

Towards a theory of optimal localisation

By JONATHAN FLOWERDEW*, *Met Office, Exeter, United Kingdom*

(Manuscript received 23 June 2014; in final form 13 January 2015)

ABSTRACT

All practical ensemble-based atmospheric data assimilation (DA) systems use localisation to reduce the damaging impact of spurious long-range correlations arising from the finite ensemble size. However, the form of the localisation function is generally ad-hoc, and requires expensive tuning to optimise the system. For the case of a single observation and known true background error correlation, we derive an expression for the localisation factor that minimises the expected root-mean-square (RMS) analysis error. Idealised tests show this formulation performs well for multiple observations provided their density is not too high. The width of the optimal localisation function scales with the width of the underlying correlation, but does not have the same shape. The optimal observation-space localisation for a single spatially integrating observation depends on the observation-to-gridpoint background error correlation, making it broader than the optimal localisation for point observations and potentially competitive with model-space localisation. A new form of hybrid DA is proposed in which localisation damps the sample correlations towards their climatological mean rather than zero, reducing the RMS error and potentially improving the dynamic balance of the analysis. The presence of variance errors causes the optimal localisation factor to depend on the ratio of observation to background error variance, and raises the possibility that a small amount of variance damping may be beneficial. For dense observations, a more elaborate theory is required, which will almost certainly depend on the observation network. We present some preliminary analysis of the features of the multi-observation problem, which for instance suggests that the optimal solution may involve different localisation factors in the numerator and denominator of the Kalman filter equation. We note that even optimal localisation gives an expected RMS error which exceeds that of perfect DA, contrary to the assumption made by ‘deterministic’ ensemble filters.

Keywords: ensemble data assimilation, sampling error, sample correlation, Schur product, integrating observations, hybrid data assimilation

1. Introduction

Almost all practical atmospheric data assimilation (DA) systems are based on the Kalman filter (KF) update equation (Kalman, 1960):

$$\begin{aligned}\mathbf{x}^a &= \mathbf{x}^f + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}^f) \\ \mathbf{K} &= \mathbf{P}^f \mathbf{H}^T (\mathbf{H} \mathbf{P}^f \mathbf{H}^T + \mathbf{R})^{-1}.\end{aligned}\quad (1)$$

This updates a prior (background) estimate $\mathbf{x}^f$ of the system state, which is subject to errors with covariance matrix $\mathbf{P}^f$, using observations $\mathbf{y}$ which are given by $\mathbf{H}$ times the true state, plus noise with covariance matrix $\mathbf{R}$. Equation (1) then provides the analysis $\mathbf{x}^a$ which is the optimal linear combination of the observations and background forecast, minimising the root-mean-square (RMS) error with respect

to the unknown true state. Under stronger assumptions, when the system is linear and all the errors are Gaussian, $\mathbf{x}^a$ is also the maximum likelihood estimate of the true state and the filter gives the exact probability distribution function (PDF) of the state conditioned on the prior and observations.

Ensemble data assimilation (EnDA) systems estimate $\mathbf{P}^f$ using a finite number, N , of samples which are presumed to have been drawn from it. This estimate $\hat{\mathbf{P}}^f$ is subject to two kinds of error. The first concerns errors in the PDF from which the members are drawn, such as model bias, inaccurate representation of forecast model uncertainty, failure to correctly sample errors in boundary conditions, and simplifications within the DA system. Most of these would harm even an infinite-sized ensemble. Second, there are statistical errors arising from the finite ensemble size: even if the system samples from the correct PDF, the sample means, covariances, and quantities such as the Kalman gain $\mathbf{K}$ which are derived from them will be distributed about their true

*Correspondence.
email: jonathan.flowerdew@metoffice.gov.uk

values with a PDF which is a function of the true statistics and the ensemble size. It is these ‘sampling errors’ which are the focus of this paper.

The most damaging impact of sampling error arises when the true correlation between two quantities is small. For a given ensemble size, there is a threshold below which the sampling error on a correlation is larger than the true value. Assimilation increments based on these sample correlations will have a greater contribution from the sampling noise than the signal provided by the true correlation, and will therefore harm rather than improve upon the background forecast (Hamill et al., 2001). The analysis also remains suboptimal for all locations at which the sampling error is non-negligible in comparison to the true correlation.

Practical ensemble systems ubiquitously apply ‘localisation’ to reduce the harm caused by sampling error. Clearly, the required damping is related to the magnitude of the noise in comparison to the true correlation, but localisation is typically applied in an ad-hoc manner. Houtekamer and Mitchell (1998) first introduced the concept by simply excluding observations at large distances, where the signal-to-noise ratio of the sample correlations is likely to be poor. Many ensemble systems follow Houtekamer and Mitchell (2001) in multiplying each sample correlation by a localisation factor which is a function of the distance between the two points involved. This can be written as a ‘Schur product’ between the sample covariance and a corresponding matrix of localisation factors. The localisation function is nearly always the Gaspari and Cohn (1999) approximation to the Gaussian, which has the advantage of falling to zero at finite distance and thus allowing the assimilation of distant observations to proceed in parallel. Other authors have explored localisation in spectral space, which is related to spatial averaging (Buehner and Charron, 2007; Buehner, 2011), or ‘adaptive’ localisation which depends upon the sample correlations (Bishop and Hodyss, 2009; Anderson, 2012). Anderson (2007) used a ‘hierarchical’ ensemble filter to explicitly sample the uncertainty in how observation-space updates should be mapped to model-space fields, whilst Anderson and Lei (2013) derive Empirical Localisation Functions for groups of single observations based on the relationship between background error and unlocalised increments in Observing System Simulation Experiments.

The lack of more fundamental theory of localisation makes it hard to predict in advance which ensemble system design will perform best. Parameters such as the localisation radius are expensive to tune and even then may not deliver optimum performance if the localisation scheme does not vary appropriately with the system parameters. The expanding field of convective-scale EnDA emphasises questions such as how localisation should adapt to variables with different spatial scales, or to systems which

combine errors on both large and small scales. Since any given DA cycle involves a single true $\mathbf{P}^f$, independent of the new observation network, one might imagine that there was also a single best localisation, dependent on $\mathbf{P}^f$ and N but independent of $\mathbf{H}$ and $\mathbf{R}$. However, Dong et al. (2011) found that, when assimilating both sparse surface observations and dense radar observations in a fairly idealised context, the performance was optimised by using two different localisation radii that reflected the densities of the two different observation sources. More recently, Kirchgessner et al. (2014) have experimentally studied the relationship between localisation radius, observation spacing and ensemble size for relatively dense observations in a hierarchy of models.

Another issue relates to observations such as satellite radiances which integrate over multiple model variables. It is clear that these require broader localisation than point observations, since the observation has non-negligible correlations with a greater range of heights. One approach is to come up with one good localisation in model space, and then map the localised covariances to observation space using the linearised observation operator. Campbell et al. (2010) argue it is hard to even define an equivalent localisation directly in observation space. This question interacts with the choice of DA method, since variational systems naturally use model-space localisation (Lorenc, 2003), whilst ensemble filter approaches typically use observation-space quantities to average over non-linearities and avoid the need for a linearised observation operator.

The aim of this paper is to gain a more fundamental understanding of what the localisation function ‘should’ be in such cases. Ultimately, we would like to be able to *derive* the optimal localisation as a function of the system parameters, rather than having to guess a functional form and experimentally tune its parameters. Like the original KF derivation, we define the statistically optimal localisation as that which minimises the expected RMS analysis error, but now considering an assimilation based on noisy sample covariances rather than the true covariances. This is inspired by Anderson (2007, 2012), where localisations are chosen to minimise related metrics.

It is important to distinguish localisation from the related problem of inflation. Whilst both of these techniques can be motivated as addressing deficiencies arising from sampling error, they do so on different time scales. This paper focusses on localisation as the problem of getting the best central analysis given background perturbations which are sampled from the true $\mathbf{P}^f$. To run the next set of ensemble forecasts, the ensemble perturbations need to be updated in a consistent manner – many methods have been developed to do this (e.g. Anderson, 2001; Houtekamer and Mitchell, 2001; Whitaker and Hamill, 2002; Hunt et al., 2007), but the details are beyond the scope of this paper. The residual

sampling error and simplifications within the DA system tend to make the resulting analysis ensemble spread systematically too small in comparison to analysis error. This will be exacerbated by insufficient perturbation growth and insufficient representation of model error in the forecast phase, such that the background perturbations for the next cycle would not appear to be drawn from the correct $\mathbf{P}^f$ without some form of inflation to restore the missing spread. Various different methods (e.g. Wang and Bishop, 2003; Zhang et al., 2004; Hamill and Whitaker, 2005; Anderson, 2009; Flowerdew and Bowler, 2013) have been proposed depending on the target forecasting system and dominant causes of underspread.

The rest of this paper is laid out as follows. Section 2 derives the form of the optimal localisation factor for a single observation, focussing on the case where variance errors are neglected. Section 3 tests this theory in an idealised context, showing that it performs well for multiple observations provided they are not too dense. Section 4 considers the implications of the theory for the optimal localisation of observations which integrate over multiple model variables. One feature of traditional localisation is that the localised correlations are biased low with respect to the true correlations. Section 5 develops an alternative formulation which avoids this problem, thereby reducing RMS error and potentially improving the dynamic balance of the analysis. Section 6 extends the analysis to include the impact of *variance* sampling errors, which provides the simplest case in which the optimal localisation depends on the characteristics of the observations in addition to those of the background. Finally, Section 7 begins the process of extending the theory to multiple observations. The general solution appears complex, but we provide evidence of how and why it will differ from the single-observation results. A summary and discussion are given in Section 8.

2. Optimal localisation for a single observation

2.1. Basic theory

In the case of a single observation y with error variance σ_o^2 , the ensemble approximation to the KF eq. (1) simplifies to:

$$x_2^a - x_2^f = \alpha \hat{b}(y - x_1^f), \quad \hat{b} = \frac{\hat{\sigma}_1 \hat{\sigma}_2 \hat{r}}{\hat{\sigma}_1^2 + \sigma_o^2}. \quad (2)$$

Here, superscript f indicates the background forecast, superscript a the analysis, x_2 is the gridpoint being updated, x_1 is the model equivalent of the observation, r is the correlation between their background errors, and σ_2^2 and σ_1^2 are the corresponding background error variances. EnDA systems replace the latter three quantities by their ensemble-

based estimates, denoted by carets. For now, we follow the standard approach of multiplying the estimated gain $\hat{b}$ by a localisation factor α to damp the impact of spurious correlations at long distances.

Inspired by Anderson (2012), we choose α to minimise:

$$J = \int (\alpha \hat{b} - b)^2 P(\hat{b}|\boldsymbol{\theta}) d\hat{b}, \quad (3)$$

where $\boldsymbol{\theta} = (r, \sigma_1, \sigma_2, \sigma_o)$ contains the true values of the statistical parameters. This expression is the mean square difference between the optimal gain $b(\boldsymbol{\theta})$ and the localised ensemble estimate $\alpha \hat{b}$, taking into account the PDF $P(\hat{b}|\boldsymbol{\theta})$ of gain estimates expected from an ensemble of size N which samples from the true σ_1 , σ_2 and r . Zhen and Zhang (2014; their eq. (3)) have very recently made a similar proposal, although they operate on the whole state with a pre-specified localisation function shape, whereas we derive the functional form by considering each gridpoint separately. The form of $P(\hat{b}|\boldsymbol{\theta})$ will be discussed in Section 2.2; for now we take it as an input from statistical theory. Assuming the errors on the innovation $(y - x_1^f)$ are statistically independent of the errors on $\hat{b}$, eq. (2) implies that J is proportional to the mean square difference between the optimal increments and those estimated from the ensemble. [Note it is not enough for the errors to be uncorrelated, since eq. (2) involves a product rather than a sum, and thus its square involves higher-order moments, although the two conditions coincide for a multivariate Gaussian]. Since the background state is common, and any departure from the optimal analysis constitutes additional error, this in turn implies that J is proportional to the additional error variance of the ensemble-based analysis over and above that of perfect DA based on the true covariances. In other words, minimising J minimises the RMS analysis error. The multi-observation equivalent of this link is discussed in Section 7 [eq. (18)]. The recent independent proposal of Periañez et al. (2014) is also based on minimising an expression for analysis error. Our two approaches are complementary: whereas Periañez et al. derive an optimal localisation *radius* by high-level heuristic arguments for a specified uniform observation density, we derive the complete localisation *function* for a single observation from the underlying KF equation and sampling error statistics.

Requiring $dJ/d\alpha = 0$ yields:

$$\int (\alpha \hat{b}^2 - b\hat{b}) P(\hat{b}|\boldsymbol{\theta}) d\hat{b} = \alpha(\bar{\hat{b}}^2 + \sigma_{\hat{b}}^2) - b\bar{\hat{b}} = 0, \quad (4)$$

where $\bar{\hat{b}}$ and $\sigma_{\hat{b}}^2$ are respectively the mean and variance of the sample gain, and $\langle \hat{b}^2 \rangle$ has been expanded using the general relation:

$$\sigma_x^2 = \langle (x - \bar{x})^2 \rangle = \langle x^2 \rangle - \bar{x}^2. \quad (5)$$

Thus, we obtain the result:

$$\alpha = \frac{b}{\bar{b}} \frac{\bar{b}^2}{\bar{b}^2 + \sigma_{\bar{b}}^2}, \quad (6)$$

where the first ratio can be interpreted as a multiplicative correction for any bias in the sample gain (as illustrated in Section 2.2 for sample correlations). In the case where bias is negligible, eq. (6) simplifies to:

$$\alpha = \frac{b^2}{b^2 + \sigma_b^2} = \frac{Q^2}{1 + Q^2}, \quad Q = b/\sigma_b. \quad (7)$$

The quantity Q can be interpreted as a signal-to-noise ratio: the ‘signal’ is how far the true correlation departs from zero, and the ‘noise’ is how much the ensemble estimate varies about the true correlation. When Q passes through unity, the optimal localisation factor passes through 0.5.

Equation (7) has the same form as Anderson (2012), eq. (9). However, the variables involved have different meanings: Anderson (2012) substitutes a mean and variance of true correlations given a particular sample correlation (the ‘adaptive’ localisation problem), whereas here we consider the PDF of sample correlations for a specified true correlation (which we have shown determines the optimal ‘static’ localisation). The relationship between static and adaptive localisation will be discussed further in Section 8.

2.2. Sampling error PDF

The sample gain $\hat{b}$ defined by eq. (2) includes sampling error on both the variances σ_1^2 and σ_2^2 and the correlation r . The standard deviation of the variance estimated from N independent samples of a Gaussian variable with true variance V is (Casella and Berger, 2002, example 7.3.3, p. 331):

$$\sigma_{\hat{V}} = V\sqrt{2/(N-1)}. \quad (8)$$

This is quite small for typical ensemble sizes (20–100) and is smaller than the variance itself provided $N > 3$. For most of this paper, we will therefore follow common practice (e.g. Anderson, 2012) and assume the sample variances are indistinguishable from their true values. This is consistent with the typical approach of applying unit localisation at zero distance, as in Houtekamer and Mitchell (2001). When variance errors are neglected, the only sampling error within the gain $\hat{b}$ [eq. (2)] for a single observation comes from the sample correlation $\hat{r}$, for which a simple analytic form is available as discussed shortly. The sample standard deviations within $\hat{b}$ are replaced by their true values, making both b and $\hat{b}$ proportional to $\sigma_1\sigma_2/(\sigma_1^2 + \sigma_2^2)$. This common factor cancels out of the minimisation, reducing the mean-square error (MSE) on the gain, eq. (3), to the

MSE on the localised correlation. The signal-to-noise ratio, Q , in eq. (7) similarly simplifies from b/σ_b to r/σ_r . These simplifications make the result usable in a wider range of EnDA systems – including those such as variational assimilation where the sample gain is not formed explicitly – provided the sample correlation can be appropriately manipulated. The additional impact of sampling error on variances is explored more thoroughly in Section 6.

The PDF of a sample correlation estimated from N samples of a Gaussian variable with true correlation r is approximately Gaussian in the variable s defined by the Fisher transform (Fisher, 1921):

$$\sigma_s = 1/\sqrt{N-3}, \quad s = \frac{1}{2} \ln \frac{1+r}{1-r}, \quad r = \tanh(s), \quad (9)$$

with similar relations between $\hat{r}$ and $\hat{s}$. This is illustrated in Fig. 1 for three representative true correlations. The standard deviation in correlation space, σ_r , is largest for small r and exceeds r for a region near $r=0$. These are the spurious long-range correlations which are particularly damaging to EnDA. The distribution narrows for large r , becoming a delta function at $r = \pm 1$ where the variables are completely dependent. The requirement that $|\hat{r}| \leq 1$ leads to a skewed distribution at large r , with $|\hat{r}|$ biased low with respect to $|r|$. This leads to the surprising result (not shown) that the optimal localisation factor provided by eq. (6) exceeds unity by a small amount for correlations just below one, where the first ratio ($b/\hat{b}$) is larger than the reciprocal of the second. However, we will generally neglect this bias,

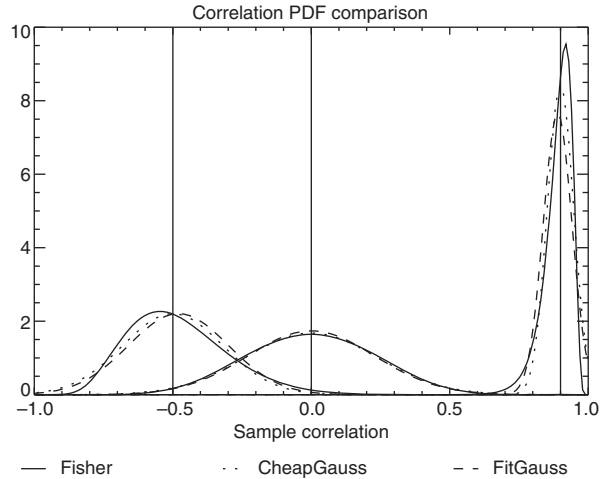


Fig. 1. Different approximations to the distribution of sample correlations for 20 ensemble members given true correlations of 0.0, -0.5 and $+0.9$ (indicated by vertical lines). The curves show the PDF implied by the Fisher transform (solid), a Gaussian with mean and standard deviation given by the cheap approximation introduced in Section 2.2 (dotted), and a Gaussian whose parameters are obtained by numerical integration of the Fisher transform (dashed).

and approximate $\sigma_{\tilde{r}}$ as half the difference between the correlations corresponding to $s \pm \sigma_s$. This is much cheaper than evaluating $\tilde{r}$ and $\sigma_{\tilde{r}}$ by numerical integration of the full PDF, and simplifies the interpretation. In practice it tends to perform as well or better than the full integrals in assimilation tests where the assumptions are not fully satisfied, for example, due to the presence of errors in the variances. As discussed in Section 6, this appears to be because the bias and variance error have opposite effects on the optimal localisation factor, which tend to cancel out.

2.3. Sample localisation functions

Figure 2 illustrates the steps of the simplified optimal localisation calculation for three scenarios (columns) and a range of ensemble sizes (rows). The left-hand column shows the universal form as a function of the true correlation [for clarity, only positive correlations are shown, since the combination of eqs. (7) and (9) is symmetric about $r=0$]. The

approximated standard deviation of sample correlations is shown by the dashed line, falling from approximately $1/\sqrt{N}$ at $r=0$ to zero at $r = \pm 1$ as discussed in the previous section. The signal-to-noise ratio Q (dotted) rises with true correlation because this both directly increases its numerator and decreases the sampling error in the denominator. The optimal localisation factor α (solid) is then determined from Q through the non-linear mapping provided by eq. (7). The optimal localisation factor is fairly similar to the true correlation (double-dot-dashed) for $N=5$, but becomes increasingly broad, trusting samples from a wider range of true correlations, as ensemble size increases. This universal form is very similar to the results obtained by Lei and Anderson (2014) from both the hierarchical filter and Empirical Localisation Functions applied to a bivariate normal distribution (their Fig. 1; the small differences between the methods are probably due to the differing treatment of variance errors).

Both correlations and localisation functions are often assumed to be Gaussian-shaped as a function of distance.

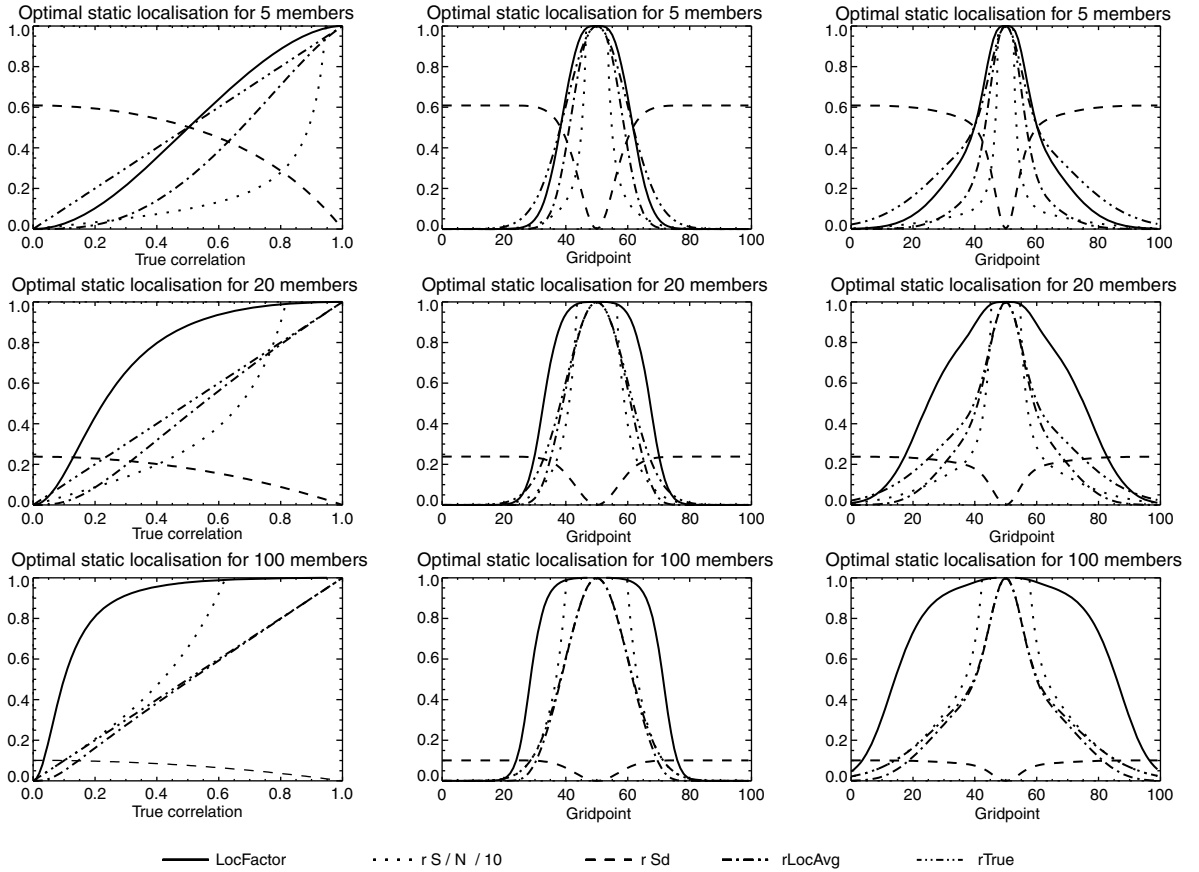


Fig. 2. The steps of the simplified optimal localisation calculation for 5 (top), 20 (middle) and 100 (bottom) ensemble members, showing the universal function of the true correlation (left), and results as a function of gridpoint where the true correlation varies as a Gaussian with half-width 10 gridpoints (middle) or an average of Gaussians with half-widths of 5 and 20 gridpoints (right). The lines show the true correlation (double-dot-dashed), approximated standard deviation of sample correlations (dashed), signal-to-noise ratio Q (dotted, divided by 10 for convenience of plotting), calculated optimal localisation factor (solid) and mean localised correlation (dot-dashed).

The middle column of Fig. 2 examines the optimal localisation for a true correlation which falls off as a Gaussian with a half-width of 10 gridpoints. The rising true correlation (double-dot-dashed), declining standard deviation (dashed), and non-linearity in the $Q \rightarrow \alpha$ relationship all combine to give the optimal localisation function (solid) a broader peak and tighter tails than the original Gaussian, particularly at large ensemble sizes. Since α is a function of r , the optimal localisation function directly replicates any overall horizontal stretching of the true correlation function. This implies that meteorological variables with larger length scales should have correspondingly broader localisation functions.

The right-hand column illustrates the optimal localisation for a more complicated true correlation, which is the sum of two Gaussians with half-widths of 5 and 20 gridpoints. This is intended to represent the multi-scale correlations which might apply in convective-scale DA. The calculated localisation is neither Gaussian nor much like the underlying correlation. Indeed, the shape at moderate correlations changes from concave to convex as the ensemble size increases.

Since the localisation factor is fixed for any given true correlation, the mean localised correlation is just the localisation factor multiplied by the expected sample correlation (again approximated as equal to the true correlation, i.e. ignoring biases). This is shown by the dot-dashed lines in Fig. 2. This highlights the important fact that the standard approach to localisation embodied by eq. (2) reduces noise at the expense of the mean value: at larger distances the average increment is smaller than it should be, in the proportions given by the difference between the dot-dashed and double-dot-dashed lines. This motivates an improved form of hybrid localisation, as described in Section 5 below.

2.4. Non-trivial prior PDF

The combination of eqs. (7) and (9) gives the statistically optimal factor α for localisation applied in the traditional way via eq. (2), as a function of the true correlation r and ensemble size N . This is valuable in itself for understanding what form localisation functions should take as a function of the system parameters. However, it is also somewhat unsatisfying, since if the true correlation was known, it would be better to just use it directly rather than the noisy ensemble estimate. A number of extensions are possible to make the scenario more realistic. In the simplest approach, each localisation factor might be derived from an estimate of the ‘typical’ value of the corresponding correlation, hoping that this allows useful case-specific information to be extracted from the ensemble. To progress further, the optimisation needs to be redefined so that the true parameters vary from case to case, with some overall PDF $P(\theta)$.

We assume the ensemble samples from the correct parameters in each case, and balance this signal against the noise arising from the finite ensemble size. Equation (3) acquires an outer integral over $P(\theta)$, to identify the single localisation factor that optimises performance over the long term. Neglecting as before any bias in the sample gain, we obtain a new eq. (7) in which the numerator and denominator are separately integrated over $P(\theta)$:

$$\alpha = \frac{\int b^2 P(\theta) d\theta}{\int (b^2 + \sigma_b^2(\theta)) P(\theta) d\theta} = \frac{\bar{b}^2 + \sigma_b^2}{\bar{b}^2 + \sigma_b^2 + \langle \sigma_b^2(\theta) \rangle}. \quad (10)$$

This can again be written as $\alpha = Q^2/(1 + Q^2)$ if $Q^2 = (\bar{b}^2 + \sigma_b^2)/\langle \sigma_b^2(\theta) \rangle$: the ‘signal’ now includes both the mean $\bar{b}$ and variance σ_b^2 of the true gain, whilst the ‘noise’ is the mean over $P(\theta)$ of the sampling variance $\sigma_b^2(\theta)$. Equation (10) provides the optimal α when localisation in the form of eq. (2) is applied to a system whose true parameters vary over a known climatology. However, the same information can be used to construct a better solution if the localisation is allowed to include a constant as well as the linear damping term, as explored in Section 5.

3. Idealised localisation experiments

A series of idealised experiments have been performed to test the calculated localisation functions and compare their performance to a traditional Gaussian-shaped localisation with tuned width. The scenario considers a (non-cyclic) line of 100 points. The true $\mathbf{P}^f$ is specified analytically. It has unit variances, correlations as described later for each scenario, and is given to the optimal localisation calculation. The ‘truth’ and five background ensemble members are obtained as random samples from $\mathbf{P}^f$ using an eigenvector decomposition. The very small ensemble size was chosen to emphasise the impact of sampling errors in this small and simple system, although it will exaggerate the influence of variance errors compared to typical meteorological ensembles. An observation is provided at the first gridpoint and every ‘ObSpc’ gridpoints thereafter, with errors drawn from a Gaussian distribution with standard deviation ‘ObSd’. The analysis is obtained by direct application of the KF eq. (1), modified to use the localised sample covariance $\tilde{\mathbf{P}}^f$:

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

where L_{ij} is the calculated localisation factor between points i and j (with $L_{ii} = 1$). The scenario is repeated 10^4 times to estimate the RMS analysis error, which is measured with respect to the true state (not the noisy observations). The same set of random numbers is used to test each localisation

option for a given observation density; this greatly reduces the sampling error on the difference between these configurations, in the same way as testing two different Numerical Weather Prediction systems on the same set of dates. Repeat runs with different random number seeds have been used to confirm that the key results are robust.

This scenario is chosen to focus on the statistical aspects of the DA problem, where all of the assumptions can be satisfied and each parameter is directly controlled. There are no explicit model dynamics: just a background state whose error is drawn from the specified $\mathbf{P}^f$, fitting the assumptions of the KF equation. This separates the problem of updating the ensemble mean from consistently updating the perturbations, and isolates the problem of localisation from that of inflation. The ensemble is ‘perfect’ in that it samples from $\mathbf{P}^f$; the only imperfections arise from the finite ensemble size, and how the resulting sampling error is managed by the localisation procedure. The assumptions of the optimal localisation scheme are also satisfied, with two important exceptions. First, most of the tests use unmodified samples from $\mathbf{P}^f$, and thus include sampling errors in both variances and correlations, whereas the optimal localisation calculation only considers the sampling error on correlations. In a few cases, we also consider results in which the variance error has been artificially removed by extracting the matrix of sample correlations and multiplying each element by the corresponding pair of true standard deviations. Second, the system assimilates multiple observations, whereas the localisation derivation assumed a single observation. The localisation of sample correlations in the numerator of eq. (11) fits the assumptions of the derivation, but the corresponding elements in the denominator are new since the denominator of the single observation eq. (2) only involved variances.

Figure 3a shows results for the simplest scenario in which the true correlation has a Gaussian shape with a half-width of five gridpoints, similar to the middle column of Fig. 2 above. Each line has been normalised by its minimum value to emphasise the impact of the alternative localisations despite the large difference in RMS error between the different observation densities. A half-width of about six gridpoints is found to be optimal for a traditional Gaussian-shaped localisation function, with little dependence on observation density. The theoretical localisation matches the best tuned localisation for large observation spacings, but has trouble once the observation spacing falls below the half-width of the true correlation. This is the point at which there start to be significant correlations between the background errors for the different observations, so that the denominator of eq. (11) no longer separates into the single-observation form assumed by the derivations in Section 2.

Figure 3b considers the same scenario as the right column of Fig. 2, in which the true correlation is a sum

of Gaussians with half-widths of 5 and 20 gridpoints. Now, the optimal half-width for a Gaussian-shaped localisation function varies with the observation density, from about six gridpoints at $\text{ObSpc} = 2$ to double that at $\text{ObSpc} = 40$ (the observation error variance has been halved to emphasise this effect). This is consistent with the Dong et al. (2011) result that different localisation radii can be optimal for observations with different densities, and confirms that this behaviour can arise directly from the statistics of idealised EnDA, without requiring the complications of non-linear atmospheric models or complex observation types. In a real system, the density of particular observations might vary between land and sea, as a function of height, or following the distribution of precipitation in the case of radar radial velocity data. Some variables can be observed by radar at high density, others by surface observations at lower density, and others by radiosondes at very low density. A theoretical localisation that either provided a shape appropriate to a range of observation densities or could adapt to the observation density could therefore have an important advantage. In these experiments, the single theoretical localisation again matches the best tuned localisation for low to moderate observation densities, but performs poorly at $\text{ObSpc} = 2$. This degradation can be mostly but not completely eliminated by artificially removing the variance errors as described above (not shown). This suggests that accounting for variance errors (Section 6) could improve the performance of the theoretical localisation, and it makes sense that this would have the greatest impact for short-range covariances, where the correlation errors are small. Equally, the fact that removing variance errors does not eliminate the performance gap confirms there is more to the dense observations problem than just variance errors, as explored further in Section 7.

Figure 3c considers a situation in which the true correlation varies across the domain. Such variations might occur in real systems as a function of latitude, height, variable, season, or as a result of the spatial distribution of observations in previous assimilation cycles. To make a legitimate correlation matrix, the half-width is specified as a function of the location halfway between each pair of points, rising linearly from 5 to 20 gridpoints across the domain. No single localisation function can be optimal for all gridpoints in this scenario. Whereas in most other tests the theoretical localisation has only matched the best tuned localisation, in this case its spatial adaptivity allows it to achieve an RMS error that is 2–7% better than the best single Gaussian localisation, with particular advantage at large observation spacings. On the other hand, the theoretical localisation again performs poorly for $\text{ObSpc} = 2$ (not shown).

The problems with dense observations have been linked to the denominator of the KF eq. (11). This suggests the situation should improve in the limit of ‘weak assimilation’,

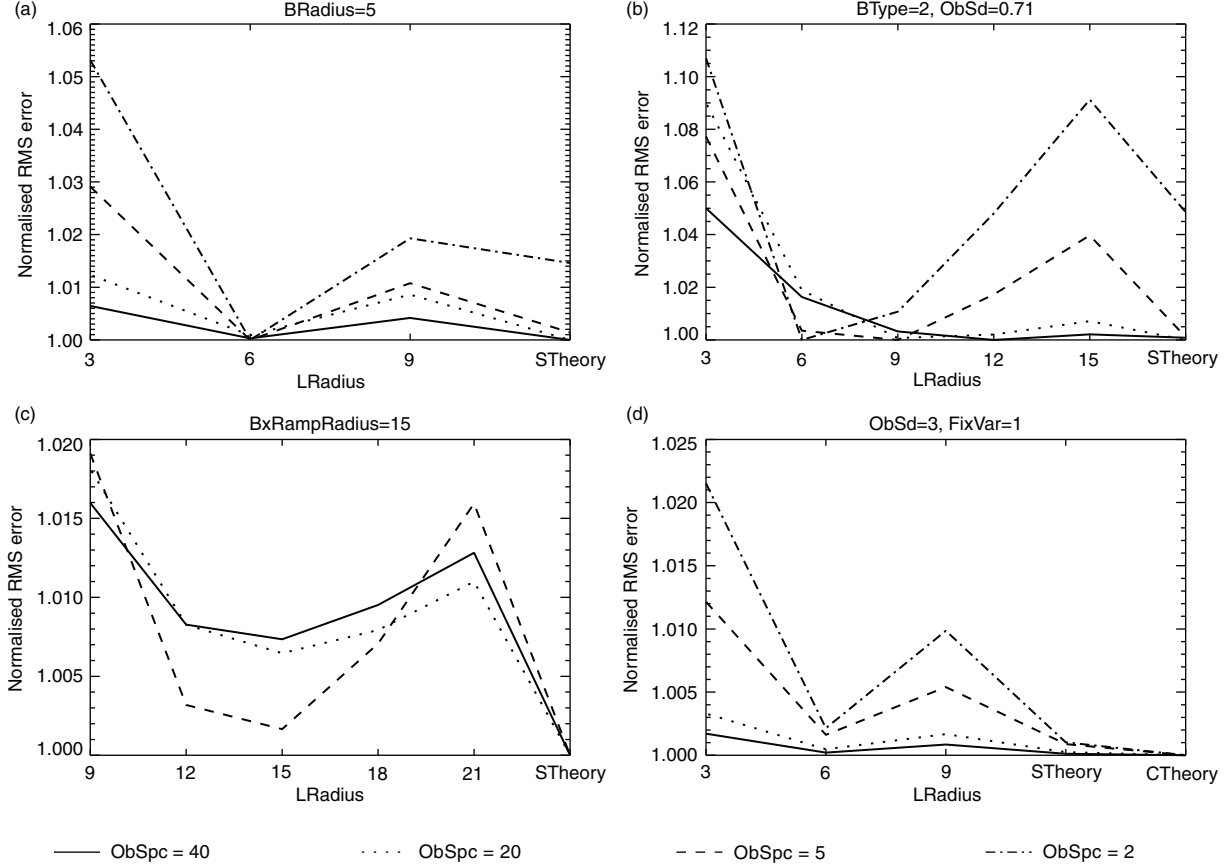


Fig. 3. RMS analysis error as a function of the half-width (in gridpoints) of a Gaussian-shaped localisation function, and also for the simplified optimal localisation (‘STheory’), and, in one case, optimal localisation with numerical evaluation of the full Fisher integrals (‘CTheory’). In (a) the true correlation is a Gaussian with a half-width of five gridpoints; in (b) it is the average of Gaussians with half-widths of 5 and 20 gridpoints, with observation error variance halved; in (c) the half-width varies linearly from 5 to 20 gridpoints across the domain; (d) is like (a) except the observation standard deviation is trebled and variance sampling errors are artificially removed to match the assumptions of the optimal localisation calculation. Results are shown for observation spacings of 40 (solid), 20 (dotted), 5 (dashed) and 2 [dot-dashed, omitted for case (c)] gridpoints. All tests use five ensemble members.

where $\|\mathbf{R}\| \gg \|\mathbf{H}\mathbf{P}^f\mathbf{H}^T\|$ so that localisation has negligible impact on the denominator as a whole. This does indeed occur, although the best performance is only obtained when variance errors are artificially removed as discussed above. The results are shown in Fig. 3d, for a scenario with the same $\mathbf{P}^f$ as Fig. 3a but observations whose errors have three times the standard deviation (nine times the variance) of the background errors. The simplified optimal localisation (‘STheory’) now matches or beats the best tuned localisation for *all* observation densities, and the more expensive form which evaluates $\hat{b}$ and σ_b^2 by numerical integration of eq. (9) (‘CTheory’) has a slight further advantage.

4. Integrating observations

Observations which average over more than one model value (i.e. have non-trivial $\mathbf{H}$) are a prime example of where a

more fundamental understanding of localisation is required to establish what the ‘correct’ approach might be. A satellite radiance which averages over many model levels clearly provides information on a wider range of locations than a point observation; however, when the localisation of a point observation is defined as an ad-hoc function of distance, it is unclear how this should be modified for integrating observations. The very notion of physical distance is problematic because such observations do not have a well-defined ‘location’. One suggested resolution (Lorenc, 2003; Campbell et al., 2010) is to define a localisation once-and-for-all in model space, then map the localised covariances to observation space using the linearised observation operator, $\mathbf{H}$, as occurs naturally in variational methods. For reasons of cost and simplicity, ensemble filters typically work directly in observation space. This appears to place them at a disadvantage, since a localisation applied in observation

space does not have access to all the model-space variables, and so cannot precisely replicate a model-space localisation for arbitrary background perturbations.

The theory presented above clarifies the options available. Consider first the observation-space approach typically used in ensemble filters. The key insight is the primacy of the true gridpoint-to-observation background error correlation, r , in determining the signal and noise in eq. (2) and thus the optimal localisation factor for assimilation based on samples from r . These correlations are just as legitimate as those in model space: there is no fundamental difference between model prognostic variables and observed quantities, since we can regard them both as parts of the same joint state PDF (Anderson, 2003). The definition of a covariance implies that, if $\mathbf{P}^f$ is the true background error covariance in model space, then $\mathbf{H}\mathbf{P}^f\mathbf{H}^T$ is the error covariance between the background predictions of the observed quantities, and $\mathbf{P}^f\mathbf{H}^T$ the corresponding covariance between the observed and modelled variables. Dividing these cross-covariances by the relevant standard deviations from $\mathbf{P}^f$ and $\mathbf{H}\mathbf{P}^f\mathbf{H}^T$ gives the required correlations. It is these true correlations which provide the desired measure of ‘distance’ between any pair of model variables and/or observed quantities. Practically, these correlations could be estimated either directly from historic ensemble data or by applying $\mathbf{H}$ to a climatological covariance model.

A common approach in ensemble filters is to ascribe a ‘location’ to each observation (e.g. based on the peak of the weighting function, as in Houtekamer et al., 2005) and then apply the model-space localisation function using distances calculated with respect to this location. The framework presented in Section 2 confirms that this approach is sub-optimal, since the same localisation function is used for both point observations (where the optimal localisation is a function of $\mathbf{P}^f$) and integrating observations (where the optimal localisation depends on the broader correlations implied by $\mathbf{P}^f\mathbf{H}^T$). This provides the mathematical explanation for the broadening of the optimal model-to-observation-space localisation observed empirically by Anderson and Lei (2013; their section 4c). Another notable feature of this process is that the peak value of the localisation function is less than unity, which is now seen to be a direct consequence of the fact that no gridpoint has unit background error correlation with the model equivalent of the observation.

A theoretical comparison between observation-space and model-space localisation is complicated by the fact that they involve different functional forms, with different optimal results. The observation-space approach takes the sample covariance between the gridpoint and observed quantity, and multiplies this by a single localisation factor whose optimal value is given by the theory presented above. The model-space localisation uses a sum of individually localised

model-space sample covariances ($\tilde{\mathbf{P}}^f\mathbf{H}^T$), where the localisation factors might for instance take the values which are optimal for the assimilation of point observations. The only way this could beat observation-space localisation is if the model-space approach provides a better partition between signal and noise in the sample covariances. As a concrete example, consider an observation which averages over two gridpoints with background error correlation r_m and unit variances. If $r_m=0$, the model-space approach perfectly partitions the correlation: there is no noise on the sample correlation of each gridpoint with itself, so it requires no localisation ($\alpha=1$), whilst there is no useful signal in the sample correlations with the other gridpoint, so the associated sampling error can be completely damped ($\alpha=0$). For this special case, the localised correlations are always equal to the true correlations, giving a perfect assimilation if the sampling error on variances can be neglected. The observation-space approach, on the other hand, is based on the intermediate correlation $r = \sqrt{(1+r_m)/2} \sim 0.707$. Whilst not a perfect separation of signal and noise, this still gives a relatively large $\alpha \sim 0.971$ using the simplified theory for 20 ensemble members, as can be read off the middle left panel of Fig. 2. This implies the assimilation is able to use most of the signal in the sample correlation, and cannot be suffering too badly from the modest sampling noise which arises from the relatively high value of r . For larger r_m , the model-space approach has to localise an intermediate correlation between the gridpoints, whilst the observation-space method benefits from an even larger value of r .

Figure 4 compares the mean localised covariances ($\alpha\tilde{\mathbf{P}}^f\mathbf{H}^T$ versus $\tilde{\mathbf{P}}^f\mathbf{H}^T$) produced by the two approaches, as a function of r_m . The results are similar for $r_m > 0$, suggesting they give similar assimilation performance. Larger discrepancies emerge for negative r_m at small ensemble sizes (Fig. 4a), because the model-space localisation performs a separate signal-to-noise trade-off for each model-space covariance. It is unaware that the overall signal r becomes smaller, not larger, as $r_m \rightarrow -1$, and thus probably lets through too much sampling noise.

The issues with negative r_m are one example of a more general problem with model-space localisation: $\tilde{\mathbf{P}}^f\mathbf{H}^T$ is a sum of separately localised covariances. This introduces some of the same mathematical complications found with dense observations: indeed eq. (21) from the Appendix demonstrates that the localisation factors required to optimise the sum can differ from those required to optimise the assimilation of point observations. Whilst the observation-space approach is restricted to a single overall localisation factor, there is at least a simple expression for its optimal value in the case of sparse observations. The ‘perfect’ partition provided by the model-space approach when $r_m=0$ is also open to question: it is particularly unrealistic to neglect variance errors in this case, since they are part of

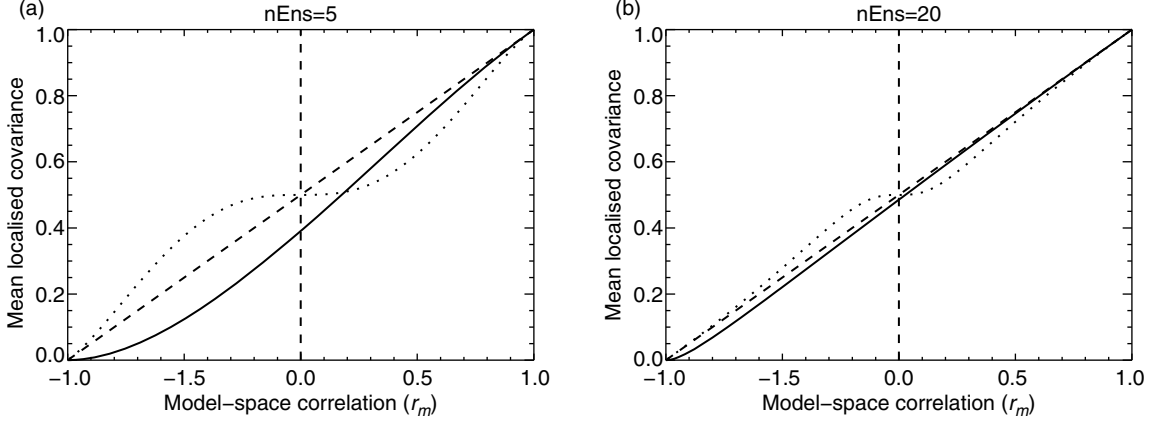


Fig. 4. Mean localised covariances between modelled and observed quantities for two gridpoints with model-space correlation r_m , $\mathbf{H}=(0.5, 0.5)$, and ensembles with (a) 5 and (b) 20 members, resulting from optimal observation-space localisation ($\alpha \mathbf{P}^f \mathbf{H}^T$, solid) and the model-space localisation ($\mathbf{P}^f \mathbf{H}^T$, dotted) that is optimal for point observations. The dashed diagonal lines show the true covariance, $\mathbf{P}^f \mathbf{H}^T$, and the simplified optimal localisation calculation is used throughout.

the sum which leads to the sampling error on the model-to-observation correlation. Finally, it should be noted that model-space localisation systematically underestimates the observation-space variance $\mathbf{H} \mathbf{P}^f \mathbf{H}^T$ in the denominator of the KF eq. (1), since $\mathbf{H}$ integrates over sample covariances which are damped towards zero. By contrast, the observation-space method simply uses the sample variance $\hat{\sigma}_1^2$ of the ensemble equivalents of the observation, which is unbiased.

A more detailed comparison of the performance of the different approaches to the localisation of integrating observations is left to future work.

5. A new form of hybrid DA

Figure 2 highlighted the fact that the traditional form of localisation, in which sample correlations are multiplied by a damping factor between 0 and 1, leads to mean localised correlations (the dot-dashed lines) which are systematically biased low with respect to the true correlations (the double-dot-dashed lines). In this formulation, the benefits of damping noise in the sample correlation come at the price of also damping not only the case-specific variations predicted by the ensemble but also the climatological mean correlation itself. In the extreme case as $\alpha \rightarrow 0$ at moderate distances, this bias is reflected in the fact that the additional error variance $J \rightarrow b^2 > 0$ from eq. (3), even though there is no longer any sampling noise in the localised gain $\alpha \hat{b}$. Traditional hybrid DA approaches (e.g. Lorenc, 2003) using a fixed combination of ensemble and climatological covariance reduce but do not eliminate this bias.

These considerations suggest an alternative formulation in which the localisation factor α damps not the sample

gain itself, but the departure of the sample gain from our prior estimate (its climatological mean value). For simplicity, we assume throughout this section that all sample statistics are unbiased. The first part of eq. (2) becomes:

$$x_2^a - x_2^f = \tilde{b}(y - x_1^f), \quad \tilde{b} = \bar{b} + \alpha(\hat{b} - \bar{b}). \quad (12)$$

Taking expectations, the average localised gain, $\langle \tilde{b} \rangle$ reproduces the climatological mean $\bar{b}$ regardless of α , eliminating the bias in mean localised correlations observed in Fig. 2. To make the result non-trivial, it is necessary to assume a PDF $P(\boldsymbol{\theta})$ of true gain factors, as discussed in Section 2.4. The penalty eq. (3) becomes:

$$J = \int \int (\bar{b} + \alpha(\hat{b} - \bar{b}) - b)^2 P(\hat{b}|\boldsymbol{\theta}) d\hat{b} P(\boldsymbol{\theta}) d\boldsymbol{\theta}. \quad (13)$$

Now, in the case where $\alpha \rightarrow 0$, the only penalty arises from the variance of b about $\bar{b}$ (i.e. the fact that $\bar{b}$ provides only a climatological, not case-specific, estimate of the gain), not the mean gain itself. One can also start from a general $\tilde{b} = \alpha \hat{b} + \beta$. The requirement $dJ/d\beta = 0$ gives $\beta = \bar{b}(1 - \alpha)$, confirming that eq. (12) is indeed the optimal linear estimator, with lower RMS error than the standard approach given by eq. (2).

Setting $dJ/d\alpha = 0$ gives:

$$2 \int \int (\alpha(\hat{b} - \bar{b})^2 - (\hat{b} - \bar{b})(b - \bar{b})) P(\hat{b}|\boldsymbol{\theta}) d\hat{b} P(\boldsymbol{\theta}) d\boldsymbol{\theta} = 0.$$

Consider first the inner integration over $\hat{b}$ for each value of b . The term in $\langle \hat{b}^2 \rangle$ can again be expanded using eq. (5). Combined with the assumption that $\langle \hat{b} \rangle = b$, this yields:

$$\alpha \int (\sigma_b^2(\boldsymbol{\theta}) + b^2 - 2\bar{b}b + \bar{b}^2) P(\boldsymbol{\theta}) d\boldsymbol{\theta} = \int (b - \bar{b})^2 P(\boldsymbol{\theta}) d\boldsymbol{\theta}.$$

Recognising a similar expansion of b^2 and the cancellation of several terms in $\bar{b}^2$ finally gives:

$$\alpha = \frac{\sigma_b^2}{\langle \sigma_b^2(\boldsymbol{\theta}) \rangle + \sigma_b^2}. \quad (14)$$

This is similar in form to eq. (7), but the ‘signal’ which the ensemble samples from is now the variance σ_b^2 of the true gain about its overall mean $\bar{b}$, whilst the noise is the mean over $P(\boldsymbol{\theta})$ of the sampling variance $\sigma_b^2(\boldsymbol{\theta})$. If there is a wide range of true correlations (making the climatological $\bar{b}$ a poor predictor of b) and/or the ensemble sampling error is small, $\alpha \rightarrow 1$. In the opposite case, $\alpha \rightarrow 0$. In the extreme case where the true correlation is fixed, $\sigma_b = 0$ and thus $\alpha = 0$ regardless of the true correlation: this matches the suggestion made above that there is no point using a noisy ensemble estimate if you know the case-specific true correlation in advance.

The mean true gain $\bar{b}$ does not appear explicitly in eq. (14), due to cancellations which arise from the localisation eq. (12) being unbiased, and therefore no longer having to trade the desirable damping of the sampling noise against undesirable damping of the mean correlation. As a result, the method is able to use systematically tighter localisation factors than the simple damping approach applied to the same scenario, eq. (10), and thus attain the lower RMS error noted above. The mean correlation does still play an indirect role, since it sets the magnitude of the mean sampling variance $\langle \sigma_b^2(\boldsymbol{\theta}) \rangle$ for a given ensemble size, as shown in Fig. 1. The optimal localisation is sensitive to this because the penalty is defined in terms of analysis error, which is proportional to correlations rather than the transformed variable s for which the sampling distribution has fixed width for a given ensemble size.

Whereas eq. (7) required only a climatological mean correlation to estimate the optimal localisation, eq. (14) requires the whole $P(r)$, or at least its mean and variance. The Fisher transform, eq. (9), provides one way to estimate $P(r)$, since it says that $P(\hat{s}) = \int P(\hat{s}|s)P(s)ds$ is the convolution of $P(s)$ with a Gaussian whose width depends only on the ensemble size. If $P(\hat{s})$ can be estimated from a long-term archive of ensemble correlations, then $P(s)$ might in turn be estimated by deconvolution, for example, using the ‘one-parameter method’ of Nahman and Guillaume (1981).

The unbiased nature of eq. (12) has one further potential advantage. The standard approach to localisation based on simple damping by a fixed ad-hoc function is known to harm the dynamic balance between model variables (Kepert, 2009). Firstly, one might hope that a localisation derived from the characteristics of each variable (including their respective length scales) might better preserve the relationship between them. Second, the correlation between two model variables is subject to sampling error in the same

way as the correlation between two locations: the RMS error should again be minimised by applying a localisation factor which depends on $P(r)$. However, traditional localisation damps the mean correlation between the variables, so that it no longer reflects the mean of the underlying balance. The unbiased eq. (12) avoids this problem. Detailed analysis of the balance properties of this revised localisation, and indeed how it might be applied in practical ensemble systems, is left to future work.

6. Sampling error on variances

Equation (6) expresses the optimal localisation factor for a single observation as a function of the mean and variance of the error on the sample gain, $\hat{b}$, defined in eq. (2). The derivation presented in Section 2.2 followed common practice in focussing on the sample correlations, neglecting the sampling error on variances. One might imagine that variance errors just add a bit more noise everywhere, without any systematic changes as a function of distance. However, results in Section 3 involving multi-scale covariances and weak assimilation showed better performance when variance errors were artificially excluded, suggesting their role may be worth further examination. Authors such as Raynaud et al. (2009) and Bonavita et al. (2012) have proposed spectral filtering of ensemble variances to reduce the RMS error of analyses based upon them.

Closer inspection highlights the fact that there are *two* variances involved in the single-observation Kalman gain eq. (2). The increment scales directly with the standard deviation σ_2 of the target gridpoint, whilst the standard deviation σ_1 of the model equivalent of the observation appears in both the numerator and denominator. Furthermore, there is a correlation between the sampling errors on these variances, which turns out to be proportional to r^2 , the square of the underlying background error correlations (Raynaud et al., 2009; their eq. (5)). Thus, $\hat{\sigma}_1$ and $\hat{\sigma}_2$ move from having uncorrelated sampling errors at $r = 0$ to having perfectly correlated errors at $r = \pm 1$.

The net effect of these sampling errors depends on the balance of terms in the denominator of $\hat{b}$, and thus on the ‘strength’ of the assimilation. This is the simplest analytic example of how the optimal localisation factor can depend on parameters relating to the observations in addition to those of the background. In the case of strong assimilation ($\sigma_o/\sigma_1 \rightarrow 0$), as may be applicable at convective scales, $\hat{b} \rightarrow \hat{\sigma}_2 \hat{r} / \hat{\sigma}_1$. As $r \rightarrow 1$ at short distances, the two sample variances become identical, cancel out, and satisfy our original assumption that the sampling error comes from correlations alone. In the case of weak assimilation ($\sigma_o/\sigma_1 \rightarrow \infty$), as may be applicable at global scales (Bowler et al., 2013),

$\hat{b} \rightarrow \hat{\sigma}_1 \hat{\sigma}_2 \hat{r} / \sigma_o^2$. In this limit, $\hat{b}$ follows the sampling error of the covariance $\hat{c} = \hat{\sigma}_1 \hat{\sigma}_2 \hat{r}$, which is approximately:

$$\sigma_{\hat{c}} \approx \sigma_1 \sigma_2 \sqrt{(1 + r^2)/N}, \quad (15)$$

adapting eq. (6) of Hamill et al. (2001). Whereas the sampling error on correlations decreased as $r^2 \rightarrow 1$, the sampling error on the covariance increases, and is non-zero at $r = \pm 1$. This implies, through eq. (6), that the optimal localisation factor may be slightly less than unity even for points which are directly observed.

The full behaviour of the sample gain is illustrated in Fig. 5 for a range of assimilation strengths (left to right) and ensemble sizes (top to bottom). These results have been obtained by constructing ensembles of the specified size where each member is a pair of variables sampled from a joint Gaussian distribution with the specified correlation. The sample gain $\hat{b}$ is calculated for each ensemble using eq. (2). The process is repeated 10^4 times for each

true correlation so that the mean (solid) and standard deviation (dotted) of the estimates can be plotted as a function of r .

The results are consistent with the expectations outlined above. The standard deviation generally lies between the theoretical estimates for correlations (dashed) and covariances (dot-dashed) alone, except for the smallest ensemble size where the approximations underlying the theoretical estimates start to break down. For strong assimilation (left), the standard deviation is slightly larger than for correlations alone, whilst for weak assimilation (right) it follows the result for a pure covariance, with a gradual transition between these two regimes. Whilst increased noise above that predicted for pure correlations will decrease the optimal localisation factor produced by eq. (6), this is counterbalanced by the $b/\hat{b}$ multiplicative correction for the low bias which appears in Fig. 5 at small ensemble sizes for intermediate assimilation strengths. Furthermore, the non-linear form of the damping factor $Q^2/(1+Q^2)$ gives it less sensitivity to

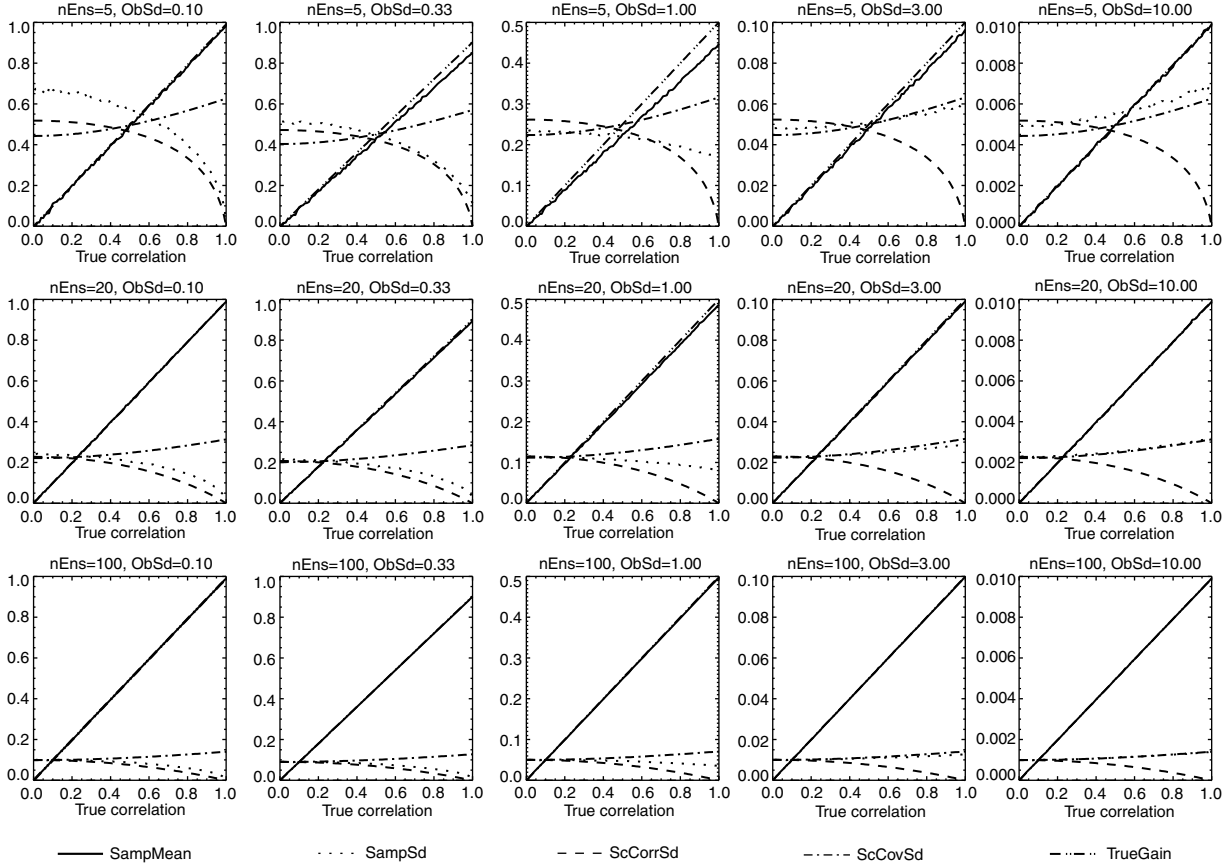


Fig. 5. The mean (solid) and standard deviation (dotted) of the sample gain calculated from ensembles of 5 (top), 20 (middle) or 100 (bottom) members, as a function of the true correlation (r) between the two points, with unit background error variances and observation standard deviations of 0.1, 0.33, 1.0, 3.0 and 10.0 (left to right). Additional lines indicate the gain implied by the true covariances (double-dot-dashed), and the standard deviation expected when the sampling error is associated with the correlation alone [dashed, by numerical integration with $P(\hat{r}|r)$ derived from eq. (9)], or the covariance alone [dot-dashed, using eq. (15)].

small changes in the signal-to-noise ratio Q once r and N combine to take Q above about 2.

It is unclear how localisation factors which allow for variance errors should be applied in a real system without a full multi-observation solution. When variance errors are neglected, eq. (3) amounts to minimising the element-wise mean square difference between $\tilde{\mathbf{P}}^f$ and $\mathbf{P}^f$, since the sampling errors in the denominator of $\hat{b}$ in eq. (2) are neglected. It seems reasonable to apply the resulting $\tilde{\mathbf{P}}^f$ in both the numerator and denominator of eq. (11). Once variance errors are included, the localisation applies to the whole single-observation gain $\hat{b}$ in eq. (2), so it is not clear what corresponding operation should be applied to the denominator of eq. (11). More detailed investigation is left to future work.

7. Dense observations

The theory developed thus far has been based on balancing signal and noise in the assimilation of a single observation where the Kalman gain is estimated from ensemble data. To extend the theory to high observation densities, we need to include within the balance the extra terms which arise in the denominator of $\tilde{\mathbf{K}}$, eq. (11), for the ensemble-estimated background error covariances between the observations. It turns out these are mixed up and combined in much more complicated ways than the sampling errors in the numerator, primarily due to the structure of the matrix inverse. This problem is not fully solved, and indeed the full analytic solution would probably be too expensive to implement. The purpose of this section is instead to illustrate some basic features of the multi-observation problem, and how it differs from the single-observation situation.

The equation we wish to solve is deceptively simple. In its most basic model-space form, we seek the localisation matrix $\mathbf{L}$ that, when multiplied element-wise by the sample covariance $\tilde{\mathbf{P}}^f$ and passed through eq. (11), minimises the error covariance $\mathbf{P}^a$ of the resulting analysis:

$$\frac{\partial \text{Tr}(\mathbf{P}^a)}{\partial \mathbf{L}} = 0. \quad (16)$$

Assuming as before that the sampling errors are statistically independent of the innovation errors, it can be shown that the posterior error covariance takes the familiar form:

$$\mathbf{P}^a = \langle (\mathbf{I} - \tilde{\mathbf{K}}\mathbf{H})\mathbf{P}^f(\mathbf{I} - \tilde{\mathbf{K}}\mathbf{H})^T \rangle + \langle \tilde{\mathbf{K}}\mathbf{R}\tilde{\mathbf{K}}^T \rangle, \quad (17)$$

where the expectations are required because the sampling errors make $\tilde{\mathbf{K}}$ a stochastic variable. Equation (16) is hard to solve because $\mathbf{L}$ appears in multiple places to various powers, including inside a matrix inverse, and we require the average over such terms. These interactions change the resulting optimal localisation factors compared to the

single-observation case; the Appendix illustrates this algebraically for the sum of two covariances.

In the single-observation case (Section 2.1), it was noted that our ultimate aim of minimising the analysis error variance is equivalent to eq. (3), which minimises the mean square difference between the optimal gain and the localised ensemble estimate. In the multi-observation case, the localisation-dependent part of the analysis error variance eq. (17) can be rearranged as:

$$J = \text{Tr} \langle (\tilde{\mathbf{K}} - \mathbf{K})\mathbf{S}(\tilde{\mathbf{K}} - \mathbf{K})^T \rangle, \quad (18)$$

where $\mathbf{S} = \langle \mathbf{z}\mathbf{z}^T \rangle = \mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}$ is the innovation covariance. Were the innovations completely uncorrelated, this would simply be a weighted mean of the MSE of the gain for each observation. The cross terms between the gains for different observations only affect the trace in so far as they are brought into it by correlations within the innovations.

One consequence of Bayes Theorem is that observations with independent observation errors can be assimilated sequentially without changing the result, at least in the absence of localisation, provided the posterior PDF is fully updated each time (Anderson, 2003, Section 2a). This suggests the multiple observation problem might be avoided by sequential assimilation. However, the assimilation of each observation changes the true background error covariance for the next observation, and thus the optimal localisation. Without some high-rank method (such as a larger ensemble) to track these covariance changes, it could be hard to keep this process optimal. Here we focus on analysing the impact of dense observations in a single assimilation step, where only one initial climatological covariance is required to establish the localisation.

7.1. Empirically determined localisation factors

The multi-observation localisation problem is awkward to solve analytically, but the above considerations suggest there could be important interactions between the terms, particularly those which arise from the denominator of the approximated KF eq. (11). To complement the theoretical results and develop intuition, the optimal localisation factor(s) have been found experimentally for a range of 2–3-gridpoint, 1–2 observation problems. A set of possible values is specified for each localisation factor, all combinations are tested, and the one giving the lowest RMS error over a large number of assimilation trials (here 10^5) is reported. Each complete experiment has been repeated multiple times with different random number seeds to confirm the key results are robust, although it should be noted that the RMS error is more sensitive to some parameters than others. To simplify the problem, variance errors have been artificially removed as described in Section 3. This avoids

the need to consider filtering the sample variances, and provides closer contact with most of the theory presented in this paper. Some tests have been performed with variance errors included, but are not described in detail; whilst these give different numerical results, they demonstrate many of the behaviours seen in the tests presented here.

Table 1 presents results for the two-gridpoint cases, where either one or both points are observed. The single-observation case matches the scenario considered in Section 2, and the results are consistent with the theoretical predictions of 0.44, 0.91 and 1.00 obtained with numerical integration of eq. (9). The addition of a second observation tends to reduce the optimal localisation factor, with stronger reductions of around 0.1 in tests with five ensemble members (not shown). This is consistent with the idea that, when dense observations are available, it is better to get more of the spatial structure of the correction to the background forecast from the observations themselves and less from the noisy ensemble-estimated correlations, which should therefore be localised more tightly in well-observed regions.

For the final rows of Table 1, the system is allowed to choose separate covariance localisation factors for the numerator and denominator of eq. (11). This reduces the optimal value of both localisation factors compared to a single choice, though this effect is less prominent when variance errors are included (not shown). The optimal localisation factor for the denominator is lower than that for the numerator. This is consistent with the suggestions made above, since the sample covariances from the denominator are involved in complicated sums, whereas those from the numerator remain simple multiplying factors.

Table 2 considers a more general two-observation scenario in which a third, unobserved, gridpoint is also updated. The correlation between the observed gridpoints is now called r_1 . For simplicity, the unobserved gridpoint

Table 1. The empirically determined optimal covariance localisation factor(s) between two gridpoints with background error correlation r and unit variances, where either one (nObs = 1) or both (nObs = 2) points are directly observed

nObs	Factor	r		
		0.20	0.50	0.75
1	α	0.45	0.90	1.00
2	α	0.45	0.85	0.95
2	α_u	0.25	0.70	0.90
	α_d	0.05	0.45	0.75

For nObs = 2, the system can be forced to use a single localisation factor (α), or allowed to choose separate values for the numerator (α_u) and denominator (α_d) of eq. (11). The listed values achieved the lowest RMS error from the set 0, 0.05, ..., 1 with 10^5 runs each, 20 ensemble members, and variance errors artificially removed.

Table 2. The empirically determined optimal covariance localisation factors between three gridpoints, two of which are observed, with background error correlation r_1 between the observed points, r_2 between the observed and unobserved points, and unit variances

Factor	r_2	r_1		
		0.20	0.50	0.75
α_{u1}	0.20	0.2	0.6	0.9
	0.50	0.4	0.6	0.9
	0.75	0.6	0.7	0.9
α_{u2}	0.20	0.4	0.4	0.4
	0.50	0.9	0.8	0.8
	0.75	1.0	0.9	0.9
α_{d1}	0.20	0.0	0.3	0.7
	0.50	0.4	0.3	0.7
	0.75	0.8	0.5	0.7

The system is allowed to choose separate localisation factors for r_1 (α_{u1}) and r_2 (α_{u2}) in the numerator and r_1 in the denominator (α_{d1}) of eq. (11). The listed values achieved the lowest RMS error from the set 0, 0.1, ..., 1 with 10^5 runs each, 20 ensemble members, and variance errors artificially removed.

has the same correlation, r_2 , with both observed gridpoints; the values have been chosen to ensure $\mathbf{P}^f$ is a legitimate (positive-definite) covariance matrix. The same scenario is considered in Section 6 of Bowler et al. (2013). Three separate localisation factors, α_{u1} , α_{u2} and α_{d1} , are optimised simultaneously. These are respectively applied to $\hat{r}_1$ in the numerator of eq. (11), $\hat{r}_2$ (which is only used in the numerator), and $\hat{r}_1$ in the denominator.

The localisation α_{u2} to the unobserved gridpoint takes fairly similar values as a function of r_2 to the single-observation case in Table 1, consistent with the fact that $\hat{r}_2$ simply scales the increment from each observation. The optimal α_{u2} is almost independent of the correlation between the observations, the slight reduction with increasing r_1 perhaps reflecting the increasingly correlated noise in the sample correlations with the two observed gridpoints.

The optimal numerator localisation between the observed gridpoints, α_{u1} , rises as usual with their true correlation r_1 , but also rises with the correlation r_2 to the unobserved gridpoint, particularly for low values of r_1 . This effect is also prominent when variance errors are included (not shown), yet it is hard to explain directly, since α_{u1} only affects the increment applied to the observed points, which does not involve $\hat{r}_2$. The denominator localisation, α_{d1} , also rises with r_2 at low r_1 , and this is easier to explain because the sample correlations from the denominator combine in various sums and products, which we have already seen can change the optimal localisation factor. It then seems plausible that α_{u1} has to make a corresponding adjustment in the numerator to stay closer to the cancellations and relationships which would arise in the perfect KF based on the true covariances.

In most localisation schemes, the same α_1 would be used for both numerator and denominator, particularly for variational methods where the denominator of eq. (11) is not represented explicitly. However, Table 2 again suggests the ideal α_{dl} is typically smaller than α_{ul} , except for the odd but apparently robust 0.8 at the bottom left of the table.

8. Discussion

Ensemble DA systems estimate flow-dependent background error statistics using a sample of possible states. Even if this sample is drawn from the correct PDF, the covariances will be subject to sampling errors arising from the finite size of the ensemble. The need for some form of ‘localisation’ to reduce damaging increments from spurious long-range sample correlations has long been recognised, but there has been little theoretical analysis of how sampling error changes the optimal estimation problem or what form a statistically optimal localisation might take.

In this paper, we have defined the statistically optimal localisation as that which minimises the analysis error variance. This widely used measure of proximity is convenient for initial investigation and matches the simplest KF derivation, but more sophisticated measures such as maximum likelihood could of course be explored in future, along with issues such as non-linearity. Minimising analysis error variance is equivalent to minimising the mean square difference between the ensemble estimate of the gain and its unknown true value, weighted by the innovation covariance.

In the case of a single observation, with variance errors neglected, the optimal localisation factor is a universal function of ensemble size and the true correlation between the background errors on the observed quantity and the gridpoint being updated. A cheap approximation of this result performed well in idealised tests for sparse observations, where the off-diagonal terms in the denominator of the KF equation can be neglected. This implies that localisation functions should vary with the length scale of the affected variables, and that correlations between variables (or times) should be localised like any other relationship. It also implies that the mapping from correlation to localisation should only be applied at the final step: even if horizontal and vertical correlations were separable, for example, the corresponding optimal localisation would not be, since it is a non-linear function of the final correlation.

This correlation-centric approach casts new light on the localisation of spatially integrating observations: the correct measure of ‘distance’ between a model quantity and a satellite radiance is their true background error correlation. This might be estimated by a climatological average of ensemble data, or by applying a linearised observation operator to climatological estimates of model-space correlations. The resulting localisation function is broader than

that between gridpoint quantities, with a peak value of less than unity. A detailed comparison with model-space localisation is left to future work, but we note that the model-space localisation which is optimal for the assimilation of point observations is generally not optimal for integrating observations.

The analysis highlights the fact that the traditional form of localisation, in which sample covariances are damped towards zero, leads to a low bias in the magnitude of the localised covariance and an unnecessary penalty in the analysis error variance. This suggests an improved form of hybrid DA in which localisation damps not the whole sample covariance, but merely its departure from the climatological mean. Such a formulation may also improve the dynamical balance of the increments, since the mean relationship between model variables is preserved.

The optimal localisation is derived by minimising a penalty J which is proportional to the mean square additional error of the EnDA over and above the perfect KF based on the true covariances. This emphasises the fact that even the optimal value of J usually exceeds zero, because the localisation is mixing two sources of information which both have errors. In the best case where the localisation eq. (12) damps departures from the climatological mean gain, a fraction α of the localised gain comes from the ensemble, and is thus subject to sampling error, whilst the fraction $(1-\alpha)$ comes from the climatological mean, about which the true case-specific gain has standard deviation σ_b . Whilst this paper has focussed on the central analysis, the update of the ensemble perturbations needs to be consistent with the errors in that analysis. This creates a potential advantage for ‘stochastic’ filters which directly sample the posterior PDF, even for suboptimal $\tilde{\mathbf{K}}$ [eq. (17)], over ‘deterministic’ filters, which are based on the simplified equation $\mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^f$ that assumes you have used the optimal gain based on the true covariances.

The fractional sampling error on variances is much smaller than the worst case for correlations. However, closer examination reveals a distance-dependent effect, arising from the correlation between the background error variance sampling errors for the observed and updated quantities. Variance errors are theoretically interesting because they demonstrate how, even for a single observation, the optimal localisation factor depends on the characteristics of the observations (specifically the ‘strength’ of the assimilation) in addition to those of the background. In practice, the largest impacts are in situations where the signal-to-noise ratio is already high, and there is some cancellation between the additional sampling error and low biases in the sample gain, so that the simplified optimal localisation (which neglects both effects) performs reasonably well.

The theoretical framework and idealised experimental results both show that a more sophisticated solution is

required in the case of dense observations. The minimisation of analysis error variance can in principle be applied to the full multi-observation KF, but the problem is difficult and almost certainly too computationally expensive to solve in its full generality. However, it is clear that the traditional notion of a single ‘best’ localisation which leads to a unique ‘localised covariance’ only applies in the limit of sparse observations, with variance errors neglected. In the simplest case, with a single localisation matrix applied through eq. (11), $\mathbf{L}$ turns out to be the solution of a seventh-order matrix polynomial defined by eqs. (16) and (17). Since this involves both products and sums of localised covariances, neither the traditional nor unbiased optimal localisation factors pass through unchanged. The empirically determined optimal localisation factors for small test problems support these ideas and suggest that analysis error variance can be further reduced by allowing the denominator to use tighter localisation factors than the numerator.

These considerations emphasise the fact that optimal EnDA may not actually take the precise form of a KF. The KF is derived as a balance between signal and noise in the background and observations assuming the error statistics are perfectly known. The solution happens to take the form of a ratio of covariance matrices. EnDA generalises this problem to include sampling error on the estimated background error statistics. This new problem leads to a new solution, which need not take the precise form of eq. (1); we could in principle get any function of the fundamental inputs (the ensemble members, $\mathbf{H}$ and $\mathbf{R}$). Even when we assume a single localisation matrix $\mathbf{L}$, eq. (6) can yield localisation factors greater than one due to biases in the sample gain. The non-linear way in which the optimal localisation factor depends on the true correlation, illustrated in Fig. 2, may well yield localisation matrices which are not positive-definite, given their tendency to combine shallow gradients at short distances with steep gradients at long distances, particularly for larger ensemble sizes. All of these features can be accommodated in ensemble filter formulations which manipulate the sample covariances directly, but may be difficult or impossible to achieve without further approximation in variational schemes, where the covariances are implicit, the KF denominator is not explicitly represented, and it is assumed that localisation is performed by a positive-definite correlation matrix.

This paper has focussed on methods which would generally be classed as ‘static’ localisation, where a fixed localisation matrix is applied regardless of the sample covariances. One awkward feature is that these derivations require us to propose a functional form for the generalised KF, and then optimise the coefficients; we do not optimise the functional form itself, and thus do not know that we have chosen the most efficient estimator for the analysis. A link with ‘adaptive’ methods, where the localisation factor

depends on the sample correlation, is potentially useful here. Section 2.4 extended the static localisation analysis to allow for a PDF $P(r)$ of true correlations. This allows both static and adaptive localisation to be seen as different approaches to the same underlying problem: choosing a localised gain based on sample covariances and a prior PDF of true covariances. This link is particularly direct with Anderson (2012), who performs an offline Monte Carlo calculation to find the distribution $P(\hat{r}|r)$ of sample correlations for an ensemble of size N about each true correlation. Inverting the table gives the Bayes’ Theorem $P(r|\hat{r})$ of true correlations when the ensemble predicts a particular sample correlation. However, we disagree with Anderson (2012) regarding the relationship between these results and the optimal adaptively localised correlation. Generalising from correlation to the gain for a single observation, the correct analogue of eq. (3) is to choose the localised gain $\tilde{b}$ which minimises the mean square difference between $\tilde{b}$ and each of the possible values b of the true optimal gain:

$$J = \int (\tilde{b} - b)^2 P(b|\hat{b}) db. \quad (19)$$

This principle is used in Zhen and Zhang (2014; their eq. (5)) to adaptively estimate the radius of a prescribed localisation function. Note crucially that eq. (19) integrates over the PDF of b for the measured $\hat{b}$, whereas eq. (3) integrated over the PDF of $\hat{b}$ for a given set of parameters $\boldsymbol{\theta}$ (which imply b). Setting $dJ/d\tilde{b} = 0$, we obtain $\tilde{b} = \int b P(b|\hat{b}) db$, which is just the standard result that the mean is the value with minimum RMS difference from the distribution of values which it characterises. This mean is already the optimal adaptively localised correlation, without any further damping. The Bayesian conversion from $P(\hat{b}|\boldsymbol{\theta})$ to $P(b|\hat{b})$ has removed the need to assume a functional form for the optimal gain, and since it gives the full conditional PDF we know there cannot be a more efficient estimator based on the same predictors. One could also avoid the assumption that the sample gain is the best predictor by constructing the table based on, say, the sample variances and correlation. This approach may be helpful for exploring the form of the optimal gain for a single observation in the presence of variance error, but the extension to multiple observations would require a multi-dimensional Bayesian inversion, which would be prohibitively expensive in general. This analysis also highlights the fact that, since they solve the same underlying problem, ‘adaptive’ localisation methods are likely to need at least as strong a prior (e.g. specified as a function of distance) as static localisation in order to achieve competitive performance.

The work presented in this paper could be taken forward in a variety of ways. A method has been presented for calculating the statistically optimal localisation function for sparse observations, but without a simpler theory for

dense observations, some tuning and ad-hoc tightening will still be required in practical systems. Simpler high-level approximations for the dependence of analysis error on localisation radius and observation density, such as developed by Periañez et al. (2014) may be helpful for exploring this problem. The main benefit of the present work is improving fundamental intuition about how the form of the optimal localisation function is affected by the various system parameters. At the Met Office, these insights will be used to guide the localisation choices in the forthcoming development of a prototype convective-scale EnDA system for the UK.

9. Acknowledgements

The author thanks Neill Bowler and Adam Clayton for helpful discussions during the development of this work. Comments from Gordon Inverarity and two anonymous reviewers helped to improve the quality of the manuscript.

10. Appendix

Section 7 demonstrates that the multi-observation problem is hard to solve explicitly, but the error variance to be minimised involves sums of products of the individual covariances. An alternative approach is therefore to ask how the localisation factors which optimise a sum or product of covariances are related to those which optimise the individual contributions. This analysis is also relevant to model-space localisation of integrating observations (Section 4), since that involves a sum of covariances, with weights given by $\mathbf{H}$.

When traditional damping localisation is applied to a sum of covariances, the penalty eq. (3) becomes:

$$\begin{aligned} J &= < ((\alpha\hat{a} + \beta\hat{b}) - (a + b))^2 > \\ &= < (\alpha\hat{a} - a)^2 + (\beta\hat{b} - b)^2 \\ &\quad + 2(\alpha\delta_a + (\alpha - 1)a)(\beta\delta_b + (\beta - 1)b) >, \end{aligned} \quad (20)$$

writing $\delta_a = \hat{a} - a$ (which is unbiased for sample covariances). The final term averages to $2\alpha\beta C_{ab} + 2(\alpha - 1)(\beta - 1)ab$, where $C_{ab} = < \delta_a \delta_b >$ is the covariance of the sampling errors. Setting $dJ/d\alpha = 0$ yields:

$$\begin{aligned} \alpha(a^2 + \sigma_a^2) - a^2 + (\beta - 1)ab + \beta C_{ab} &= 0 \\ \Rightarrow \alpha &= \frac{a^2 + (1 - \beta)ab - \beta C_{ab}}{a^2 + \sigma_a^2}, \end{aligned} \quad (21)$$

with a similar equation for β . These equations are coupled, so neither localisation factor can be found without considering the statistics relevant to the other, and the substitution to eliminate β will generally lead to a higher-order polynomial in α . The numerator contains two additional terms compared to eq. (7). The factor $(1 - \beta)b$ is the

magnitude of the bias introduced into $\tilde{b} = \beta\hat{b}$ by the use of a damping localisation; the penalty is sensitive to the product of these biases, increasing the optimal value of α . This correction will be largest when b is moderately large in comparison to a , but not so large that its localisation factor β is close to unity. The final term in the numerator of eq. (21) reduces α in situations (such as short distances) where β and C_{ab} are large. The presence of both observations has changed the signal-to-noise balance so that slightly more information on the spatial variation of forecast error is taken from the observations as opposed to the noisy ensemble covariances.

Corresponding results exhibiting similar issues can be derived for the unbiased localisation factors introduced in Section 5, and for the optimal localisation of a product of covariances.

References

- Anderson, J. L. 2001. An ensemble adjustment Kalman filter for data assimilation. *Mon. Weather Rev.* **129**, 2884–2903.
- Anderson, J. L. 2003. A local least-squares framework for ensemble filtering. *Mon. Weather Rev.* **131**, 634–642.
- Anderson, J. L. 2007. Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. *Physica D* **230**, 99–111.
- Anderson, J. L. 2009. Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus* **61A**, 72–83.
- Anderson, J. L. 2012. Localization and sampling error correction in ensemble Kalman filter data assimilation. *Mon. Weather Rev.* **140**, 2359–2371.
- Anderson, J. and Lei, L. 2013. Empirical localisation of observation impact in ensemble Kalman filters. *Mon. Weather Rev.* **141**, 4140–4153.
- Bishop, C. H. and Hodyss, D. 2009. Ensemble covariances adaptively localized with ECO-RAP. Part I: tests on simple error models. *Tellus* **61A**, 84–96.
- Bonavita, M., Isaksen, L. and Holm, E. 2012. On the use of EDA background error variances in the ECMWF 4D-Var. *Q. J. R. Meteorol. Soc.* **138**, 1540–1559.
- Bowler, N. E., Flowerdew, J. and Pring, S. 2013. Tests of different flavours of EnKF on a simple model. *Q. J. R. Meteorol. Soc.* **139**, 1505–1519.
- Buehner, M. 2011. Evaluation of a spatial/spectral covariance localisation approach for atmospheric data assimilation. *Mon. Weather Rev.* **140**, 617–636.
- Buehner, M. and Charron, M. 2007. Spectral and spatial localisation of background-error correlations for data assimilation. *Q. J. R. Meteorol. Soc.* **133**, 615–630.
- Campbell, W. F., Bishop, C. H. and Hodyss, D. 2010. Vertical covariance localization for satellite radiances in ensemble Kalman filters. *Mon. Weather Rev.* **138**, 282–290.
- Casella, G. and Berger, R. L. 2002. *Statistical Inference*. Cengage Learning, Boston, MA, USA.
- Dong, J., Xue, M. and Droegemeier, K. 2011. The analysis and impact of simulated high-resolution surface observations in

- addition to radar data for convective storms with an ensemble Kalman filter. *Meteorol. Atmos. Phys.* **112**, 41–61.
- Fisher, R. A. 1921. On the ‘probable error’ of a coefficient of correlation deduced from a small sample. *Metron* **1**, 3–32.
- Flowerdew, J. and Bowler, N. E. 2013. Online calibration of the vertical distribution of ensemble spread. *Q. J. R. Meteorol. Soc.* **139**, 1863–1874.
- Gaspari, G. and Cohn, S. E. 1999. Construction of correlation functions in two and three dimensions. *Q. J. R. Meteorol. Soc.* **125**, 723–757.
- Hamill, T. M. and Whitaker, J. S. 2005. Accounting for the error due to unresolved scales in ensemble data assimilation: a comparison of different approaches. *Mon. Weather Rev.* **133**, 3132–3147.
- Hamill, T. M., Whitaker, J. S. and Snyder, C. 2001. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Weather Rev.* **129**, 2776–2790.
- Houtekamer, P. L. and Mitchell, H. L. 1998. Data assimilation using an ensemble Kalman filter technique. *Mon. Weather Rev.* **126**, 796–811.
- Houtekamer, P. L. and Mitchell, H. L. 2001. A sequential ensemble filter for atmospheric data assimilation. *Mon. Weather Rev.* **129**, 123–137.
- Houtekamer, P. L., Mitchell, H. L., Pellerin, G., Buehner, M., Charron, M. and co-authors. 2005. Atmospheric data assimilation with an ensemble Kalman filter: results with real observations. *Mon. Weather Rev.* **133**, 604–620.
- Hunt, B. R., Kostelich, E. J. and Szunyogh, I. 2007. Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter. *Physica D* **230**, 112–126.
- Kalman, R. E. 1960. A new approach to linear filtering and prediction problems. *J. Basic Eng.* **82**, 35–45.
- Kepert, J. D. 2009. Covariance localisation and balance in an ensemble Kalman filter. *Q. J. R. Meteorol. Soc.* **135**, 1157–1176.
- Kirchgessner, P., Nerge, L. and Bunse-Gerstner, A. 2014. On the choice of an optimal localisation radius in ensemble Kalman filter methods. *Mon. Weather Rev.* **142**, 2165–2175.
- Lei, L. and Anderson, J. L. 2014. Comparisons of empirical localisation techniques for serial ensemble Kalman filters in a simple atmospheric general circulation model. *Mon. Weather Rev.* **142**, 739–754.
- 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.
- Nahman, N. S. and Guillaume, M. E. 1981. *Deconvolution of Time Domain Waveforms in the Presence of Noise*. National Bureau of Standards Tech. Note 1047.
- Periáñez, Á., Reich, H. and Potthast, R. 2014. Optimal localization for ensemble Kalman filter systems. *J. Met. Soc. Japan.* **62**, 585–597.
- Raynaud, L., Berre, L. and Desroziers, G. 2009. Objective filtering of ensemble-based background-error variances. *Q. J. R. Meteorol. Soc.* **135**, 1177–1199.
- Wang, X. and Bishop, C. H. 2003. A comparison of breeding and ensemble transform Kalman filter ensemble forecast systems. *J. Atmos. Sci.* **60**, 1140–1158.
- Whitaker, J. S. and Hamill, T. M. 2002. Ensemble data assimilation without perturbed observations. *Mon. Weather Rev.* **130**, 1913–1924.
- Zhang, F., Snyder, C. and Juang, S. 2004. Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Weather Rev.* **132**, 1238–1253.
- Zhen, Y. and Zhang, F. 2014. A probabilistic approach to adaptive covariance localisation for serial ensemble square-root filters. *Mon. Weather Rev.* **142**, 4499–4518.