with different observation error or density need different localization length scales.

Keywords  ensemble Kalman filter; localization; adaptive methods

1. Introduction
Ensemble data assimilation techniques, such as the ensemble Kalman filter (EnKF) (Evensen 1994; Burgers et al. 1998), are types of state estimation procedures, which are based on a Monte Carlo approach of the classic Kalman filter (Kalman 1960), that use ensemble forecasts to estimate the flow-dependent background error covariance matrix and other quantities needed to update the analysis.

For the numerical weather prediction (NWP) applications, the ensemble size is normally orders of magnitude smaller than the number of atmospheric observations available and the number of grid points in the model. Too small ensembles leads to undersampling problems, which can produce additional difficulties such as underestimation of covariance, filter
divergence, or the development of spurious correlations (Petrie et al. 2010). To mitigate these problems, covariance inflation and localization were developed. Multiple ensemble data assimilation systems have localization (Houtekamer et al. 1998; Houtekamer et al. 2001; Whitaker et al. 2002; Ott et al. 2004; Anderson 2001; Janjić et al. 2011) as a fundamental requirement. Computations are then performed in a local region instead of the global space (Szunyogh et al. 2005; Szunyogh et al. 2008) allowing feasibility in real time for operational applications on one hand, and improving the quality of the analysis on the other. Localization is used for both avoiding spurious correlations between different and spatially separated events along the atmosphere and increasing the rank of the covariance matrix because the ensemble size is small as compared with the number of observations to be processed and the rank of the background error covariance matrix. Localization is crucial for ensemble Kalman filter methods to provide quality assimilation in large systems. In particular, vertical localization for radiances (Campbell et al. 2010; Miyoshi et al. 2007) is especially challenging because radiances provide integrated information and they are sensitive to a broad layer of the atmosphere (Houtekamer et al. 2005; Whitaker et al. 2004). The implementation of the correct localization length scale for every case is an open research field towards which we conduct the work presented here.

In our study, we are interested in the local versions of the ensemble Kalman filters, and we focus on its computationally more efficient version developed by Hunt et al. LETKF, the local ensemble transformed Kalman filter (Hunt et al. 2007) that will be used in the numerical experiment examples in Section 3.3. A detailed review about EnKF from the deterministic and Bayesian perspectives can be found in (Freitag et al. 2013).

Although there are several examples of successful applications of localized data assimilation systems (even at operational level), a full basic knowledge of localization is still missing. The purpose of this article is to contribute to a deeper understanding of localization and its limitations. This work is inspired by the deterministic error estimates in relation to localization derived in (Perianez et al. 2014), which includes convergence limits of localized EnKF. For homogeneous distribution of measurements, we propose an analytical formula for an optimal localization radius depending on the approximation error, observation error, and density of data. We present this theoretical approach with a couple of toy examples with least squares analysis and LETKF as data assimilation schemes.

The organization of the paper is as follows. First, we give a brief introduction about EnKF, LETKF, and localization techniques. In Section 3 the optimal localization expression is explained. The experiments used to test the validity of the theory are also presented in this section. Some conclusions and future steps are discussed in Section 4.

2. Methods

2.1. Ensemble Kalman Filter

The iterative process of the assimilation of observations consists of two stages: the forecast and the analysis. In the first step, the background state $x_b$ evolves in time with the dynamical model until the time of the observations. In the analysis phase, the analysis state $x_a$ is improved using the Kalman gain, which uses a ratio between the background error covariance matrix $B$ and the observation error matrix $R$ to determine the degree of adjustment between data and background information. The background error covariance matrix $B$ will be implicitly flow dependent. In contrast to the classical Kalman Filter, where a single state estimate is taken into account to produce the analysis, the ensemble version of the Kalman Filter, described above, will use an ensemble, i.e., a statistical sample of state estimates, to compute an analysis state for each member. Hence, the sample mean and error covariance are calculated from the background state ensemble to produce a single Kalman gain. We can write the analysis update with the ensemble Kalman filter as

$$x_a = x_b + B H^T (R + H B H^T)^{-1} (y - H x_b),$$

(1)

where $H$ is the linearized observation operator used to convert from the model space to observation space. Then,

$$B = \frac{1}{K - 1} X_b X_b^T$$

(2)

with the background ensemble matrix $X_b$ being the perturbations with respect to the mean of the different $k$ ensemble members

$$X_b = (x_1 - \bar{x}^b, ..., x_k - \bar{x}^b),$$

(3)

where $\bar{x}^b$ is the first guess mean, and $k$ is the number of ensemble members.

There are several techniques to update the ensemble members and we will be focusing on the one we use in the next section: LETKF.
2.2. LETKF

The experiments conducted in Section 3.3 were performed with LETKF (Hunt et al. 2007) as the data assimilation system. The basic idea of the LETKF is to compute the analysis in the space of the ensemble perturbations. This is computationally much more efficient, and as other EnKF methods, also restricts corrections to the subspace spanned by the ensemble. Because of the limited ensemble size compared with the real observation space, an explicit localization is necessary. Through this procedure, in both the covariance and observation space, we increase the effective size of the ensemble space and we can also select the data that are more relevant for the analysis update. This means computing a separate analysis at every grid point, where only certain observations are chosen. Thus, the analysis ensemble members are a local linear combination of first guess ensemble members.

The analysis mean is generated using

$$x_a = y_b + X_b \tilde{P}_a (Hx_b)^T R^{-1} (y - y_b),$$  \hspace{1cm} (4)$$

with $\tilde{P}_a$ being the analysis error covariance matrix in the ensemble space determined by

$$\tilde{P}_a = [(k - 1)I + (Hx_b)^T R^{-1} Hx_b]^{-1}. \hspace{1cm} (5)$$

The analysis ensemble is obtained through

$$X_a = X_b [(k - 1)\tilde{P}_a]^{1/2} = X_b W_a. \hspace{1cm} (6)$$

We are working in a low-dimensional ensemble subspace and the analysis computations are done independently for all grid points, which permit an efficient parallelization. These two features allow for real time computations NWP applications, where the $B$ matrix has $N \times N$ dimension (with $N$ the model size of order of magnitude approximately $10^8$).

2.3. Localization

Let us first describe the current localization techniques more popular in the data assimilation community. Covariance Localization, also called B-localization in (Greybush et al. 2011), is a common localization method primarily introduced in (Houtekamer et al. 2001; Hamill et al. 2001) to mitigate the negative influence of long range spurious correlations, i.e., correlations between variables located at grid points that are not close to one another, and therefore, not expected to be physically related. These correlations appear because of the approximation problem of working in a lower dimensional ensemble space and we know that the true correlation between different state variables decreases with the distance. Covariance localization is the procedure of cutting these remote correlations by multiplying, with a Schur product, the localization function matrix with the $B$ matrix.

The localization function depends on the distance between the grid points and the localization scale we want to employ (distance at which these correlations will be reduced to zero). The update with covariance localization can be expressed as

$$x^{(a)} = x^{(b)} + (\rho_B \circ B)H^T (R + H(\rho_B \circ B)H^T)^{-1} (y - Hy^{(b)}),$$  \hspace{1cm} (7)$$

where $\circ$ means the Schur product or element wise multiplication and $\rho_B$ is the covariance localization matrix (Garpari et al. 1999). Using this procedure, the rank of the background error covariance matrix $B$ is increased.

Spatial Localization, or R-localization (Greybush et al. 2011), is also very popular, and as we are using LETKF for the two-dimensional experiments, it is the method used in Section 3.3. It was first introduced in (Houtekamer et al. 1998) and in this case the localization is performed in the observation space and applied to the $R$ matrix instead of background error covariance matrix. Data that fall out of a certain local region, defined by a radius $\rho$, will not be considered for the update. The observation error covariance matrix is multiplied by a distance-dependent function. Hence, distant observations, beyond the limit determined by the localization radius, will not be included in the analysis computation. We can write the update formula as

$$x^{(a)} = x^{(b)} + BH^T (\rho_R \circ R + HBH^T)^{-1} (y - Hy^{(b)}),$$  \hspace{1cm} (8)$$

where $\rho_R$ is the diagonal space localization matrix. Its non-diagonal elements are zero, taking an approximation only valid for uncorrelated observation errors, and the diagonal ones are the inverse of the same elements used in the $\rho_B$. This means that distant observations will have infinite error and will not be considered for the analysis update. In the LETKF scheme explained above, the R-localization procedure differs from the first approach shown in (Houtekamer et al. 1998), because the weight of the data continuously decreases with the distance to zero, in contrast to the first strategy where all the measurements had the same weight.
3. Optimal localization

3.1. A formula for an optimal localization radius

Finding the optimal localization for every grid point is a challenging goal. Previous attempts in adaptive localization, like the ones performed by Anderson (Anderson 2012) and his sophisticated hierarchical filter (Anderson 2007), or the methods developed by Bishop and Hodyss (Bishop et al. 2007) and the ECO-RAP (Bishop et al. 2009a; Bishop et al. 2009b), are different approaches to the problem for determining the optimal localization flow dependency for different observations and situations of the atmosphere at every time step. Additional works about information-based localization methods (Migliorini 2013; Zupanski et al. 2007) and optimal localization (Kirchgesner 2014) schemes have been developed.

We estimate the optimal localization radius \( \rho_{\text{loc}} \) as a function of the observation error \( \sigma_{\text{obs}} \), which for this work is considered uncorrelated, and the data density \( \mu \).

To be able to model the optimal area for localizing in every case, we minimize the analysis error due to two error contributions in the update process: the effective observation error and the approximation error. The approximation error, also known as undersampling error (Oke et al. 2006), appears because we are working in the ensemble space and the analysis performance depends on the size of the ensemble we use. The ensemble must be statistically representative and span the space of the physical model we approximate. Increasing the size of the ensemble would help to get a better description of the model space into the ensemble one but it would have an expensive computational cost. Applying localization to the EnKF we increase the effective size of the ensemble space; therefore, approximation error will decrease with smaller values of the localization radius \( \rho \). To determine the convergence speed of \( \rho \), we should calculate the parameter \( p \) in \( \rho^p \), which represents the relation between the size of the ensemble space (degrees of freedom) and the truth to approximate. The Taylor expansion method has been used to compute \( p \); hence, we approximate a higher order function with lower order ones (details can be found in traditional books of approximation theory).

The effective observation error \( \delta \) is influenced by the process of sampling a function \( f(x) \), i.e., measuring it at different points \( x_1, x_2, x_3, \ldots, x_N \), and with stochastic error with variance \( \sigma \) in the measurement of the value \( f(x_j) \) at a point \( x_j \). By standard stochastic estimates, it is well known that the error \( \delta \) is proportional to \( \sigma/\sqrt{N} \) and it tends to zero when the sample size tends to infinity. With an homoge-neous observation density \( \mu \) the number of observations \( N \) can be expressed as

\[
N = \int_V \mu dx = c \mu \rho^d,
\]

with some constant \( c \), where \( d \) is the space dimension under consideration. Then the effective observation error decreases with larger values of \( \rho \), as a larger number of observations gives a better statistical estimate.

Considering the two error contributions described before, they lead to the following error asymptotics

\[
E = \alpha \rho^p + \frac{\beta \sigma_{\text{obs}}}{\sqrt{\mu}} \rho^{-\frac{d}{2}}, \quad \rho > 0,
\]

with constants \( \alpha \) and \( \beta \), \( d \) is the dimension of our system and \( p \) the convergence parameter mentioned before. The first term in (3.2) refers to the approximation error and the second one to the effective observation error. The minimum of \( E(\rho) \) as a function of the localization radius \( \rho \) can be obtained by setting the derivative of this equation equal to zero, leading to

\[
\rho_{\min} = c \left( \frac{d}{p} \right)^{\frac{2}{d+2p}},
\]

with some constant \( c \) depending on \( \alpha \), \( \beta \), \( p \), \( \sigma_{\text{obs}} \) and \( \mu \). Both \( \alpha \) and \( \beta \) are tuning parameters with different values for the two systems we use in the numerical experiments in Section 3.2 and Section 3.3, respectively. For the one-dimensional study, \( \alpha = 0.016 \) and \( \beta = 0.75 \). In Fig. 1, we can check the sensitivity plot between the two parameters \( \alpha \) and \( \beta \) for this case. For the two-dimensional example, included in Section 3.3, the tuning parameters are \( \alpha = 1 \) and \( \beta = 30 \).

In Figs. 2 and 2a, we display the curve (3.2) for \( d = 1 \) and \( p = 2 \), where the optimal radius is found by the value

\[
\rho_{\min} = \left( \frac{\beta \sigma_{\text{obs}}}{\sqrt{\mu}} \frac{1}{4\alpha} \right)^{\frac{1}{2}}.
\]

Experiments performed to obtain the numerical results for a single analysis update in Fig. 2b, will be described in Section 3.2. Note that only the unregularized experiments, without taking into account the background information, were used to validate the mathematical expression 3.2. Similar results were found in Figure. 4 of (Greybush et al. 2011), where localization radius and error were also plotted to obtain analogous curves, in a more realistic set-up including other variables, and in this case also, balancing aspects were taken into account. In our experiments, they were not being considered because we have a one-variable model.
In the second example developed in a two-dimensional space, i.e., with $d = 2$ and $p = 1$, with LETKF as data assimilation method, we find Fig. 3b, where the corresponding theoretical radius that minimizes the error analysis is determined by

$$\rho_{\text{min}} = \left( \frac{\beta \sigma_{\text{obs}} }{\sqrt{\mu} \alpha} \right)^{\frac{1}{2}}.$$

The numerical experiments, which were performed for a single analysis cycle, used to obtain Fig. 3b, are explained in detail in Section 3.3. We find some small discrepancies between experimental and theoretical curves, which are because of the fact that the analytical expression (3.2) has been developed on the basis of asymptotic estimates for $\rho \to 0$ and $N \to \infty$. For examples, Figs. 2b and 3b show an approximation to these estimates.
3.2. A one-dimensional example with Least Squares Analysis

The EnKF analysis carries out a least squares problem calculation in the ensemble space $U^{(L)}$. Here, we study the behavior of such problems in an interval $[0, A]$ with $A \in \mathbb{R}$ when the ensemble space is given by linear functions determined by the parameters $a$ and $b$, i.e.,

$$U^{(L)} = \{a + bx : a, b \in \mathbb{R}\} \subset L^2([0, A]).$$

(14)

We study the truth $x^{(\text{true})}$ given by a quadratic function

$$x^{(\text{true})}(x) = \lambda (x - \gamma)^2, \quad x \in [0, A],$$

with constants $\lambda, \gamma \in \mathbb{R}$. Observations are drawn from a Gaussian distribution with variance $\sigma_{\text{obs}}$. We use $N_{\text{tot}}$ observations on $[0, A]$. Here, we carry out localization by a decomposition of $[0, A]$ into $q \in \mathbb{N}$ subsets $[A_j, A_{j+1}]$ where

$$A_j := \frac{j A}{q}, \quad j = 0, ..., q.$$

(16)

Then, on each subset, the analysis is carried out by solving the least squares problem in $U^{(L)}_{[A_j, A_{j+1}]}$ with the observations that are supported in this subset $[A_j, A_{j+1}]$. The localization radius here corresponds to half the interval length, i.e., $\rho = A/2q$.

We remark that the unregularized analysis is the plain regression fit by linear functions on the intervals $[A_j, A_{j+1}]$. The result are the coefficients $a_j^{(a)}$ and $b_j^{(a)}$ such that the best fit is given by

$$x^{(a)}(x) = a_j^{(a)} + b_j^{(a)} x, \quad \text{for } x \in [A_j, A_{j+1}].$$

(17)

Let $x_{\xi}, \xi = 1, ..., N_j$ be the measurement points in the interval $[A_j, A_{j+1}]$, then the matrix $G$ and the measured data vector $f$ are defined by

$$G_j := \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_{N_j} \end{pmatrix}, \quad f_j := \begin{pmatrix} f(x_1) \\ \vdots \\ f(x_{N_j}) \end{pmatrix}. \quad (18)$$

Further, let the observation error covariance matrix be given by $R = \sigma_{\text{obs}}^2 I$ (a $N_j \times N_j$ matrix) and the background error covariance matrix $B = \sigma_{\text{bg}}^2 I$ with dimension $2 \times 2$. Then, the least squares solution is given by

$$\begin{pmatrix} a_j^{(a)} \\ b_j^{(a)} \end{pmatrix} = (G_j^T G_j)^{-1} G_j^T f_j. \quad (19)$$

When a piecewise linear background term with coefficients $a_j^{(b)}$ and $b_j^{(b)}, j = 1, ..., q,$ is employed, if we substitute these values in the original update formula (2.1), the result of the analysis is given by

$$\begin{pmatrix} a_j^{(a)} \\ b_j^{(a)} \end{pmatrix} = \begin{pmatrix} a_j^{(b)} \\ b_j^{(b)} \end{pmatrix} + \sigma_{\text{bg}}^2 (\sigma_{\text{obs}}^2 + \sigma_{\text{bg}}^2 G^T G)^{-1} G^T (f_j - G \begin{pmatrix} a_j^{(b)} \\ b_j^{(b)} \end{pmatrix}).$$

(20)
To compare with the EnKF update, we are assuming the observation operator $H$ (called $G$ in this LSA example) is the identity matrix, and the $B$ and $R$ matrices are diagonal.

We now investigate particular generic situations reflected by special choices of parameters. For the observation error, we employ $\sigma_{\text{obs}} \in \{0.0005, 0.05, 0.5\}$. The true function $x^{(\text{true})}$ is given by (3.7) with $\lambda = 0.05$ and $\gamma = 4.5$, where the interval is defined by $A = 9$. We choose a total of $N = 96$ observations and use the localization radii $\rho$ in $\{A/2 = 4.5, A/4 = 2.25, A/6 = 1.5, A/8 = 1.125, A/12 = 0.75, A/16 = 0.5625, A/30 = 0.375, A/48 = 0.1875, A/96 = 0.09375\}$.

In Figs. 4–6 we demonstrate the result of the analysis for the above choices of $\rho$ using $A/q = 2\rho$ in (3.8) and for the different choices of the observation error given by $\sigma_{\text{obs}}$. The blue line is the truth, whereas the observations are represented by the blue circles. The green line is the (global) background, the red line shows the unregularized analysis (not considering the background information to compute the analysis), and the black line is the regularized analysis (where the influence of the background information is used).

Starting with a very small observation error $\sigma_{\text{obs}} = 0.0005$, shown in Fig. 4, we find very similar results for the unregularized and the regularized analysis. This can be expected from the EnKF update equations for $a^{(a)}$, $b^{(a)}$, which tends to $a^{(a)}$, $b^{(a)}$ for $\sigma_{\text{obs}} \to 0$. The optimal result (which minimizes the error analysis) is obtained for $\rho = 0.1875$.

For a medium value of $\sigma_{\text{obs}} = 0.05$, Fig. 5, a small difference between the unregularized analysis and the regularized analysis becomes visible, but the $l^2$-differences remain to be comparable in size. The optimal value, for the unregularized analysis, of $\rho$ shifts to larger values, i.e., $\rho = 1.125$.

With a large observation error, $\sigma_{\text{obs}} = 0.05$, as shown in Fig. 6, pronounced differences between the unregularized analysis and the regularized analysis are visible; now, we obtain better results for the regularized analysis, especially for small localization radii. This behavior corresponds to the regularizing effect in D systems where the background stabilizes the analysis in the case of noisy data. The optimal value of $\rho$ shifts to even larger values,, i.e., $\rho = 2.25$.

### 3.3. Estimating fronts in a two-dimensional example

Now, we demonstrate the behavior of the LETKF (2.6) in the analysis step by a two-dimensional example, where a curved front is estimated given some particular ensemble space. Again, we will focus on the approximation properties that arise when the front is not in the ensemble space, i.e., we try to approximate the truth, which has a lot more of degrees of freedom, using a lower dimensional ensemble.

Here, a simple setup is given by studying an
ensemble space where the fronts have no curvature compared to reality. We use straight-line type of fronts defined by

\[ x_t(x) := \begin{cases} 
1, & x \cdot v_l \leq \rho_l \\
0, & \text{otherwise,}
\end{cases} \]  

(21)
on $x \in D := [a_1, b_1] \times [a_2, b_2]$ with constants $a_1, a_2, b_1,$ and $b_2,$ where $v_l \in \mathbb{R}^2$ and $\rho_l \in \mathbb{R}$ parametrize the straight-line front of the ensemble member to build the ensemble space $U^{(2)}$, as shown in Fig. 7. The true front given by a function $x^{(true)}$ is shown in Figs. 8a or 9a, see also the 6th subfigure in Fig. 7. For the experiments, 10000 (200 × 50) grid points were defined with one observation per grid point.

Fig. 5. Observation error takes an intermediate value in this example with $\sigma_{obs} = 0.05$. As in the previous Fig. 4, no localization is applied for (a). The same values of $\rho$, as in Fig. 4, are used in (b), (c), and (d) with localization radii being progressively reduced. Note that in all subfigures we work with dimensionless quantities.

Fig. 6. An example with a higher observation error, $\sigma_{obs} = 0.5$, is provided. As in previous Figs. 4 and 5 the same values of $\rho$ were selected, the analysis in (a) is computed with $\rho = 0,$ being progressively smaller in (b), (c) and (d) cases. Note that in all subfigures we work with dimensionless quantities.
When we study the analysis error in $L^2(D)$, the approximation order that can be achieved comes from local approximations of a quadratic front by straight-line fronts, which leads to $\rho = 1$ and $d = 2$. The dependence of the error on $\rho$ as modeled in (3.2) is shown in Fig. 3.

Figures 8b and 9b show the best approximation in ensemble space when no localization is employed. The terrace-like appearance comes from the discretization that is enforced by the ensembles under consideration, but the effect shown here will also appear when other ensembles are used. Localization allows a much better approximation of the front by the terraced functions in small environments of some points, leading to an analysis as shown in Figs. 8c and 8d, where Fig. 8c has been obtained with a localization radius $\rho = 15$, while Fig. 8d employs $\rho = 5$. Here, adding random error, we obtain analogous results similar to Figs. 8b–d comparing to Figs. 9b–d.

4. Discussion and conclusions

To achieve high quality assimilation with EnKF, the adequate localization distance must be selected for each different observation configuration and particular atmospheric situation. If the distance we choose is too large compared with the degrees of freedom that the ensemble members allow, it would be difficult to approach the truth with the analysis. On the other hand, if the localization length scale is too small the observation error will dominate and we will not obtain a good analysis performance because noisy observations will overfit and the subsequent forecast will be affected. We would need a larger region to properly estimate the analysis and remove the influence of the observation errors.

Given a homogeneous distribution of observations, it is possible to determine the optimal localization length scale for the LETKF in terms of the observation error (background error is kept fixed) and the density of measurements, as it is described in Section 3.2. We are obtaining analogous results in experiments developed with the Lorenz-95 model with LETKF, which will be included in our next work. We plan to test our theoretical results about localization in inhomogeneous environments, which are more realistic, like the model based on the shallow water equations developed by (Würsch et al. 2013), or the high-resolution model with radar observations set-up created by (Lange and Craig 2014). We will analyze how localization behaves with different degrees of inhomogeneous distributions of observations to adapt the expression (3.2) to these cases. Similarly, it should be very interesting to compare our analytical expression for the optimal $R$ localization distance in the observation space with its equivalent derived in the context of covariance localization optimum length scale (Yoon et al. 2010). The localization must be large enough to assimilate the most relevant data and small enough to avoid long-range
spurious correlations.

At the DWD (German Weather Service), we are testing the so-called sequential or serial assimilations using real NWP models. We are interested in studying the theoretical expression for the optimal localization for every step analysis to minimize the total error analysis. In the case of two-step assimilation system, the latter is described as

\[
\begin{align*}
    x_1^{(a)} &= x^{(b)} + BH_T^T (\rho R_1 \circ R_1 + H_1 H_T)^{-1} (y_1 - H_1 x^{(b)}), \\
    x_2^{(a)} &= x_1^{(a)} + B_2 H_T^T (\rho R_2 \circ R_2 + H_2 H_T)^{-1} (y_2 - H_2 x_1^{(a)}), \\
    x_{\text{total}}^{(a)} &= x_2^{(a)}.
\end{align*}
\]

which consists of processing several groups of observations for a certain time in several steps. The analysis obtained for the assimilation of the first group of

Fig. 8. Example for \( \sigma_{\text{obs}} = 0.1 \). The truth (front) and observations (red crosses) are represented in (a). The truth in (a) is approximated without any localization procedure in (b). Different localization radii are applied to perform the analysis: in (c) with \( \rho = 15 \) grid points, with \( \rho = 5 \) grid points in (d), and \( \rho = 3 \) grid points in (e). Note that in all subfigures we work with dimensionless quantities.
observations is going to be used for the following one as background. Equation (4.1) was first described in the data assimilation context by (Houtekamer et al. 1998) and (Whitaker et al. 2002). Additional efforts include the successive covariance localization (Zhang et al. 2009) and the multi-scale localization approach (Miyoshi et al. 2013; Kondo et al. 2013). More information about comparing simultaneous and sequential assimilation can be found in (Holland et al. 2013).

Fig. 9. Example for $\sigma_{obs} = 0.5$. As in Fig. 8, in (a) the truth and observations are shown. The analysis is approximated without any localization in (b). A localization radius of $\rho = 15$ grid points is applied in (c), $\rho = 5$ grid points in (d), and $\rho = 1$ grid points in (e). Note that in all subfigures, we work with dimensionless quantities.

**References**


Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998:


Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data


