Tellus
SERIES A
DYANAMIC
METEOROLOGY
AND OCEANOGRAPHY
PUBLISHED BY THE INTERNATIONAL METEOROLOGICAL INSTITUTE IN STOCKHOLM

Eigenvector-spatial localisation

By TRAVIS HARTY[1]*, MATTHIAS MORZFELD[2], and CHRIS SNYDER[3], [1]*Program in Applied Mathematics, University of Arizona, Tucson, AZ, USA;* [2]*Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California, San Diego, CA, USA;* [3]*National Centre for Atmospheric Research, Boulder, CO, USA*

(Manuscript Received 5 November 2020; in final form 3 March 2021)

ABSTRACT

We present a new multiscale covariance localisation method for ensemble data assimilation that is based on the estimation of eigenvectors and subsequent projections, together with traditional spatial localisation applied with a range of localisation lengths. In short, we estimate the leading, large-scale eigenvectors from the sample covariance matrix obtained by spatially smoothing the ensemble (treating small scales as noise) and then localise the resulting sample covariances with a large length scale. After removing the projection of each ensemble member onto the leading eigenvectors, the process may be repeated using less smoothing and tighter localizations or, in a final step, using the resulting, residual ensemble and tight localisation to represent covariances in the remaining subspace. We illustrate the use of the new multiscale localisation method in simple numerical examples and in cycling data assimilation experiments with the Lorenz Model III. We also compare the proposed new method to existing multiscale localisation and to single-scale localisation.

Keywords: localisation, multiscale, data assimilation, Kalman filter, ensemble

1. Introduction

The use of an ensemble of forecasts to estimate the required instantaneous covariances, as first proposed in the ensemble Kalman filter (EnKF) of Evensen (1994), has fundamentally reshaped geophysical data assimilation schemes over the last two decades. Whether applied to EnKFs or to ensemble-variational schemes (EnVar; Lorenc, 2003), an essential element of ensemble-based data assimilation in all but the simplest applications is localisation (Houtekamer and Mitchell, 1998, 2001; Hamill et al., 2001; Ott et al., 2004). In essence, localisation enforces on the sample covariance from an ensemble the assumption that the covariance between two locations will be small for sufficiently large separations. It is often implemented by multiplying the sample covariance by a scalar that decreases to zero as the separation increases. We will term the functional dependence of that scalar on separation the localisation function; more details appear in Section 2.3.

A very active area of research is to generalise localisation to systems with a range of spatial scales by relaxing the requirement of a single localisation function everywhere and for all variables. The present paper contributes a novel approach to localisation in multiscale systems and presents initial tests in an idealised system.

Various ideas related to such generalisations have been put forward, with the specifics of the approach often tied to the the sort of ensemble assimilation technique employed. In one thread, the algorithms decompose ensemble members into "scales", typically by applying bandpass filters in Fourier space, and then localise differently for different scales (Buehner, 2012; Miyoshi and Kondo, 2013; Buehner and Shlyaeva, 2015; Lorenc, 2017) This thread is specifically aimed at systems where correlation lengths vary spatially and also considers only the state covariance matrix, independent of observations. Our method, while different, shares this spatial focus.

A second thread are techniques that seek to optimise a localisation function, depending not only on spatial separation but also on other quantities such as location, temporal separation, type of state variable, or aspects of the observing network. For serial EnKFs, it is natural to choose localisation to minimise difference of the resulting Kalman gain from the optimal value or, equivalently, the expected squared analysis error (Anderson and Lei, 2013; Zhen and Zhang, 2014; Flowerdew, 2015; De La

*Corresponding author. e-mail: travisharty@gmail.com

Citation: Tellus A: 2021, **73**, 1903692, https://doi.org/10.1080/16000870.2021.1903692

Chevrotière and Harlim, 2017). For EnVar, it is more convenient to minimise the expected squared error in estimates of the state covariance matrix (Ménétrier et al., 2015). In all these techniques, it is necessary to aggregate information in some way, to get robust estimates in the presence of sampling error. When the aggregation is spatial, the result is a spatially varying localisation function (Ménétrier et al., 2015).

Our approach begins from the notion that a few leading eigenvectors and eigenvalues can provide an efficient estimate of a covariance matrix. Because leading eigenvectors typically represent large spatial scales in atmospheric and oceanic assimilation applications, we seek to estimate the leading eigenvectors from a localised sample covariance matrix based on a large-scale localisation function and some spatial smoothing applied to the ensemble members. The idea of using smoothing of the ensemble is not new and is also used, for example, in Bishop and Hodyss (2007, 2011), and in multi scale waveband localisation (Buehner, 2012; Buehner and Shlyaeva, 2015), which we will discuss in more detail further below. We approximate the covariance in the remaining "small scale" subspace complementary to the leading eigenvectors by projecting each member onto the complementary subspace and forming a localised sample covariance matrix using a narrower localisation function. Though we do not present results here, this process could be continued, computing additional eigenvectors, projecting out from each member, and localising the remaining sample covariance with a still narrower localisation function. We term our approach "eigenvector-spatial localisation" (ESL).

The paper proceeds as follows. Section 2 gives further background and notation, including for localisation, that is helpful in presenting ESL, while Section 3 presents the ESL algorithm in detail. We then assess ESL in two idealised examples: first for a state with specified Gaussian statistics in a one-dimensional spatial domain (Section 4), and then in cycled data assimilation experiments using model III of Lorenz (2005) in Section 5. A summary and conclusions appear in Section 6.

2. Background

We briefly review background materials on data assimilation, ensemble Kalman filtering, localisation and inflation. We also define the root mean squared error and spread, which are used to compare several localisation techniques in numerical examples. In the process, we set up the notation used in this paper.

2.1. Data assimilation

The goal in data assimilation (DA) is to merge information from a computational model with information from observations of a geophysical process. We denote the system state by $\mathbf{x}$ and the observations by $\mathbf{y}$. Throughout this paper, we denote the dimension of a vector by N with subscript, i.e. N_x and N_y are the dimensions of the $\mathbf{x}$ and $\mathbf{y}$. We assume that the model state and observations satisfy

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \varepsilon, \quad \varepsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{R}), \tag{1}$$

where, $\mathbf{H}$ is a $N_y \times N_x$ matrix and ε is a N_y-dimensional Gaussian random vector with mean zero and covariance matrix $\mathbf{R}$. The observation equation (1) defines the likelihood $p_l(\mathbf{y}|\mathbf{x})$. The model defines a forecast distribution $p_0(\mathbf{x})$. In the Bayesian framework, one can view the forecast distribution as a prior distribution and define the posterior distribution

$$p(\mathbf{x}|\mathbf{y}) \propto p_0(\mathbf{x}) p_l(\mathbf{y}|\mathbf{x}). \tag{2}$$

All numerical methods for DA use the posterior distribution to compute state estimates: Variational methods compute the maximiser of the posterior distribution and ensemble methods approximate the posterior distribution by Monte Carlo. One of the main difficulties in DA in geophysical problems is the large dimension of $\mathbf{x}$. In global numerical weather prediction (NWP), for example, the dimensions of observation and state vectors are $N_y = O(10^7)$ and $N_x = O(10^8)$. Geophysically relevant DA techniques must be able to operate in these high-dimensional spaces.

2.2. Ensemble Kalman filter

The basic idea of the ensemble Kalman filter (Evensen, 1994) is to use a Monte Carlo approximation within the Kalman filter framework. Given a forecast ensemble with N_e ensemble members $\mathbf{x}_j, j = 1, ..., N_e$, the goal is to update the forecast in view of the observation $\mathbf{y}$. The EnKF performs this update as follows. Define a $N_x \times N_e$ matrix that contains the normalised forecast perturbations

$$\mathbf{X} = \frac{1}{\sqrt{N_e - 1}} \cdot [\mathbf{x}_1 - \bar{\mathbf{x}}, ..., \mathbf{x}_{N_e} - \bar{\mathbf{x}}], \tag{3}$$

where $\bar{\mathbf{x}} = (1/N_e) \sum_{i=1}^{N_e} \mathbf{x}_i$ is the mean of the forecast ensemble. The forecast covariance is then estimated by

$$\hat{\mathbf{P}} = \mathbf{X}\mathbf{X}^{\mathrm{T}}, \tag{4}$$

where the super-script T denotes a transposed. The analysis ensemble is defined by applying the Kalman filter formulas to each ensemble member:

$$\mathbf{x}_i^a = \mathbf{x}_i + \hat{\mathbf{K}}(\mathbf{y} - (\mathbf{H}\mathbf{x}_i + \varepsilon_i)), \tag{5}$$

where ε_i is a realisation of ε (see (1)) and where the Kalman gain is estimated using $\hat{\mathbf{P}}$,

$$\hat{\mathbf{K}} = \hat{\mathbf{P}}\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\hat{\mathbf{P}}\mathbf{H}^{\mathrm{T}} + \mathbf{R}\right)^{-1} \tag{6}$$

Thus, the analysis corrects the forecast based on the differences between the observations and the forecast, with the degree of correction defined by the Kalman gain.

The EnKF we describe is called the "stochastic" EnKF (Burgers et al., 1998) and has the property that the analysis ensemble is distributed according to the posterior distribution if (i) the model and observation equations are linear; (ii) the model and observation errors are Gaussian; and (iii) the ensemble size goes to infinity, see, e.g. Mandel et al. (2011). There are many other implementations of the EnKF, see, e.g. Tippet et al. (2003); Hunt et al. (2007); Buehner et al. (2017). In the numerical examples below, we test new and current localisation methods on the stochastic EnKF, but the methods are more general and can be applied to other implementations of the EnKF. Specifically, application to high-dimensional problems requires that all methods that are used are scalable to a high dimension. The stochastic EnKF in its simplest form is not scalable, but can be made scalable by using a "variational" implementation as described by Buehner et al. (2017) (see also Section 3.4), or ensemble transforms as described, e.g. in Tippet et al. (2003); Hunt et al. (2007).

2.3. Localisation and inflation

In high-dimensional problems, the ensemble size is typically limited to the order of tens or hundreds owing to the computational costs of the ensemble forecasts in the EnKF. As a specific example, consider again global NWP, where the dimension of the model is $N_x \approx 10^8$ and the number of observations is $N_y \approx 10^7$. A small ensemble size of $N_e \approx 100$, causes large errors in the sample estimates of the high-dimensional forecast covariance matrix, which result in unrealistic long-range correlations. The essential idea of localisation is to reduce sampling error by deleting spurious long-range correlations (Hamill et al., 2001; Houtekamer and Mitchell, 2001). In this way, localisation increases the accuracy of ensemble covariances to exceed what one should expect based on the small ensemble size.

Though there are a variety of approaches to localisation, see, e.g. Shlyaeva et al. (2019), in this paper we will be concerned with localisation of the sample forecast covariance using Schur (or elementwise) multiplication by a localisation matrix $\mathbf{L}$:

$$\hat{\mathbf{P}}_{\mathrm{loc}} = \mathbf{L} \circ (\mathbf{X}\mathbf{X}^{\mathrm{T}}). \tag{7}$$

Recall that if $\mathbf{L}$ is positive definite (and thus full rank), then $\hat{\mathbf{P}}_{\mathrm{loc}}$ is also positive definite and full rank (see, e.g. Horn, 1990). Typically, $\mathbf{L}$ is taken to be defined by a homogeneous, isotropic, and non-negative correlation function, with correlations decreasing monotonically with distance (e.g. Gaspari and Cohn, 1999). The length scales of the correlation function, which control how heavily long-range sample covariances are damped in (7), are tuned to maximise forecast skill. For a given localisation function, localisation of the stochastic EnKF amounts to replacing $\hat{\mathbf{P}}$ by $\hat{\mathbf{P}}_{\mathrm{loc}}$.

A variety of techniques, typically termed "inflation," are used to compensate for the tendency of EnKFs to generate analysis ensembles whose covariance underestimates actual errors, see, e.g. Kotsuki et al. (2017). Here, we employ multiplicative covariance inflation (Anderson and Anderson, 1999), but other techniques can also be used, see, e.g. Luo and Hoteit (2011). In multiplicative inflation, the inflated ensemble is defined by

$$\mathbf{X}_{\mathrm{infl}} = (1+\alpha)(\mathbf{X} - \bar{\mathbf{X}}) + \bar{\mathbf{X}}, \tag{8}$$

where $\alpha \geq 0$ defines the amount of inflation, and where the columns of the $N_x \times N_e$ matrix $\bar{\mathbf{X}}$ are the forecast mean $\bar{\mathbf{x}}$. In an inflated EnKF, the inflated forecast ensemble (8) replaces the forecast ensemble. The inflation parameter α is usually tuned to maximise forecast skill.

2.4. Multiscale waveband localisation

Geophysical systems often span a range of spatial and temporal scales, but localisation and inflation, as described above, do not take multiscale characteristics into account. The idea of multiscale localisation is to address multiscale behaviour during the process of localisation. An example of a multiscale method is multiscale waveband localisation, which we refer to as "waveband localisation" (Buehner, 2012). Waveband localisation uses spectral filters to decompose the state into separate scales, and then applies different localisation matrices to each of the resulting sample covariance matrices. The following summarises the algorithm described in Buehner and Shlyaeva (2015), which extends the method presented in Buehner (2012).

One first defines N_s spectral filters $\mathbf{\Psi}^k, k = 1, ..., N_s$, with the property that $\sum_{k=1}^{N_s} \mathbf{\Psi}^k = \mathbf{I}$. Buehner and Shlyaeva (2015) use spectral filters whose coefficients decay sinusoidally and we also make this choice (see below for more details). The filters are applied to each member of the forecast ensemble

$$\mathbf{x}_i^k = \mathbf{\Psi}^k \mathbf{x}_i. \tag{9}$$

After filtering, one thus ends up with N_s ensembles, where each ensemble corresponds to a different scale,

because the filters $\mathbf{\Psi}^k$ isolate spectral wavebands (scales). For each ensemble, we define normalised forecast perturbations

$$\mathbf{X}_k = \frac{1}{\sqrt{N_e - 1}} \left[\mathbf{x}_1^k - \bar{\mathbf{x}}^k, ..., \mathbf{x}_{N_e}^k - \bar{\mathbf{x}}^k \right], \tag{10}$$

where the bar denotes the sample average. The corresponding forecast covariance matrices are given by

$$\hat{\mathbf{P}}_{k_1,k_2} = \mathbf{X}_{k_1} \mathbf{X}_{k_2}^{\mathrm{T}}. \tag{11}$$

Note that $\hat{\mathbf{P}}_{k_2,k_1} = (\hat{\mathbf{P}}_{k_1,k_2})^{\mathrm{T}}$, so that one obtains $N_s(N_s+1)/2$ covariance matrices. To implement a scale-dependent localisation, one applies localisation matrices $\mathbf{L}_{k_1,k_2}$ to the covariance matrices $\hat{\mathbf{P}}_{k_1,k_2}$:

$$\hat{\mathbf{P}}_{k_1,k_2}^{\mathrm{loc}} = \mathbf{L}_{k_1,k_2} \circ \hat{\mathbf{P}}_{k_1,k_2}. \tag{12}$$

Here, $\mathbf{L}^{k_1,k_2}$ is a localisation matrix as before if $k_1 = k_2$, and if $k_1 \neq k_2$, it is defined by

$$\mathbf{L}_{k_1,k_2} = \mathbf{L}_{k_1,k_1}^{1/2} \mathbf{L}_{k_2,k_2}^{\mathrm{T}/2}, \tag{13}$$

where $\mathbf{L}^{1/2}$ denotes a matrix square root such that $\mathbf{L} = \mathbf{L}^{1/2}\mathbf{L}^{\mathrm{T}/2}$. Note that the localisation varies as a function of the waveband indices k_1 and k_2. The sum of all localised covariance matrices across the various wavebands defines the localised forecast covariance matrix:

$$\hat{\mathbf{P}}_{\mathrm{loc}} = \sum_{k_1=1}^{N_s} \sum_{k_2=1}^{N_s} \hat{\mathbf{P}}_{k_1,k_2}^{\mathrm{loc}}. \tag{14}$$

Note that we recover the "usual" localisation as described in Section 2.3, if $\mathbf{L}_{k_1,k_2} = \mathbf{L}$, i.e. if the same localisation is applied to all covariances in (11) (Buehner and Shlyaeva, 2015). In general, the number of spectral filters and which wavebands are filtered, is problem dependent and is tuned.

In the numerical examples below, we use waveband localisation with two scales, which means that we use two spectral filters. Following Section 3 of Buehner and Shlyaeva (2015), we define one filter by a cosine function that decays to zero at a specified wavenumber. We tune this wavenumber to obtain minimum forecast/analysis errors. The second filter follows from the property that both filters add up to one.

2.5. Assessing the skill of ensemble DA

To assess the skill of a DA technique, we compute the root mean squared error (RMSE) and the spread for forecasts and analyses. Forecast RMSE is defined by

$$\mathrm{RMSE} = \left(\frac{1}{N_x} \sum_{i=1}^{N_x} ([\bar{\mathbf{x}}]_i - [\mathbf{x}^t]_i)^2 \right)^{\frac{1}{2}}, \tag{15}$$

where $\bar{\mathbf{x}}$ is the forecast mean and $\mathbf{x}^t$ is the "true" state at forecast time; here and below, we denote the elements of a vector $\mathbf{x}$ by $[\mathbf{x}]_i, i = 1, ..., N_x$. The *forecast* spread is defined by

$$\mathrm{spread} = \left(\frac{\mathrm{trace}(\hat{\mathbf{P}})}{N_x} \right)^{\frac{1}{2}} \tag{16}$$

where $\mathrm{trace}(\hat{\mathbf{P}}) = \sum_{i=1}^{N_x} [\hat{\mathbf{P}}]_{i,i}$ and $[\hat{\mathbf{P}}]_{i,j}, i,j = 1, ..., N_x$ are the elements of $\hat{\mathbf{P}}$. Analysis RMSE and spread are defined as above but with $\bar{\mathbf{x}}$ replaced by the analysis mean, $\mathbf{x}^t$ replaced by the true state at analysis time, and $\hat{\mathbf{P}}$ replaced by the sample covariance of the analysis ensemble. For both forecasts and analyses, a well-tuned EnKF should have the property that RMSE≈spread, so that expected errors are comparable to actual errors.

We use the Frobenius norm to measure error between two matrices. The Frobenius norm of an $N \times N$ matrix $\mathbf{A}$ is defined as

$$\|\mathbf{A}\| = \left(\sum_{i=1}^{N} \sum_{j=1}^{N} [\mathbf{A}]_{i,j}^2 \right)^{\frac{1}{2}} = \left(\sum_{i=1}^{N} \sigma_i^2(\mathbf{A}) \right)^{\frac{1}{2}}, \tag{17}$$

where $\sigma_i(\mathbf{A})$ are the singular values of the matrix $\mathbf{A}$. Note that all singular values of a matrix contribute to the Frobenius norm, whereas the spectral norm of an $N \times N$ matrix $\mathbf{A}$, defined as $\|\mathbf{A}\|_2 = \max_i \sigma(\mathbf{A})_i$, only measures the largest singular value.

3. Eigenvector-spatial localisation

At the core of the localisation method are projections of the ensemble onto eigenvectors of covariance matrices. Recall that the eigenvectors $\mathbf{q}_i$ of a $N_x \times N_x$ covariance matrix $\mathbf{P}$ are orthogonal: $\mathbf{q}_i^{\mathrm{T}}\mathbf{q}_j = 0$ if $i \neq j$ and $\mathbf{q}_i^{\mathrm{T}}\mathbf{q}_i = 1$. Generically, a projection matrix $\mathbf{\Pi}$ has the properties that $\mathbf{\Pi}^2 = \mathbf{\Pi} = \mathbf{\Pi}^{\mathrm{T}}$. The N_x eigenvectors of a covariance matrix thus define N_x projections

$$\mathbf{\Pi}_i = \mathbf{q}_i \mathbf{q}_i^{\mathrm{T}}. \tag{18}$$

We now construct a localisation technique that separates large-scales from small scales by projections of the ensemble onto eigenvectors of a localised, large-scale covariance. We call this method "eigenvector-spatial localisation" (ESL).

3.1. Construction of a localised large-scale covariance

We expect that large-scale features reside in a subspace whose dimension is much less than the state dimension. The reasons why the dimension of this subspace is low can be illustrated by simple examples. Consider a covariance matrix defined over a Discretised spatial domain with elements $[\mathbf{P}]_{i,j} = \exp(-0.5(z_{i,j}/l)^2)$, where l is a

length scale and $z_{i,j}$ is the distance between positions i and j. The eigenvalues of $\mathbf{P}$ decay exponentially and the rate of decay is controlled by the length scale l. For large l, the eigenvalues decay quickly. By setting small eigenvalues equal to zero, one obtains a low-rank approximation of the covariance $\mathbf{P}$, using the first few eigenvalues and eigenvectors.

We wish to construct such a low-rank approximation of the large-scale covariance from an ensemble that contains all scales. If small scales do not immediately affect large-scale structures, we can treat them as noise and remove them by smoothing, or equivalently the use of a large-scale waveband filter as in (9). Applying smoothing to all N_e ensemble members yields the smoothed ensemble

$$\mathbf{x}_i^{\mathrm{s}} = \mathrm{smooth}(\mathbf{x}_i). \tag{19}$$

In the numerical examples we implement the smoothing using a spatial average with weights that decrease with separation following a Gaussian. We view the spatial smoothing of the ensemble as a practical step that has the benefit of reducing the ensemble size needed for reasonable estimates, discussed below, of the leading eigenvectors of large scale covariances.

The smoothed ensemble defines normalised forecast perturbations

$$\mathbf{X}_{\mathrm{s}} = \frac{1}{\sqrt{N_e-1}}\left[\mathbf{x}_i^{\mathrm{s}} - \bar{\mathbf{x}}^{\mathrm{s}}, ..., \mathbf{x}_{N_e}^{\mathrm{s}} - \bar{\mathbf{x}}^{\mathrm{s}}\right], \tag{20}$$

and we localise the resulting ensemble covariance by a symmetric positive definite localisation matrix $\mathbf{L}_{\mathrm{s}}$:

$$\hat{\mathbf{P}}_{\mathrm{s}} = \mathbf{L}_{\mathrm{s}} \circ \left(\mathbf{X}_{\mathrm{s}}\mathbf{X}_{\mathrm{s}}^{\mathrm{T}}\right). \tag{21}$$

The localised covariance $\hat{\mathbf{P}}_{\mathrm{s}}$ is positive semi-definite by construction and, since small scales have been removed via smoothing, its eigenvalues decay.

We can now compute a large-scale covariance by projecting the ensemble onto the N_{lg} leading eigenvectors of $\hat{\mathbf{P}}_{\mathrm{s}}$. We write the projections as

$$\mathbf{\Pi}_i = \mathbf{q}_i\mathbf{q}_i^{\mathrm{T}}, \quad i = 1, ..., N_{\mathrm{lg}}, \tag{22}$$

where $\mathbf{q}_i, i = 1, ..., N_{\mathrm{lg}}$ are the leading eigenvectors of $\hat{\mathbf{P}}_{\mathrm{s}}$ in (21). The large-scale covariance is defined by the sum of N_{lg} projections of the ensemble $\mathbf{X}$:

$$\hat{\mathbf{P}}_{\mathrm{lg}} = \sum_{i=1}^{N_{\mathrm{lg}}} (\mathbf{\Pi}_i\mathbf{X})(\mathbf{\Pi}_i\mathbf{X})^{\mathrm{T}}. \tag{23}$$

Note that we do not apply localisation to $\hat{\mathbf{P}}_{\mathrm{lg}}$ directly; instead, it reflects the effects of the localisation applied to $\hat{\mathbf{P}}_{\mathrm{s}}$, through its dependence on the eigenvectors $\mathbf{q}_i$.

To see that $\hat{\mathbf{P}}_{\mathrm{lg}}$ is indeed low-rank, substitute the definition of $\mathbf{\Pi}_i$ into (23) to find

$$\hat{\mathbf{P}}_{\mathrm{lg}} = \sum_{i=1}^{N_{\mathrm{lg}}} s_i\left(\mathbf{q}_i\mathbf{q}_i^{\mathrm{T}}\right), \tag{24}$$

where $s_i = \mathbf{q}_i^{\mathrm{T}}(\mathbf{X}\mathbf{X}^{\mathrm{T}})\mathbf{q}_i$ are scalars. The matrices $\mathbf{q}_i\mathbf{q}_i^{\mathrm{T}}$ are rank-one and have orthogonal ranges. It follows that $\mathrm{rank}(\hat{\mathbf{P}}_{\mathrm{lg}}) = N_{\mathrm{lg}}$. Since we assume that $N_{\mathrm{lg}} \ll N_x$, $\hat{\mathbf{P}}_{\mathrm{lg}}$ is low-rank.

3.2. Construction of a localised small-scale covariance

We construct a localised covariance for the small scales by first projecting the ensemble perturbations onto the space orthogonal to the space spanned by the N_{lg} leading eigenvectors $\mathbf{q}_i, i = 1, ..., N_{\mathrm{lg}}$, of $\hat{\mathbf{P}}_s$ in (21). The required projection matrix is given by

$$\mathbf{\Pi}_{\perp} = \mathbf{I} - \mathbf{Q}\mathbf{Q}^{\mathrm{T}}, \tag{25}$$

where the $N_x \times N_{\mathrm{lg}}$ matrix $\mathbf{Q}$ has the N_{lg} leading eigenvectors $\mathbf{q}_i$ as its columns:

$$\mathbf{Q} = \left[\mathbf{q}_1, ..., \mathbf{q}_{N_{\mathrm{lg}}}\right]. \tag{26}$$

With these definitions, we define the localised, small-scale covariance by

$$\tilde{\hat{\mathbf{P}}}_{\mathrm{sm}} = \mathbf{L}_{\mathrm{sm}} \circ \left(\mathbf{\Pi}_{\perp}(\mathbf{X}\mathbf{X}^{\mathrm{T}})\mathbf{\Pi}_{\perp}\right), \tag{27}$$

where $\mathbf{L}_{\mathrm{sm}}$ is a positive definite localisation matrix for the small scales. Note that $\mathbf{L}_{\mathrm{sm}}$ is chosen independently of the large-scale localisation $\mathbf{L}_{\mathrm{lg}}$, so that we can apply a tighter localisation to the small scales than to the large scales.

Owing to the localisation with $\mathbf{L}_{\mathrm{sm}}$, the eigenvectors of $\tilde{\hat{\mathbf{P}}}_{\mathrm{sm}}$ are not orthogonal to the eigenvectors of $\hat{\mathbf{P}}_{\mathrm{lg}}$. To keep all eigenvectors orthogonal, we apply the projection once more and obtain

$$\hat{\mathbf{P}}_{\mathrm{sm}} = \mathbf{\Pi}_{\perp}\tilde{\hat{\mathbf{P}}}_{\mathrm{sm}}\mathbf{\Pi}_{\perp}. \tag{28}$$

The rank of $\hat{\mathbf{P}}_{\mathrm{sm}}$ will in general be high because of the localisation. In particular, $\tilde{\hat{\mathbf{P}}}_{\mathrm{sm}}$ will be full rank when $\mathbf{L}_{\mathrm{sm}}$ is full rank and thus $\hat{\mathbf{P}}_{\mathrm{sm}}$ will have the same rank, $N_x - N_{\mathrm{lg}}$, as $\mathbf{\Pi}_{\perp}$. Thus, while we leverage low-rank approximations in the large-scale covariances, we do not rely on low-rank approximations when constructing small-scale covariances.

3.3. Construction of a localised multiscale covariance

The localised multiscale covariance is simply the sum of the localised small and large-scale covariances:

$$\hat{\mathbf{P}}_{\mathrm{loc}} = \hat{\mathbf{P}}_{\mathrm{lg}} + \hat{\mathbf{P}}_{\mathrm{sm}}. \tag{29}$$

Because the eigenvectors of $\hat{\mathbf{P}}_{\mathrm{lg}}$ and $\hat{\mathbf{P}}_{\mathrm{sm}}$ are orthogonal the rank of $\hat{\mathbf{P}}_{\mathrm{loc}}$ is:

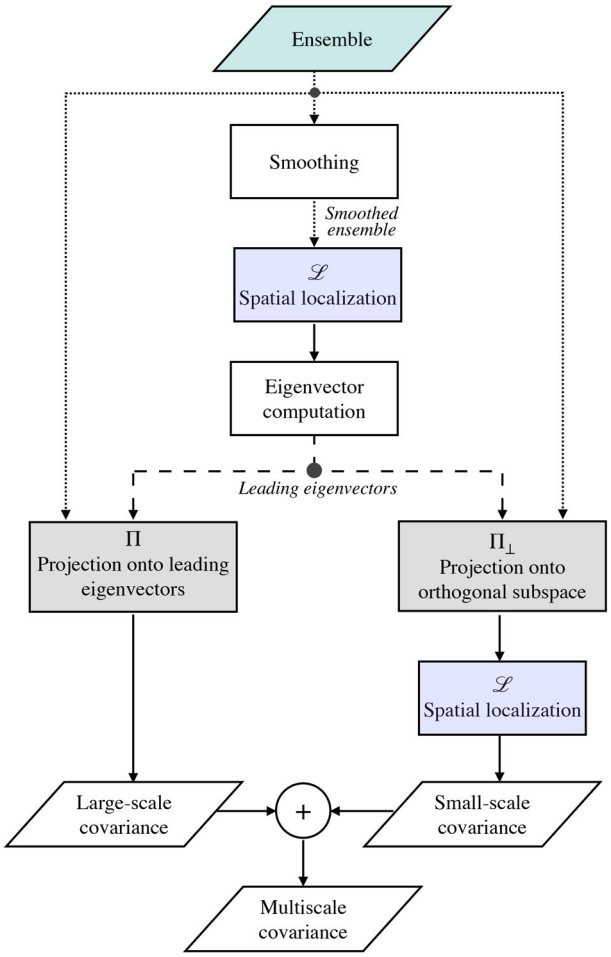

Fig. 1. Flowchart of ESL. Rounded boxes are operations and trapezoids are inputs and outputs (some intermediate inputs and outputs are omitted for clarity). Dotted lines represent ensembles, solid lines represent covariance matrices, and dashed lines represent a collection of eigenvectors.

$$\operatorname{rank}(\hat{\mathbf{P}}_{\text{loc}}) = \operatorname{rank}(\hat{\mathbf{P}}_{\text{lg}}) + \operatorname{rank}(\hat{\mathbf{P}}_{\text{sm}}). \tag{30}$$

Since the rank of $\hat{\mathbf{P}}_{\text{sm}}$ is high, the rank of the multiscale covariance is also high, as is necessary because it covers all scales. Indeed, $\hat{\mathbf{P}}_{\text{sm}}$ is full rank when the small-scale localisation matrix $\mathbf{L}_{\text{sm}}$ in (27) is full rank, following the arguments in the previous subsection.

A flowchart summarising ESL is shown in Fig. 1. The top of the chart illustrates how the eigenvectors are computed from a smoothed ensemble, and then the calculation splits into two threads to compute large-scale and small-scale covariance matrices. The two covariance matrices are then added to give the multiscale covariance matrix of ESL.

Figure 2 gives an example of the covariance matrices at various stages in ESL in the cases of spatially homogeneous (top row) and heterogeneous (bottom row) multiscale covariances, which will be specified in Section 4. The true covariance matrices $\mathbf{P}^t$ in each case are shown on the left (column (a)). Proceeding to the right, the raw sample covariances (column (b)) exhibit considerable sampling noise. The large-scale covariance $\hat{\mathbf{P}}_{\text{lg}}$ (column (c)) captures large-scale aspects but retains some sampling error (note the variation along the diagonal even in the homogeneous case). The small-scale covariance $\hat{\mathbf{P}}_{\text{sm}}$ (column (d)), is the result of the processes described in Section 3.2. The sum, $\hat{\mathbf{P}}_{\text{lg}} + \hat{\mathbf{P}}_{\text{sm}}$ (column (e)), greatly improves over the raw sample covariance in estimating $\mathbf{P}^t$.

3.4. Computational considerations and connections to waveband localisation

As outlined in Section 2.1, ESL is intended for use when the dimensions N_x and N_y of the state and observation spaces are large, at least of the order of millions. This

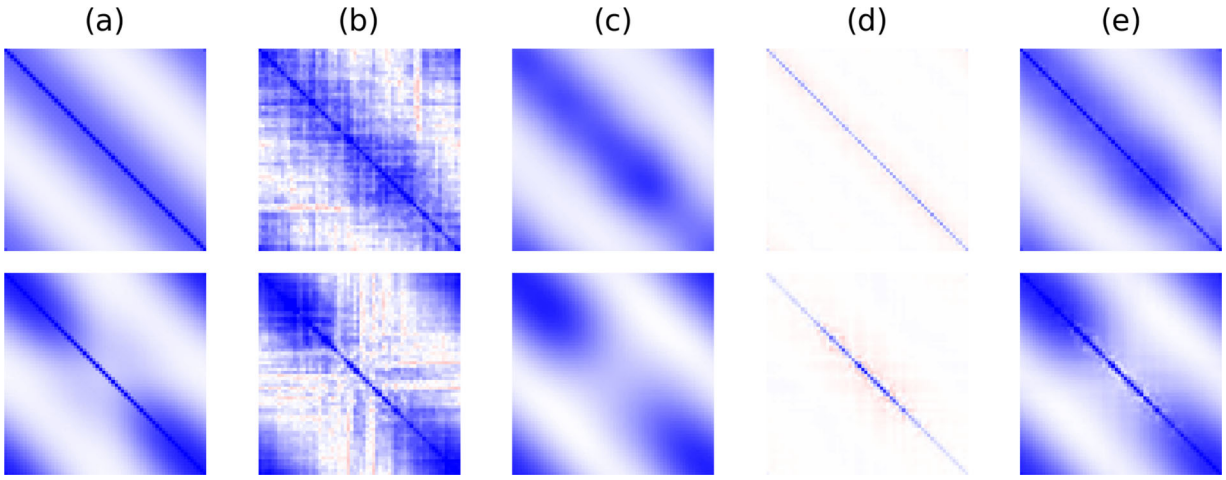

Fig. 2. Illustration of the homogeneous (top row) and heterogeneous (bottom row) systems (see also the numerical illustrations in Section 4). The colour indicates the value of the elements of a covariance matrix after being clipped between -1 and 1, where blue shades are positive and red shades are negative. Column (a): true covariance matrix $\mathbf{P}^t$. Column (b): sample covariance matrix $\hat{\mathbf{P}}$ of an ensemble of size 40. Column (c): large-scale covariance matrix $\hat{\mathbf{P}}_{\text{lg}}$ (Section 3.1). Column (d): small-scale covariance matrix $\hat{\mathbf{P}}_{\text{sm}}$ (Section 3.2). Column (e): multiscale covariance matrix $\hat{\mathbf{P}}_{\text{loc}}$.

immediately means that a practical implementation cannot construct covariance matrices such as $\hat{\mathbf{P}}_{\text{loc}}$ explicitly, but must represent their action on vectors implicitly, using sequences of operations that are each computationally feasible. Here we sketch how ESL would work computationally for large N_x and N_y, without attempting to detail an optimised implementation.

First note that, in practice, ESL would be used for localisation in an EnVar scheme or in "variational" implementations of the stochastic EnKF (Bowler et al., 2017; Buehner et al., 2017), rather than using the simple stochastic EnKF outlined in Section 2.2. EnVar or variational stochastic EnKF schemes compute analysis increments by minimising a cost fuction rather than by solving (5) and the minimisation involves multiplication of vectors by the background covariance matrix, which for ESL would be $\hat{\mathbf{P}}_{\text{loc}}$ from (30).

To calculate the action of $\hat{\mathbf{P}}_{\text{loc}}$ on a vector when N_x is large, there are three parts of ESL to consider: computing leading eigenvectors of covariance matrices, projecting arbitrary vectors onto the subspace of leading eigenvectors or its complement [e.g. $\mathbf{\Pi}_\perp$ defined by (25)], and computing the action of localised sample covariance matrices such as $\tilde{\hat{\mathbf{P}}}_{\text{sm}}$ from (27) on arbitrary vectors. Leading eigenvectors of a large covariance matrix can be computed by several techniques that require only the ability to multiply vectors by the matrix. Next, the necessary projections involve only dot products with the eigenvectors that are retained and thus are feasible as long as the number of eigenvectors retained is not too large. Finally, the action of localised sample covariance matrices on vectors can be affordably computed by adapting the approach of Lorenc (2003) and writing, e.g.

$$\tilde{\hat{\mathbf{P}}}_{\text{sm}}\mathbf{x} = \mathbf{L}_{\text{sm}} \circ \left(\mathbf{\Pi}_\perp(\mathbf{X}\mathbf{X}^{\mathrm{T}})\mathbf{\Pi}_\perp\right)\mathbf{x} = \sum_{j=1}^{N_e} \mathbf{v}_j \circ \left(\mathbf{L}_{\text{sm}}(\mathbf{v}_j \circ \mathbf{x})\right),$$

where $\mathbf{v}_j = \mathbf{\Pi}_\perp(\mathbf{x}_j - \bar{\mathbf{x}})$. Each of the terms following the final equality can be easily computed, as long as applying the localisation matrix is affordable.

We do not use the scalable implementation of the EnKF or ESL in the numerical illustrations below (see Sections 4 and 5), because the problems are simple enough (low-dimensional) to construct all required matrices directly. Nonetheless, the dimension of the problem is high enough to require localisation and all problems are multi-scale, in the sense that a single scale technique leads to unsatisfactory results. We view the numerical illustrations as a "proof of concept" with application of ESL to more realistic or real scenarios postponed to future work.

The numerical experiments do indicate that tuning of ESL (or other multiscale localisation techniques) is more cumbersome than tuning localisation with a single scale. Localising with a single scale requires that one tunes *one* length scale, which is relatively straightforward. ESL requires that we determine how many scales are present in the system, how many eigenvectors should be kept to represent each scale, and the amount of smoothing applied at each scale.

ESL and waveband localisation are similar in that they both "separate" ensemble perturbations into contributions from different scales. The details of that separation, however, differ between the two techniques. In waveband

localisation, the contributions at each scale are defined by applying bandpass filters centred at different wavenumbers to each ensemble perturbation, without requiring orthogonality between the results. ESL, in contrast, first computes the leading eigenvectors of the localised sample covariance matrix, using a large length scale for localisation. The contribution at each scale from a given ensemble perturbation is then defined by projecting onto the subspace of leading eigenvectors and its orthogonal complement. ESL as implemented in the experiments that follow employs spatial smoothing of the ensemble members before the initial step of computing leading eigenvectors. Such smoothing is a low-pass filter, like that used to define large-scale contributions to the covariance with waveband localisation. In ESL, however, the smoothing does not define the large scales but rather is simply a practical technique to improve the accuracy of estimates of the leading eigenvectors, under the assumption that the leading eigenvectors are largely independent of smaller scales.

3.5. Extension to multiple scales

We have focussed on two scales (small and large), but ESL can be extended to multiple scales. We now describe this process as a recursion over N_s scales. Starting with the largest scales, we initialise the process as in Section 3.1, which gives the localised covariance $\hat{\mathbf{P}}_1$ (indexing the covariances by numbers, rather than by "lg" and "sm" as before).

To calculate the localised covariance matrix for the jth scale, $j = 2, ..., N_s$, we begin by calculating

$$\tilde{\hat{\mathbf{P}}}_j = \mathbf{L}_j \circ \left(\mathbf{\Pi}_{\perp,j}\left(\mathbf{X}_j\mathbf{X}_j^{\mathrm{T}}\right)\mathbf{\Pi}_{\perp,j}\right), \tag{31}$$

where $\mathbf{X}_j$ are normalised ensemble perturbations, smoothed by an appropriate amount for the jth scale. To keep eigenvectors orthogonal, we apply the projection again and obtain

$$\hat{\mathbf{P}}_j = \mathbf{\Pi}_{\perp,j}\tilde{\hat{\mathbf{P}}}_j\mathbf{\Pi}_{\perp,j}. \tag{32}$$

The required orthogonal projection matrices are defined recursively by

$$\mathbf{\Pi}_{\perp,j} = (\mathbf{I}-\mathbf{Q}_{j-1}\mathbf{Q}_{j-1}^{\mathrm{T}})\mathbf{\Pi}_{\perp,j-1}, \tag{33}$$

where the matrix $\mathbf{Q}_{j-1}$ contains the N_{j-1} leading eigenvectors of $\hat{\mathbf{P}}_{j-1}$ as its columns. The recursion is initialised by $\mathbf{\Pi}_{\perp,0} = \mathbf{I}$. The N_j leading eigenvectors $\mathbf{q}_i^j$ of $\hat{\mathbf{P}}_j$ define the N_j projections $\mathbf{\Pi}_i^j = \mathbf{q}_i^j(\mathbf{q}_i^j)^{\mathrm{T}}$ that are used to obtain the localised covariance:

$$\hat{\mathbf{P}}_j^{\mathrm{loc}} = \sum_{i=1}^{N_j}\left(\mathbf{\Pi}_i^j\mathbf{X}\right)\left(\mathbf{\Pi}_i^j\mathbf{X}\right)^{\mathrm{T}}, \quad j = 1, ..., N_s-1. \tag{34}$$

For the smallest scale, we have $\hat{\mathbf{P}}_{N_s}^{\mathrm{loc}} = \hat{\mathbf{P}}_{N_s}$, in analogy to how we compute the small-scale covariance in the two-scale scenario. The multiscale covariance is defined, as before, as the sum of the covariances for the various scales:

$$\hat{\mathbf{P}}_{\mathrm{loc}} = \sum_{j=1}^{N_s}\hat{\mathbf{P}}_j^{\mathrm{loc}}. \tag{35}$$

This multiscale covariance matrix is the subsequently used within the EnKF.

Finally, we note that it is not fully understood if it is indeed important in ESL to be careful about orthogonality of the subspaces (e.g. Equations (28) and (32)). It might be worthwhile to relax the orthogonality requirement to save computation time, but from a mathematical standpoint, ensuring orthogonality seems natural. For example, ensuring orthogonality allows us to make definite statements about the rank of the multiscale covariance (see (30)), but the rank of $\hat{\mathbf{P}}_{\mathrm{loc}}$ may still be large *without* insisting on orthogonality. We found in numerical experiments that keeping the subspaces orthogonal leads to slightly better results (smaller analysis/forecast errors), but the generality of this idea is not understood.

3.6. Construction of a localised large-scale covariance using a coarse ensemble

The smoothing step in ESL may cause errors if the system features abrupt changes in correlation structure. In this case, one may want to avoid smoothing. We present one way of constructing a large-scale covariance without a smoothing step. Instead of smoothing, we assume that the underlying dynamics are solved on a fine and a coarse grid. Computations on the coarse grid are less expensive, because if each dimension (spatial and temporal) is coarsened by a factor of 2 then the computational cost is reduced by a factor of approximately 2^4 (in a 3 D system). One can leverage the computational advantages of the coarse grid and generate a coarse grid ensemble whose ensemble size is large enough to make the smoothing step obsolete. Once the covariance on the coarse grid is constructed and interpolated appropriately onto the fine grid, the small-scale covariance is obtained by projection as before.

The two grids (coarse and fine) are as follows. On the fine grid, the state dimension is N_x and we are given an ensemble of size $N_e \ll N_x$ to form the $N_e \times N_x$ matrix $\mathbf{X}$ of normalised ensemble perturbations. On the coarse grid, the state dimension is $N_{x,c} \ll N_x$ and we are given a coarse ensemble of size $N_{e,c} \gg 1$. The normalised ensemble perturbations of the coarse ensemble are $\mathbf{X}_c$ (a $N_{e,c} \times N_{x,c}$ matrix). We use the coarse ensemble to construct a large-scale covariance on the coarse grid

$$\hat{\mathbf{P}}_{\mathrm{c}} = \mathbf{L}_{\mathrm{c}} \circ \left(\mathbf{X}_{\mathrm{c}}\mathbf{X}_{\mathrm{c}}^{\mathrm{T}}\right), \tag{36}$$

where $\mathbf{L}_c$ is a positive definite localisation matrix. We compute the leading N_{lg} eigenvalues λ_i^c and corresponding eigenvectors $\mathbf{q}_i^c$, $i = 1, ..., N_{\mathrm{lg}}$, of $\hat{\mathbf{P}}_{\mathrm{c}}$.

To move from the coarse ensemble to the fine ensemble, we interpolate the coarse grid eigenvectors onto the fine grid:

$$\hat{\mathbf{q}}_i = \mathrm{interp}(\mathbf{q}_{i,c}). \tag{37}$$

This interpolation implies that $\hat{\mathbf{q}}_i$ are no longer normalised (norm equal one) or orthogonal. We form a new set of N_{lg} orthogonal, unit length vectors $\mathbf{q}_i$, $i = 1, ..., N_{\mathrm{lg}}$, each of size N_x, by Gram-Schmidt (QR decomposition). The localised, large-scale covariance is now constructed in analogy to (24):

$$\hat{\mathbf{P}}_{\mathrm{lg}} = \sum_{i=1}^{N_{\mathrm{lg}}} s_i(\mathbf{q}_i\mathbf{q}_i^{\mathrm{T}}), \quad s_i = \frac{\lambda_i^c}{||\hat{\mathbf{q}}_i||^2}, \tag{38}$$

where, $||\hat{\mathbf{q}}_i|| = \sqrt{\mathbf{q}_i^{\mathrm{T}}\mathbf{q}_i}$ is the Euclidean norm. Note that the eigenvalues of the covariance on the coarse grid are scaled by the inverse of the length of the interpolated eigenvectors, which is needed to maintain the magnitude of the coarse ensemble covariance after interpolation. We also emphasise that $\hat{\mathbf{P}}_{\mathrm{lg}}$ is constructed solely from the coarse ensemble; the ensemble on the fine grid is never used.

With $\hat{\mathbf{P}}_{\mathrm{lg}}$ and $\mathbf{q}_i$, defined on the fine grid, the construction of the small-scale covariance is as in Section 3.2. Specifically, we use (27) and (28), with $\mathbf{X}$ being the ensemble on the fine grid, to construct the small-scale covariance. The multiscale covariance is the sum of the large and small-scale covariances (see Section 3.3).

A coarse-resolution model will generally be less accurate, if cheaper, than the original, fine-resolution model, and there is no guarantee that the resulting coarse ensemble will provide useful covariances. Our approach, however, relies on the coarse ensemble only for the large-scale covariances, where we have more confidence, though again no guarantee, that a reduced-resolution model can be useful. Supporting evidence for the atmosphere at least is the now routine use of reduced-resolution ensemble forecasts to provide covariances in operational atmospheric data assimilation (e.g. Clayton et al., 2013).

4. Numerical illustration 1: Gaussians with two-scale covariance matrices

We illustrate ESL on examples with a known, multiscale covariance, similar to the numerical experiments in Lorenc (2017). This set of numerical experiments does not feature a dynamical model or a cycling data assimilation.

Consider a system whose state is a periodic function of a single spatial coordinate. We assume the the elements of the Discretised state vector $\mathbf{x}$ correspond to $N_x = 64$ equally spaced points and that the true state $\mathbf{x}^t$ has a Gaussian prior distribution with mean $\mathbf{0}$ and forecast covariance matrix $\mathbf{P}^f, \mathbf{x}^t \sim \mathcal{N}(\mathbf{0}, \mathbf{P}^f)$. We suppose also that we are given realisations of observations related to the state by

$$\mathbf{y} = \mathbf{H}\mathbf{x}^t + \varepsilon, \quad \varepsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{R}). \tag{39}$$

In what follows, we choose $\mathbf{H}$ to select every 8th position of the state, so that we have $N_y = 8$ observations, and we take $\mathbf{R} = \mathbf{I}$. We have done a number of numerical experiments with different observation configurations (more observations, or fewer observations) and these experiments lead to qualitatively similar results as those discussed below.

We evaluate ESL (and other localisation approaches) in this simple context in two ways: the accuracy of its estimate of $\mathbf{P}^f$ from a finite ensemble, using the Frobenius norm (17), and the analysis RMSE (15) of the state update. In each realisation of the experiment, a true state $\mathbf{x}^t$ and an ensemble of size N_e are drawn with distribution $\mathcal{N}(\mathbf{0}, \mathbf{P}^f)$, and observations are generated from (39) with independent draws of ε. All results are averaged over 1000 realisations. The comparisons below include two other quantities computed from the experiments. The Kalman filter (using the correct mean and covariance) gives the minimum RMSE that can be achieved by any update. An EnKF that employs $\mathbf{P}^f$ in the gain for each member gives the lower bound for RMSE that can be achieved by ESL or other localizations, as even with perfect covariances the EnKF retains sampling error in the prior mean.

We consider single-scale localisation, waveband localisation and ESL. Each localisation method is tuned. For single-scale localisation, we tune the localisation length scale; for waveband localisation, we tune the spectral filter, and the large and small-scale localisation length; for ESL we tune the smoothing length scale and the large- and small-scale localisation length. We consider only two scales (small and large), this means that we use two wavebands (waveband localisation) and two sets of eigenvectors (ESL). Tuning over three scales (not shown) did not result in a significant difference in analysis RMSE. We tune inflation only on the single-scale localisation and use the resulting inflation parameter in the multi-scale methods.

We consider two multiscale covariances, one spatially homogeneous and the other heterogeneous. The corresponding covariance matrices $\mathbf{P}^f$ appear in column (a) of Figure 2. Explicit formulae for the covariances, along with results for ESL, are given in sections 4.1 and 4.2

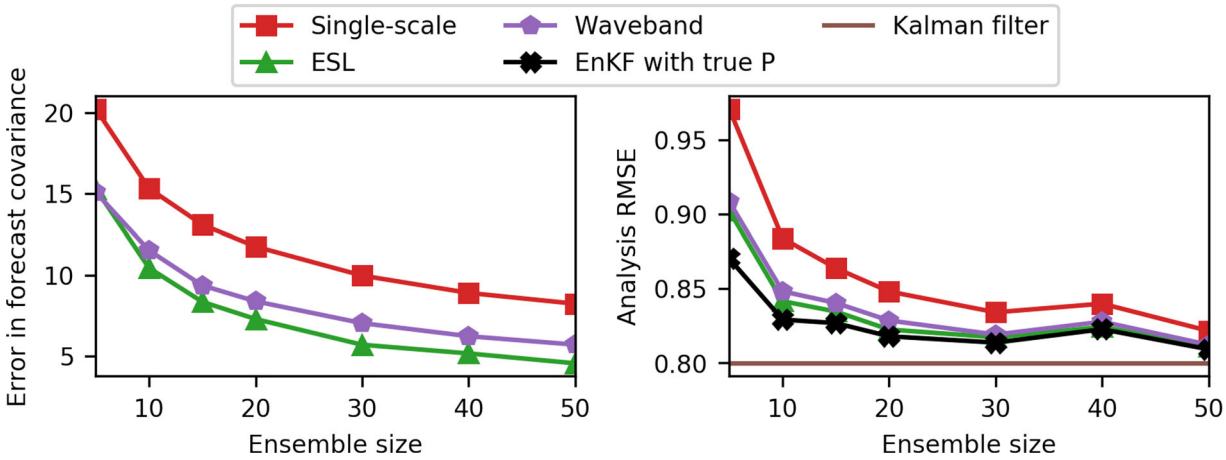

Fig. 3. Results for the case of an isotropic homogeneous forecast covariance. Left: Error in forecast covariance matrix (Frobenius norm) as a function of ensemble size. Right: Analysis RMSE as a function of ensemble size. Green triangles: ESL. Purple hexagons: waveband localisation. Red squares: single-scale localisation. Black x's (right panel): EnKF using true forecast covariance. Brown line (right panel): average analysis standard deviation of the Kalman filter.

below. In the remainder of this section, we examine ESL results using a coarse ensemble in the estimation of the large-scale covariance rather than spatial smoothing of the ensemble members (Section 3.6), and discuss and summarise our numerical findings (Section 4.4).

4.1. Isotropic homogeneous forecast covariance

We apply localisation (single-scale, ESL and waveband localisation) to a problem defined on a periodic domain, Discretised on 64 regularly spaced points. The forecast covariance is isotropic and homogeneous, and is given by the weighted sum:

$$\mathbf{P}^t = 0.6\mathbf{P}_1 + 0.4\mathbf{P}_2. \tag{40}$$

The elements of $\mathbf{P}_1$ and $\mathbf{P}_2$ are defined by

$$[\mathbf{P}_k]_{i,j} = \exp\left(-\left(\frac{z_{i,j}}{\sqrt{2}l_k}\right)^2\right), \quad i,j = 1,...,64, k = 1,2, \tag{41}$$

and where $z_{i,j}$ is the distance between $[\mathbf{x}]_i$ and $[\mathbf{x}]_j$ in the periodic domain. The length scales for $\mathbf{P}_1$ and $\mathbf{P}_2$ are $l_1 = 0.2$ and $l_2 = 0.01$ respectively and, thus, $\mathbf{P}_1$ describes large scales and $\mathbf{P}_2$ describes correlations across the small scales. We note that the variance structure is constant throughout the domain and that the eigenvectors of $\mathbf{P}_1$ and $\mathbf{P}_2$ and, therefore of $\mathbf{P}^t$, are Fourier modes, which naturally describe different "scales." This means that this example satisfies all assumptions required for ESL and we expect that the method will do well on this example.

Figure 3 shows the results of our numerical experiments. The left panel shows the Frobenius norm of the difference of the estimated forecast covariance matrices and the "true" forecast covariance, $||\mathbf{P}^{\text{est}} - \mathbf{P}^t||_F$, as a function of the ensemble size. We note that the multiscale techniques lead to smaller errors than single-scale localisation and that ESL leads to slightly smaller errors than waveband localisation. We also note that the rate with which the errors decay are comparable for all three methods. The right panel of Figure 3 shows the analysis RMSE as a function of ensemble size. Again, we note that the multiscale methods lead to smaller errors than the single-scale method and that ESL has a slight edge over waveband localisation. It is also important to point out that ESL leads to errors that are comparable to the errors of the EnKF, using the true forecast covariance (no sampling error in the forecast covariance). This means that ESL leads to errors that are nearly as small as possible with any localisation scheme.

Figure 4 shows the 32nd column of the localised forecast covariance for three localisation methods. The solid lines (often hidden) are the average variances and covariances over the 1000 trials and the shaded regions are defined by the average covariance plus or minus one standard deviation. The solid black line corresponds to the corresponding variance and covariances of the true forecast covariance (no sampling error). We note that single-scale localisation decreases the covariances too quickly at large distances. The multiscale methods describe the true covariances more accurately, and ESL leads to the most accurate estimates.

4.2. Heterogeneous forecast covariance

We apply localisation (single-scale, ESL and waveband localisation) to a problem with a heterogeneous forecast covariance given by

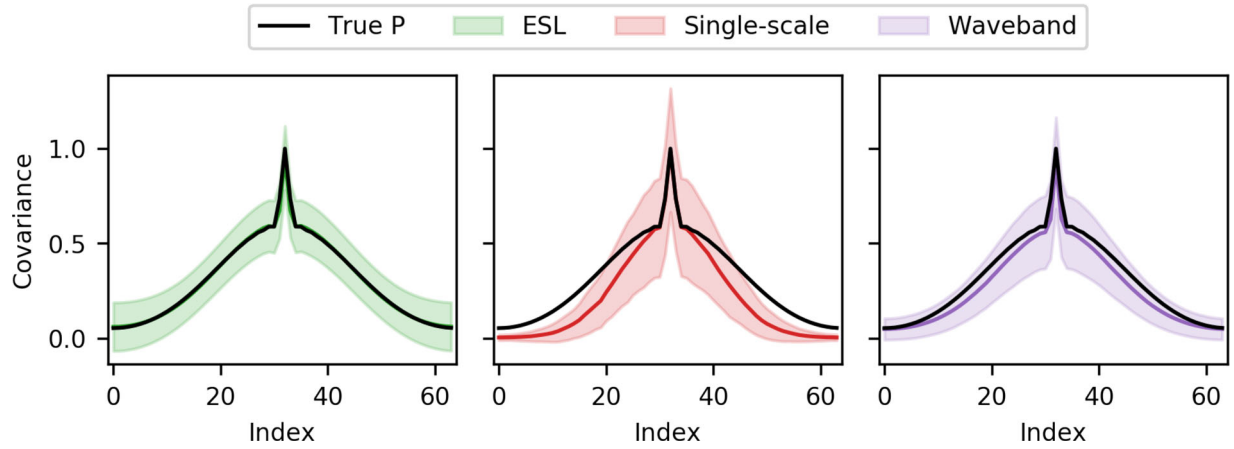

Fig. 4. Results for the case of an isotropic homogeneous forecast covariance. Shown is the covariance with index 32, which falls roughly in the centre of the domain. Left (green): ESL. Centre (red): single-scale localisation. Right (purple): Waveband localisation. The solid lines show the covariance averaged over 1000 trials, and the shaded region shows this average plus and minus one standard deviation. In all three panels, the true covariance is shown with a solid black line.

$$\mathbf{P}^f = \mathbf{D}_1\mathbf{P}_1\mathbf{D}_1 + \mathbf{D}_2\mathbf{P}_2\mathbf{D}_2, \tag{42}$$

where $\mathbf{P}_1$ and $\mathbf{P}_2$ are defined in the previous section, and where $\mathbf{D}_1$ and $\mathbf{D}_2$ are diagonal matrices with positive diagonal elements. The matrices $\mathbf{D}_1$ and $\mathbf{D}_2$ change the variance of $\mathbf{P}$ within the domain. To define $\mathbf{D}_1$ and $\mathbf{D}_2$, we first define a vector of length N_x with elements:

$$[\tilde{\mathbf{v}}]_i = 1 - \exp\left(-(i-31.5)^2/(2\cdot 96)\right). \tag{43}$$

We then rescale $\tilde{\mathbf{v}}$ so that its largest element is equal to 0.9 and its smallest element is equal to 0.2:

$$\mathbf{v} = \frac{\tilde{\mathbf{v}} - \min(\tilde{\mathbf{v}})}{\max(\tilde{\mathbf{v}}) - \min(\tilde{\mathbf{v}})} \cdot (0.9 - 0.2) + 0.2. \tag{44}$$

The resulting vector $\mathbf{v}$ is used to define $\mathbf{D}_1$ and $\mathbf{D}_2$:

$$[\mathbf{D}_1]_{i,i} = \sqrt{[\mathbf{v}]_i}, \quad [\mathbf{D}_2]_{i,i} = \sqrt{1 - [\mathbf{v}]_i}. \tag{45}$$

Since the variance structure changes within the domain, the smoothing step in ESL may artificially increase weak long distance correlations, because locations with stronger long distance correlations are averaged (smoothed) with locations with weak long distance correlations.

Figure 5 summarises the results of our numerical experiments. The left panel shows the Frobenius norm of the difference of the estimated covariance and the "true" covariance matrices as a function of the ensemble size. The right panel of Figure 5 shows the analysis RMSE as a function of ensemble size. As in the previous example, we note that the multiscale methods lead to smaller errors than the single-scale localisation, but the differences in errors between the two multiscale methods are even less pronounced than in the previous example. This is not so surprising because the eigenvectors of the forecast covariance matrix of this example do not separate large and small scales from one another, so that some of the assumptions made in ESL are not met. This renders the technique less effective than in the first example, where essentially all required assumptions are met.

4.3. Heterogeneous covariance – Large-scale covariance estimated from a coarse ensemble

We apply ESL, but using a coarse ensemble to estimate large-scale covariances, to the same setup and heterogeneous forecast covariance matrix as in Section 4.2. The coarse ensemble members have a dimension of 16 which is derived from the fine ensemble of dimension 64 by retaining every 4th position. Quadratic interpolation is used to convert the coarse eigenvectors to the fine grid. The ensemble size of the coarse ensemble is set to 200.

Figure 6 shows the covariance of grid cells distanced five (fine-scale) grid points apart (ensemble size on the fine grid is $N_e = 15$). The left panel, in green, shows the covariance obtained by ESL with smoothing. The right panel, in pink, shows the covariance obtained by ESL using the coarse ensemble. The solid lines show the covariance averaged over the 1000 trials and the shaded regions indicate the region within one standard deviation of this average. We note that the coarse ensemble leads to more accurate estimates of the large-scale covariance. Even more significantly, the coarse ensemble reduces the standard deviation, because a larger ensemble size is used in the estimation of large-scale covariances.

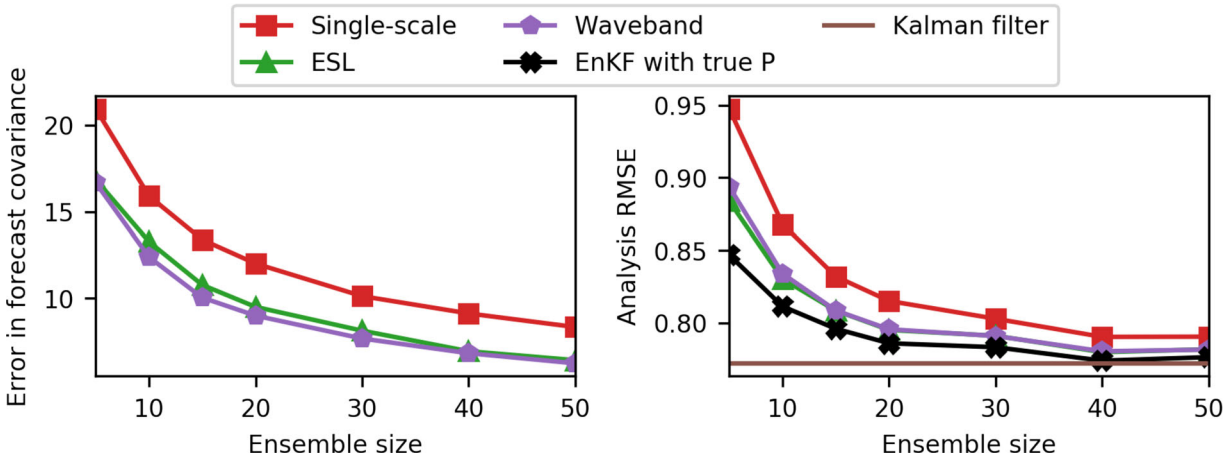

Fig. 5. Results for the case of a heterogeneous forecast covariance. Left: Error in forecast covariance matrix (Frobenius norm) as a function of ensemble size. Right: Analysis RMSE as a function of ensemble size. Green triangles: ESL. Purple hexagons: waveband localisation. Red squares: single-scale localisation. Black x's (right panel): EnKF using true forecast covariance. Brown line (right panel): average analysis standard deviation of the Kalman filter.

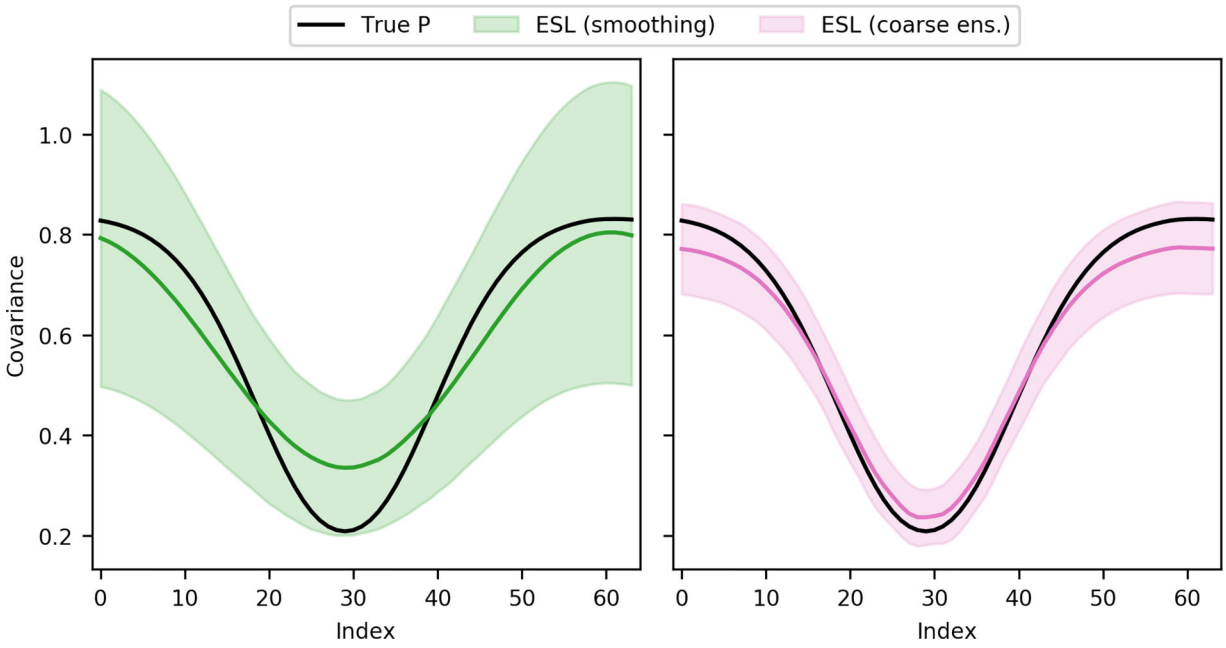

Fig. 6. Comparison of ESL using smoothing (green, left) and a coarse ensemble (pink, right). Shown is the large-scale covariance of grid cells distanced five (fine-scale) grid points apart. The shaded areas represent the average plus and minus one standard deviation. Black: true covariance.

Table 1 shows the analysis RMSE and the error in the covariance matrices localised by ESL, using smoothing or a coarse ensemble, as a function of the ensemble size on the fine grid (the ensemble size of the coarse ensemble is fixed to 200). We note that using the coarse ensemble leads to smaller errors in the covariance matrix and analysis RMSE, independently of the ensemble size. The reduction in analysis RMSE, however, is minor. This can be explained by considering errors in the Kalman gain, which is a *nonlinear* function of the forecast covariance and the observation network. The error in the Kalman gain is also shown in Table 1 and we note that the errors in the Kalman gain are comparable for the two variants of ESL. Since the analysis RMSE is determined primarily by errors in the Kalman gain, rather than the forecast covariance, this explains why the two techniques show

Table 1. Errors in forecast covariance and analysis RMSE for ESL using smoothing or a coarse ensemble.

N_e	5	10	15	20	30	40	50
				Error in forecast covariance			
Smoothing	19.5	14.0	11.8	10.7	8.57	7.33	6.92
Coarse Ens.	6.16	5.83	5.29	5.15	5.14	5.57	5.05
				Error in Kalman gain			
Smoothing	1.19	0.886	0.776	0.704	0.605	0.548	0.500
Coarse Ens.	0.685	0.586	0.552	0.527	0.507	0.486	0.480
				RMSE			
Smoothing	0.883	0.826	0.810	0.797	0.791	0.795	0.782
Coarse Ens.	0.858	0.814	0.802	0.792	0.789	0.793	0.780

significant differences in the errors in the covariance matrices, but nearly identical analysis RMSE.

4.4. Discussion

In the above experiments multiscale localisation results in more accurate estimates of forecast covariance and, hence, smaller analysis RMSEs than single-scale localisation. This is especially apparent at lower ensemble sizes where localisation has the biggest impact. This result, however, is not surprising, because the problem is set up to be multiscale. ESL performs similarly, though marginally better, than waveband localisation in terms of analysis RMSE. The fact that ESL and waveband localisation perform similarly is not surprising since both methods are constructed using similar ideas (see Section 3.4).

Because localisation is particularly important when the ensemble size is small (in these examples, 15 or fewer ensemble members), we discuss ESL in more detail for such small ensembles. In this case, our tuning of ESL results in a large smoothing length scale and a tight localisation, when calculating the large-scale covariance matrix. The reason is that a small ensemble contains only a limited amount of information and may not fully resolve all relevant details (the aspects of the dynamics that are relevant to the large-scale dynamics). A large amount of smoothing results in a covariance matrix that approaches a constant (in the limit of infinite smoothing). After localisation, we are then essentially calculating the leading eigenvectors of the localisation matrix which, in our examples, returns the leading Fourier modes. This means that for a small ensemble, ESL is similar to waveband localisation with no large-scale localisation and no cross scale covariance. In this context, we recall that ESL does not apply spatial localisation directly (to the large scales). The localisation is done indirectly via projection of the ensemble onto the eigenvectors of a smoothed and localised sample covariance (see Section 3.4).

Table 2. Parameters of the LM3 model.

Parameter	N_z	K	I	F	b	c
This work	960	32	12	14	1	0.37
Lorenz (2005)	960	32	12	15	1	2.5

ESL with large-scale covariances computed from a coarse ensemble avoids the smoothing step. The simple heterogenous example illustrates this point, and we find that the coarse ensemble leads to more accurate covariance estimates than the smoothing technique. This reduction in the error of the forecast covariances, however, does not necessarily result in a significant reduction in the analysis RMSE. The reduction in analysis RMSE depends mostly on errors in the Kalman gain, which is a nonlinear function of the forecast covariance and the observation network, and in the example the coarse ensemble leads to comparable analysis RMSEs as the smoothing technique. We also note that the coarse ensemble size used here is very large and even larger than the dimension of the coarsened state ($N_{e,c} = 200$, $N_{x,c} = 64$). This may not be feasible in realistic scenarios, which could imply that the coarse ensemble techniques may have little to no advantage over the smoothing method.

5. Illustration 2: Cycling DA with Lorenz model III

We apply single- and multiscale localisation methods to the third model of Lorenz (2005) and compare the results of each method in terms of forecast RMSEs and spread. In our numerical experiments, we keep the observation network fixed (we do not vary the frequencies of observations in space or time). We have, however, performed several numerical experiments with varying observation networks and these experiments all led to results that are qualitatively similar to the results we show below.

5.1. Lorenz model III

The Lorenz model III (LM3, Lorenz, 2005) is the differential equation

$$dZ_n/dt = [X,X]_{K,n} + b^2[Y,Y]_{1,n} + c[Y,X]_{1,n} - X_n - bY_n + F, \tag{46}$$

where $n = 1,...,N_z$ is an integer, N_z is the state dimension, and K, b, c, and F are parameters. The square brackets are defined as,

$$[X,Y]_{K,n} = \sum_{j=-J}^{J}{}' \sum_{i==-J}^{J}{}' (X_{n-2K-i}Y_{n-K-j} + X_{n-K+j-i}Y_{n+K+j})/K^2, \tag{47}$$

where,

$$J = \begin{cases} K/2, & \text{if } K \text{ is even} \\ (K-1)/2, & \text{if } K \text{ is odd} \end{cases}. \tag{48}$$

Here, Σ' is the usual sum if K is odd, but if K is even, the first and last summands are weighted by 1/2. The variables X_n and Y_n are defined as,

$$X_n = \sum_{i=-I}^{I}{}' (\alpha - \beta|i|)Z_{n+i}, \quad Y_n = Z_n - X_n, \tag{49}$$

where,

$$\alpha = (3I^2+3)/(2I^3+4I), \quad \beta = (2I^2+1)/(I^4+2I^2), \tag{50}$$

and where I is a parameter. Note that the intermediate variable X is a smoothed version of Z, in which small scales are averaged out and that the variable Y contains the small scales. The various parameters have physical interpretations which we summarise briefly: K determines the number of waves in X; b defines the amplitude and speed of fluctuations in Y; c affects the coupling between X and Y; and F is a forcing term that determines the chaotic behaviour of the system; I is important for the smoothness of X, with larger values leading to a smoother X.

We adjust the parameter values of Lorenz (2005) so that the resulting system is characterised by forecast covariance matrices with a more pronounced multiscale structure. We achieve this by increasing the amplitude of Y, while simultaneously decreasing its time scale. The parameter values we use, and those of Lorenz (2005) are listed in Table 2. The error growth of LM3 with our choice of parameter values and with the parameter values of Lorenz (2005) is illustrated in Figure 7. The errors are averaged over 100 different realisations of error growth derived from 10 different initial conditions each with 10 different random perturbations drawn from a normal distribution with identity correlation and variance equal to 10^{-10}. We note that errors grow exponentially in both cases (both LM3 models are chaotic), but with our choices of parameter values, errors in Y grow more slowly and saturate at a higher level than when using the original LM3 parameter values. A typical state, in terms of Z, X and Y of LM3 with our adjusted parameters is shown in Figure 8. We note that the small scales (Y) and the large scales (X) are comparable in size. This is not the case with the classical parameter values reported in Lorenz (2005).

We emphasise that we make adjustments to the parameters of LM3 to force the model to generate more pronounced multi scale characteristics. We ran all numerical experiments shown below also for LM3 with "classical" parameters and found that multi scale localisation is not required, because a single scale localisation results in nearly the same forecast errors. The reason is that, with classical parameters, the large scale is dominant (much larger in magnitude than the small scale), so that any technique that captures the large scale leads to satisfactory results. With our modified parameters, the large and small scales are more comparable in size so that multi scale localisation leads to significantly smaller forecast errors than single scale localisation (see below). Our adjustments are thus justified in order to illustrate how ESL (and other multi scale localisation) may perform on a multiscale system, in which both scales are comparable and significant.

5.2. Twin experiment setup

We follow Rainwater and Hunt (2013) in the design of twin experiments with LM3. To generate a (random) initial condition for Z, we first generate an initial $\tilde{Y}_0$ by drawing N_z values from a uniform distribution on $[-6,6]$. To obtain an initial $\tilde{X}_0$, we draw N_z/K numbers from a uniform distribution on $[-9,14]$ and then interpolate onto N_z points using a quadratic spline. We then set $\tilde{Z}_0 = \tilde{X}_0 + \tilde{X}_0/\tilde{X}_{\max}\tilde{Y}$ where $\tilde{X}_{\max}$ is the largest element (absolute value) of $\tilde{X}_0$. We then evolve $\tilde{Z}_0$ for 24 time units to obtain an initial condition Z_0. As suggested in Rainwater and Hunt (2013), we use a fourth order Runge-Kutta scheme with a time step $\Delta t = 0.05/24$ to numerically integrate LM3.

We generate a ground truth by integrating a random initial condition Z_0 for 125 time units. The first 25 time units are used for spin up, and the subsequent 100 time units are used to calculate errors. The initial ensemble is computed as follows. We evolve a random initial Z_0 for 4000 time units, and save a series of states, each 0.1 time units apart. Ensemble members are randomly selected from this collection of states. The ensemble size is $N_e = 20$ in all experiments.

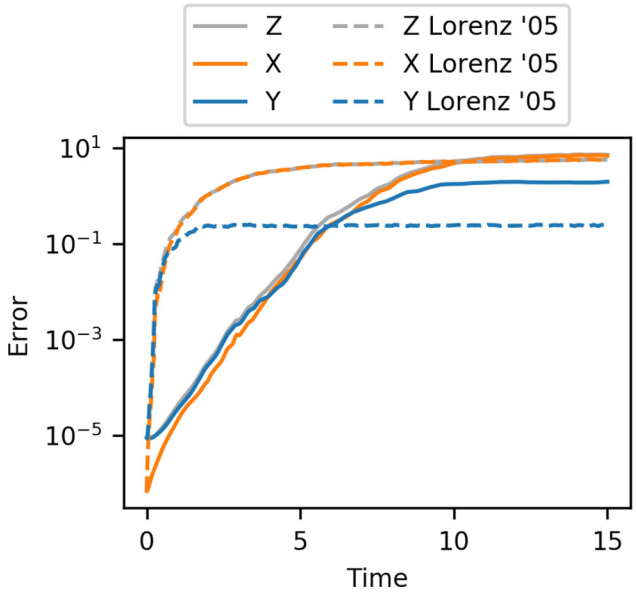

Fig. 7. Error growth in Z (multiscale, grey), X (large scale, orange) and Y (small scale, blue). Solid lines correspond to LM3 with adjusted parameters. Dashed lines correspond to LM3 with parameters as in Lorenz (2005).

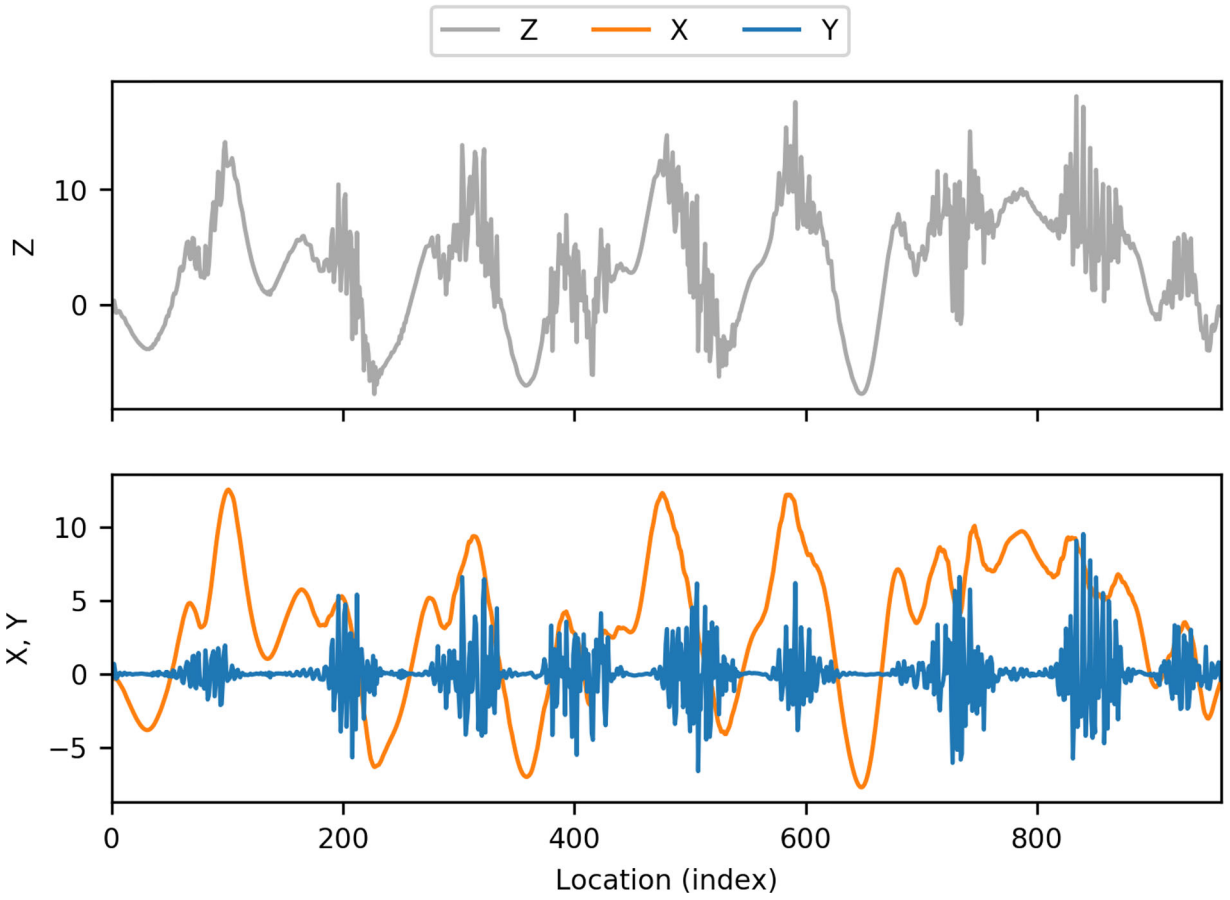

Fig. 8. An example of a state of LM3 (adjusted parameters). Top (blue): multiscale variable Z. Bottom: large-scale variable X (orange) and small-scale variable Y (blue).

Table 3. Time averages of forecast RMSE and spread (tuned) for three localisation techniques.

	RMSE	Spread
Single-scale	1.00	0.47
Waveband	0.64	0.63
ESL	0.61	0.50

Observations of every 4th component of Z are assimilated every 0.05 time units, which is approximately equal to 1/10 of the e-folding time. The observation errors are independent with an observation error variance equal to one.

5.3. Tuning of localisation and inflation

To tune the localisation methods, we run DA experiments for a range parameters that define the various localisation methods. We repeat this procedure for a range of inflation parameters and call the localisation/inflation that leads to the smallest forecast RMSE "optimal."

Specifically, for the single-scale localisation, we vary the localisation length (from 5 to 40 with a step size of 5), and the inflation values (from $\alpha = 0$ to $\alpha = 0.3$ with a step size of 0.05). For ESL, we consider only three inflation parameters ($\alpha = 0$, i.e. no inflation), $\alpha = 0.05$, and $\alpha = 0.1$). For each inflation, we vary the small-scale localisation lengths (1, 2, 4, 6, and 8), the large-scale localisation length (30, 45, and 60), and the smoothing distances (8, 10, 12, and 14). The parameters that lead to the smallest forecast RMSE are the optimal values. In all experiments, we pick $N_{lg} = 40$. This is motivated by the fact that $N_z/K = 30$, which approximates the dimension of the single-scale Model I that the X variable is meant to represent (Lorenz, 2005). We use $N_{lg} = 40$, rather than $N_{lg} = 30$, to account for possibly underestimating an appropriate large-scale dimension. We do not use a coarse ensemble to estimate large-scale the covariance because LM3 has no natural mechanism to coarsen the resolution. Tuning waveband localisation is analogous to the tuning of ESL, but the smoothing parameters are replaced by spectral filter widths, with wave numbers of 25 through 150 with a step size of 25 (see Section 2.4).

5.4. Results and discussion

Table 3 lists the (tuned) forecast RMSE and spread (for Z) of the three localisation methods. The multiscale methods lead to smaller RMSEs than than the single scale method, which, again, is not surprising because this problem is characterised by two scales. We also note that ESL leads to the smallest RMSE and largest spread, but this may be due to insufficient tuning (see below).

Forecast RMSE and spread, as a function of time, are shown in Figure 9, for single-scale localisation, ESL and waveband localisation. The figure reiterates that, on average, multiscale localisation leads to smaller errors than single-scale localisation. While RMSE occasionally grows and can be much larger than the spread, the growth in RMSE associated with the multiscale methods is less severe than in the case of single-scale localisation. We also note that single-scale localisation leads to an RMSE considerably larger than the spread. This can be remedied by increasing the inflation parameter, but a higher inflation subsequently causes even larger RMSE. Forecast RMSE and spread of the multiscale methods are comparable on average, but occasionally RMSE grows larger than the spread.

While we note that ESL leads to smaller forecast RMSE than waveband localisation, a more extensive tuning may make the errors more comparable and we do not claim that ESL outperforms waveband localisation in this example. Rather, we are satisfied with demonstrating that ESL may be competitive with waveband localisation in a non-trivial multiscale and cycling DA system.

6. Summary and conclusions

We described a new multiscale localisation method based on the estimation of eigenvectors and subsequent projections, together with traditional spatial localisation applied with a range of localisation lengths. We call this method eigenvector-spatial localisation (ESL). The basic idea is to estimate eigenvectors of the covariance matrices pertaining to different scales and then project the ensemble onto these eigenvectors. Specifically, the eigenvectors associated with large-scale covariances are estimated from the ensemble by applying smoothing and localisation with a large length scale, while eigenvectors associated with small-scale covariance are estimated by applying little or no smoothing and a smaller localisation parameter. As an alternative to using smoothing for the estimation of eigenvectors associated with large-scale covariances, one can also make use of a coarse ensemble, and we described this process in detail. We focussed in our presentation on two scales, but the methodology naturally extends to more than two scales.

We illustrated the use of ESL in two sets of numerical experiments, and multiscale localisation (ESL or waveband) leads to a reduction in errors when compared to single-scale localisation in all examples we considered. This is perhaps not surprising, since all numerical examples are set up to have multiscale characteristics, but it also suggests that multiscale localisation may bring about significant error reductions in realistic or real scenarios, which are multiscale. In the first set of experiments, we consider Gaussian forecast distributions, in which the "true" forecast covariances are known. The second set of

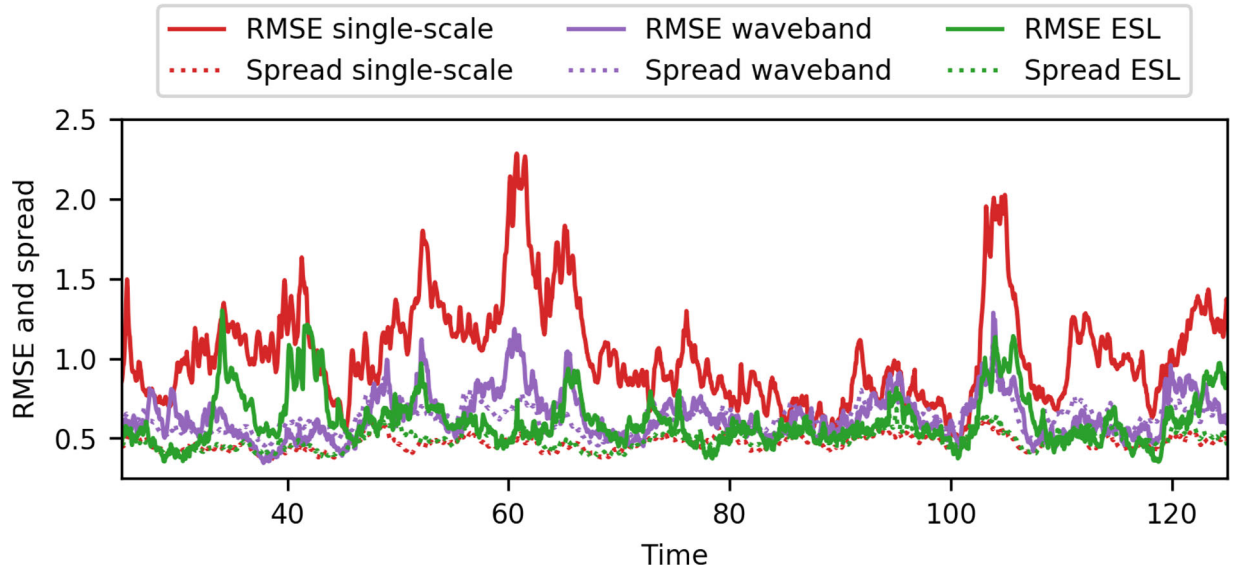

Fig. 9. RMSE and spread in the multiscale variable Z of LM3 (after a 25 cycle spin up period). Solid lines: RMSE. Dashed lines: Spread. Red: single-scale localisation. Purple: waveband localisation. Green: ESL.

experiments consists of cycling DA experiments with the Lorenz Model III (Lorenz, 2005). The numerical examples illustrate the use of ESL, but they are too simple to fully understand the advantages of using a coarse ensemble over smoothing. The Lorenz Model III has no natural way of coarsening the resolution, so that we only implement ESL using smoothing. The numerical experiments with the Gaussian forecast distribution are low dimensional ($N_x = 64$) and in particular the ensemble size of the coarse ensemble is larger than the coarse state dimension. In practice, this may not be the case and more numerical experiments are needed to investigate and understand the usefulness of this variant of ESL.

We compare ESL to waveband localisation (following Buehner, 2012; Buehner and Shlyaeva, 2015). The two methods rely on smoothing (waveband filtering) the ensemble, but *how* the smoothed ensemble is used for localisation is rather different. ESL uses the smoothed ensemble to compute orthogonal subspaces onto which the ensemble is projected and the (large scale) localisation is implicitly done via these projections. Waveband localisation uses smoothing (waveband filtering) more directly and computes (cross-)covariances directly from the filtered ensemble members. In all numerical examples, we found that ESL leads to similar or slightly smaller errors than waveband localisation. We emphasise that multiscale localisation requires a significant amount of tuning, and it is possible that our tuning is insufficient to rigorously conclude that ESL is more appropriate than waveband localisation, even in the limited set of numerical examples we consider. Our numerical examples, however, do suggest that ESL can be competitive with waveband localisation, in the sense that a modest amount of tuning leads to similar results for both methods. An advantage of ESL over waveband localisation may be its flexibility. ESL can be implemented in various ways and we outlined two (smoothing and coarse ensemble). One can easily consider other variants, e.g. using customised filters, based on characteristics of a specific problem.

We do not address the broader question of the conditions under which multiscale localisation techniques can be expected to be effective. By construction, ESL is applicable to systems that are isotropic homogeneous, as in numerical illustration 1 (possibly extending to heterogeneous problems when a coarse ensemble can be used, as in numerical illustration 2). Few problems, if any, of practical relevance are isotropic homogeneous. The numerical experiments with Lorenz model III indicate that ESL (and waveband localisation) can be effective on a broader class of problems, but the precise conditions under which ESL (or waveband localisation) should be used remain poorly understood.

Disclosure statement

No potential conflict of interest was reported by the authors.

Author Snyder was supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under cooperative agreement 1852977. Author Harty was funded by Achievement Rewards for College Scientists Foundation and was supported by NCAR's Advanced Study Program during a collaborative visit to NCAR.

References

Anderson, J. and Lei, L. 2013. Empirical localization of observation impact in ensemble Kalman filters. *Mon. Wea. Rev.* **141**, 4140–4153. doi:10.1175/MWR-D-12-00330.1

Anderson, J. L. and Anderson, S. L. 1999. A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.* **127**, 2741–2758. doi:10.1175/1520-0493(1999)127<2741:AMCIOT>2.0.CO;2

Bishop, C. and Hodyss, D. 2007. Flow-adaptive moderation of spurious ensemble correlations and its use in ensemble-based data assimilation. *Q. J. R. Meteorol. Soc.* **133**, 2029–2044. doi:10.1002/qj.169

Bishop, C. and Hodyss, D. 2011. Adaptive ensemble covariance localization in ensemble 4D-VAR state estimation. *Mon. Wea. Rev.* **139**, 1241–1255. doi:10.1175/2010MWR3403.1

Bowler, N. E., Clayton, A. M., Jardak, M., Lee, E., Lorenc, A. C. and co-authors. 2017. Inflation and localization tests in the development of an ensemble of 4d-ensemble variational assimilations. *Q. J. R. Meteorol. Soc.* **143**, 1280–1302. doi:10.1002/qj.3004

Buehner, M. 2012. Evaluation of a spatial/spectral covariance localization approach for atmospheric data assimilation. *Mon. Wea. Rev.* **140**, 617–636. doi:10.1175/MWR-D-10-05052.1

Buehner, M., McTaggart-Cowan, R. and Heilliette, S. 2017. An ensemble Kalman filter for numerical weather prediction based on variational data assimilation: Varenkf. *Mon. Wea. Rev.* **145**, 617–635. doi:10.1175/MWR-D-16-0106.1

Buehner, M. and Shlyaeva, A. 2015. Scale-dependent background-error covariance localisation. *Tellus A: Dyn. Meteorol. Oceanogr.* **67**, 28027. doi:10.3402/tellusa.v67.28027

Burgers, G., van Leeuwen, P. J. and Evensen, G. 1998. Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.* **126**, 1719–1724. doi:10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2

Clayton, A. M., Lorenc, A. C. and Barker, D. M. 2013. Operational implementation of a hybrid ensemble/4d-var global data assimilation system at the met office. *Q. J. R. Meteorol. Soc.* **139**, 1445–1461. doi:10.1002/qj.2054

De La Chevrotière, M. and Harlim, J. 2017. A data-driven method for improving the correlation estimation in serial ensemble Kalman filters. *Mon. Wea. Rev.* **145**, 985–1001. doi:10.1175/MWR-D-16-0109.1

Evensen, G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99**, 10143–10162. doi:10.1029/94JC00572

Flowerdew, J. 2015. Towards a theory of optimal localisation. *Tellus A: Dyn. Meteorol. Oceanogr.* **67**, 25257. doi:10.3402/tellusa.v67.25257

Gaspari, G. and Cohn, S. E. 1999. Construction of correlation functions in two and three dimensions. *Q. J. R. Meteorol. Soc.* **125**, 723–757. doi:10.1002/qj.49712555417

Hamill, T. M., Whitaker, J. S. and Snyder, C. 2001. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.* **129**, 2776–2790. doi:10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2

Horn, R. A. 1990. The hadamard product. *Proc. Symp. Appl. Math.* **40**, 87–169. doi:10.1090/psapm/040

Houtekamer, P. L. and Mitchell, H. L. 1998. Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.* **126**, 796–811. doi:10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2

Houtekamer, P. L. and Mitchell, H. L. 2001. A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.* **129**, 123–137. doi:10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2

Hunt, B., Kostelich, E. and Szunyogh, I. 2007. Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D.* **230**, 112–126. doi:10.1016/j.physd.2006.11.008

Kotsuki, S., Ota, Y. and Miyoshi, T. 2017. Adaptive covariance relaxation methods for ensemble data assimilation: experiments in the real atmosphere. *Q. J. R. Meteorol. Soc.* **143**, 2001–2015. doi:10.1002/qj.3060

Lorenc, A. C. 2003. The potential of the ensemble kalman filter for NWP—A comparison with 4d-var. *Q. J. R. Meteorol. Soc.* **129**, 3183–3203. doi:10.1256/qj.02.132

Lorenc, A. C. 2017. Improving ensemble covariances in hybrid variational data assimilation without increasing ensemble size. *Q. J. R. Meteorol. Soc.* **143**, 1062–1072. doi:10.1002/qj.2990

Lorenz, E. N. 2005. Designing chaotic models. *J. Atmos. Sci.* **62**, 1574–1587. doi:10.1175/JAS3430.1

Luo, X. and Hoteit, I. 2011. Robust ensemble filtering and its relation to covariance inflation in the ensemble Kalman filter. *Mon. Wea. Rev.* **139**, 3938–3953. doi:10.1175/MWR-D-10-05068.1

Mandel, J., Cobb, L. and Beezley, J. 2011. On the convergence of the ensemble Kalman filter. *Appl. Maths (Prague)* **56**, 533–541.

Miyoshi, T. and Kondo, K. 2013. A multi-scale localization approach to an ensemble kalman filter. *SOLA* **9**, 170–173. doi:10.2151/sola.2013-038

Ménétrier, B., Montmerle, T., Michel, Y. and Berre, L. 2015. Linear filtering of sample covariances for ensemble-based data assimilation. Part I: Optimality criteria and application to variance filtering and covariance localization. *Mon. Wea. Rev.* **143**, 1622–1643. doi:10.1175/MWR-D-14-00157.1

Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J. and co-authors. 2004. A local ensemble kalman filter for atmospheric data assimilation. *Tellus A* **56**, 415–428. doi:10.3402/tellusa.v56i5.14462

Rainwater, S. and Hunt, B. 2013. Mixed-resolution ensemble data assimilation. *Mon. Wea. Rev.* **141**, 3007–3021. doi:10.1175/MWR-D-12-00234.1

Shlyaeva, A., Whitaker, J. and Snyder, C. 2019. Model-space localization in serial ensemble filters. *J. Adv. Model. Earth Syst.* **11**, 1627–1636. doi:10.1029/2018MS001514

Tippet, M., Anderson, J., Bishop, C., Hamill, T. and Whitaker, J. 2003. Ensemble square root filters. *Mon. Wea. Rev.* **131**, 1485–1490. doi:10.1175/1520-0493(2003)131<1485:ESRF>2.0.CO;2

Zhen, Y. and Zhang, F. 2014. A probabilistic approach to adaptive covariance localization for serial ensemble square root filters. *Mon. Wea. Rev.* **142**, 4499–4518. doi:10.1175/MWR-D-13-00390.1