

A regulated localization scheme for ensemble-based Kalman filters

L. Nerger*, Tijana Janjić, Jens Schröter, and Wolfgang Hiller

Alfred Wegener Institute, Bremerhaven, Germany

*Correspondence to: Lars Nerger, Alfred Wegener Institute, Am Handelshafen 12, D-27570 Bremerhaven, Germany.

E-mail: Lars.Nerger@awi.de

Localization is an essential element of ensemble-based Kalman filters in large-scale systems. Two localization methods are commonly used: Covariance localization and domain localization. The former applies a localizing weight to the forecast covariance matrix while the latter splits the assimilation into local regions in which independent assimilation updates are performed. The domain localization is usually combined with observation localization, which is a weighting of the observation error covariance matrix, resulting in a similar localization effect to that of covariance localized filters. It is shown that the use of the same localization function in covariance localization and observation localization results in distinct effective localization length scales in the Kalman gain. In order to improve the performance of observation localization, a regulated localization scheme is introduced. Twin experiments with the Lorenz-96 model demonstrate that the regulated localization can lead to a significant reduction of the estimation errors as well as increased stability of the assimilation process. Copyright © 0000 Royal Meteorological Society

Key Words: data assimilation; ensemble Kalman filter; localization

Received ...

Citation: ...

1. Introduction

Ensemble-based Kalman filter algorithms have evolved significantly since the introduction of the original so-called Ensemble Kalman Filter (EnKF, Evensen 1994). Among the recent developments are ensemble square-root Kalman filters like the Ensemble Transform Kalman Filter (ETKF, Bishop *et al.* 2001) the Ensemble Adjustment Kalman Filter (EAKF, Anderson 2001), and the Ensemble Square-root Kalman filter with sequential processing of observations (EnSRF, Whitaker and Hamill 2002). These algorithms avoid the need to generate an ensemble of perturbed observations required in the EnKF (Burgers *et al.* 1998; Houtekamer and Mitchell 1998). Instead, the analysis equation of the Kalman filter is expressed in a square-root form and combined with an explicit transformation of the state ensemble (see Tippett *et al.* 2003). Similar computations are performed by the Singular “Evolutive” Interpolated Kalman (SEIK) filter (Pham *et al.* 1998; Pham 2001).

The computation time of a data assimilation application using an ensemble-based Kalman filter is dominated by the time integration of the ensemble of model states. To keep the computation time low, the ensemble is typically chosen to be small, even for large scale models. Small ensembles, however, will lead to significant sampling errors of the estimated error covariance matrix in particular for long-range covariances. This sampling error can lead to a divergence of the filter in which the state estimate diverges from the true state without accurately estimating the error (Houtekamer and Mitchell 1998). This finding has motivated the localization of covariance matrices, such that long-distance covariances are damped or neglected. In addition, the localization increases the rank of the forecast covariance matrix and increases the local number of degrees of freedom for the analysis.

Houtekamer and Mitchell (1998, 2001) applied localization to the forecast covariance matrix. The method is

denoted *covariance localization* (CL) and uses an element-wise (i.e. Schur/Hadamard) product of the ensemble covariance matrix with a chosen correlation matrix of compact support. Frequently, a 5-th order polynomial function, which mimics a Gaussian function but has compact support (Gaspari and Cohn 1999), below referred to as ‘GC function’, is used for the localization. CL is only possible if the forecast covariance matrix, or its projection onto observation space, is explicitly computed. Although this is the case for the EnSRF, this matrix is never computed in the ETKF and the SEIK filter (for a discussion of this issue see Janjić *et al.* 2011). To enable localization in these filters, so-called *domain localization* (DL) is applied (e.g. Ott04a, Nerger06a, Hunt07a). Here a sequence of local analyses is performed in which disjoint domains in the physical space are updated independently by the filter analysis. For each local analysis, observations within some defined cut-off radius are considered. The observational domains can be larger than the local analysis domains, which ensures some smoothness of the state analysis estimate. This localization method was standard in Optimal Interpolation (see, e.g. Cohn *et al.* 1998) and was also used for the EnKF (Haugen and Evensen 2002; Brusdal *et al.* 2003).

The method of *observation localization* (OL) was introduced (Hunt *et al.* 2007; Nerger and Gregg 2007) to obtain with DL a similar localization effect to CL for a general localization function. In OL the inverse of the observation error covariance matrix corresponding to a local analysis domain is Schur-multiplied with a chosen localization matrix that is constructed using correlation functions of compact support. Thus, the weight of observations is reduced as a function of their distance from the local analysis domain by increasing their assumed error variance. Miyoshi and Yamane (2007) discussed that the effect of OL is similar to CL but generally results in a weaker localization. The relation of OL and CL has been studied in detail by Sakov and Bertino (2011), Greybush *et al.* (2011) as well as Janjić *et al.* (2011). Janjić *et al.*

(2011) found, by using twin experiments with the Lorenz-96 model, that CL leads to smaller estimation errors if the observation error variance is smaller than the initial estimate of forecast variance. If observation errors dominate, both localization methods showed analogous performance. A similar behavior was described by Sakov and Bertino (2011). Greybush *et al.* (2011) describe that the optimal localization length is wider for CL than for OL.

In this study, the relation of the localization effects of CL and OL is utilized to formulate a scheme for OL that computes a regulated localization function. The method adaptively regulates the width of the localization function based on the estimated variances of the observations and the forecast state. Hence, it will be denoted 'regulated OL' below. The regulated localization method introduced here only aims at avoiding the disadvantage of the commonly used fixed OL. Thus, its motivation is different from adaptive localization methods (e.g., Anderson 2007; Bishop and Hodyss 2009).

The regulated localization scheme will be examined here in the context of the domain-localized SEIK filter (Nerger *et al.* 2006). However, it can be applied analogously in domain-localized ensemble square-root filters like the LETKF (Hunt *et al.* 2007). The SEIK filter and an EnKF square-root formulation will be reviewed in section 2. Section 3 discusses the common localization methods. Subsequently, the regulated localization will be formulated in section 4. The influence of regulated localization will be studied in twin assimilation experiments with the Lorenz-96 model in section 5. Finally, conclusions are drawn in section 6.

2. Filter algorithms

As a prototype of a filter algorithm that applies DL combined with OL, the SEIK filter is considered. CL is commonly applied with the EnSRF and the EAKF. However, as the EnSRF and EAKF operate sequentially on the observations, their result will depend on the order of the observations (Whitaker *et al.* 2008, discuss a scheme

to sort the observations for optimal performance). Here, a very simple square-root formulation of an ensemble Kalman filter (following Sakov and Bertino 2011), denoted EnKF-sqrt, is considered that allows to apply CL when assimilating all observations at once.

As localization is an additional feature that can be imposed onto a filter algorithm, the global formulations of the SEIK filter and the EnKF-sqrt are discussed here before the localization methods are discussed in section 3.1.

In filter methods based on the Kalman filter, the state vector \mathbf{x}_k^a of dimension n at some time t_k estimates the true state of a physical system, such as the ocean or the atmosphere. The corresponding covariance matrix \mathbf{P}_k^a represents the error estimate of the state vector. The superscript 'a' denotes the analysis, while 'f' denotes the forecast. In ensemble based filters these quantities are represented by an ensemble of N vectors $\mathbf{x}^{a(\alpha)}$, $\alpha = 1, \dots, N$, of model state realizations

$$\mathbf{X}_k^a = [\mathbf{x}_k^{a(1)}, \dots, \mathbf{x}_k^{a(N)}]. \quad (1)$$

The state estimate is given by the ensemble mean

$$\overline{\mathbf{x}}_k^a = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_k^{a(i)}, \quad \overline{\mathbf{X}}_k^a = [\overline{\mathbf{x}}_k^a, \dots, \overline{\mathbf{x}}_k^a] \quad (2)$$

and \mathbf{P}_k^a is approximated by the ensemble covariance matrix

$$\hat{\mathbf{P}}_k^a = \frac{1}{N-1} \mathbf{X}'_k{}^a (\mathbf{X}'_k{}^a)^T \approx \mathbf{P}_k^a. \quad (3)$$

Here, $\mathbf{X}' = \mathbf{X} - \overline{\mathbf{X}}$ denotes the matrix of ensemble perturbations.

2.1. The SEIK filter

The SEIK filter is presented here shortly, following the formulation used by Nerger and Gregg (2008). As all operations of the analysis step are performed at the time t_k , we omit the time index k . $\hat{\mathbf{P}}^f$ can be computed from the

forecast ensemble \mathbf{X}^f according to

$$\hat{\mathbf{P}}^f = \mathbf{LGL}^T \quad (4)$$

with

$$\mathbf{L} = \mathbf{X}^f \mathbf{T}, \quad \mathbf{G} = (N-1)^{-1} (\mathbf{T}^T \mathbf{T})^{-1}. \quad (5)$$

The matrix \mathbf{G} has size $(N-1) \times (N-1)$ and \mathbf{T} is a matrix of size $N \times (N-1)$ with all entries being equal to $-N^{-1}$ except for those in the diagonal, which are equal to $1 - N^{-1}$. Matrix \mathbf{T} has zero column sums and implicitly subtracts the ensemble mean when computing $\hat{\mathbf{P}}^f$.

The analysis update of the state estimate is given by

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^f + \mathbf{L}\mathbf{a} \quad (6)$$

where the vector \mathbf{a} of size $N-1$ is

$$\mathbf{a} = \mathbf{U}(\mathbf{HL})^T \mathbf{R}^{-1} (\mathbf{y}^o - \mathbf{H}\bar{\mathbf{x}}^f), \quad (7)$$

$$\mathbf{U}^{-1} = \rho \mathbf{G}^{-1} + (\mathbf{HL})^T \mathbf{R}^{-1} \mathbf{HL}. \quad (8)$$

Here, \mathbf{H} is the observation operator. \mathbf{y}^o denotes the vector of observations of size m with observation error covariance matrix \mathbf{R} . ρ with $0 < \rho \leq 1$, is denoted forgetting factor. It is the inverse of the covariance inflation factor used, for example, in the ETKF. The analysis covariance matrix is given by $\hat{\mathbf{P}}^a = \mathbf{LUL}^T$, but does not need to be explicitly computed.

The analysis ensemble \mathbf{X}^a is obtained by transforming the forecast ensemble such that it represents $\bar{\mathbf{x}}^a$ and $\hat{\mathbf{P}}^a$. The transformation is performed by a $N \times (N-1)$ random matrix $\mathbf{\Omega}$ that is generated from uniformly distributed random numbers. The columns of $\mathbf{\Omega}$ are constrained to be orthonormal and orthogonal to the vector $(1, \dots, 1)^T$, which implies that each column has zero mean and is a vector of unit norm. The analysis ensemble is

$$\mathbf{X}^a = \bar{\mathbf{x}}^a + \sqrt{N-1} \mathbf{L}\mathbf{C}^T \mathbf{\Omega}^T \quad (9)$$

where a Cholesky decomposition is applied to the matrix \mathbf{U}^{-1} to obtain $\mathbf{C}^{-1}(\mathbf{C}^{-1})^T = \mathbf{U}^{-1}$.

2.2. An EnKF square-root formulation

The EnKF-sqrt is a simple formulation of a square-root filter that allows to apply CL (see Sakov and Bertino 2011). The update of the state estimate is performed according to

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^f + \mathbf{K} (\mathbf{y}^o - \mathbf{H}\bar{\mathbf{x}}^f) \quad (10)$$

with the Kalman gain \mathbf{K} given by

$$\mathbf{K} = \hat{\mathbf{P}}^f \mathbf{H}^T (\mathbf{H}\hat{\mathbf{P}}^f \mathbf{H}^T + \mathbf{R})^{-1}. \quad (11)$$

The ensemble transformation is performed by multiplying the forecast ensemble perturbations \mathbf{X}^f from the left according to

$$\mathbf{X}^{fa} = \tilde{\mathbf{C}} \mathbf{X}^f \quad (12)$$

with

$$\tilde{\mathbf{C}} = (\mathbf{I} + \hat{\mathbf{P}}^f \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1/2}. \quad (13)$$

The square-root in Eq. (13) is computed as the symmetric square-root. Under these conditions, the ensemble transformation preserves the ensemble mean. A forgetting factor $\tilde{\rho}$ can be applied in this algorithm by dividing \mathbf{X}^f by $\tilde{\rho}$ before computing the analysis update. The matrix of analysis ensemble perturbations can be multiplied by a random rotation matrix $\tilde{\mathbf{\Omega}}$ similar to matrix $\mathbf{\Omega}$ used in the SEIK filter.

The ensemble transformation according to Eqns. (12) and (13) is only applicable if the matrix $\mathbf{I} + \hat{\mathbf{P}}^f \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}$ is symmetric. This will be the case if the state is fully observed and if \mathbf{R} is diagonal with equal variance for all observations. Under these constraints, the EnKF-sqrt method allows to access the influence of CL in small scale systems like the Lorenz-96 Model used in section 5.

3. Localization methods

3.1. Domain localization

Domain localization is discussed here shortly for the SEIK filter. More details can be found in Nerger *et al.* (2006). The localization method is similar to that applied in the LETKF (Hunt *et al.* 2007) and analogous to the practical implementation discussed by Miyoshi *et al.* (2007).

For the DL, the operations of the analysis and the ensemble transformation are organized in a loop through disjoint local analysis domains of the model grid. For simplicity, a local analysis domain can be a single vertical column of a 3-dimensional model or a single grid point. This reordering of the operations will not change the results of the analysis and ensemble transformation steps as long as all globally available observations are considered in the analysis step and if OL (see section 3.2) is not applied.

For each local analysis domain, the analysis is performed using only observations within a prescribed influence distance from the local analysis domain. Let the subscript σ denote a local analysis domain. The domain of the corresponding observations is denoted by the subscript δ . Then, the equations for the local SEIK analysis and ensemble transformation can be written analogously to the global analysis equations (6 – 9) as

$$\mathbf{x}_\sigma^a = \overline{\mathbf{x}}_\sigma^f + \mathbf{L}_\sigma \mathbf{a}_\delta, \quad (14)$$

$$\mathbf{a}_\delta = \mathbf{U}_\delta (\mathbf{H}_\delta \mathbf{L})^T (\mathbf{R}_\delta)^{-1} (\mathbf{y}_\delta^o - \mathbf{H}_\delta \overline{\mathbf{x}}^f), \quad (15)$$

$$\mathbf{U}_\delta^{-1} = \rho_\delta \mathbf{G}^{-1} + (\mathbf{H}_\delta \mathbf{L})^T (\mathbf{R}_\delta)^{-1} \mathbf{H}_\delta \mathbf{L}, \quad (16)$$

$$\mathbf{X}_\sigma^a = \overline{\mathbf{X}}_\sigma^a + \sqrt{N-1} \mathbf{L}_\sigma (\mathbf{C}_\delta)^T \boldsymbol{\Omega}^T \quad (17)$$

where $\mathbf{C}_\delta^{-1} (\mathbf{C}_\delta^{-1})^T = \mathbf{U}_\delta^{-1}$. \mathbf{H}_δ is the observation operator that projects a global state vector onto the local observation domain. Thus, it combines the operation of a global observation operator with the restriction of the observation vector to the local observation domain. \mathbf{R}_δ is the observation error covariance matrix on the local observation domain. ρ_δ denotes the local forgetting factor, which can

vary for different local analysis domains. For the ensemble transformation, the same matrix $\boldsymbol{\Omega}$ has to be used for each local analysis domain to ensure consistent transformations throughout all local domains.

3.2. Observation localization

OL is commonly described as an addition to DL. Thus, OL always implies DL. With OL, each local observation error covariance matrix is weighted such that the influence of observations is reduced with increasing distance from the local analysis domain (Hunt *et al.* 2007; Nerger and Gregg 2007). The localization is performed by a Schur product of the inverse observation error covariance matrix \mathbf{R}_δ^{-1} with a localization matrix $\tilde{\mathbf{D}}$. Hence, equations (15) and (16) are rewritten as

$$\mathbf{a}_\delta = \mathbf{U}_\delta (\mathbf{H}_\delta \mathbf{L})^T (\tilde{\mathbf{D}} \circ \mathbf{R}_\delta^{-1}) (\mathbf{y}_\delta^o - \mathbf{H}_\delta \overline{\mathbf{x}}^f) \quad (18)$$

$$\mathbf{U}_\delta^{-1} = \rho_\delta \mathbf{G}^{-1} + (\mathbf{H}_\delta \mathbf{L})^T (\tilde{\mathbf{D}} \circ \mathbf{R}_\delta^{-1}) \mathbf{H}_\delta \mathbf{L}. \quad (19)$$

Here \circ denotes the Schur product. $\tilde{\mathbf{D}}$ is usually constructed using correlation functions of compact support. Possible choices are, for example, an exponential decrease or the GC function. Under the common assumption that the matrix \mathbf{R} is diagonal, $\tilde{\mathbf{D}}$ can be a diagonal matrix with elements varying according to the distance of an observation from the local analysis domain.

An alternative to the localization of \mathbf{R}_δ^{-1} was used by Nerger and Gregg (2008). Here, the covariance localization was performed by a Schur product of a localization matrix with the matrix $\mathbf{H}_\delta \mathbf{L}$. Sakov and Bertino (2011) stressed that the localizations of $\mathbf{H}_\delta \mathbf{L}$ and \mathbf{R}_δ^{-1} are equivalent if the Schur product with the localization matrix is also applied to the residual $\mathbf{y}_\delta^o - \mathbf{H}_\delta \overline{\mathbf{x}}^f$.

3.3. Covariance localization

In the EnKF-sqrt method as well as the EnSRF and the original EnKF CL can be directly applied to the forecast error covariance matrix. In practice, the matrices $\mathbf{P}^f \mathbf{H}^T$ and $\mathbf{H} \mathbf{P}^f \mathbf{H}^T$ are Schur-multiplied with correlation matrices

\mathbf{W}_{PH} and \mathbf{W}_{HPH} of compact support. Thus, the localized matrices

$$(\mathbf{P}^f \mathbf{H}^T)_{loc} = \mathbf{W}_{PH} \circ (\mathbf{P}^f \mathbf{H}^T) \quad (20)$$

and

$$(\mathbf{HP}^f \mathbf{H}^T)_{loc} = \mathbf{W}_{HPH} \circ (\mathbf{HP}^f \mathbf{H}^T) \quad (21)$$

are used. If observations are processed sequentially as in the EnSRF, only the part of the localized matrices that corresponds to each single observation needs to be computed. The matrix \mathbf{W}_{HPH} in Eq. (21) reduces to a scalar in this case.

Analogously to OL, the application of CL to the matrices $\mathbf{P}^f \mathbf{H}^T$ and $\mathbf{HP}^f \mathbf{H}^T$ relates to the observation space. The difficulties to define distance for vertically integrated measurements have been pointed out by Campbell *et al.* (2010). These difficulties exist analogously for the OL and CL. In this study, the observations are assumed to be defined on grid points. A distance measure is thus well defined.

4. Regulated localization

4.1. Effective localization of the Kalman gain

Janjić *et al.* (2011) discussed on the example of the Lorenz-96 model that CL can lead to superior assimilation estimates compared to OL. Also the experiments discussed by Greybush *et al.* (2011) exhibited a slightly better performance of CL than OL. These studies motivate to examine the reason for the different performances of CL and OL.

Miyoshi and Yamane (2007) pointed out that OL is not equivalent to CL. The different influences can be seen on the Kalman gain \mathbf{K} of the traditional analysis equations (10) and (11) of the Kalman filter. For a single observation with variance σ_R^2 the matrix $\mathbf{HP}^f \mathbf{H}^T$ becomes a scalar value, which is denoted HPH^T . The gain \mathbf{K} as well as $\mathbf{P}^f \mathbf{H}^T$ are now of size $n \times 1$. Below, only the i 'th element of these vectors is considered. In this case, the matrix

\mathbf{W}_{PH} in Eq. (20) reduces to a scalar that is given by the correlation function $w_{(i)}^{CL}$ whose value is defined by the distance between location of the i 'th element and the observation. The matrix \mathbf{W}_{HPH} in Eq. (21) becomes a scalar with value one. The i 'th element of the localized gain for CL is now (see Miyoshi and Yamane 2007)

$$\mathbf{K}_{(i)}^{CL} = \frac{w_{(i)}^{CL}}{HPH^T + \sigma_R^2} (\mathbf{P}^f \mathbf{H}^T)_{(i)}. \quad (22)$$

The localized gain for OL is obtained by dividing the observation error variance in the gain by $w_{(i)}^{OL}$. Then it is

$$\mathbf{K}_{(i)}^{OL} = \frac{w_{(i)}^{OL}}{w_{(i)}^{OL} HPH^T + \sigma_R^2} (\mathbf{P}^f \mathbf{H}^T)_{(i)}. \quad (23)$$

To exemplify the different effects of both localization methods, the following example is considered: w^{CL} and w^{OL} are identical and given by a Gauss function with variance 1000, while HPH^T and $(\mathbf{P}^f \mathbf{H}^T)_{(i)}$ are set to one. Figure 1 shows the value of i 'th element of the Kalman gains in (22) and (23) as a function of the distance of the observation from the analysis location for this example. Three values of σ_R^2 are considered: 10, 1, and 0.1.

Figure 1 allows to discuss the effective localization length scale of the analysis, which is the localization length scale of the gain. For simplicity, the length scale is considered to be the distance at which the gain or the correlation function is one half of the value at zero distance. The length scale of the localization in the gain for CL is always equal to the length scale of the function w^{CL} . In contrast, the localization length scale in the gain for OL is distinct from the length scale of the function w^{OL} . For $\sigma_R^2 = 10$, i.e. ten times the value of HPH^T , the gains for CL and OL are almost identical. Thus, the effect of the localization schemes should be nearly indistinguishable. For decreasing ratio of σ_R^2 to HPH^T , the difference between the length scales increases. While w^{OL} and w^{CL} remain identical, the localization length scale in the gain becomes wider for OL. Thus, while the function w^{OL} might

indicate a very small influence of a distant observation, its weight in the analysis might be much larger because of a larger effective localization weight in the gain. To obtain comparable localization length scales one would need to decrease the length scale for w^{OL} . However, the gains in Eqs. (22) and (23) are distinct functions of the distance. For comparable length scale, the gain with OL will be larger for short distances and smaller for long distances than with CL.

The effective localization length scale discussed above provides an explanation for the findings of Greybush *et al.* (2011). They found on the basis of assimilation experiments that for an optimal assimilation performance a smaller localization radius is required for OL than for CL (Fig. 4 of Greybush *et al.* 2011). In addition, the OL resulted in slight larger errors than CL. Considering the effective localization length, the smaller localization radius is required because of the longer effective localization length scale of OL. The better performance of CL might be caused by the different shape of the effective localization functions for comparable localization length scales. Similarly a different effect of the CL and OL on imbalance (Fig. 5 and 6 of Greybush *et al.* 2011) can be attributed to the different effective localization length scales. In addition, the different shapes of the effective localization functions for comparable length scales can lead to different levels of imbalance.

The dependence of the effective localization length of OL on the relative size of the forecast error variance to the observation error variance can also be relevant during the initial transient phase of a data assimilation experiment. Typically, the initial errors of the state estimate are large. They are reduced during the initial transient phase of the data assimilation sequence until they reach some asymptotic level. In contrast, the errors of the assimilated observations are independent of the transient phase. Frequently, the initially estimated variance of the state is of the same order as the observation error variance or larger than this. If w^{OL} is identical to w^{CL} , the assimilation with OL will start with a significantly longer effective localization length than with CL. Thus, observations at an intermediate distance

will have a larger influence in the analysis. However, if the correlation function w^{OL} has compact support, the effective localization function reaches zero at the same distance as the prescribed function w^{OL} . In this case, the total number of observations that are used in the local analysis remains constant.

During the transient phase the effective localization length will become shorter until it reaches an asymptotic level. In general, one could choose the support radius for OL such that the effective localization width is comparable to that of CL when the asymptotic phase is reached. However, in the numerical experiments discussed below, the initially large effective localization length led to instabilities during the transient phase of the assimilation process.

4.2. Regulating the localization width

To avoid a long effective localization length, one can adjust the width of the effective localization that depends on the ratio of the observation variance to the forecast state error variance. This adjustment is achieved by the regulated localization function that is derived in this section.

For the regulated localization method, the single-observation example of the previous section is considered again. The same effective localization length for OL and CL can be obtained by requiring that right hand sides of Eqs. (22) and (23) are equal. This condition leads to the equation for the regulated weight w^{OLR} as a function of w^{CL} :

$$w^{OLR} = \frac{w^{CL}\sigma_R^2}{HPH^T + \sigma_R^2} \left(1 - \frac{w^{CL}HPH^T}{HPH^T + \sigma_R^2} \right)^{-1} \quad (24)$$

Using Eq. (24) for OL will result in identical effective localizations of the gain for OL and CL. Further, w^{OLR} is a correlation function as long as w^{CL} is a correlation function.

The regulated localization function w^{OLR} is exemplified in Fig. 2 for three values of σ_R^2 (10, 1, and 0.1). As in Fig. 1, w^{CL} is chosen to be a Gaussian function with variance 1000. While for $\sigma_R^2 = 10$ both weight functions

lie on top of each other, w^{OLR} narrows with decreasing σ_R^2 to keep the effective localization length of the gain constant.

Eq. (24) for the regulated OL is only exact in case of a single observation. In general, the exact regulated function varies with the number of observations. Appendix B discusses the case of 2 observations. The computation of the exact regulated localization function becomes increasingly costly for multiple observations. However, Eq. (24) is an approximation in the case of multiple observations that reduces the variation of the effective localization length with the ratio of the error estimates.

In domain localized filters like the local SEIK filter, several observations within the influence radius around a local analysis domain are assimilated. For each observation a weight has to be computed. The matrix $\mathbf{H}_\delta \mathbf{P}^f \mathbf{H}_\delta^T$ needed to extract the term HPH^T for each observation is never explicitly computed in the analysis step of the SEIK filter. However, the matrix $\mathbf{H}_\delta \mathbf{L}$ is computed in Eq. (19), which is a square-root of the required matrix. To obtain a value for HPH^T there are two obvious possibilities: First, one can use the estimated variance from $\mathbf{H}_\delta \mathbf{L}$ that corresponds to a single observation. This will be an entry from the diagonal of $\mathbf{H}_\delta \mathbf{P}^f \mathbf{H}_\delta^T$. If the variance estimates vary strongly within the local domain, this method would result in a non-smooth weighting function w^{OLR} over all observations. This could lead to numerical instability of the data assimilation scheme. The second possibility is to use the mean variance of the covariance matrix $\mathbf{H}_\delta \mathbf{P}^f \mathbf{H}_\delta^T$. In both cases the diagonal of $\mathbf{H}_\delta \mathbf{P}^f \mathbf{H}_\delta^T$ can be computed directly from $\mathbf{H}_\delta \mathbf{L}$ without computing the full matrix.

The regulated OL method was exemplified here for the LSEIK filter. In general, it can be applied in all filter methods that apply OL, like the LETKF. The additional computational cost to compute the regulated localization from a fixed OL is generally negligible compared to the cost of the full analysis steps of the LSEIK filter and the LETKF.

5. Numerical experiments

To examine the performance of the regulated OL method, identical twin experiments are conducted using the Lorenz-96 model (Lorenz 1996; Lorenz and Emanuel 1998). This nonlinear model has been used in several studies to examine the behavior of different ensemble-based Kalman filters (e.g. Anderson 2001; Whitaker and Hamill 2002; Ott *et al.* 2004; Lawson and Hansen 2004; Sakov and Oke 2008). Our experiments use the same configuration as in Janjić *et al.* (2011) who found significant lower estimation errors for the EnSRF with CL compared to the LSEIK filter using a fixed OL for small observation errors.

5.1. Experimental setup

The Lorenz-96 model is prescribed by the non-dimensional equations

$$\frac{dx_j}{dt} = (x_{j+1} - x_{j-2})x_{j-1} - x_j + F \quad (25)$$

where $j = 1, \dots, J$ is the grid point index with cyclic boundary conditions. The common configuration with $J = 40$ and $F = 8$ is used. Time stepping is performed using a fourth-order Runge-Kutta scheme with a non-dimensional time step size of 0.05. The model as well as the filter algorithms have been implemented within the Parallel Data Assimilation Framework (PDAF, Nerger *et al.* 2005b).

For the twin experiments, a trajectory over 60000 time steps is computed from an initial state of constant value of 8.0 but $x_{20} = 8.008$, following Lorenz and Emanuel (1998). This trajectory represents the "truth". Observations of the full state are generated by disturbing the true trajectory by uncorrelated random normal noise. Three cases are examined in which the standard deviation σ_R of the observation error is 1, 0.5, and 0.1.

The initial ensemble for all experiments is generated by second-order exact sampling from the true trajectory* (see Appendix A). An ensemble of 10 members is used.

*Note, that the relative behavior of the localization methods does not depend on the choice of the method used to generate the ensemble. Tests

The assimilation is performed at each time step over 50000 time steps using observations starting from time step 1000 of the true trajectory. Because the results depend on the set of random numbers used in the ensemble generation, all experiments are repeated ten times with varying random numbers.

The experiments are performed using the LSEIK filter with regulated localization, referred to as LSEIK-reg. The estimates will be compared with estimates obtained by the EnKF-sqrt filter with CL (EnKF-sqrt) and the LSEIK filter with fixed OL (LSEIK-fix). The localization functions w^{CL} and w^{OL} are given by the compactly-supported GC function. The width of the GC function is defined by a support radius that describes the distance beyond which the function is zero.

For LSEIK-reg, two variants to compute HPH^T were described in section 4.2. Experiments were conducted using both the mean variance estimate as well as the single variable estimate at the observation location. The results obtained with both methods were not significantly different. Also, no stability problem, as discussed in section 4.2, was observed. Due to this, only results from LSEIK-reg using the mean variance estimate are discussed in the sequel.

5.2. Assimilation performance

Figure 3 shows the time-mean RMS errors averaged over each set of 10 experiments for a range of forgetting factors and support radii of the GC function. The corresponding minimum mean RMS errors and their standard deviations are shown in table I. The RMS errors are computed as the mean error over the 50000 analysis steps of each experiment and then averaged over 10 experiments. This computation includes the initial transient phase of the assimilation in which the errors are larger than during the later phase of the experiment. However, the relative behavior of the three compared methods is similar if the initial transient phase of about 2000 analysis steps is excluded. In particular, the

using an ensemble generated by random selection of states from the true trajectory, showed analogous behavior to that discussed in the text.

inferior performance of LSEIK-fix (see below) remains, as the later part of each experiment is influenced by the initial transient phase.

For the observation error of $\sigma_R = 1.0$ (top row of Fig. 3) the distributions of the errors obtained with the three localization methods are very similar. For LSEIK-reg, a particularly small mean error of 0.1988 is obtained for a forgetting factor of 0.95 and a support radius of 18 grid points. As this value is only obtained for a particular pair of these parameters, it will in practice be difficult to obtain this value. Obviously, the range of pairs of forgetting factor and support radius that provide errors close to the minimum should be as large as possible. Only in this case the chances are high that a good choice for these parameters can be obtained with a limited number of tests. Both, the EnKF-sqrt and the LSEIK-reg, show a comparable region of errors below 0.205. This region is smaller for LSEIK-fix, while the minimum error obtained with this method is statistically not different from that obtained with the EnKF-sqrt. The initial RMS error estimated by the ensemble standard deviation is about 2.5. Thus, the effective localization length of LSEIK-fix is for $\sigma_R = 1.0$ already wider than that of the EnKF-sqrt. Nonetheless, this difference appears to have only a small effect over the 50000 analysis step of each experiment.

The area of smallest errors extends from parameter-pairs with large forgetting factor but small support radius to pairs with small forgetting factor and large support radius. For very small support radii (below 8 grid points), the filter process is stable for all examined forgetting factors. However, the mean RMS errors are about twice as large as the minimum errors that can be obtained with larger support radii. It is striking that the smallest estimation errors occur close to the edge at which filter divergence happens. Directly at the edge, there are configurations at which a rather large mean RMS error is obtained. The reason for this behavior will be discussed below.

When the σ_R is reduced to 0.5, LSEIK-fix performs visibly worse than EnKF-sqrt and LSEIK-reg. This difference is statistically significant.

The difference of the three methods becomes even more obvious, when the observation error is reduced to $\sigma_R = 0.1$. The case LSEIK-fix results in significantly larger RMS errors than both EnSRF and LSEIK-reg. In addition, the parameter region where the filter process converges is much smaller for LSEIK-fix. Thus, the regulated OL can significantly improve the filter performance over that obtained with fixed OL. The error distributions obtained with EnKF-sqrt and LSEIK-reg are very similar up to a support radius of 28 grid points. Table I shows that the minimum mean RMS errors are almost identical. The divergence of the EnKF-sqrt method for support radii above 28 grid points can be attributed to a large condition number of the matrix $\mathbf{I} + \hat{\mathbf{P}}^f \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}$ that is decomposed to compute the square root in Eq. (13). For support radii above 28 grid points, the matrix exhibits very small singular values. These result in the dominance of single ensemble members and an effective rank-reduction of the ensemble.

The influence of the regulated OL is similar in less optimal cases. For example, if only each second grid point is observed, the regulated OL results in smaller mean RMS errors compared to fixed OL (not shown). The effect of smaller RMS errors and an increased stability region is also preserved when the assimilation interval is increased to 5 time steps (not shown).

Next to the minimum RMS error that can be obtained, it is important how likely it is to obtain it in a single experiment. As noted above, the assimilation result depends on the set of random numbers that is used to generate the initial ensemble. The mean RMS errors discussed above were obtained by performing ten experiments with different random numbers for each pair of forgetting factor and support radius. Figure 4 shows the corresponding base-10 logarithm of the standard deviations corresponding to the errors. For each filter method and observation error there exists a large region where the standard deviation is very small (e.g. below 10^{-3} for an $\sigma_R = 1.0$). In these regions, the RMS error varies only slightly with the random numbers. Toward the edge of filter divergence there is

generally a band of parameters where the standard deviation is high. In addition, a larger variance occurs if the forgetting factor is close to one. In these regions, the value of the estimation error from the experiment varies strongly for different sets of random numbers. In addition, the higher RMS errors are typically located in these regions. The strong variability is mostly caused by varying length of the initial transient phase. In the cases with large RMS errors, the assimilation process typically takes very long to reach the asymptotic phase. These cases are already close to the divergent cases, where the error in the state estimate remains similar to the error of the initial state estimate. In the cases with large standard deviations there is actually a chance to obtain a very small RMS estimation error for some choice of random numbers. However, there is no possibility to do this in a systematic way. Thus, the optimal choice of the pair of forgetting factor and support radius is in the region of small standard deviation where also the smallest mean RMS errors occur.

For decreasing observation error, there is a growing region at the edge to filter divergence, where the EnSRF and LSEIK-reg methods show stable behavior with a small standard deviation and small RMS error. Most striking is the behavior of LSEIK-reg and EnKF-sqrt for an observation error of 0.1. In this case, both methods shows no unstable filter processes in the region of convergence. The variance remains always below 10^{-3} . In contrast, LSEIK-fix shows unstable filter behavior for a support radius above 14. For LSEIK-reg and EnKF-sqrt, the stable behavior is combined with the strongly enlarged parameter region of convergence.

6. Conclusion

This study introduced a method to perform observation localization (OL), i.e. weighting of elements of the observation error covariance matrix by a localization function of variable width. The function is motivated by the different localizing effects of covariance localization (CL) and OL on the Kalman gain. The effective localization length scale for CL is identical to the prescribed localization

function. In contrast, for 'normal' (fixed) OL, the effective localization length of the Kalman gain grows if the observation error variance becomes smaller than the estimated error of the state. We refer to the new method as 'regulated OL' because it counters this effect by regulating the width of the localization function by the amplitudes of the estimated error variance of the state and of the observation error variance. The regulation is formulated to result in equal effective localization length scales in the Kalman gain for OL and CL in case of a single observation.

Identical twin experiments using the nonlinear Lorenz-96 model were conducted to compare the effect of regulated OL with fixed OL and CL. For CL, a simple square-root formulation was applied while for both methods of OL the LSEIK filter (Nerger *et al.* 2006) was used. The covariance inflation was varied in the experiments as was the support radius of the localization function. For cases when the observation error is comparable to the estimated error of the state, the differences between the three methods were negligible. However, regulated OL outperforms fixed OL, if the observation error is considerably smaller than the estimated error. The results from the LSEIK filter with regulated OL are almost identical to those obtained with the square-root filter with CL. When the regulated OL and CL outperform the fixed OL, smaller estimation errors are obtained. In addition, the range of parameters that result in the smallest estimation errors is increased compared to fixed OL. As these parameters are problem specific and are typically determined using experiments in order to obtain satisfactory assimilation results, it will be easier to find good parameter choices with regulated OL and CL than with CL.

The advantage of the regulated OL becomes visible when the observation error is smaller than the error estimate of the state. This situation is not untypical during the initial transient phase of an assimilation process but can also happen during the later stages of the assimilation procedure when the ensemble forecast by a nonlinear model can result in locally increased error estimates.

This study used the LSEIK filter and the simple EnKF square-root formulation to exemplify the localization methods. The findings regarding the different effective localization length scales of OL and CL are, however, independent of the particular filter method. Accordingly, the regulated OL can be applied with all filter methods that apply OL and the effect of the regulation of the localization length scale should be the same for all these filters.

The regulated OL becomes identical to the fixed OL in the case that the observation errors are much larger than the estimated state error. For smaller observation errors, the regulated OL avoids the widening of the effective localization length that can deteriorate the assimilation performance. Further, the additional computational cost of regulated OL is negligible compared to the total cost of the analysis step. For these reasons, the use of the regulated localization method should be always recommended when filter algorithms like LSEIK or LETKF are applied with OL. In this work, the effect of the regulated OL was studied only in the simplified test case of the Lorenz-96 model. Its impact in realistic assimilation applications will be examined in the future.

Appendix

A. Ensemble generation by second-order exact sampling

To initialize the filter algorithm, an ensemble of state realizations is required that represents the initial state estimate \mathbf{x}_0^a and the initial covariance matrix \mathbf{P}_0^a .

Ensemble-based filters do not base on a particular scheme to initialize the ensemble (Nerger *et al.* 2005a). In the numerical experiments discussed in section 5 the second-order exact sampling method (Pham 2001) was applied. This sampling method ensures an exact representation of a covariance matrix of given rank with an ensemble of minimum size.

Consider \mathbf{P}_0^a to be a rank- r matrix. It can be written as $\mathbf{P}_0^a = \mathbf{V}_0 \mathbf{U}_0 \mathbf{V}_0^T$ where \mathbf{U}_0 and \mathbf{V}_0 are matrices of size

$r \times r$ and $n \times r$, respectively. This decomposition can be obtained by singular value decomposition such that \mathbf{U}_0 is a diagonal matrix holding singular values. Now a random ensemble of minimum size $N = r + 1$ is generated whose statistics represent \mathbf{x}_0^a and \mathbf{P}_0^a exactly. This can be achieved by transforming the columns of matrix \mathbf{V}_0 by a $N \times r$ random matrix $\mathbf{\Omega}_0$ generated from uniformly distributed random numbers. The columns of $\mathbf{\Omega}_0$ are constrained to be orthonormal and orthogonal to the vector $(1, \dots, 1)^T$. Using $\mathbf{\Omega}_0$, the ensemble of state realizations is given by

$$\mathbf{X}_0^a = \overline{\mathbf{X}_0^a} + \sqrt{N-1} \mathbf{V}_0 (\mathbf{U}_0)^{1/2} \mathbf{\Omega}_0^T. \quad (26)$$

B. Regulated localization with multiple observations

Greybush *et al.* (2011) showed that the effect of OL in the case of two observations is different from the situation when only a single observation is considered. In this Appendix, it is examined how the regulated OL is influenced by the presence of multiple observations.

Following Greybush *et al.* (2011) we consider two grid points, indexed 1 and 2. We assume that the model variables at both locations are observed. Thus \mathbf{H} is the identity. In this case, the Kalman gain defined by Eq. (11) can be written for a diagonal matrix \mathbf{R} as

$$\mathbf{K} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \left(\begin{bmatrix} P_{11} + R_1 & P_{12} \\ P_{21} & P_{22} + R_2 \end{bmatrix} \right)^{-1} \quad (27)$$

where we dropped the index f of the forecast error covariance matrix.

For the localization, we consider the first grid point, i.e. the first row of the gain. Let α denote the localization function for CL. To obtain the CL, the off-diagonal elements of \mathbf{P}^f are multiplied by α . Thus it is

$$\mathbf{K}_1^{CL} = \begin{bmatrix} P_{11} & \alpha P_{12} \\ P_{21} & P_{22} + R_2 \end{bmatrix} \left(\begin{bmatrix} P_{11} + R_1 & \alpha P_{12} \\ \alpha P_{21} & P_{22} + R_2 \end{bmatrix} \right)^{-1} \quad (28)$$

For OL, the observation variances are multiplied by the localization function. For R_1 the weight is one, because the distance is zero. For R_2 , let β denote the localization function for OL. In this case, it is:

$$\mathbf{K}_1^{OL} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} + \beta R_2 \end{bmatrix} \left(\begin{bmatrix} P_{11} + R_1 & P_{12} \\ P_{21} & P_{22} + \beta R_2 \end{bmatrix} \right)^{-1} \quad (29)$$

The regulated localization function is derived from requiring that the elements \mathbf{K}_{12}^{CL} and \mathbf{K}_{12}^{OL} are equal. These entries specify the effect of the observation at the second grid point on the analysis update at the first grid point. After same algebra one obtains:

$$\beta^{OLR} = \frac{P_{12}R_1 - (P_{11}P_{22} + P_{22}R_1 - P_{12}P_{21})K_{12}^{CL}}{(P_{11}R_2 + R_1R_2)K_{12}^{CL}} \quad (30)$$

The regulated localization function is controlled by the variance estimates of both observations as well as all elements of the state error covariance matrix.

The effective localization function, can be visualized by plotting the elements \mathbf{K}_{12}^{CL} and \mathbf{K}_{12}^{OL} of the gains as a function of the distance of the observation from the first grid point. An example is shown in Fig. 5 for an observation variance of 0.1, forecast error variances of 1 and 0.5, and a covariance of 0.25. As in the single-observation examples discussed before, the effective localization length for OL is wider than that for CL. The effective localization function obtained using the regulated OL defined by Eq. (24) is much closer to the function for CL. However, it shows a slightly shorter effective length scale. The regulated OL for the case of two observations (Eq. 30) results in an effective localization function that is identical to CL.

This example illustrates that the regulated OL function derived for a single observation results in a similar effective localization for cases when multiple observations are assimilated. It is not exact in the case of multiple observations, but still serves to avoid the growth of the effective localization length scale if the observations are very accurate.

References

- Anderson JL. 2001. An Ensemble Adjustment Kalman Filter for data assimilation. *Mon. Wea. Rev.* **129**: 2884–2903.
- Anderson JL. 2007. Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. *Physica D* **230**: 99–111.
- Bishop CH, Etherton BJ, Majumdar SJ. 2001. Adaptive sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical aspects. *Mon. Wea. Rev.* **129**: 420–436.
- Bishop CH, Hodyss D. 2009. Ensemble covariances adaptively localized with ECO-RAP. Part 1: Tests on simple error models. *Tellus* **61A**: 84–96.
- Brusdal K, Brankart JM, Halberstadt G, Evensen G, Brasseur P, van Leeuwen PJ, Dombrowsky E, Verron J. 2003. A demonstration of ensemble based assimilation methods with a layered OGCM from the perspective of operational ocean forecasting systems. *J. Mar. Syst.* **40–41**: 253–289.
- Burgers G, van Leeuwen PJ, Evensen G. 1998. On the analysis scheme in the Ensemble Kalman Filter. *Mon. Wea. Rev.* **126**: 1719–1724.
- Campbell WF, Bishop CH, Hodyss D. 2010. Vertical covariance localization for satellite radiances in ensemble Kalman filters. *Mon. Wea. Rev.* **138**: 282–290.
- Cohn SE, Silva AD, Guo J, Sienkiewicz M, Lamich D. 1998. Assessing the effects of data selection with the DAO physical-space statistical analysis system. *Mon. Wea. Rev.* **126**: 2913–2926.
- Evensen G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99**(C5): 10 143–10 162.
- Gaspari G, Cohn SE. 1999. Construction of correlation functions in two and three dimensions. *Q. J. Roy. Meteor. Soc.* **125**: 723–757.
- Greybush SJ, Kalnay E, Miyoshi T, Ide K, Hunt BR. 2011. Balance and ensemble Kalman filter localization techniques. *Mon. Wea. Rev.* **139**: 511–522.
- Haugen VE, Evensen G. 2002. Assimilation of SLA and SST data into an OGCM for the Indian Ocean. *Ocean Dynamics* **52**: 133–151.
- Houtekamer PL, Mitchell HL. 1998. Data assimilation using an Ensemble Kalman Filter technique. *Mon. Wea. Rev.* **126**: 796–811.
- Houtekamer PL, Mitchell HL. 2001. A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.* **129**: 123–137.
- Hunt BR, Kostelich EJ, Szunyogh I. 2007. Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D* **230**: 112–126.
- Janjić T, Nerger L, Albertella A, Schröter J, Skachko S. 2011. On domain localization in ensemble based Kalman filter algorithms. *Mon. Wea. Rev.* **139**: 2046–2060.
- Lawson WG, Hansen JA. 2004. Implications of stochastic and deterministic filters as ensemble-based data assimilation methods in varying regimes of error growth. *Mon. Wea. Rev.* **132**: 1966–1981.
- Lorenz EN. 1996. Predictability - a problem partly solved. In: *Proceedings Seminar on Predictability*. ECMWF, Reading, UK, pp. 1–18.
- Lorenz EN, Emanuel KA. 1998. Optimal sites for supplementary weather observations: Simulation with a small model. *J. Atm. Sci.* **55**: 399–414.
- Miyoshi T, Yamane S. 2007. Local ensemble transform Kalman filter with an AGCM at a T159/L48 resolution. *Mon. Wea. Rev.* **135**: 3841–3861.
- Miyoshi T, Yamane S, Enomoto T. 2007. Localizing the error covariance by physical distances within a Local Ensemble Transform Kalman Filter (LETKF). *SOLA* **3**: 89–92.
- Nerger L, Danilov S, Hiller W, Schröter J. 2006. Using sea level data to constrain a finite-element primitive-equation ocean model with a local SEIK filter. *Ocean Dynamics* **56**: 634–649.
- Nerger L, Gregg WW. 2007. Assimilation of SeaWiFS data into a global ocean-biogeochemical model using a local SEIK filter. *J. Mar. Syst.* **68**: 237–254.
- Nerger L, Gregg WW. 2008. Improving assimilation of SeaWiFS data by the application of bias correction with a local SEIK filter. *J. Mar. Syst.* **73**: 87–102.
- Nerger L, Hiller W, Schröter J. 2005a. A comparison of error subspace Kalman filters. *Tellus* **57A**: 715–735.
- Nerger L, Hiller W, Schröter J. 2005b. PDAF - the Parallel Data Assimilation Framework: Experiences with Kalman filtering. In: *Use of High Performance Computing in Meteorology - Proceedings of the 11. ECMWF Workshop*, Zwiefelhofer W, Mozdzyński G (Eds). World Scientific, pp. 63–83.
- Ott E, Hunt B, Szunyogh I, Zimin AV, Kostelich EJ, Corazza M, Kalnay E, Patil DJ, Yorke JA. 2004. A local ensemble Kalman filter for atmospheric data assimilation. *Tellus* **56A**: 415–428.
- Pham DT. 2001. Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Mon. Wea. Rev.* **129**: 1194–1207.
- Pham DT, Verron J, Gourdeau L. 1998. Singular evolutive Kalman filters for data assimilation in oceanography. *C. R. Acad. Sci., Ser. II* **326**(4): 255–260.
- Sakov P, Bertino L. 2011. Relation between two common localisation methods for the EnKF. *Comput. Geosci.* **15**: 225–237.

- Sakov P, Oke PR. 2008. Implications of the form of the ensemble transformation in the ensemble square root filters. *Mon. Wea. Rev.* **136**: 1042–1053.
- Tippett MK, Anderson JL, Bishop CH, Hamill TM, Whitaker JS. 2003. Ensemble square root filters. *Mon. Wea. Rev.* **131**: 1485–1490.
- Whitaker JS, Hamill TM. 2002. Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.* **130**: 1913–1927.
- Whitaker JS, Hamill TM, Wei X, Song Y, Toth Z. 2008. Ensemble data assimilation with the NCEP global forecast system. *Mon. Wea. Rev.* **136**: 463–482.

Table I. Minimum mean RMS errors and their standard deviation over each 10 experiments using different random numbers in the initialization for three different filter configurations and three different observation errors σ_R

σ_R	EnKF-sqrt	LSEIK-fix	LSEIK-reg
1.0	0.2006 ± 0.0010	0.2025 ± 0.0021	0.1988 ± 0.0007
0.5	0.0963 ± 0.0003	0.0992 ± 0.0005	0.0951 ± 0.0005
0.1	0.0187 ± 0.0001	0.0205 ± 0.0002	0.0185 ± 0.0001

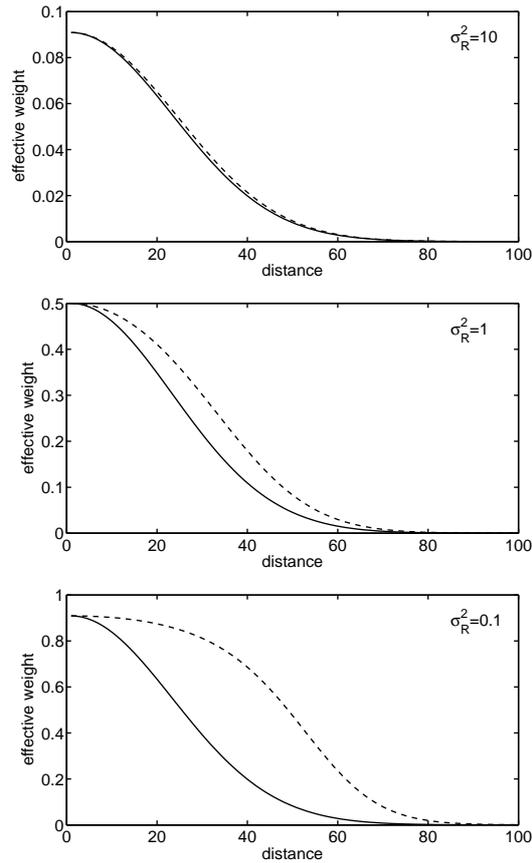


Figure 1. Effective weighting in the Kalman gain for different observation error variances σ_R^2 and state error variance 1. (Solid): Γ th element of the Kalman gain for CL (Eq. 22). (Dashed): Γ th element of the gain for OL (Eq. 23). The effective weighting is increasingly wider for observation localization for decreasing σ_R^2 .

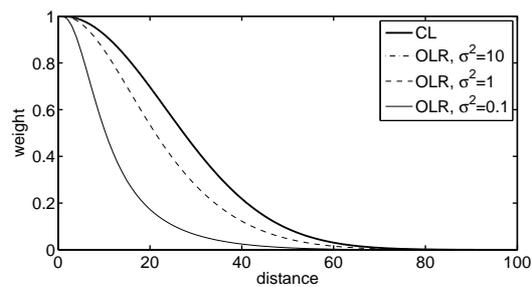


Figure 2. Gaussian weight function w^{CL} and regulated weight function w^{OLR} for three different observation error variances σ_R^2 . The curves for w^{CL} and w^{OLR} with $\sigma_R^2 = 10$ lie on top of each other.

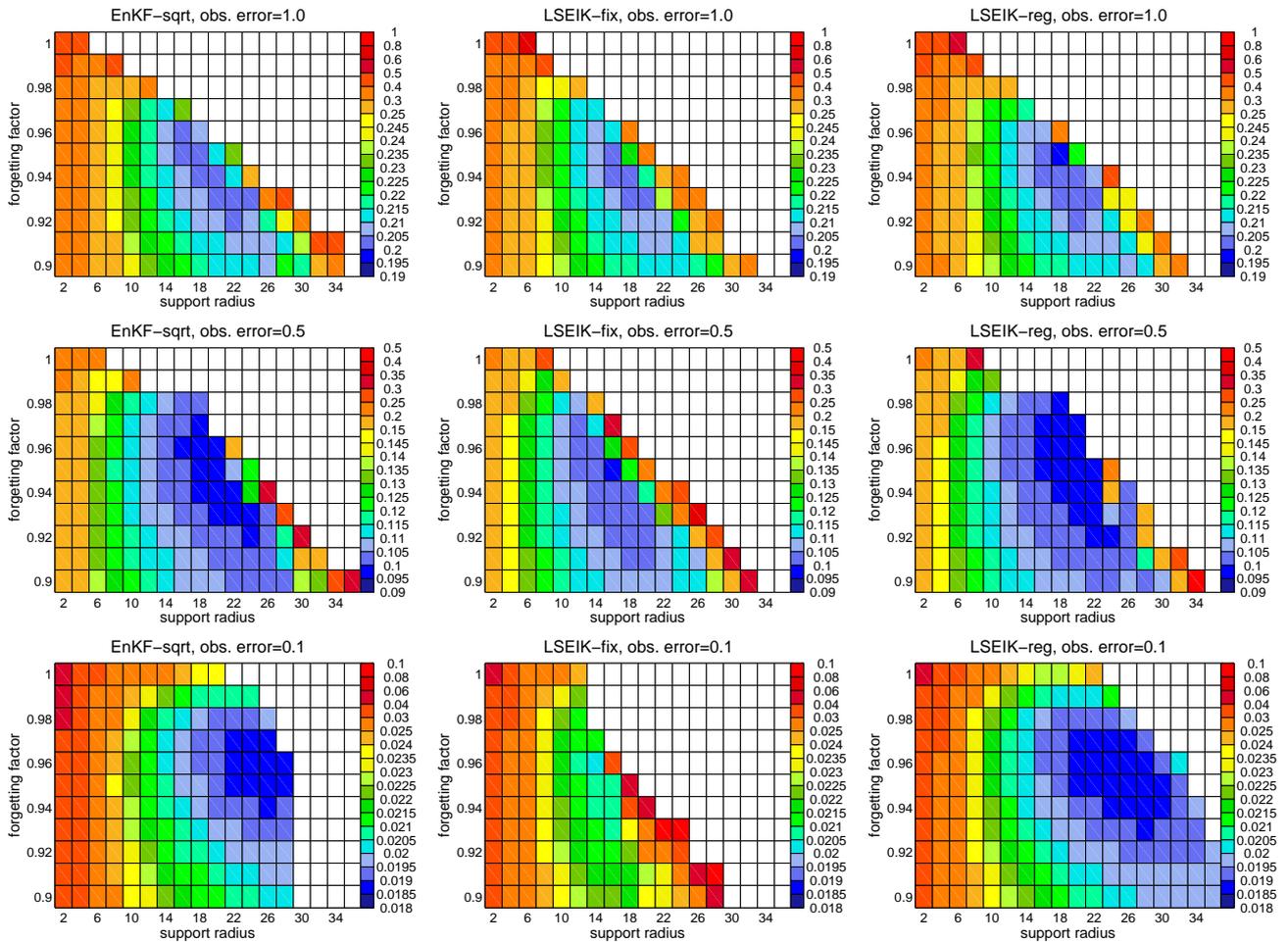


Figure 3. Mean RMS errors averaged over the full length of the experiment for the EnKF-sqrt (left), LSEIK-fix (center), and LSEIK-reg (right) for three different observational errors: 1.0 (top), 0.5 (middle), 0.1 (bottom). White fields denote filter divergence, which is defined here as the case that the RSM mean error is larger than the observational error.

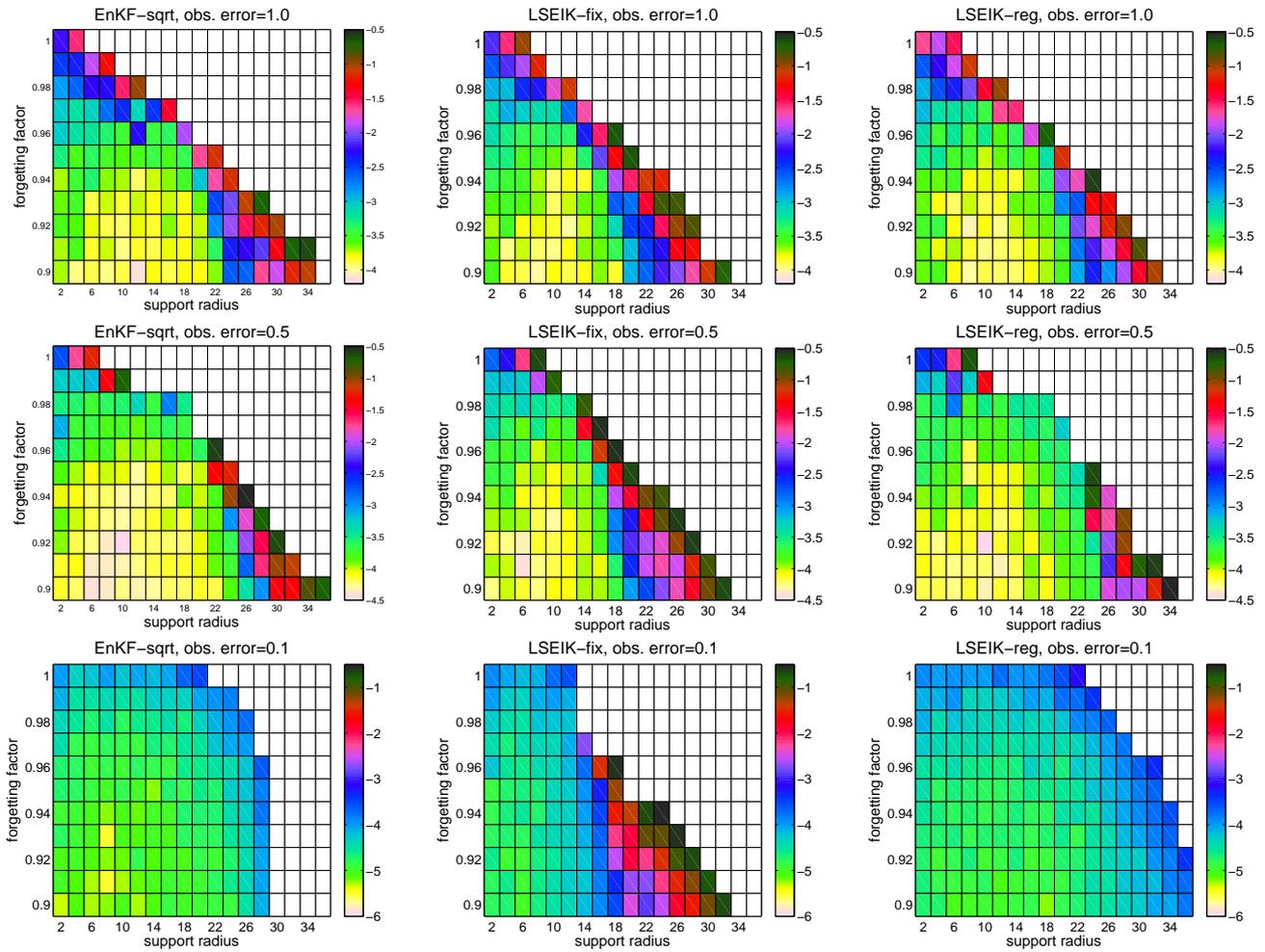


Figure 4. Logarithm of the standard deviation for the RMS errors over 10 different realizations of random numbers in the ensemble initialization for the EnSRF (left), LSEIK-fix (center), and LSEIK-reg (right) for three different observational errors: 1.0 (top), 0.5 (middle), 0.1 (bottom). As in Fig. 3, white fields denote cases in which the filter diverges.

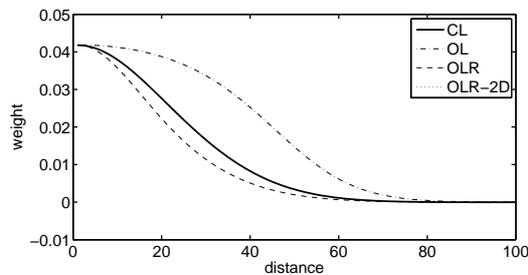


Figure 5. Effective weighting in the Kalman gain for the case of 2 observations as discussed in Appendix B. Shown is the effect of the observation at the second grid point on the analysis update at the first grid point for CL (thick solid line), fixed OL (dash-dotted) and regulated OL according to Eq. (24) (dashed). The function for 2D-regulated OL according to Eq. (30) lies on top of the function for CL.