The DOI for this manuscript is doi: 10.1175/MWR-D-13-00246.1

The final published version of this manuscript will replace the preliminary version at the above DOI once it is available.

If you would like to cite this EOR in a separate work, please use the following full citation:

Kirchgessner, P., L. Nerger, and A. Bunse-Gerstner, 2014: On the choice of an optimal localization radius in ensemble Kalman filter methods. Mon. Wea. Rev. doi:10.1175/MWR-D-13-00246.1, in press.

On the choice of an optimal localization radius in ensemble Kalman filter methods

Paul Kirchgessner, * and Lars Nerger

Alfred Wegener Institute, Helmholtz Center for Polar and Marine Research, Bremerhaven, Germany

Angelika Bunse-Gerstner

University of Bremen, Bremen, Germany

* *Corresponding author address:* Paul Kirchgessner, Alfred Wegener Institute Helmholtz Center for Polar and Marine Research, Am Handelshafen 12, D-27570 Bremerhaven

E-mail: paul.kirchgessner@awi.de

ABSTRACT

In data assimilation applications using ensemble Kalman filter methods, localization is necessary to make the method work with high-dimensional geophysical models. For ensemble square-root Kalman filters, domain localization (DL) and observation localization (OL) are commonly used. Depending on the localization method, one has to choose appropriate values for the localization parameters, such as the localization length and the weight function. Although frequently used, the properties of the localization techniques are not fully investigated. Thus, up to now an optimal choice for these parameters is a priori unknown and they are generally found by expensive numerical experiments. In this study, the relationship between the localization length and the ensemble size in DL and OL is studied using twin experiments with the Lorenz-96 model and a 2-dimensional shallow water model. For both models, it is found that the optimal localization length for DL and OL depends linearly on an effective local observation dimension that is given by the sum of the observation weights. In the experiments no influence of the model dynamics on the optimal localization length was observed. The effective observation dimension defines the degrees of freedom that are required for assimilating observations, while the ensemble size defines the available degrees of freedom. Setting the localization radius such that the effective local observation dimension equals the ensemble size yields an adaptive localization radius. Its performance is tested using a global ocean model. The experiments show that the analysis quality using the adaptive localization is similar to the analysis quality of an optimally tuned constant localization radius.

1. Introduction

In ocean modeling and weather forecasting an estimate of the current state is important to initialize forecasts of the dynamical process. In sequential data assimilation, variants of the Ensemble Kalman Filter (EnKF, Evensen 1994) are commonly used. To deal with the particular problems of the geophysical systems many improvements of the methods, e.g. covariance inflation and localization (Houtekamer and Mitchell 1998), have been introduced. Typically, the state dimension of the models is very high, but only a small ensemble is feasible to use. This introduces noise and spurious correlations in the covariance matrices and limits the degrees of freedom for the analysis, which are defined by the ensemble size. Localization is used to access the problem of spurious correlations, and increases the degrees of freedom by calculating a local analysis in every grid point. This approach is justified by the fact that dynamical systems can locally behave like a low dimensional systems (see Patil et al. 2001). The positive effect of localization for ensemble Kalman filters has recently been described for different applications in oceanography and meterology (e.g. Nerger et al. 2006; Janjić et al. 2011; Otkin 2012; Losa et al. 2012; Kang et al. 2012)).

Localization can be applied to the covariance matrices by point-wise multiplication (Houtekamer and Mitchell 2001), referred to as covariance localization (CL). Alternatively, the domain is decomposed as in domain localization (DL) and separate analysis for each subdomain are calculated (Houtekamer and Mitchell 1998). The latter method can be combined with observation localization (OL), where the observations are weighted according to their distance, as described in Hunt et al. (2007). Several studies (Miyoshi and Yamane 2007; Greybush et al. 2011; Sakov and Bertino 2011; Nerger et al. 2012) investigated the relationship between CL and OL and found that the results were comparable, even though the effective localization length is shorter for OL than for CL. The relation between different weight functions and localization radii was examined in Whitaker and Hamill (2002). They found that using a weight function similar to the Gaussian curve (see Gaspari and Cohn 1999, Eq. 4.10) produces better results than using a Heaviside step function. For a regional ocean model, the effect of different localization radii in DL was examined in Nerger et al. (2006). Yoon et al. (2010) have shown that localization improves the estimation of the covariances. According to their findings the localization radius should be chosen large enough to get most of the relevant covariances. For all of these localization methods, extensive tuning of the localization parameters is necessary to achieve the optimal results.

Recently, adaptive localization methods (Anderson 2007, 2012; Bishop and Hodyss 2007, 2009) have been developed to estimate the correlations between different variables flow-dependently. Further, information-based localization schemes have been developed (Zupanski et al. 2007; Migliorini 2013). As shown for different examples, these methods improve the assimilation results, but they still require the choice of different parameters or are computational very expensive.

Here, an alternative approach to define the localization radius is investigated. From experiments using two small models, a relationship between the ensemble size and the optimal localization radius is derived in the context of dense observations with uniform error statistics. Examples of these kind of observations are gridded satellite observations of sea surface temperature or sea surface elevation, which are frequently used in ocean data assimilation applications (see e.g. Janjić et al. 2012; Losa et al. 2012; Sakov et al. 2012). The relation is then used to define an adaptive localization method and tested using a global ocean model.

The article is structured as follows. In Section 2 the assimilation algorithm and the localization techniques are discussed. Afterwards, the models are introduced and the numerical experiments are described in Section 3. In Section 4, the results for the Lorenz-96 model are presented. The experiments using the Shallow-Water-Equations are discussed in relation to the Lorenz-96 model in Section 5. In Section 6 assimilation results using a global ocean model are discussed and conclusions are drawn in Section 7.

2. Assimilation algorithm

The assimilation experiments in this study are performed with the widely used Ensemble Transform Kalman Filter (ETKF, Bishop et al. 2001) with localization (Hunt et al. 2007). In this section, the ETKF and the localization techniques are reviewed.

a. ETKF

Data assimilation methods provide an estimate of the state of a system $\mathbf{x}_k \in \mathbb{R}^n$ at time k given the model dynamics

$$\mathbf{x}_{k+1} = \mathbf{M}(\mathbf{x}_k) + \epsilon_{\mathbf{k}} \tag{1}$$

and a set of observations $\mathbf{y}_k^o \in \mathbb{R}^p$. These are related to the model state via the observation operator $\mathbf{H}$

$$\mathbf{y}_k^o = \mathbf{H}(\mathbf{x}_k) + \eta_{\mathbf{k}}. \tag{2}$$

The errors $\epsilon \in \mathbb{R}^{\mathbf{n}}$ and $\eta \in \mathbb{R}^{\mathbf{p}}$ are assumed to be Gaussian with zero mean and covariance matrices $\mathbf{Q} \in \mathbb{R}^{n \times n}$ and $\mathbf{R} \in \mathbb{R}^{p \times p}$ respectively . Below, the time index k is omitted.

The background state $\mathbf{x}^f$ and the covariance matrix $\mathbf{P}^f$ are now represented by an ensemble of state realisations $\mathbf{x}^{f(i)}, i \in \{1, \ldots, N\}$. The matrix $\mathbf{X}^f$ denotes the matrix whose column vectors are the ensemble members, and $\mathbf{X}^{f'}$ is the matrix of ensemble perturbations. The state estimate is given by the mean of the ensemble $\bar{\mathbf{x}}$.

The idea of the ETKF is to carry out the analysis in the ensemble space and then map the corrections into the state space via the ensemble perturbations. Here, only the equations for the ETKF are given. For a detailed derivation of the filter equations see Hunt et al. (2007).

At an analysis time, an analysis weight vector $\bar{\mathbf{w}}^a$ and an analysis covariance matrix $\tilde{\mathbf{P}}^a$

are calculated in the space spanned by the ensemble perturbations:

$$\tilde{\mathbf{P}}^a = [(N-1)\mathbf{I}\rho + (\mathbf{H}\mathbf{X}^{f'})^T\mathbf{R}^{-1}\mathbf{H}\mathbf{X}^{f'}]^{-1} \tag{3}$$

$$\bar{\mathbf{w}}^a = \tilde{\mathbf{P}}^a(\mathbf{H}\mathbf{X}^{f'})^T\mathbf{R}^{-1}(\mathbf{y}^o - \mathbf{H}\bar{\mathbf{x}}^f) \tag{4}$$

The factor $\rho \geq 1$ is used to inflate the ensemble (see Hunt et al. 2007).

The forecast ensemble is then

$$\mathbf{X}^a = \bar{\mathbf{x}}^f\mathbf{1}^T + \mathbf{X}^{f'}(\bar{\mathbf{w}}^a\mathbf{1}^T + [(N-1)\tilde{\mathbf{P}}^a]^{1/2}). \tag{5}$$

During the forecast phase, the ensemble members are all moved forward in time using the full nonlinear model

$$\mathbf{x}^{f(i)} = \boldsymbol{M}(\mathbf{x}^{a(i)}) \tag{6}$$

for all $i = 1 \ldots N$.

b. Localization in ETKF

For a local analysis with the ETKF, the domain is decomposed into different local regions (Houtekamer and Mitchell 1998), e.g. every single grid point. An analysis increment is then calculated separately for every local domain. For the local analysis domains a support radius l for the observations is defined. Only observations closer than l from the analysis point will have a non-zero weight and thus influence on the local analysis. According to Hunt et al. (2007), the observations used for two neighbouring analysis regions should overlap significantly to ensure that the weights are similar and a smooth analysis is produced. Except for very small localization radii, this was ensured in the experiments.

The observations inside each observation region are weighted according to their distance to the analysis point. These weights are applied by Schur-multiplying the inverse of the observation covariance matrix $\mathbf{R}$ by a matrix constructed from a correlation function (see Hunt et al. 2007).

We examine the effect of two localization techniques, domain localization (DL) and observation localization (OL) that are characterised by their weighting functions. DL was formulated without explicit weights to the observations (see e.g. Houtekamer and Mitchell 1998; Nerger et al. 2006), but implicitly the weights

$$w_{DL}(z,l) := \begin{cases} 1 \text{ if } |z| \leq l \\ 0 \text{ else} \end{cases}$$

are used. Here, l is a predefined cut-off radius. This weighting corresponds to a unit weight inside an observation domain and zero outside.

For OL, a fifth-order polynomial (Gaspari and Cohn 1999, eq. 4.10) is used for weighting the observations. This function is very popular because its shape is similar to the probability density function of a normal distribution but has compact support. The equations can be written as

$$w_{OL}(z,l) := \begin{cases} f_1(z/2l) & \text{if } 0 \leq |z| \leq l/2 \\ f_2(z/2l) & \text{if } l/2 \leq |z| \leq l \\ 0 & \text{if } |z| \geq l \end{cases}$$

with

$$\begin{aligned} f_1(c) &= -\frac{c^5}{4} + \frac{c^4}{2} + \frac{5c^3}{8} - \frac{5c^2}{3} + 1 \\ f_2(c) &= \frac{c^5}{12} - \frac{c^4}{2} + \frac{5c^3}{8} + \frac{5c^2}{3} - 5c + 4 - \frac{2}{3c}. \end{aligned}$$

OL is the current standard scheme for localization in the LETKF (e.g. Miyoshi and Yamane 2007). DL is an older formulation (see e.g. Houtekamer and Mitchell 1998; Nerger et al. 2006) and nowadays it is unusual to use DL, because OL yields better assimilation performance. However, the constant observation weights allow to investigate the influence of localization without considering the effects of varying weights. If the results from DL are then compared to the variable weight functions of OL, the basic properties of localization become clearer.

3. Configuration of Numerical Experiments

The numerical experiments are performend with the Lorenz-96 model (Lorenz 1995) and a shallow-water model. Although being rather simple, both models exhibit strong nonlinear behaviour. For the Lorenz-96 model this was described in (Lorenz 1995). The shallow water model configuration used here can develop strongly nonlinear dynamics in the form of a meandering zonal jet and associated eddies (see Krysta et al. 2011). Since the dynamics of the models are distinct, the comparison of the results from both models provides insight to which extent the localization behaviour is independent of the model.

a. Experiments with the Lorenz-96 model

The characteristics of the localization techniques are first investigated with twin experiments using the 40-dimensional Lorenz-96 model (Lorenz 1995). For the twin experiments, the initial condition $X \in \mathbb{R}^{40}$ with $X_{20} = 8.008$ and $X_j = 8$ for all $i \neq 20$ is first integrated for 1000 time steps by using the classical forth-order Runge-Kutta scheme with a time step of 0.05. By integrating the model for another 5000 time steps, a trajectory is obtained that represents the truth. The observations are generated by adding Gaussian distributed random numbers with unit variance and zero mean to the truth. All grid points are observed. The observation error covariance matrix $\mathbf{R}$ is chosen to be diagonal with the variance of the observation error on the diagonal. A constant inflation factor of $\rho = 1.05$ is used to inflate the background covariance matrix.

The initial ensemble is generated by second-order exact sampling (Pham 2001) from a model run over 10000 time steps. The ensemble size is varied between 5 and 28. Localization radii between 0 and 20 are used for the experiments with DL, while for OL localization radii from 0 to 50 are used. All experiments are repeated ten times with different random numbers for the ensemble initialisation and observations. The ETKF as implemented in the Parallel Data Assimilation framework (PDAF, Nerger and Hiller 2013, http://pdaf.awi.de) is used for the experiments.

For evaluating the assimilation performance, the root mean squared error, averaged over the assimilation times and the repetitions is used. This quantity will be denoted as MRMSE.

b. Experiments with the Shallow water model

A 2D model using the shallow water equations (see Krysta et al. 2011) is used to asses the localization in case of a multivariate model. A detailed review of the model is given in Appendix A. The model is calculated on a regular square grid with $25km$ resolution. At each grid point, the sea surface height (h), the horizontal (u) and the vertical velocities (v) are defined. The state vector has 19380 elements, of which only the sea surface height is observed in the experiments. Both, fully observed h and partial observations of h are considered in the experiments. For the partial observations, every second and every third point in both directions is observed.

The experiment is initialised by integrating the initial state $h = 500$m and $u = v = 0$m s^{-1} for 15 years. The first 5 years are used to spin up the model state. A sample of every second day from year 6 to 15 is used to initialise the ensemble through second-order-exact sampling. Synthetic observations are generated from the sea surface height with zero mean and constant variance of 2m^2. The observation errors are assumed to be uncorrelated and are assimilated once a day.

A local analysis is calculated for every single grid point. The influence region for the observations is a circle of radius l around the analysis location. The weighting is applied according to the Euclidean distance. For the experiments, localization radii between 20km and 350km with a step size of 10km and ensemble sizes from 5 to 40 are used.

The inflation factor is set to $\rho = 1.08$. It is tuned so that the estimated and true errors are in the same order of magnitude for several converged configurations. Thus, is not tuned to achieve the minimal error, but such that the following results do not depend on the choice of the inflation factor. For the experiments, the same configuration of PDAF as in section

3a was used.

To compare the analysis quality of the different experiments, the root mean squared error (RMSE) of the height field h is examined.

4. Localization behaviour with the Lorenz-96 model

a. Optimal localization radius for DL

Figure 1 shows in the top row the MRMSE for all considered parameter values N and l for DL. The parameter region can be clearly divided into diverged and converged results. An experiment is defined as divergent, if the MRMSE of an experiment is larger than the observation error. For every ensemble of less than 21 members, filter divergence occurs when a certain localization radius is exceeded (e.g. $l = 4$ for $N = 5$). In the following, this radius is denoted by l_{div}.

For a constant localization radius, increasing the ensemble size reduces the MRMSE. However, after the most information content from the observations is extracted, very little error reduction is gained (e.g. for $N > 14$ for $l = 7$).

If the ensemble size is kept constant and the localization radius is increased, the error shrinks until an optimal localization radius, denoted by l_{opt}, is reached. Increasing l beyond this radius deteriorates the assimilation results and filter divergence can occur.

In the top panel of figure 2, l_{opt} and l_{div} as functions of the ensemble size N are shown for DL. The optimal value for l_{opt} is always close to $N/2$. Filter divergence occurs approximately if the localization radius, measured in grid points, exceeds the number of ensemble members. As long as in a local analysis not all observations are used, l_{opt} and l_{div} depend linearly on the ensemble size. For DL, the behaviour changes if the ensemble size is big enough so that the filter converges without localization. In this case, filter divergence doesn't occur anymore and the global filter produces the best results.

b. Optimal ensemble size for OL

For OL, the MRMSE for various localization radii and ensemble sizes is also divided into regions where the filter diverges or converges (Fig. 1, bottom). The assimilation converges as long as l is only slightly bigger than $2N$. Compared to DL, the convergence region in case of OL is enlarged approximately by a factor of two. A similar relationship holds for the optimal localization radius. Since more observations are assimilated, the best assimilation results for OL are more accurate than the ones for DL, even with less ensemble members. As expected, the observation weighting of OL results in a smaller error with a minimum MRMSE=0.1883 compared to MRMSE=0.1901 in case of DL.

The lower panel of Fig. 2 shows that the relationship between the optimal localization radius l_{opt} and the ensemble size N is also linear. However, with OL longer localization radii can be used than with DL. The behaviour of the optimal localization radius for $N > 20$ is not representative for OL. The reason is that l_{opt} is bounded by the largest tested localization radius. Thus, for $N > 20$ l_{opt} is likely to be larger than the radii tested here.

c. Sampling quality of the covariance matrix

The localization implicitly modifies the state covariance matrix. Here, it is examined how well the true covariance matrix is approximated with localization. The results are shown for a single ensemble size ($N = 16$), but also hold for other choices.

The true covariance matrix $\mathbf{P}^t$ is generated from a twin experiment using an Ensemble Kalman Filter with an ensemble size of 128. Since the ensemble is significantly larger than the state dimension, this covariance matrix should be close to the truth.

At the end of the assimilation experiment, the normalised difference between the true covariance matrix and the analysis ensemble covariance matrix

$$\delta(\mathbf{P}^{a_l}) := \frac{\left\|\mathbf{P}^{a_l} - \mathbf{P}^t\right\|_F}{\left\|\mathbf{P}^t\right\|_F} \tag{7}$$

is compared in the Frobenius norm $\| \ \|_F$. Here, the matrix $\mathbf{P}^{a_l}$ denotes the ensemble covariance matrix calculated from an assimilation experiment with the localization radius l using the LETKF with OL .

In the local filter, not all elements of the covariance matrix are used. To take this into account, we define the matrix $\mathbf{P}_l$ as the matrix $\mathbf{P}$ with all elements $(p)_{ij}$ set to zero that correspond to long distances beyond the localization radius i.e.

$$(p)_{ij} = \begin{cases} p_{ij} \text{ if } \|x_i - x_j\| \leq l \\ 0 \text{ else.} \end{cases} \tag{8}$$

The quantity δ_l is then defined as

$$\delta_l(\mathbf{P}^{a_l}) := \frac{\left\|\mathbf{P}_l^{a_l} - \mathbf{P}_l^t\right\|_F}{\left\|\mathbf{P}_l^t\right\|_F}. \tag{9}$$

In Fig. 3, δ and δ_l are plotted for the case of OL for $N = 16$ over all localization radii. Both curves show small errors in the covariance estimates as long as $l < 13$. Increasing l beyond 13 worsens the estimation of the covariances. If only the observation at each analysis grid point is used ($l = 0$), the estimates of the variance are even worse than in the case when all observations are assimilated at once. Despite this, the state estimation with $l = 0$ is improved over the global filter (see Fig. 1). The smallest error is obtained for the localization radius $l = 11$. This is consistent with the optimal localization radius in Section 4a. For $l > 14$ the assimilations become unstable until divergence happens.

Compared to the global estimate $\mathbf{P}^{a_l}$, the error of the local estimate $\mathbf{P}_l^{a_l}$ is always smaller for all localization radii. This shows that the neglected covariances are noisy and therefore it is beneficial to omit those noisy parts. For l between 3 and 11 the error of the local approximation has roughly the same smallest value. In this interval, the covariances are gradually improved by increasing the localization radius. The interval becomes narrower if a smaller ensemble is used. Thus, it becomes more difficult to find the optimal localization radius. Overall, this experiment shows, that a good choice of the localization radius improves the estimate of the covariance matrix $\mathbf{P}$.

d. Relation between domain- and observation localization

Domain and observations localization differ only in their weight functions. To relate the localizations of DL and OL, we define an effective observation dimension d_{W_k} for an assimilation experiment as the sum of the local weights used to compute the analysis, i.e.

$$d_{W_k} := \sum_{i=0}^{p_l} W_k(i,l) \tag{10}$$

where p_l is the number of observations in each local region, l the localization radius, and k the localization type (OL or DL). Thus, the effective observation dimension takes not only into account the number of observations but also the weights given to the observations. Because in the experiments the observations have uniform density, the effective observation dimension is identical for all grid points. It follows directly from the definition (10) that for DL the effective observation dimension $d_{W_{DL}}$ is equal to the number of observations. In Fig. 4, d_W is plotted for the optimal and divergence localization radii for both DL and OL. The optimal effective observation dimensions are in good agreement for ensemble sizes below 16 with a difference of at most one. For $16 \leq N \leq 20$ the difference gets slightly bigger. Only values up to $N = 20$ are shown, because, as noted in Section 4b, the effective observation dimension for OL is bounded by the considered localization radii for $N \geq 20$.

The effective observation dimension where divergence occurs (bottom of Fig. 4) is nearly equal for $N < 9$ for DL and OL. Above $N = 9$, the observation dimension where the analysis with OL diverges is slightly smaller that the one for domain localization. Yet, the trend for the two functions is still similar. Above $N = 17$, the filter with OL converged for all considered localization radii. The behaviour of the curves is also similar if an exponential weight function is used (not shown). Over all, by decreasing the weight of the observations, they do not constrain the ensemble as strong anymore and the number of observations that can effectively assimilated is increased.

5. Localization with the Shallow Water Equations

In this section the localization experiments are repeated using a model with different dynamics, to examine whether similar results are obtained. In addition, the shallow water model is multivariate, so an additional degree of complexity is introduced.

The MRMSEs for the experiments with the shallow water model (see Fig. 5) are qualitatively similar to the ones for the Lorenz-96 model. The ability of the filter to handle more observations with increasing ensemble size is clearly visible (e.g. the step from $l = 70$km to $l = 80$km for $N = 8$ to $N = 9$) for DL (Fig. 5, top). Compared to the experiments with the Lorenz-96 model, the convergence region is not increasing uniformly with growing ensemble size. This is due to the nonuniform increase of the number of observations in the local domains because the domain is 2-dimensional. The smallest errors for the considered ensemble sizes are achieved for localization radii between 80km and 100km. If l is increased beyond this value, the analysis quality is degraded. For OL (Fig. 5, bottom), the methods behave more uniformly, since the weighting of the observations allows a smoother increase of the observation dimension. This leads to an almost linear increase of the optimal localization radius for $N \leq 14$.

For OL, the convergence region is almost twice as large compared to DL. This occurs because the weight of distant observations is decreased so that more observations can assimilated in a beneficial way. As a consequence, the errors are also slightly reduced. The smallest MRMSE = 0.27 is obtained with a localization radius between 190km and 210km and the largest investigated ensemble size.

In Fig. 6, the effective observation dimensions for the experiments are shown. For DL, the optimal observation dimension l_{opt} is nearly a step function. This means that a much bigger ensemble is needed to assimilate the step-wise increase of observations in an optimal way. This effect does not occur for OL where the optimal observation dimension is growing at a slower rate. For $N = 15$ and $N = 28$, the optimal observation dimension for DL and OL are almost the same. In between, the optimal observation dimension increases about linearly for OL compared to the sudden step for DL. The optimal value for the effective observation dimension is slightly smaller than the ensemble size N for OL, and depends linearly on the ensemble size.

For the effective observation dimension l_{div} at which the filter diverges, the behaviour is slightly different. Divergence occurs for both weighting functions for nearly the same effective observation dimension. Again, the dependence on N is smoother for OL than for DL.

The optimal localization radii for the unobserved u and v fields are almost equal to the optimal localization radius for the height field. There is only a minor difference for DL, when the local number of observations is heavily increased (e.g. $l = 70$km to $l = 80$km). At this point the optimal localization radius is a bit smaller for the u and v fields than for the h field (not shown).

For DL, the slopes of l_{opt} and l_{div} as functions of the ensemble size are reduced compared to the experiment with the Lorenz-96 model. Nevertheless, the effective observation dimensions for DL and OL are very similar, thus the degrees of freedom for both methods are very close to each other.

If the observation density is reduced, the optimal effective observation dimension still depends linearly on the ensemble size (see. Fig. 7). The smaller the observation density, the smaller the optimal effective observation dimension becomes. Thus, if not the whole field is observed, the optimal localization radius has to be normalised by the observational density. This becomes especially an issue, if the spatial distribution of the observations is not regular. This case will be examined in future studies.

Figure 5 also allows to estimate the optimal localization radius as a function of the ensemble size. The relationship is approximately

$$l_{opt} \approx 8\sqrt{\frac{N}{40}}\, dx \tag{11}$$

where dx denotes the grid spacing. At this localization radius, the effective observation dimension is approximately equal to the ensemble size. This relation should hold in general for dense observations that are distributed in 2 dimensions and a regular orthogonal model grid.

6. Localization in a global ocean model

The experiments discussed above indicate that an optimal localization radius is obtained when the effective observation dimension is approximately equal to the ensemble size. To assess whether this localization can be applied in a realistic large-scale model, we apply it here in twin experiments using a global configuration of the finite-element sea-ice ocean model (FESOM, Danilov et al. 2004; Wang et al. 2008; Timmermann et al. 2009). The twin experiments are similar to an application of FESOM by Janjić et al. (2012) where real satellite dynamic ocean topography data was assimilated.

a. Experimental setup

FESOM is an ocean general circulation model that utilises finite elements to solve the hydrostatic ocean primitive equations. Unstructured triangular meshes are used, which allow for a varying resolution of the mesh. The configuration used here has a horizontal resolution of about 1.3° with refinement in the equatorial region. The model uses 40 vertical levels.

For the data assimilation, FESOM was coupled to the assimilation framework PDAF (Nerger et al. 2005; Nerger and Hiller 2013, http://pdaf.awi.de) into a single program. The state vector includes the sea surface height (SSH) and the 3-dimensional fields of temperature, salinity, and the velocity components. The state vector has a size of about 10 million. For the twin experiments, the model is initialised from a spin-up run and a trajectory over one year is computed. This trajectory contains the model fields at each tenth day and represents the "truth" for the assimilation experiments. An ensemble of 32 members is used, which is generated by second-order exact sampling from the variability of the true trajectory (see Pham 2001). The initial state estimate is given by the mean of the true trajectory. Pseudo observations of the SSH at each surface grid point are generated by adding uncorrelated random Gaussian noise with a standard deviation of 5 cm to the true model state. The analysis step is computed after each forecast phase of 10 days with an observation vector containing about 68000 observations. Overall, the experiments were conducted over a period of 360 days.

The experiments use the ETKF with OL. Two experiments with fixed localization radii of l=500km and l=1000km are performed. A third experiment uses the localization radius determined such that the effective observation dimension is equal to the ensemble size. The inflation factor was set to $\rho = 1.1$.

b. Assimilation performance

Figure 8 shows of the RMS errors of the sea surface height over time relative to an experiment without data assimilation for the three experiments. For the fixed radius of l=1000km, the relative RMS error is quickly reduced below 0.5, but increases again after day 150. The relative RMS errors for the fixed radius of 500km and the experiment with the localization radius based on the effective observation dimension are similar and the error generally decrease over time. However, the variable localization results in smaller RMS errors than the fixed localization radius. In the second half of the experiment, the RMS errors obtained with the variable localization radius are even smaller than those for the fixed localization radius of 1000km.

Overall, the experiments show that the effective observation dimension can be used to specify a spatially varying localization radius that yields estimates of similar quality than those produced by a fixed radius. However, while the fixed radius has to be tuned with several experiments this is not required for the variable radius.

7. Conclusion

In this study, the optimal value for the localization radius in domain localization and observation localization was examined using numerical experiments. Using the Lorenz-96 model and a nonlinear shallow-water model allowed to assess the localization behaviour with two simple nonlinear models with different dynamics. The main focus was on dense observations with uniform observational error, which are used in real assimilation applications, e.g., as gridded satellite observations of the ocean surface temperature or sea surface height. For this type of observations, it was possible to assess the relation of the localization radius to the ensemble size over the whole model domain.

The localization radius is optimal if the estimation errors are minimal. It depends on the ensemble size and varies for different weight functions. Typically, the optimal radius is determined by experimentation. Yet, one can define an effective observation dimension given as the sum of the observation weights involved in a local analysis. The optimal localization radius was obtained, if the effective observation dimension was about equal to the size of the ensemble. Moreover, the optimal value of the effective observation dimension is constant for different weighting functions. This situation can be explained by the fact that the degrees of freedom for the analysis are determined by the rank of the ensemble. The degrees of freedom are optimally utilized if the ensemble size equals the effective observation dimension. In the case of constant observation errors, the degrees of freedom are distributed over different numbers of observations for different weight functions. If the observation network is less dense, other effects, like sampling error for distant observations, become more important so that this relation is weaker. For multivariate data assimilation in the shallow water model, the optimal effective observation dimension was the same for all three model fields. If the observation density is reduced, the linear relation in the shallow water model was still conserved, but the slope was different. For both models, the optimal value of the effective observation dimension was roughly equal to the ensemble size if a field was completely observed. For dense observations that are distributed in two dimensions, a simple relation between the ensemble size and the optimal localization radius was deduced from the experiments. This relation can be used to define an adaptive localization radius that ensures that the effective observation dimension is equal to the number of ensemble members. The relation was tested using a global ocean model where synthetic observations of the sea surface height were assimilated. With the adaptive localization, without tuning, a similar error reduction as using an optimally tuned fixed localization radius was achieved. This study lead to a simple relation between the ensemble size and the localization radius that should result in the minimal analysis errors of the observed field for dense observations. However, in real applications the localization radius can be influenced by other factors. For example, it is known that localization influences balances in the model state and a longer localization radius will have a smaller impact on the balances. Accordingly, one might prefer a longer localization radius in multivariate assimilation applications. In addition, the study only considered twin experiments. When assimilating real observations one can encounter biases and the observation error covariance matrix might be incorrectly estimated. It is unclear to which extend these factors can require the adaption of the localization radius to obtain overall optimal assimilation results.

In the experiments, the optimal localization length was not influenced by the model properties. Thus, while different fields in a model can have different correlation length scales, this does not seem to influence the optimal localization radius. A reason for this finding might be that the optimal localization radius for dense observations is rather short. For example, the optimal radius was 8 grid points in the shallow water model for the largest tested ensemble of 40 members. In combination with the weighting by observation localization, observations have only an influence over a distance of a few grid points. This distance should be short enough to effectively remove spurious correlations when the real correlations are very short ranged. If the true error correlations are significant over a long range, at some point they can no longer influence the analysis, because of the limited degrees of freedom provided by the ensemble. Since it is well known that long range correlations are not well approximated with small ensembles, this might be desirable. Nevertheless, the relation between the optimal localization radius and the physical error correlation should be further investigated.

The findings of this study hold for dense observations with uniform observation errors and spatially constant inflation. The experiments with lower observation density indicate that the chosen effective localization dimension has to be smaller in this case, to account for the lack of information. This effect might be related to the sampling quality of the ensemble-estimated state error covariance matrix. When observations with spatially varying error variances and varying spatial distribution are assimilated, the global measurements of this study are no longer possible. One can expect that observations with different error variances show a varying influence on the analysis step that should be accounted for in the localization, perhaps by information-based methods (e.g. Migliorini 2013). These aspects will be investigated in a future study.

Acknowledgments.

The authors thank the reviewers and the editor for their helpful comments.

This work was funded by the SANGOMA EU project (grant FP7-SPACE-2011-1-CT-283580-SANGOMA).

APPENDIX

Appendix A

a. *The shallow-water equations*

The shallow-water model used in section 4 is similar to that used in Krysta et al. (2011). For completeness, the equations are given here. This 2-dimensional model consist of the horizontal and vertical velocities (u, v) and the water height h. The model equations are:

$$\delta_t u + u\delta_x u + v\delta_y u - fv + g^*\delta_x h = \frac{\tau_x}{\rho_0 h} - ru + \nu\Delta u$$
$$\delta_t v + u\delta_x v + v\delta_y v + fu + g^*\delta_y h = \frac{\tau_y}{\rho_0 h} - rv + \nu\Delta v$$
$$\delta_t h + \delta_x(hu) + \delta_y(hv) = 0$$

The model domain is chosen as the square domain $[0, L] \times [y_0 - L, y_0 + L]$ with length $L =$ 2000km and $y_0 = 0$. The Coriolis parameter f is approximated by a β-plane approximation

$$f(y) \approx f(y_0) + \beta(y - y_0) \tag{A1}$$

where $\beta = 2 \cdot 10^{-11}$ m^{-1} s^{-1}. The variable g^* denotes the reduced gravity, ρ_0 water density, ν diffusivity friction and r the bottom friction coefficient. The system is driven by a wind stress $\tau = (\tau_x, \tau_y)^T$, which is given by $\tau_x(y) = \tau_0 \cos[2\pi(y - y_0)/L]$ and $\tau_y = 0$. The constants are chosen as $f(0) = 7 \cdot 10^{-5}\text{s}^{-1}$, $g^* = 0.02\text{ms}^{-2}$, $\rho_0 = 10^3\text{km}^{-3}$, $\tau_0 = 0.015\text{N m}^{-2}$, $r = 5 \cdot 10^{-9}\text{s}^{-1}$ and $\nu = 9\text{m}^2\text{s}^{-1}$.

The domain is discretized on a regular Arakawa C grid with 25km resolution in both directions. For the boundary, a no-slip condition is used, i.e. $u = v = 0$. As time stepping method, a leapfrog scheme (Sadourny 1975) smoothed by the Robert-Asselin filter (Robert 1966) with $\alpha = 0.01$ and $\Delta_t = 30$min is used.

REFERENCES

Anderson, J. L., 2007: Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. *Physica D: Nonlinear Phenomena*, **230 (1-2)**, 99–111.

Anderson, J. L., 2012: Localization and Sampling Error Correction in Ensemble Kalman Filter Data Assimilation. *Mon. Wea. Rev.*, **140 (7)**, 2359–2371.

Bishop, C., B. Etherton, and S. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, **129 (3)**, 420–436.

Bishop, C. H. and D. Hodyss, 2007: Flow-adaptive moderation of spurious ensemble correlations and its use in ensemble-based data assimilation. *Quart. J. Roy. Meteor. Soc.*, **133 (629)**, 2029–2044.

Bishop, C. H. and D. Hodyss, 2009: Ensemble covariances adaptively localized with ECO-RAP. Part 1: tests on simple error models. *Tellus*, **61 (1)**, 84–96.

Danilov, S., G. Kivman, and J. Schröter, 2004: A finite-element ocean model: principles and evaluation. *Ocean Modelling*, **6 (2)**, 125–150.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10 143–10 162.

Gaspari, G. and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757.

Greybush, S. J., E. , T. Miyoshi, K. Ide, and B. R. Hunt, 2011: Balance and ensemble Kalman filter localization techniques. *Mon. Wea. Rev.*, **139**, 511–522.

Houtekamer, P. and H. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811.

Houtekamer, P. and H. Mitchell, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **129**, 123–137.

Hunt, B., E. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D*, **230**, 112–126.

Janjić, T., L. Nerger, A. Albertella, J. Schröter, and S. Skachko, 2011: On domain localization in ensemble-based Kalman filter algorithms. *Mon. Wea. Rev.*, **139**, 2046–2060.

Janjić, T., J. Schröter, R. Savcenko, W. Bosch, a. Albertella, R. Rummel, and O. Klatt, 2012: Impact of combining GRACE and GOCE gravity data on ocean circulation estimates. *Ocean Science*, **8**, 65–79.

Kang, J.-S., E. Kalnay, T. Miyoshi, J. Liu, and I. Fung, 2012: Estimation of surface carbon fluxes with an advanced data assimilation methodology. *J. Geophys. Res.*, **117 (D24)**, D24 101.

Krysta, M., E. Cosme, and J. Verron, 2011: A consistent hybrid variational-smoothing data assimilation method: Application to a simple shallow-water model of the turbulent midlatitude Ocean. *Mon. Wea. Rev.*, **139**, 3333–3347.

Lorenz, E., 1995: Predictability: A problem partly solved. *Proc. Seminar on Predictability.*, ECMWF, 1–18.

Losa, S. N., S. Danilov, J. Schröter, L. Nerger, S. Maßmann, and F. Janssen, 2012: Assimilating NOAA SST data into the BSH operational circulation model for the North and Baltic Seas: Inference about the data. *J. Mar. Syst.*, **105-108 (6)**, 152–162.

Migliorini, S., 2013: Information-based data selection for ensemble data assimilation. *Quart. J. Roy. Meteor. Soc.*, doi: 10.1002/qj.2104.

Miyoshi, T. and S. Yamane, 2007: Local ensemble transform Kalman filtering with an AGCM at a T159/L48 resolution. *Mon. Wea. Rev.*, **135 (11)**, 3841–3861.

Nerger, L., S. Danilov, W. Hiller, and J. Schröter, 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. and W. Hiller, 2013: Software for ensemble-based data assimilation systems implementation strategies and scalability. *Computers & Geosciences*, **55**, 110–118.

Nerger, L., W. Hiller, and J. Schröter, 2005: PDAF - the parallel data assimilation framework: Experiences with kalman filtering. *Use of High Performance Computing in Meteorology - Proceedings of the 11. ECMWF Workshop*, 63–66.

Nerger, L., T. Janjic, J. Schroeter, and W. Hiller, 2012: A regulated localization scheme for ensemble-based Kalman filters. *Quart. J. Roy. Meteor. Soc.*, **138**, 802–812.

Otkin, J. A., 2012: Assessing the impact of the covariance localization radius when assimilating infrared brightness temperature observations using an ensemble Kalman filter. *Mon. Wea. Rev.*, **140**, 543–561.

Patil, D., B. Hunt, E. Kalnay, J. Yorke, and E. Ott, 2001: Local low dimensionality of atmospheric dynamics. *Phys. Rev. Lett.*, **86**, 5878–5881.

Pham, D. T., 2001: Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Mon. Wea. Rev.*, **129**, 1194–1207.

Robert, A. J., 1966: The integration of a low order spectral form of the primitive meteorological equations. *J. Meteor. Soc. Japan*, **44**, 237–245.

Sadourny, R., 1975: The dynamics of finite-difference models of the shallow-water equations. *J. Atmos. Sci.*, **32**, 680–689.

Sakov, P. and L. Bertino, 2011: Relation between two common localization methods for the EnKF. *Comput. Geosci*, **15**, 225–237.

Sakov, P., F. Counillon, L. Bertino, K. a. Lisæter, P. R. Oke, and A. Korablev, 2012: TOPAZ4: an ocean-sea ice data assimilation system for the North Atlantic and Arctic. *Ocean Science*, **8 (4)**, 633–656.

Timmermann, R., S. Danilov, J. Schröter, C. Böning, D. Sidorenko, and K. Rollenhagen, 2009: Ocean circulation and sea ice distribution in a finite element global sea ice ocean model. *Ocean Modelling*, **27 (3-4)**, 114–129.

Wang, Q., S. Danilov, and J. Schröter, 2008: Finite element ocean circulation model based on triangular prismatic elements, with application in studying the effect of topography representation. *Journal of Geophysical Research*, **113 (C5)**, C05 015.

Whitaker, J. and T. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1722.

Yoon, Y.-n., E. Ott, and I. Szunyogh, 2010: On the propagation of information and the use of localization in ensemble Kalman filtering. *J. Atmos. Sci.*, **67**, 3823–3834.

Zupanski, D., a. S. Denning, M. Uliasz, M. Zupanski, A. E. Schuh, P. J. Rayner, W. Peters, and K. D. Corbin, 2007: Carbon flux bias estimation employing Maximum Likelihood Ensemble Filter (MLEF). *J. Geophys. Res.*, **112 (D17)**, D17 107.

List of Figures

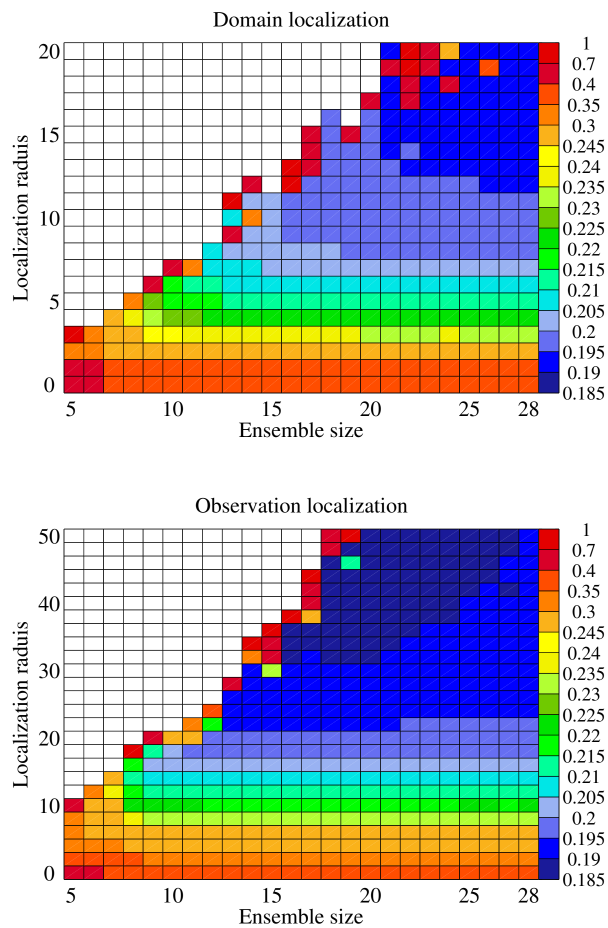
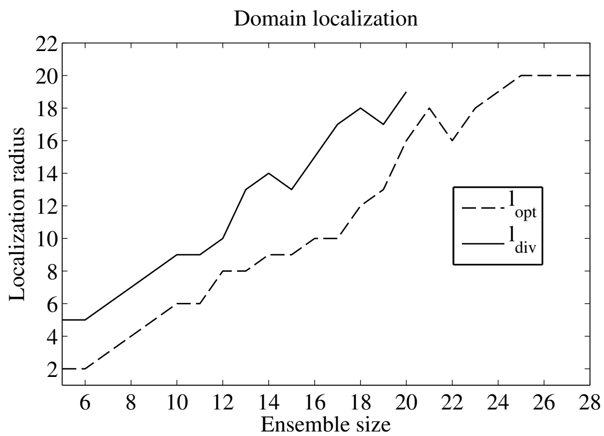


FIG. 1. MRMSE for the assimilation experiments with DL for the different parameter values (top) and for OL (bottom) with the Lorenz-96 model.



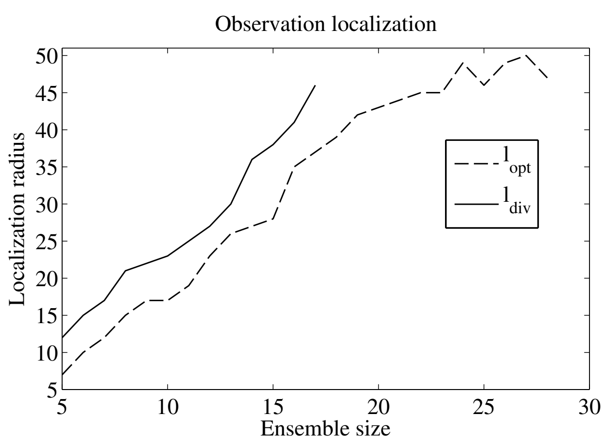


FIG. 2. The optimal and divergent localization radii for DL (top) and OL (bottom).

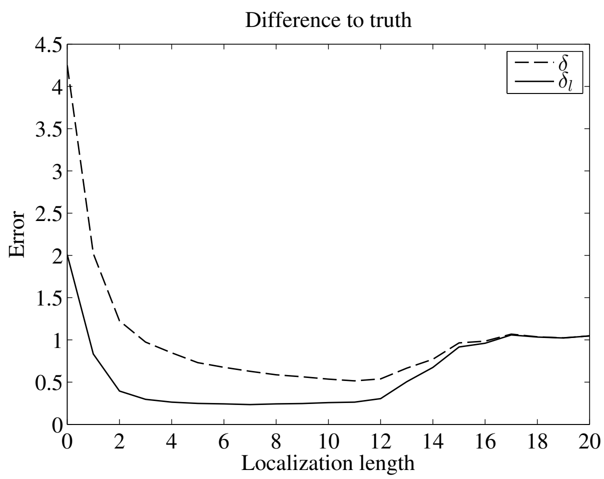


FIG. 3. The error of the global and local covariance matrix to the true covariance matrix calculated from an experiment with 128 ensemble members.

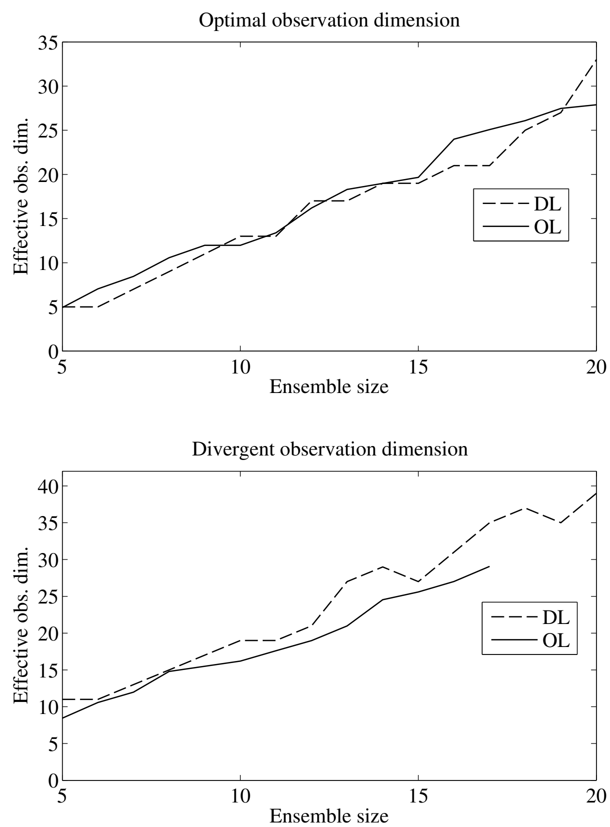


FIG. 4. Comparison of the optimal effective observation dimension (top) and the effective observation dimension where the filter on average diverges (bottom).

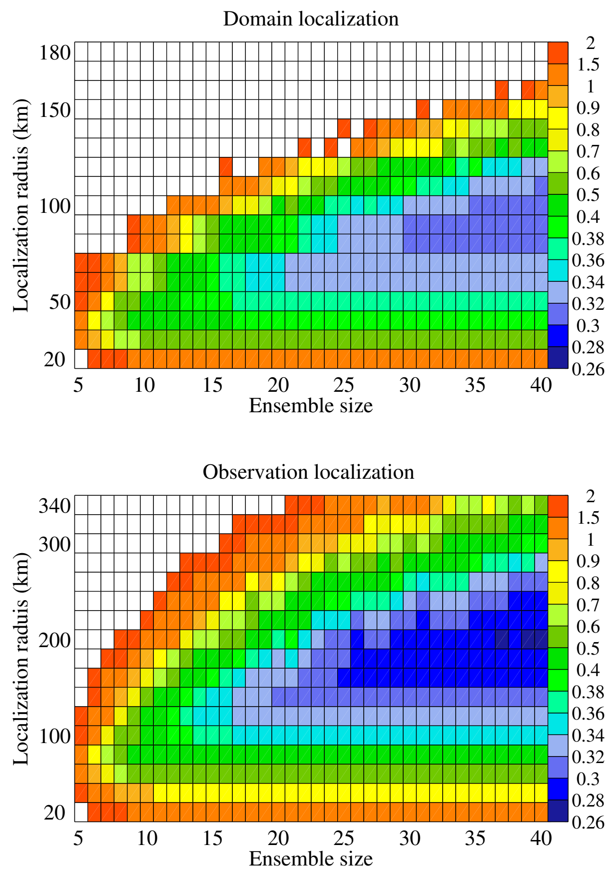


FIG. 5. MRMSE for the assimilation experiments with DL for the different parameter values (top) and for OL (bottom) with the shallow water model.

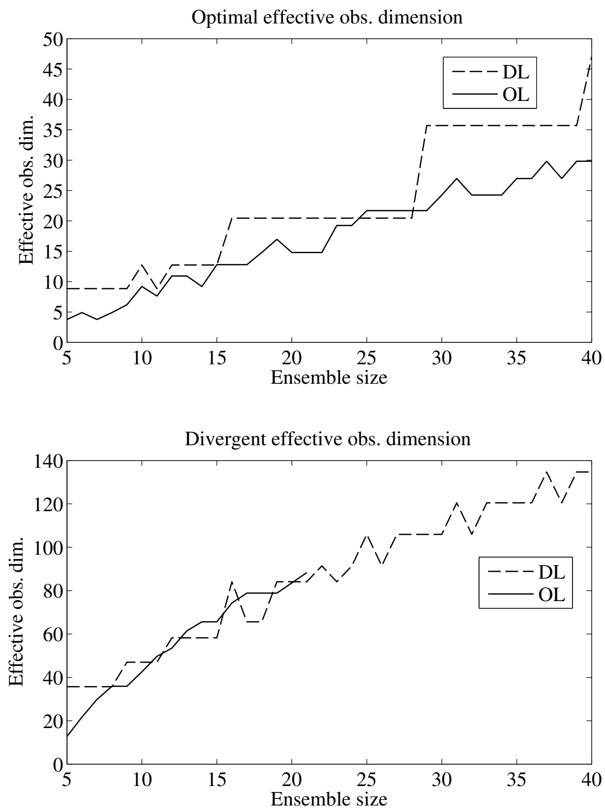


Fig. 6. The optimal and divergent observation dimensions for DL (top) and OL (bottom) for the shallow water model.

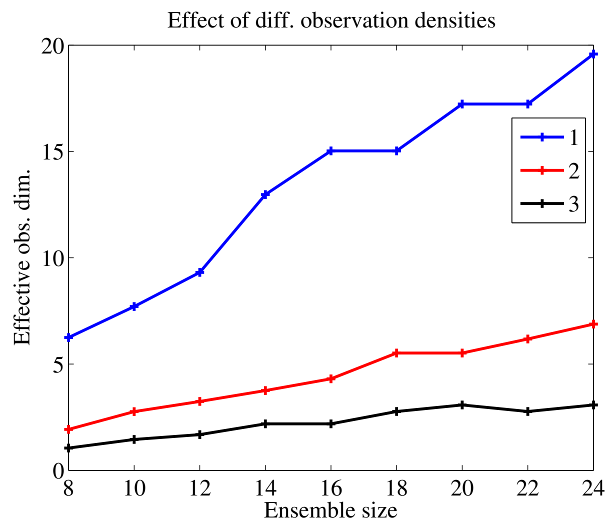


Fig. 7. The optimal effective observation dimension with observation frequency one (blue), two (green) and three (red). For each observation frequency, the optimal value depends linearly on the ensemble size. The smaller the the observation density the smaller the slope of the function.

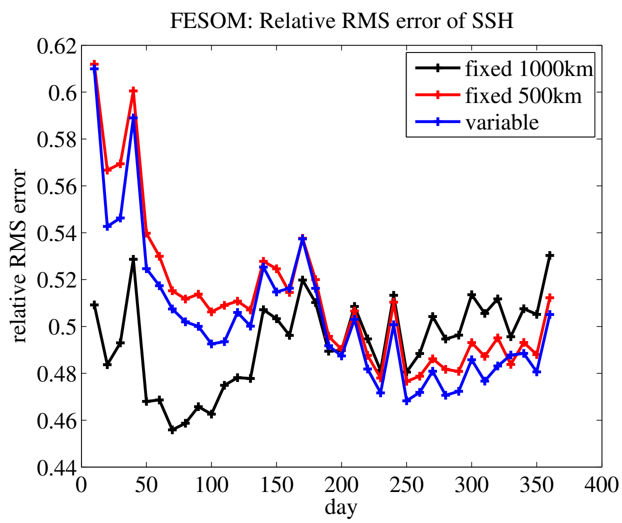


FIG. 8. RMS errors for the assimilation experiment using FESOM relative to the errors from an experiment without assimilation. Shown are the relative RMS errors for fixed localization radius of 1000km (black), 500km (red), and the variable localization derived from the effective observation dimension (blue).