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

Data compression in the presence of observational error correlations

By A.M. FOWLER[1,2]*, [1]*Department of Meteorology, University of Reading, Reading, UK;* [2]*National Centre for Earth Observation, University of Reading, Reading, UK*

(Manuscript received 20 August 2018; in final form 10 June 2019)

ABSTRACT

Numerical weather prediction (NWP) models are moving towards km-scale (and smaller) resolutions in order to forecast high-impact weather. As the resolution of NWP models increase the need for high-resolution observations to constrain these models also increases. A major hurdle to the assimilation of dense observations in NWP is the presence of non-negligible observation error correlations (OECs). Despite the difficulty in estimating these error correlations, progress is being made, with centres around the world now explicitly accounting for OECs in a variety of observation types. This paper explores how to make efficient use of this potentially dramatic increase in the amount of data available for assimilation. In an idealised framework it is illustrated that as the length-scales of the OECs increase the scales that the analysis is most sensitive to the observations become smaller. This implies that a denser network of observations is more beneficial with increasing OEC length-scales. However, the computational and storage burden associated with such a dense network may not be feasible. To reduce the amount of data, a compression technique based on retaining the maximum information content of the observations can be used. When the OEC length-scales are large (in comparison to the prior error correlations), the data compression will select observations of the smaller scales for assimilation whilst throwing out the larger scale information. In this case it is shown that there is a discrepancy between the observations with the maximum information and those that minimise the analysis error variances. Experiments are performed using the Ensemble Kalman Filter and the Lorenz-1996 model, comparing different forms of data reduction. It is found that as the OEC length-scales increase the assimilation becomes more sensitive to the choice of data reduction technique.

Keywords: observation network design, data assimilation, degrees of freedom for signal, mutual information

1. Introduction

In numerical weather prediction (NWP), the assimilation of observations and prior information (commonly referred to as the background) is performed routinely to find the most probable state of the atmosphere represented on the model grid. Data assimilation (DA) has proven to be essential for accurate weather forecasting by providing the initial conditions for NWP (Rabier et al., 2000; Rawlins et al., 2007).

With increasing computer power, the resolution of NWP models are increasing with the aim of improving the accuracy of high-impact small-scale events. In order to constrain these models there is an increasing need for higher spatial and temporal resolution observations. These observational needs may potentially be met by the next generation of geostationary satellites (e.g. Advanced Himawari Imager onboard Himawari-8, The Infrared Sounder onboard MTG-S) and phased array weather Radars (Miyoshi et al., 2016). However, there are currently many factors limiting the use of the large datasets produced by these observations. These include issues with data storage, computational time and complications caused by correlated errors, which are difficult to estimate and complicate the DA algorithms (Liu and Rabier, 2003). To address these issues observations are commonly thinned (both spatially, temporally and spectrally) (e.g. Rabier et al., 2002; Dando et al., 2007; Migliorini, 2015; Fowler, 2017) and 'super-obbed' (Berger and Forsythe, 2004; van Leeuwen, 2015). Variances are also commonly inflated (Hilton et al., 2009) to reduce the impact of the sub-optimal use of observations that are known to have correlated errors not represented during the assimilation.

*Corresponding author. e-mail: a.m.fowler@reading.ac.uk

Citation: Tellus A: 2019, **71**, 1634937, https://doi.org/10.1080/16000870.2019.1634937

The main hurdle to the use of observation error correlations (OECs) in DA is the difficulty in estimating them as they are often attributed to uncertainty in the comparison of the observations to the model variables, known as representation error, rather than instrument noise (see Janjic et al., 2017 and references there in). As such they can be state and model dependent (Waller et al., 2014) and are predominantly found in observation types with complex observation operators (such as satellite radiances Bormann and Bauer, 2010; Stewart et al., 2014; Waller et al., 2016a; Bormann et al., 2016; Campbell et al., 2017), observation types measuring features with natural length- and time-scales that are different to those resolved by the model (e.g. Doppler radial winds Waller et al., 2016b), and high level observation products which go through a large amount of preprocessing (e.g. atmospheric motion vectors (AMVs) derived from satellite radiances Bormann et al., 2003; Cordoba et al., 2017). However, progress is being made in estimating and allowing for the presence of correlated observation errors in DA operationally (e.g. Weston et al. 2014; Bormann et al., 2016; Campbell et al., 2017), potentially facilitating the assimilation of much denser observations. This barrier to the use of dense observations is therefore reducing, however issues with computational effort still remain.

Instead of regular thinning, the compression of the data may allow for much of the information within the observations to be retained while still reducing the computational cost of assimilating the data. Data compression, via principle component analysis, has previously been suggested and tested on hyper-spectral satellite instruments such as AIRS and IASI (Collard et al., 2010). It is generally found that the atmospheric signal contained in the original spectrum of a few thousand radiances can be represented by about 200 globally generated principal components (Goldberg et al., 2003). However, the observations that provide the greatest constraint for estimating the initial state depend not only on the observations and their uncertainty but also on the information already provided by the prior. This is acknowledged in channel selection methods, which use metrics such as mutual information and degrees of freedom for signal to choose a subset of the most informative channels for selection based on a typical prior error covariance matrix (e.g. Rabier et al. 2002; Migliorini, 2015; Fowler, 2017). However, the prior error covariance matrix can be highly flow dependent, as such the information content of the observations will also be flow dependent. This flow-dependency is naturally represented by DA methods based on the Ensemble Kalman Filter.

Other methods for intelligent network design include adaptive methods such as targeted observations. These methods depend not only on the observation and prior uncertainty but also on the error growth in the forecast model. This allows regions to be identified where additional observations would reduce the forecast error. A variety of techniques have been developed to measure the forecast sensitivity to perturbations to the initial conditions, including those based on the forecast adjoint model, singular vectors and ensemble techniques (see review by Majumdar, 2016).

For high-resolution forecasting, which is the focus of this paper, nested grids are often used with lead-times of $0-12$ h. For example, the UK Met Office runs a nested grid over the UK at a resolution of 1.5 km concentrating on precipitation forecasts out to 6 h (Li et al., 2018). Rapid up-date forecasting systems are also under development at centres such as RIKEN-Center for Computational Science, that perform DA at resolutions up to 100 m on a local domain in order to produce 30 min forecasts (Miyoshi et al., 2016). At these shorter lead times, the influence of the observation on the analysis, rather than the forecast, becomes more insightful in defining an optimal data compression strategy.

This work demonstrates how spatially correlated observation errors affect the influence of the observations on the analysis. In turn this is used to explain how spatially correlated observation errors affect the use of data reduction techniques within the ensemble Kalman Filter.

It is known that observations with correlated errors provide more information about small-scale features than observations with uncorrelated error (Seaman, 1977; Rainwater et al., 2015; Fowler et al., 2018). As such the use of dense observations is shown in this manuscript to be much more beneficial when the errors are correlated than when they are uncorrelated. This contradicts previous results of Liu and Rabier (2002) and Bergman and Bonner (1976), which both showed that even when OECs are correctly modelled, as the density of observations with correlated error is increased the reduction in analysis root mean square error (RMSE) is smaller than when the observation errors are uncorrelated. In Fowler et al. (2018), this result was shown to depend upon not only the OECs but how close the likelihood probability distribution function (PDF) structure (representing the uncertainty in the observations) is to the structure of the prior PDF. It is only when increasing the OECs brings the likelihood PDF more in line with the prior PDF that the observations have a reduced ability to correct the analysis RMSE. In addition to this it was shown in Fowler et al. (2018) that the analysis RMSE (or analysis error variance) only gives a partial measure of the effect of allowing for OECs in the assimilation, and as will be demonstrated further here, is more sensitive to large scale corrections than small-scale corrections.

This manuscript is organised as follows. In Section 2, the DA theory at the heart of the ensemble Kalman filter (EnKF) is presented. In Section 3, a method of optimal data compression, based on that of Xu (2007) and Migliorini (2013), is presented in which the observations retained have maximum information content. In Section 3.1, the impact of the OEC length-scale on the data compression is explored in a simplified framework. Then in Section 3.1, experiments are performed with the Lorenz 1996 model using the EnKF, comparing different strategies for data reduction.

2. Ensemble Kalman filter

Flavours of the EnKF, which merges Kalman filter theory with Monte Carlo estimation methods (Evensen, 1994), are currently in use at many operational centres (see review by Houtekamer and Zhang, 2016).

The EnKF makes the assumption that both the background, $\mathbf{x}^{\mathrm{b}} \in \mathbb{R}^n$, and observation, $\mathbf{y} \in \mathbb{R}^p$, errors are unbiased and Gaussian.

$$\mathbf{x}^{\mathrm{b}} - \mathbf{x}^{\mathrm{t}} \sim N(0, \mathbf{B}), \tag{1}$$

$$\mathbf{y} - h(\mathbf{x}^{\mathrm{t}}) \sim N(0, \mathbf{R}), \tag{2}$$

where $\mathbf{x}^{\mathrm{t}} \in \mathbb{R}^n$ represents the truth in state space. $\mathbf{B} \in \mathbb{R}^{n\times n}$ and $\mathbf{R} \in \mathbb{R}^{p\times p}$ are the prior and observation error covariance matrices, respectively. $h : \mathbb{R}^n \to \mathbb{R}^p$ is the observation operator (the mapping from state to observation space). These assumptions lead to an analytical form for the analysis, $\mathbf{x}^{\mathrm{a}}$ (the state that maximises the posterior probability)

$$\mathbf{x}^{\mathrm{a}} = \mathbf{x}^{\mathrm{b}} + \mathbf{K}\big(\mathbf{y} - h(\mathbf{x}^{\mathrm{b}})\big). \tag{3}$$

The matrix $\mathbf{K} \in \mathbb{R}^{n\times p}$ is known as the Kalman gain and can be given by

$$\mathbf{K} = (\mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}, \tag{4}$$

where $\mathbf{H} \in \mathbb{R}^{p\times n}$ is the observation operator linearised about the best estimate of the state.

The analysis error covariance matrix, $\mathbf{P}^{\mathrm{a}} \in \mathbb{R}^{n\times n}$, can be shown to be the inverse of the sum of the background and observation accuracies in state space ($\mathbf{B}^{-1}$ and $\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}$, respectively),

$$\mathbf{P}^{\mathrm{a}} = (\mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H})^{-1}. \tag{5}$$

See chapter 5 of Kalnay (2003) for a derivation of Eqs. (3)–(5).

In the EnKF the prior uncertainty is represented by an ensemble of model realisations, $\mathbf{x}^{\mathrm{f}} \in \mathbb{R}^{n\times N}$ where N is the size of the ensemble. The ensemble mean, $\bar{\mathbf{x}}^{\mathrm{f}} \in \mathbb{R}^n$, is computed as $\bar{\mathbf{x}}^{\mathrm{f}} = \frac{1}{N}\sum_{j=1}^{N}\mathbf{x}_j^{\mathrm{f}}$. The perturbation matrix is given by $\mathbf{X}^{\mathrm{f}} = \mathbf{x}^{\mathrm{f}} - \mathbf{x}^{\mathrm{f}}\mathbf{1}_N$, where $\mathbf{1}_N$ is a row vector of length N with each element given by 1.

An approximation to the background error covariance matrix is then given by $\mathbf{B} \approx \frac{1}{N-1}\mathbf{X}^{\mathrm{f}}(\mathbf{X}^{\mathrm{f}})^{\mathrm{T}}$. Note that if $N \leq n$ then this is rank deficient and non-invertible. In an attempt to regularise the ensemble estimate of $\mathbf{B}$ (and in turn expand the directions that the observations can update the prior estimate) ad-hoc localisation and inflation are often performed (Hamill et al., 2001; Oke et al., 2007). For the idealised experiments shown here, these techniques are not necessary and a study of their impact will be left for future work.

At the time of the observations, the ensemble is updated given the observation values and the observation error covariance matrix. The updated ensemble mean is a linear combination of the prior mean and the observations, analogous to (3),

$$\bar{\mathbf{x}}^{\mathrm{a}} = \bar{\mathbf{x}}^{\mathrm{f}} + \mathbf{K}\big(\mathbf{y} - h(\bar{\mathbf{x}}^{\mathrm{f}})\big). \tag{6}$$

The Kalman gain is now approximated as

$$\mathbf{K} = \mathbf{X}^{\mathrm{f}}\Big((N-1)\mathbf{I}_N + (\mathbf{Y}^{\mathrm{f}})^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{Y}^{\mathrm{f}}\Big)^{-1}(\mathbf{Y}^{\mathrm{f}})^{\mathrm{T}}\mathbf{R}^{-1}, \tag{7}$$

where $\mathbf{Y}^{\mathrm{f}} = \mathbf{H}\mathbf{X}^{\mathrm{f}}$ is the perturbation matrix transformed to observation space.

The updated ensemble perturbation matrix can be given by Hunt et al. (2007)

$$\mathbf{X}^{\mathrm{a}} = \sqrt{(N-1)}\mathbf{X}^{\mathrm{b}}\Big((N-1)\mathbf{I}_N + (\mathbf{Y}^{\mathrm{f}})^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{Y}^{\mathrm{f}}\Big)^{-1/2}. \tag{8}$$

An approximation to the analysis error covariance matrix (5) is then given by $\mathbf{P}^{\mathrm{a}} \approx \frac{1}{N-1}\mathbf{X}^{\mathrm{a}}(\mathbf{X}^{\mathrm{a}})^{\mathrm{T}}$. Equations (6) and (8) allow the updated ensemble to be reconstructed $\mathbf{x}^{\mathrm{a}} = \mathbf{X}^{\mathrm{a}} + \bar{\mathbf{x}}^{\mathrm{a}}\mathbf{1}_N$. The updated ensemble is then propagated forward using the model equations and stochastic forcing (to represent model error) to the time of the next assimilation step.

3. Information content of observations and data compression

The information content of the observations at each assimilation time can be related to the sensitivity of the analysis (the updated ensemble mean) to the observations. This is given by the influence matrix, $\mathbf{S} \in \mathbb{R}^{p\times p}$,

$$\mathbf{S} = \frac{\partial h(\mathbf{x}^{\mathrm{a}})}{\partial \mathbf{y}} = \mathbf{K}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}} \tag{9}$$

Cardinali et al. (2004).

A scalar summary of the influence matrix is then commonly given by two quantities; the degrees of freedom for signal, *DFS*, and entropy reduction, *ER* (also known as mutual information, and Shannon information content

(e.g. Huang and Purser, 1996; Rodgers, 2000; Fowler et al., 2018).

The *DFS* is commonly defined as $E[(\mathbf{x}^{\mathrm{a}}-\mathbf{x}^{\mathrm{b}})^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}^{\mathrm{a}}-\mathbf{x}^{\mathrm{b}})]$, where $E[\chi]$ is the expectation of χ. In an optimal system this is equivalent to

$$DFS = trace(\mathbf{S}) = \sum_{k=0}^{p-1}\lambda_k^{\mathrm{s}}, \tag{10}$$

where λ_k^{s} is the kth eigenvalue of $\mathbf{S}$ (Rodgers, 2000).

The *ER* is defined as the reduction in entropy from the prior to the posterior, where entropy of a PDF is given by

$$H(\mathbf{x}) = -\int P(\mathbf{x})\ln\left(P(\mathbf{x})\right)\mathbf{x}. \tag{11}$$

When $\mathbf{x}$ follows an unbiased Gaussian distribution, $N(0,\mathbf{\Sigma})$, the entropy can be evaluated as

$$H(\mathbf{x}) = \frac{1}{2}\ln|2\pi e\mathbf{\Sigma}|. \tag{12}$$

Therefore, in an optimal system *ER* is given by

$$\begin{aligned} ER &= H(\mathbf{x}) - H(\mathbf{x}|\mathbf{y}) \\ &= 0.5\ln det\left(\mathbf{B}(\mathbf{P}^{\mathrm{a}})^{-1}\right) \\ &= -0.5\ln det(\mathbf{I}-\mathbf{S}) \\ &= -0.5\sum_{k=0}^{p-1}\ln\left(1-\lambda_k^{\mathrm{s}}\right). \end{aligned} \tag{13}$$

From equations (10) and (13) we see that, the observations associated with the largest eigenvalues of $\mathbf{S}$ have the greatest information content as measured by both *DFS* and *ER*.

The number of non-zero eigenvalues of $\mathbf{S}$ will be bounded above by $min(N,n,p)$. In most practical applications this will be given by the ensemble size N (Migliorini, 2013). The use of localisation could help to increase this bound by increasing the rank of the ensemble estimate of $\mathbf{B}$.

Let $\mathbf{M} = \mathbf{R}^{-1/2}\mathbf{H}\mathbf{B}^{1/2} \approx \frac{1}{\sqrt{(N-1)}}\mathbf{R}^{-1/2}\mathbf{Y}^{\mathrm{f}}$. It can be shown that

$$DFS = trace\left(\mathbf{M}\mathbf{M}^{\mathrm{T}}(\mathbf{I}+\mathbf{M}\mathbf{M}^{\mathrm{T}})^{-1}\right), \tag{14}$$

and

$$ER = 0.5\ln det(\mathbf{I}+\mathbf{M}\mathbf{M}^{\mathrm{T}}). \tag{15}$$

Therefore, the left singular vector of $\mathbf{M}$ can be used to compress the observations with minimum information loss (Xu et al., 2009).

Let the compression matrix $\mathbf{C} \in \mathbb{R}^{p_{\mathrm{c}}\times p}$ be given by

$$\mathbf{C} = \mathbf{I}^{\mathrm{c}}\mathbf{U}^{\mathrm{T}}\mathbf{R}^{-1/2}, \tag{16}$$

where $\mathbf{U}$ is a matrix whose columns contain the eigen vectors of $\mathbf{M}\mathbf{M}^{\mathrm{T}} = \mathbf{U}\mathbf{\Lambda}^{M^2}\mathbf{U}^{\mathrm{T}}$ ordered with respect to the magnitude of the eigenvalues. $\mathbf{I}^{\mathrm{c}} \in \mathbb{R}^{p_{\mathrm{c}}\times p}$ is a matrix of zeros except for the elements $\mathbf{I}^{\mathrm{c}}_{kk}, \forall k = 1,\ldots,p_{\mathrm{c}}$, which are equal to one. The size of p_{c} can be chosen to fulfil some criterion of desirable information loss. For example, In Migliorini (2013), a threshold on the singular values of $\mathbf{M}$ (a measure of the signal-to-noise ratio) was used to decide the number of compressed observations to be assimilated.

The observation operator that transforms the state to the space of the compressed observations then becomes

$$h^{\mathrm{c}}(\mathbf{x}) = \mathbf{C}h(\mathbf{x}). \tag{17}$$

The assimilated compressed observations are

$$\mathbf{y}^{\mathrm{c}} = \mathbf{C}\mathbf{y}. \tag{18}$$

The error covariance matrix for the compressed observations is

$$\mathbf{R}^{\mathrm{c}} = \mathbf{C}\mathbf{R}\mathbf{C}^{\mathrm{T}}, \tag{19}$$

which in this case is $\mathbf{I}^{\mathrm{c}}(\mathbf{I}^{\mathrm{c}})^{\mathrm{T}} = \mathbf{I}_{p_{\mathrm{c}}}$.

3.1. Idealised circulant framework

A natural framework for understanding the effect of correlated observation error on the optimal compression of data makes use of circulant matrices. That is the error covariances are assumed to be homogeneous and isotropic. This allows the correlation structure to be described by a single correlation function. In this case matrices of the same dimension have common eigenvectors given by the discrete Fourier basis, $\mathbf{F}$, and the eigenvalues are ordered according to wavenumber (Gray, 2006).

Let us assume that the state is directly observed such that $\mathbf{H} = \mathbf{I}_n$ and $\mathbf{B} = \beta\mathbf{F}\mathbf{\Gamma}\mathbf{F}^{\mathrm{T}}, \mathbf{R} = \rho\mathbf{F}\mathbf{\Psi}\mathbf{F}^{\mathrm{T}}$, where $\mathbf{\Gamma}$ and $\mathbf{\Psi}$ are diagonal matrices containing the eigenvalues of the prior and observation error correlations respectively. Then we can express the eigenvalues of $\mathbf{M}\mathbf{M}^{\mathrm{T}}$ as

$$\lambda_k^{M^2} = \beta\gamma_k/\rho\psi_k, \tag{20}$$

where $\beta\gamma_k$ and $\rho\psi_k$ are the kth eigenvalue of $\mathbf{B}$ and $\mathbf{R}$, respectively. The eigenvectors of $\mathbf{M}\mathbf{M}^{\mathrm{T}}$ are given by $\mathbf{F}$.

If instead of ordering the eigenvalues with respect to wavenumber we order them in descending order of magnitude of λ^{M^2} then we can express the *DFS* and *ER* in terms of a truncation of the first $p_{\mathrm{c}}<p$ compressed observations. Let these be referred to as DFS^{c} and ER^{c}.

$$DFS^{\mathrm{c}} = \sum_{k=1}^{p_{\mathrm{c}}}\lambda_k^{M^2}/\left(1+\lambda_k^{M^2}\right), \tag{21}$$

$$ER^{\mathrm{c}} = \sum_{k=1}^{p_{\mathrm{c}}}\ln\left(1+\lambda_k^{M^2}\right)^{1/2}. \tag{22}$$

As the ratio β/ρ is independent of the k^{th} wavenumber, the choice of data compression will depend only on the spectrum of the error correlations and not on their variances.

Similarly we can give an expression for $trace(\mathbf{B}-\mathbf{P}^{\mathrm{a}})$ (approximately the reduction in ensemble spread) arising

from assimilating the first p_c compressed observations,

$$trace(\mathbf{B}-\mathbf{P}^a)^c = \sum_{k=1}^{p_c} \frac{\beta\gamma_k \lambda_k^{M^2}}{1+\lambda_k^{M^2}}. \quad (23)$$

In this simple framework, we see that compressing the observations with respect to the largest eigenvalues of $\mathbf{MM}^T$ (or equivalently $\mathbf{S}$) will not necessarily ensure that the observations that have the greatest effect on reducing the ensemble spread are also selected. This will depend on the eigenvalues of $\mathbf{B}$. In the case of uncorrelated observation errors, this discrepancy between the observations with the maximum information and those that minimise the analysis error variances is not present, as in this case $\psi_k = 1, \forall k$ and so $\lambda_k^{M^2} \propto \gamma_k$.

3.2. Simple numerical illustration

In the following experiments, a simple circular domain is considered of length 64π, discretised into 32 evenly spaced points. The state is observed directly at each grid point. The circulant $\mathbf{R}$ and $\mathbf{B}$ matrices are characterised by a SOAR (Second-Order Auto-Regressive) function with length-scales L_R and L_B, respectively. The SOAR function is defined as

$$c_m = (1 + r_m/L)e^{-r_m/L}, \quad (24)$$

where r_m is the distance between two points and L is the correlation length-scale. The error variances are assumed to be 1 for both the prior and observations.

In each of the following experiments, $L_B = 5$. This results in correlations greater than 0.2 out to six grid points and an entropy of the prior, given by $0.5\ln|2\pi e\mathbf{B}|$, of 36.1 (to 3 significant digits). The top panels of Figs. 1–3 show the eigenvalues of $\mathbf{R}$ (blue triangles) and $\mathbf{HBH}^T$ (red circles) as a function of wavenumber. $L_R = 0.1$ (Fig. 1), $L_R = 5$ (Fig. 2) and $L_R = 10$ (Fig. 3). This range of length-scales in $\mathbf{R}$ in relation to the length-scales in $\mathbf{B}$ could be representative of different observations. For example, whilst radiosonde observations may be thought to have spatially uncorrelated errors, AMVs could have much larger spatially correlated length-scales in their errors than in $\mathbf{B}$ (see e.g. Cordoba et al., 2017). The ratio of the length-scales in $\mathbf{R}$ to those in $\mathbf{B}$ could also vary as balances in the model breakdown. For example during a convective event, the

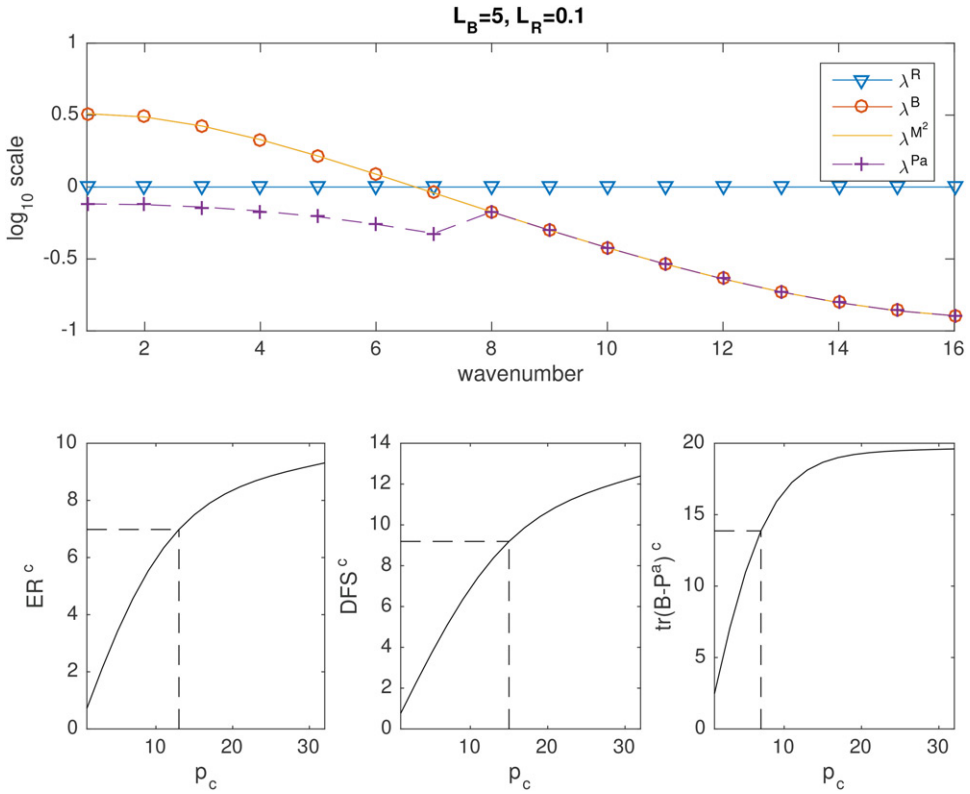

Fig. 1. Top: Eigenvalues of $\mathbf{MM}^T$ (20) (yellow line), $\mathbf{R}$ (blue triangles) and $\mathbf{B}$ (red circles) for the case of SOAR correlation functions with $L_B = 5$ and $L_R = 0.1$. Also plotted are the eigenvalues of $\mathbf{P}^a$ (purple crosses) when compressed observations retaining 75% of the total ER are assimilated (13 compressed observations in this case). Bottom: ER^c (left), DFS^c (middle) and $trace(\mathbf{B}-\mathbf{P}^a)^c$ (right) as a function of the number of compressed observations ordered according to the eigenvalues of $\mathbf{MM}^T$. The dashed lines indicate the number of compressed observations needed to achieve 75% of the total value. These numbers are also given in Table 1.

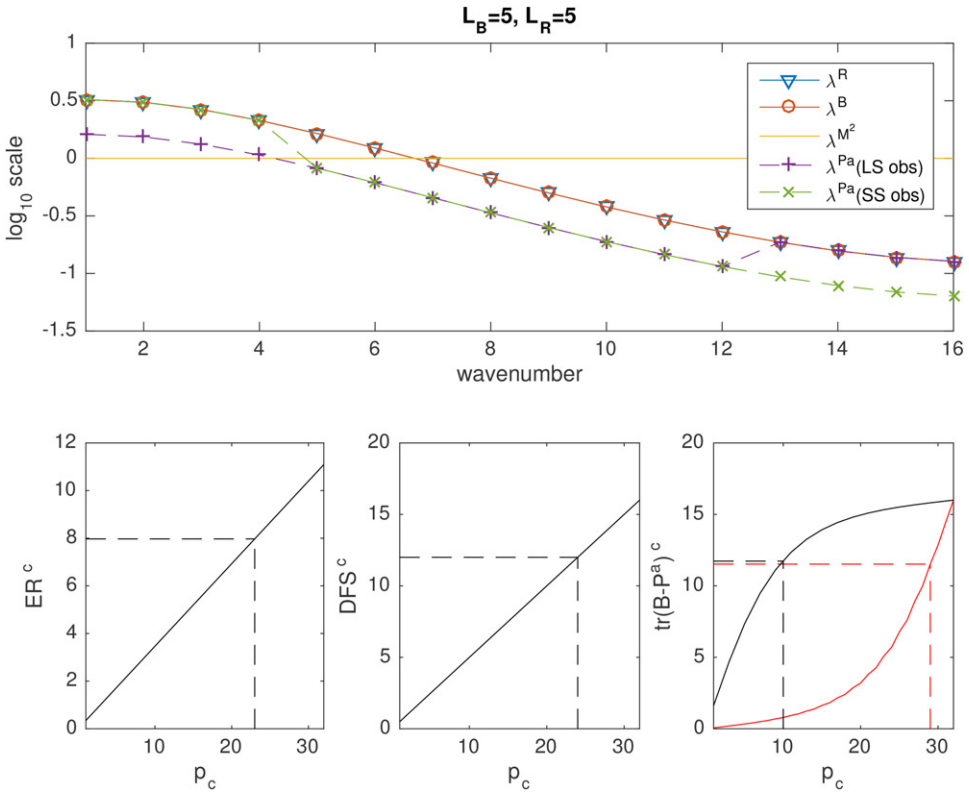

Fig. 2. As in Fig. 1 but for $L_R = 5$. The black line of the last panel shows $tr(\mathbf{B}-\mathbf{P}^a)^c$ as a function of the number of compressed observations when the assimilation is performed using large-scale observations first, and the red line when the assimilation is performed using small-scale observations first.

prior error correlations length-scales represented by the ensemble may reduce such that they are shorter than those in **R**.

When $L_R = 0.1, \mathbf{R} \approx \mathbf{I}$ and we see that the observation uncertainty is 1 at all scales. As the length-scale in **R** increases the uncertainty of the observations increases at small wave numbers and decreases at large wave numbers. This is consistent with observations with correlated errors providing more information about smaller scales and less about larger scales than observation without correlated errors (Seaman, 1977; Rainwater et al., 2015; Fowler et al., 2018).

Also plotted in the top panels of Figs. 1–3 are the eigenvalues of $\mathbf{M}\mathbf{M}^T$ (yellow line). When $L_R < L_B$ (Fig. 1) we see that the eigenvalues of $\mathbf{M}\mathbf{M}^T$ are greater at small wave numbers and hence large-scale features will be favoured by the data compression. However, when $L_R > L_B$ (Fig. 3) we see that the eigenvalues of $\mathbf{M}\mathbf{M}^T$ are greater at large wave numbers and hence small-scale features will be favoured by the data compression. When $L_R = L_B$ (Fig. 2) the eigenvalues of $\mathbf{M}\mathbf{M}^T$ are constant (i.e. the observations provide the same information about all scales). Therefore, the scales chosen by the data compression are arbitrary.

In the bottom panels of Figs. 1–3, ER^c (left) and DFS^c (middle) are shown as a function of the number of compressed observations ordered according to the magnitude of λ^M. The dashed lines indicate the number of compressed observations needed to achieve 75% of the total value. These numbers are also given in Table 1.

If we first compare the values of ER^c and DFS^c when all 32 observations are assimilated we see that both *ER* and *DFS* increase as the length-scales in **R** increases. This is explained in detail in Fowler et al. (2018).

When $L_R \neq L_B$, the information content of the observations initially rises quickly as the number of compressed observations are increased before slowly plateauing as additional observations bring little new information. This means that in these cases less than 75% of the compressed observations are needed to achieve 75% of the information of the total observations. This is obviously not the case when $L_R = L_B$, in which case the information content is proportional to the number of compressed observations.

In the last panel of Figs. 1–3, $tr(\mathbf{B}-\mathbf{P}^a)$ is shown as a function of the number of compressed observations, again ordered according to the magnitude of λ^M. Again the

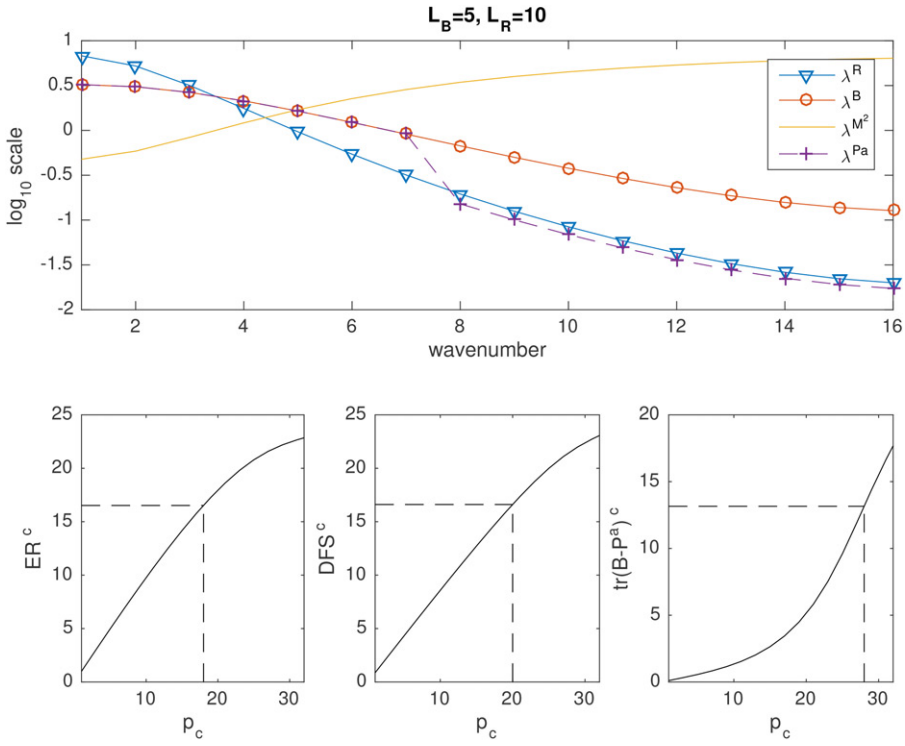

Fig. 3. As in Fig. 1 but for $L_R = 10$.

dashed lines indicate the number of compressed observations needed to achieve 75% of the total reduction in error variance (numbers are given in Table 1).

Let us first compare the values of $tr(\mathbf{B}-\mathbf{P}^a)^c$ when all 32 observations are assimilated. We see that unlike *ER* and *DFS*, $tr(\mathbf{B}-\mathbf{P}^a)$ is at a minimum when $L_R = L_B$. This is because both the observations and prior are informative about the same directions of state space, whereas when $L_R \neq L_B$ the observations and prior have more complimentary information and are more able to reduce the analysis error variances (Fowler et al., 2018).

In the case when $L_R < L_B$ compressing the observations to maximise information content is also seen to result in the maximum reduction in analysis error variance. In this case, 75% of the reduction in analysis error variance is achievable with just the first 8 compressed observations.

The eigenvalues of the analysis error covariance matrix as a function of wavenumber are plotted in the top panel of Fig. 1 (purple crosses) when the compressed observations responsible for 75% of the entropy reduction are assimilated. The reduction in the analysis uncertainty at large scales compared to the background uncertainty is clearly seen.

In the case when $L_R = L_B$, we saw previously that it does not matter which of the compressed observations are used to maximise the information content. However, this is clearly not the case if we are interested in reducing the analysis error variance. The black line of the last panel of Fig. 2 shows $tr(\mathbf{B}-\mathbf{P}^a)$ as a function of the number of compressed observations when the assimilation is performed using large-scale observations first, and the red line when the assimilation is performed using small-scale observations first. It is clear that a few observations of the large-scales (where the background uncertainty is greatest) results in a much smaller analysis error variance than many small-scale observations.

The eigenvalues of the analysis error covariance matrix as a function of wavenumber, for the two different orderings of the compressed observations, are plotted in the top panel of Fig. 2. When 24 compressed observations (chosen as the number needed to retain 75% of the information content) of the large scales are assimilated (purple crosses), the reduction in the analysis uncertainty at large scales compared to the background uncertainty is clearly seen. However, when the first 24 small-scale observations are assimilated (green crosses), the reduction in the analysis uncertainty at small scales compared to the background uncertainty is only visible due to the log scale. This is because the background is already relatively accurate at the small scales.

Despite the analysis error variances being smaller when the large-scale observations are assimilated, it is clear that the entropy of the posterior is the same irrespective of the order the compressed observations are assimilated. This is

Table 1. 75% of the value of *ER*, *DFS*, and $tr(\mathbf{B}-\mathbf{P})$ when all observations are assimilated for different values of L_R. In brackets are the number of compressed observations need to achieve these values.

	Criterion		
	$0.75 \times ER^{all}$	$0.75 \times DFS^{all}$	$0.75 \times tr(\mathbf{B}-\mathbf{P}^a)^{all}$
$L_R = 0.1$	7.08 (13 obs)	9.48 (15 obs)	14.5 (8 obs)
$L_R = 5$	8.32 (23 obs)	12.0 (24 obs)	12.0 (10 large-scale obs, 29 small-scale obs)
$L_R = 10$	16.9 (18 obs)	17.2 (20 obs)	13.4 (27 obs)

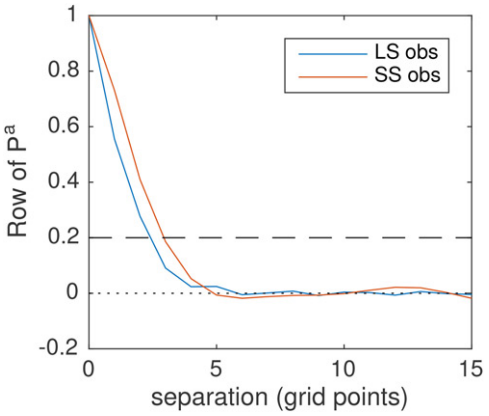

Fig. 4. Analysis error correlation structure for the case illustrated in Fig. 2. The number of compressed observations assimilated are chosen to conserve 75% of *ER*. Blue: compressed observations favour large scales and red: compressed observations favour small scales.

because assimilating the small-scale observations has an important effect on the analysis error correlations, not measured by the trace of the analysis error covariance matrix, and hence the overall entropy is the same in the two cases. The correlation structure for the circulant analysis error covariance matrices are shown in Fig. 4. We see that the correlation length-scales are longer when the small scale observations are assimilated, implying that in this case the relative accuracy of the analysis at small scales to large scales is much greater. In both cases the entropy of $\mathbf{P}^a$ is 28.2.

Lastly, when $L_R > L_B$ (Fig. 3) it is the small-scale observations which have the greatest information content. However, much more than 75% of these compressed observations are necessary to achieve 75% of the reduction in analysis error variance.

From these simple experiments it can be concluded that as the length-scales in $\mathbf{R}$ increase and the accuracy of the observations at the smallest resolved scales becomes greater than the background, that there is greater benefit to having a resolution of observations that matches that of the model. Assimilating this quantity of data is however unfeasible for many observation types. To reduce the amount of data for assimilation, data compression may be used to ensure that the most informative combination of observations are retained. It is possible that the optimal combination of observations for retaining the maximum information will not be the same as giving the greatest reduction in analysis error variance, especially when the small scales are favoured over the large scales. This discrepancy between information content and reduction in analysis error variance can be understood by comparing the equation for *ER* (13) with $tr(\mathbf{B}-\mathbf{P}^a)$. *ER* is sensitive to how the determinant of the error variance matrix changes due to the assimilation of the observations rather than just the trace. *ER* is therefore much more sensitive to corrections to the relatively accurate small-scales. The relevance of accurately constraining the small-scales is becoming increasingly important for high-resolution forecasting of high-impact weather.

Experiments were performed in which observations are available at a higher resolution than that resolved by the model but in all cases the information content at these unresolved scales was less than at those resolved. As such these unresolved scales would not be chosen during the data compression, despite having some information.

4. Application to the Lorenz 1996 model

In this section, the data compression method described in Section 3 is applied to the Lorenz 1996 model (Lorenz, 1995) using the EnKF (described in Section 2) to assimilate the observations.

The Lorenz 1996 model solves the following set of equations

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

where $j = 1, \ldots, n$, with $n = 40$ and $F = 8$, assuming a cyclic domain. Following Lorenz and Emanuel (1998), a fourth-order Runge-Kutta scheme is used with a time step of 0.05. This toy model represents the evolution of an arbitrary 'atmospheric' variable in n sectors on a latitude circle. With the forcing term $F = 8$, the model

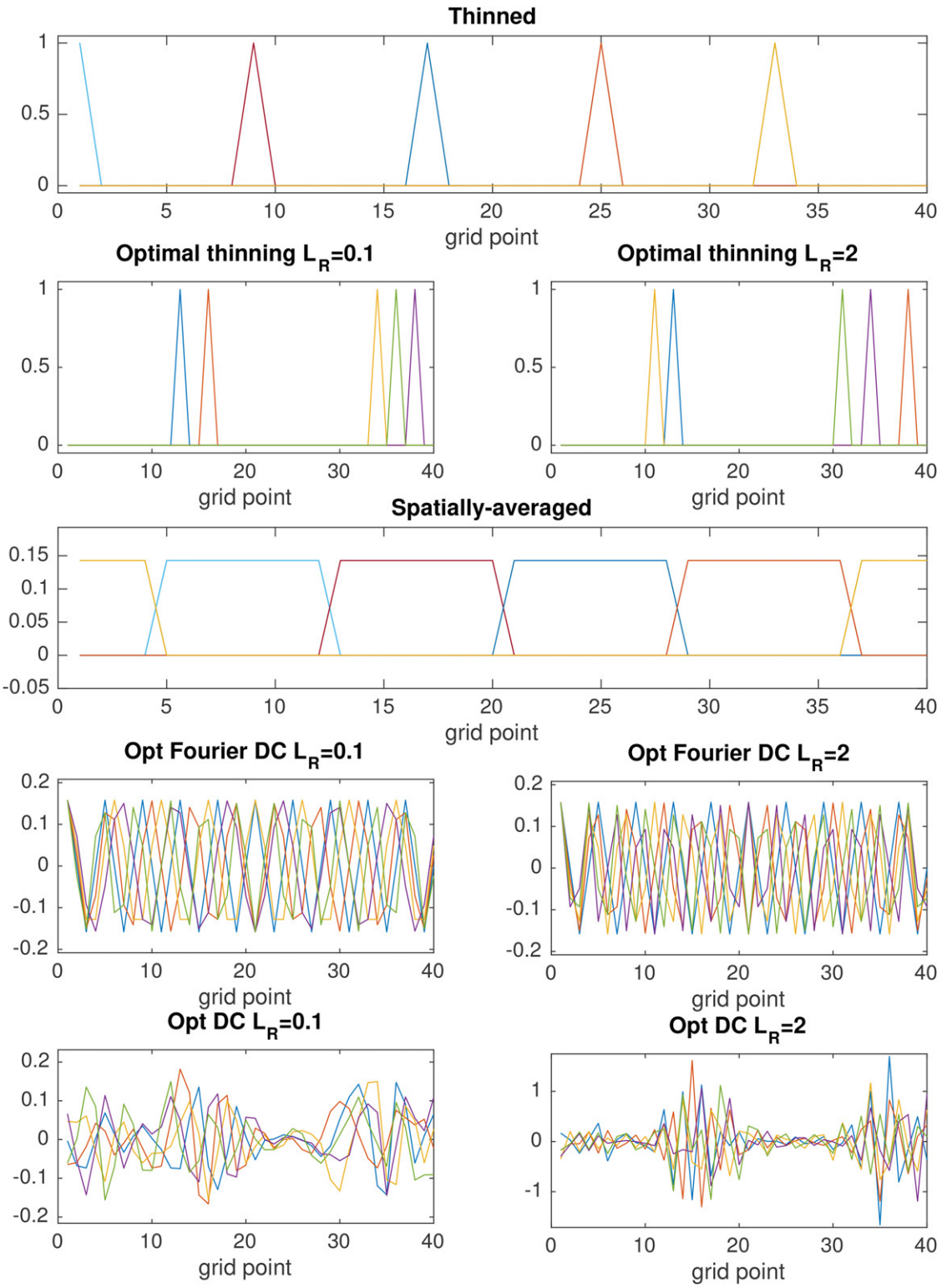

Fig. 5. Rows of the observation operator matrix for the five strategies for reducing the observation data detailed in Section 4. The optimal strategies are illustrated for the first observation time.

exhibits chaotic behaviour. The model naturally represents scales associated with the complex sinosoid with the 20 different frequencies that can be represented on the 40 grid points of the model. The largest spatial scale is the average of the variables across the circular domain and the smallest is the complex sinusoid with wavelength of twice the grid spacing.

A simulated true trajectory is generated with initial conditions given by $x_j^{\mathrm{t}}(t=0) = 2\sin(2\pi j/10)$. An initial ensemble is generated from $\mathbf{x}^{\mathrm{f}(k)}(t=0) = \mathbf{x}^{\mathrm{t}}(t=0) + \eta_k$, where $\eta_k \sim N(0, \mathbf{B}), k = 1, \ldots, N$. In the following experiments $\mathbf{B}$ is given by a circulant matrix, with a SOAR correlation function (24) with length-scale $L_{\mathrm{B}} = 2$ and variance 5. The ensemble size is $N = 100$. The ensemble is then propagated in time using the Lorenz 1996 model with stochastic forcing at every grid point and time step drawn from a Normal distribution with variances given by 0.01.

40 evenly distributed observations are simulated from the truth run every 20 time steps starting 100 time steps into the simulation (allowing for spin up of the ensemble). $\mathbf{y}(\tau) = \mathbf{x}^{\mathrm{t}}(t = t_\tau) + \boldsymbol{\epsilon}_\tau$, where $\boldsymbol{\epsilon}_\tau \sim N(0, \mathbf{R})$. 20 time steps can be thought of as equivalent to 5 days comparing the error doubling time in the Lorenz 1996 model to the real atmosphere (assuming the latter is 2 days) (Khare and Anderson, 2006). This frequency of observations is chosen to make the effect of the observations at each assimilation time clearer.

Two observation error covariance matrices are considered given by a circulant matrix with a SOAR correlation function (24) and variance 5. In the first $L_{\mathrm{R}} = 0.1$ (effectively uncorrelated observation errors). In the second $L_{\mathrm{R}} = 2$ (giving correlations greater than 0.2 out to about 7 grid points).

Experiments were performed with other parameters for the respective covariance matrices and sampling frequencies, however, there was little sensitivity to these in the qualitative results shown below and the conclusions made are unchanged.

Five methods of data reduction are compared. Those referred to as 'optimal' imply that they are dependent upon an objective criterion based on the information content of the observations. In each case only five pieces of observational data are retained for assimilation from the original 40 observations. This dramatic reduction in the data is chosen to highlight the differences between the different methods of data reduction within the numerical experiments. In practice a criteria for retaining a certain proportion of the information of the total observations may be used for choosing the number of compressed observations (as in Section 3.2), or a threshold on the information content of each compressed observation assimilated may be used as in Migliorini (2013). The five methods of data reduction are

1. *Thinning:* Observations are thinned to every 8th grid point.
2. *Optimal thinning:* Observations are chosen corresponding to the 5 largest diagonal values of $\mathbf{S}$ measuring the greatest values of $\frac{\partial h(\mathbf{x}^{\mathrm{a}})_k}{\partial \mathbf{y}_k}$.
3. *Spatial averaging:* Observations are averaged over 8 grid-points centred on every 8th grid point.
4. *Optimal Fourier Data Compression (DC):* Observations are compressed using a Fourier transform with wavelengths chosen corresponding to the 5 largest diagonal values of $\mathbf{FSF}^{\mathrm{T}}$.
5. *Optimal DC:* Observations are compressed using the method described in Section 3, again assimilating just the 5 most informative observations.

For each of these methods a compression matrix, $\mathbf{C} \in \mathbb{R}^{5\times 40}$, is defined and the new observation operators, observations assimilated and their error covariances are computed using (17) to (19). Note that in this case the original observation operator, $\mathbf{H} \in \mathbb{R}^{40\times 40}$ was simply the identity matrix.

The observation operators, $\mathbf{H}^{\mathrm{c}} \in \mathbb{R}^{5\times 40}$, corresponding to these five compression strategies are shown in Fig. 5. The coloured lines represent the 5 different rows of the $\mathbf{H}^{\mathrm{c}}$ matrix. In the first and third compression strategies, the observation operator is independent of the observation error covariance matrix and time step. In the 'optimal' second, fourth and fifth compression strategies, the observation operator is dependent upon the ensemble representation of the prior uncertainties and the full observation error covariance matrix. These are illustrated for the first observation time, so that the prior uncertainty (approximated by the ensemble) is the same in each case. It is seen that when the observation errors are correlated, the optimal data compression retains smaller scale features of the observations, wave-numbers selected by the optimal Fourier decomposition are also shown in Table 2. In comparison, the effect of spatial averaging is a severe smoothing out of the small-scale information in the observations.

Table 2. The wave-numbers selected by the optimal Fourier data compression method at each observation time for the two different types of observations. The first observation time corresponds to the observation operator plotted in Fig. 5.

Ob time	$L_R = 0.1$	$L_R = 2$
1	10, 11, 9, 12, 8	12, 13, 11, 10, 15
2	10, 11, 12, 13, 9	14, 12, 13, 20, 15
3	12, 9, 11, 10, 13	15, 14, 17, 20, 18
4	12, 10, 11, 9, 8	17, 19, 18, 16, 20
5	13, 14, 11, 12, 10	19, 17, 15, 18, 16

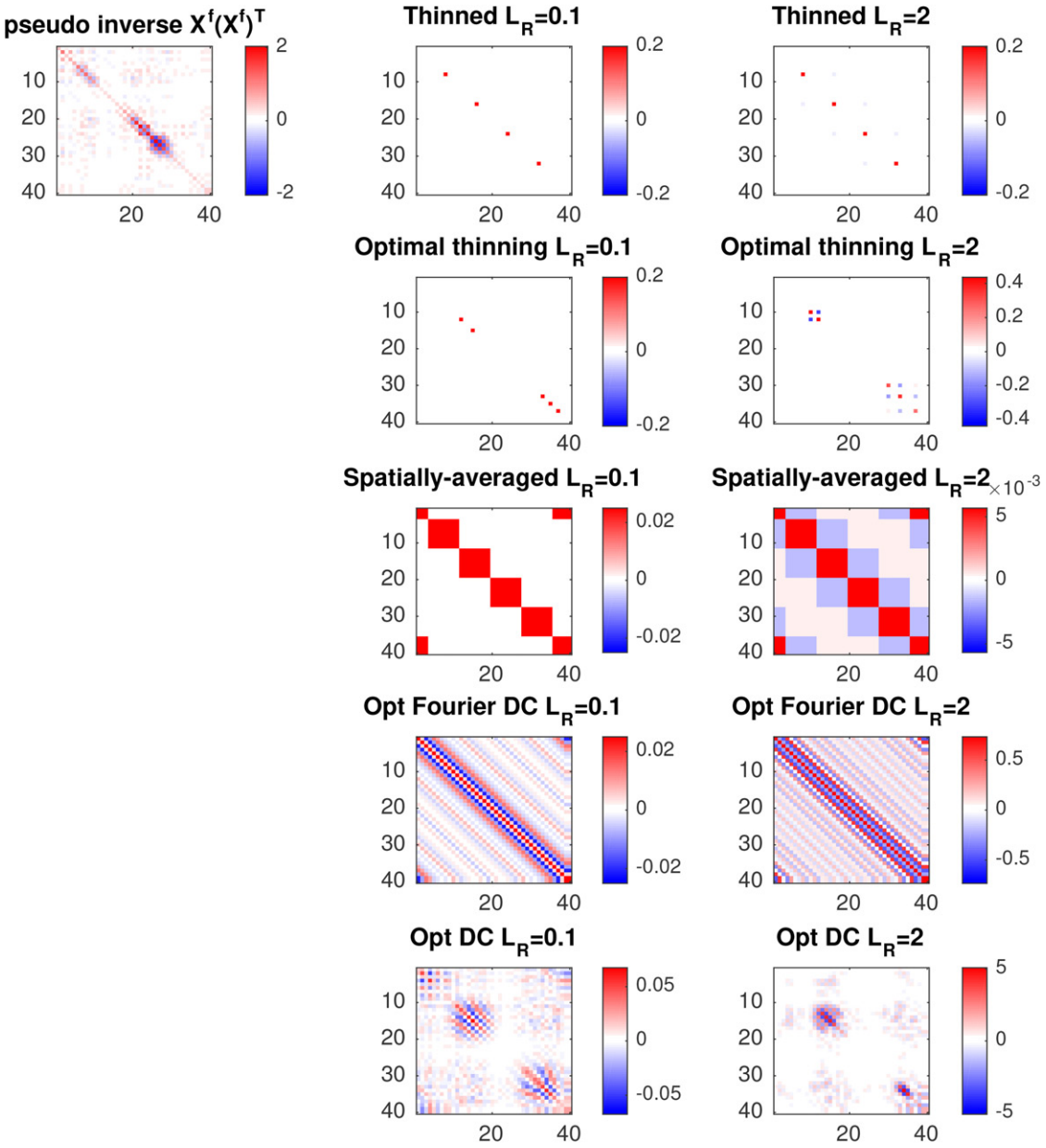

Fig. 6. Top left: pseudo inverse of $\mathbf{X}^{\mathrm{f}}(\mathbf{X}^{\mathrm{f}})^{\mathrm{T}}$ illustrated for the first observation time. Other panels: $\mathbf{H}_{\mathrm{c}}^{\mathrm{T}}\mathbf{R}_{\mathrm{c}}^{-1}\mathbf{H}_{c}$ (the accuracy of the compressed observation in state space) for the 5 different thinning strategies when observation errors are uncorrelated (middle panels) and correlated (right panels) illustrated for the first observation time.

In Fig. 6, the pseudo inverse of $\mathbf{X}^{\mathrm{f}}(\mathbf{X}^{\mathrm{f}})^{\mathrm{T}}$ and $\mathbf{H}_{\mathrm{c}}^{\mathrm{T}}\mathbf{R}_{\mathrm{c}}^{-1}\mathbf{H}_{c}$ for the different data reduction strategies are shown for the first observation time. Note that, unlike in Section 3.1, $\mathbf{P}^{\mathrm{f}}$ is no longer circulant. These represent the accuracy of the prior and assimilated observations in state space, respectively. It is seen that the 'optimal' thinning and data compression methods choose observations in regions where the prior accuracy is low (i.e. the prior uncertainty is large). When the observation errors are uncorrelated the trace of $\mathbf{H}_{\mathrm{c}}^{\mathrm{T}}\mathbf{R}_{\mathrm{c}}^{-1}\mathbf{H}_{c}$ is approximately equal to 1 irrespective of data compression strategy. When the observation errors are correlated the trace of $\mathbf{H}_{\mathrm{c}}^{\mathrm{T}}\mathbf{R}_{\mathrm{c}}^{-1}\mathbf{H}_{c}$ is greatest when observations are compressed using the optimal DC method described in Section 3, and the smallest when they are spatially averaged (47.0 compared to 0.227).

The resulting ensemble spread, entropy (estimated from (11) using the ensemble approximation of the error

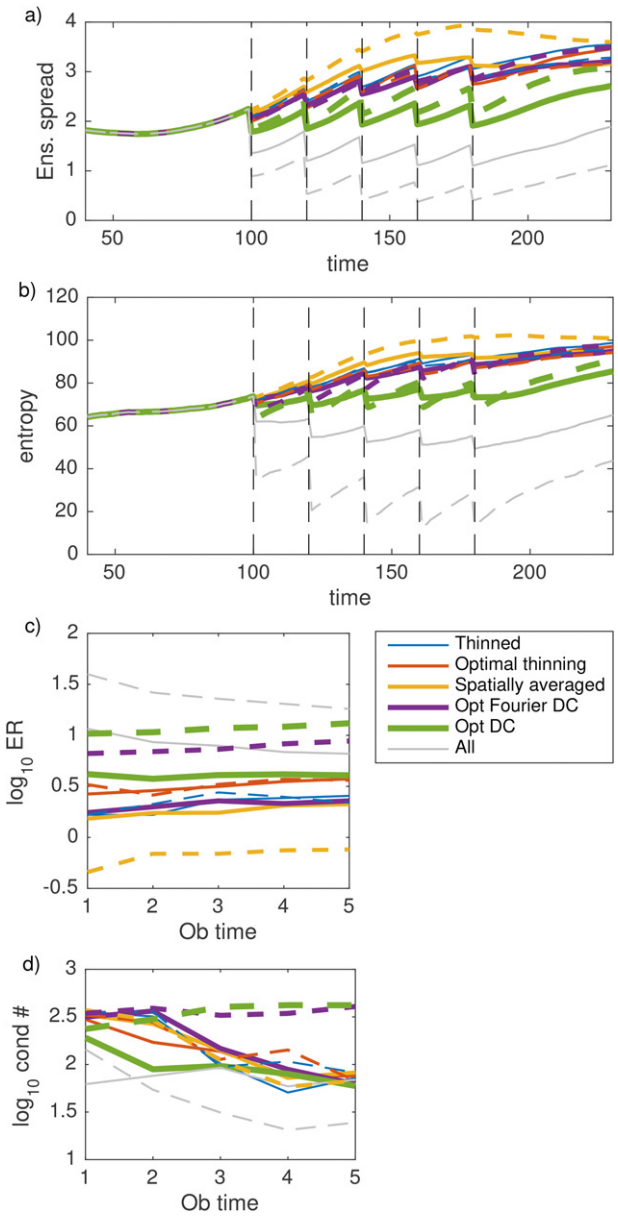

Fig. 7. (a) Ensemble spread, (b) entropy computed using the ensemble estimate of (11), (c) $log_{10}ER$, and (d) log_{10} of the condition number of the analysis error covariance matrix for the five different thinning strategies detailed in Section 3.1. Solid lines represent results when the observation errors are uncorrelated, dashed lines represent results when the observation errors are correlated. Results are averaged over 200 experiments with different realisations of the observation and model error.

covariance matrix) and *ER* for the thinning experiments are shown in Fig. 7a–c. In these figures, the values have been averaged over 200 random realisations of the observation and model error. Also plotted in Fig. 7d is the condition number of the analysis error covariance matrix, which is related to how sensitive the analysis is to perturbations in the data. The closer this is to 1 the better the conditioned the inverse problem (Golub and Van Loan, 1996). The solid lines are when the simulated observations have uncorrelated error and the dashed lines are when the observation errors are correlated.

Let us first compare the case when all 40 observations are assimilated (grey lines) every 20 time steps. We see that the observations with correlated errors result in a smaller ensemble spread and entropy than observations with uncorrelated errors. It is interesting to note that after each assimilation time the entropy represented by the ensemble increases most rapidly when the observation

errors are correlated. Recall that both the ensemble spread and entropy measure the uncertainty in the ensemble, however, the entropy is sensitive to changes in the correlations in the ensemble perturbations as well as their magnitudes, and hence more representative of the uncertainty at all scales. Therefore, the greater increase in entropy with time for the correlated errors is indicative of the small-scale corrections dissipating during the forecast period between analysis times.

Observations with correlated error are also seen to have significantly more information, as the reduction in entropy is greater at each analysis time. This reduces in time due to the reduction in the ensemble spread (increase in the prior accuracy). Initially, the conditioning is worse when the observation errors are correlated as would be expected (Tabeart et al., 2018), however, the opposite is true after the 2nd observation time. This is consistent with the ensemble spread reducing in time and the influence of the observations also reducing (Haben et al., 2011).

When the observations are thinned to every 8th grid point (blue lines) the performance of the DA is degraded as would be expected using only 12.5% of the original observations. The difference between the lines is negligible for the different observation error types. This is because the thinning distance of the observations is practically beyond the correlation length-scale.

When the observation errors are uncorrelated, there appears to be little effect on the *ER* (Fig. 7c) when the observations are thinned (blue line), spatially averaged (yellow line) or compressed using the optimal Fourier wavelengths (purple line), with numbers ranging between 1.5 and 2.6. There is a slight increase in *ER* when the optimal thinning (red line) is used (up to approximately 3.8). When the optimum compression is used (green line) there is an increase in *ER* (increasing to approximately 4.08) and a clear reduction in ensemble spread. When the observation errors are uncorrelated, there is also little difference between the condition number for the five different data reduction strategies.

When the observation errors are correlated the choice of data reduction on the results is much more significant. Compared to thinning the observations (blue dashed-line), spatial averaging (yellow dashed line) is seen to be significantly detrimental to the *ER* (0.46 compared to 2.8) and ensemble spread. This is because spatial averaging smooths out the highly accurate small scale information of the observations leaving just the less accurate large scale information. Spatial averaging is less detrimental when observation errors are uncorrelated as in this case the observations provide the same information at all scales (see Fig. 1), and averaging enables a reduction in the noise. Again there is a slight increase in *ER* when the optimal thinning (red dashed line) is used (3.7 compared to 2.8).

Using the Fourier compression or the optimal data compression, when the observation errors are correlated is seen to lead to a large increase in *ER* (greater, in many cases, than using the full 40 observations when the errors are uncorrelated). In Fig. 5, we saw that both these methods of data compression are selecting the fine scale information in the observations for assimilation.

The reason that the ensemble spread is smaller when the observations have uncorrelated error compared to correlated error with the optimal DC may be related to the results shown in Section 3.2. It was seen here that using observations to correct the small scales was less able to reduce the analysis error variance, which is approximated by the ensemble spread, despite the information content of the observations being greater. This hypothesis is supported by the entropy time series where it is seen that at each observation time the entropy of the posterior is actually *smaller* when assimilating the optimally compressed observations with correlated error rather than uncorrelated error. Increasing the number of compressed observations (and hence the number of scales retained by the data compression) sees the ensemble spread quickly reduce. For example, when just 8 of the compressed observations are retained, the ensemble spread when the observations have correlated errors is smaller than the ensemble spread when the observation errors are uncorrelated by the fifth observation time (results not shown).

In the case when the observation errors are correlated, the condition number is much larger when using these more optimal methods of data compression. This could potentially have an undesirable impact on the sensitivity of the analysis to perturbations in the data.

5. Summary and conclusions

Recent advances in the estimation and inclusion of OECs in DA means that we are getting closer to be able to assimilate denser observations with an accurate description of their uncertainty. As observations with correlated error are known to have greater information at small scales than observations with uncorrelated error this may be crucial for high resolution forecasting in which accurate forecasts of small-scale high impact weather are sought.

This may seem contradictory to previous studies of Liu and Rabier (2002) and Bergman and Bonner (1976) that concluded that increasing the density of observations with correlated error is not as beneficial as if the observations had uncorrelated error, due to the reduction in analysis RMSE not being as great. This is, in part, due to

the effect of increasing the correlation length-scales of the observation errors bringing the likelihood PDF more in line with the prior PDF (Fowler et al., 2018). Once the OEC length-scales are increased beyond those of the prior, the reduction in analysis RMSE with increasing observation density is seen to increase once again.

The potential large increase in the number of observations available for assimilation carries a large computational and storage burden with it. It is therefore important to justify any increase in the amount of data assimilated and give careful thought to the design of the observing network. One way to potentially reduce the computational burden is to reduce the data using a compression technique based on retaining the maximum information content of the observations.

In an idealised circulant framework it was shown how the scales that are retained by the data compression depends upon the structure of the prior and OECs. When the length-scales in **R** are smaller than those in **B**, observations of large-scale features are retained for assimilation. When the length-scales in **R** are greater than those in **B**, observations of small-scale features are retained for assimilation. However, the greater the length-scales in **R**, the greater the total information content of those observations.

It was shown that, despite possibly having greater information content, many more small-scale observations are needed than large-scale observations to achieve the same level of reduction in the analysis error variances. This highlights a potential discrepancy between information content (a relative quantity) and reduction in the analysis error variances (an absolute quantity). The importance of constraining the analysis at small-scales depends upon the aim of performing the assimilation. For the increasingly important applications of high-impact forecasting that aim to produce limited area forecasts out to short lead times, corrections to the small-scales are arguably more important than correcting the large-scales. In these cases the large-scales are largely constrained by the boundary conditions provided by the global model. It is only when forecasting at longer lead times that the small scale corrections in the analysis dissipate and become invaluable. In addition to this, large scale information may be available from other observations. Therefore, to understand the full potential of observations with correlated errors it is important not to concentrate only on the effect on the reduction in the analysis error variances or analysis RMSE.

In Section 3.1, the compression of the observations was then applied to the EnKF using the Lorenz 1996 model, for observations with correlated and uncorrelated errors. The data compression method was compared to other methods to reduce the amount of data. These included regular thinning, optimal thinning (in which the observations with the greatest analysis sensitivities where selected), spatial averaging, and compression using a Fourier transform based on the wavelengths with the greatest analysis sensitivity.

It was found that when the observation errors are uncorrelated, the information content of the reduced data set is largely insensitive to the strategy used. However, when the observation errors are correlated, it is found that the information content of the reduced data set is highly sensitive to the strategy used. In particular using a data compression method based on maximising the information content of the compressed observations can increase the entropy reduction of the reduced observations by up to 420% compared to regular thinning. Whereas observations reduced using spatial averaging have only 16% of the entropy reduction of observations reduced using regular thinning.

This implies that when the observations have correlated error it is more beneficial to have high-density observations. However, in order to reduce the amount of data, the use of spatial averaging can be more detrimental than regular thinning and instead data compression based on retaining the small-scale information should be used. The need for this to be performed on-line depends on how quickly the length scales of the correlations represented by the ensemble are evolving.

In the Lorenz 1996 example, although the optimal methods were adaptive with the changing flow of the ensemble, the compression was actually relatively static (see Table 2 for an example of the how the most important wave numbers change for the 5 observation times). Future work will look further at the effect of the flow dependent estimate on the data compression using a more physically realistic model in which prior error correlations are more dynamic. For example, using the multivariate modified shallow water model of Kent et al. (2017) which represents simplified dynamics of cumulus convection and associated precipitation, and the corresponding disruption to large-scale balances. Methods for reducing the computational cost of on-line data compression will also be investigated, as well as the possibility of retaining some base line scales.

In practice the use of data compression can be expected to be sensitive to the accuracy of the specified error characteristics. This will be the focus of future work. The results could also be sensitive to the way the ensemble Kalman filter is implemented (e.g. ensemble size, localisation, inflation, etc.). Experiments were also performed with a small ensemble size of $N = 10$ with little effect on the results, largely because, even in this case, the number of compressed observations ($p_c = 5$) was less

than N. The sensitivity to the ensemble parameters can be expected to increase as p_c increases.

Acknowledgements

I would like to thank J. Eyre (UK Met Office), R. N. Bannister and J. A. Waller (Uni. Reading) for providing valuable feedback and discussion.

Funding

This study was funded as part of NERC's support of the National Centre for Earth Observation, contract number PR140015.

References

Berger, H. and Forsythe, M. 2004. Satellite wind superobbing. *Met Office Forecasting Research Technical Report* No. **451.** http://research.metoffice.gov.uk/research/nwp/publications/papers/technical_reports

Bergman, K. H. and Bonner, W. D. 1976. Analysis error as a function of observation density for satellite temperature soundings with spatially correlated errors. *Mon. Wea. Rev.* **104**, 1308–1316. doi:10.1175/1520-0493(1976)104<1308:AEAAFO>2.0.CO;2

Bormann, N. and Bauer, P. 2010. Estimates of spatial and interchannel observation-error characteristics for current sounder radiances for numerical weather prediction. I: methods and application to ATOVS data. *Q. J. R. Meteorol. Soc.* **136**, 1036–1050. doi:10.1002/qj.616

Bormann, N., Saarinen, S., Kelly, G. and Thépaut, J.-N. 2003. The spatial structure of observation errors in atmospheric motion vectors from geostationary satellite data. *Mon. Wea. Rev.* **131**, 706–718. doi:10.1175/1520-0493(2003)131<0706:TSSOOE>2.0.CO;2

Bormann, N., Bonavita, M., Dragani, R., Eresmaa, R., Matricardi, M. and co-authors. 2016. Enhancing the impact of IASI observations through an updated observation-error covariance matrix. *Q. J. R. Meteorol. Soc.* **142**, 1767–1780. doi:10.1002/qj.2774

Campbell, W. F., Satterfield, E. A., Ruston, B. and Baker, N. L. 2017. Accounting for correlated observation error in a dual formulation 4D-variational data assimilation system. *Mon. Wea. Rev.* **145**, 1019. doi:10.1175/MWR-D-16-0240.1

Cardinali, C., Pezzulli, S. and Andersson, E. 2004. Influence-matrix diagnostics of a data assimilation system. *Q. J. R Meteorol. Soc.* **130**, 2767–2786. doi:10.1256/qj.03.205

Collard, A. D., McNally, A. P., Hilton, F. I., Healy, S. B. and Atkinson, N. C. 2010. The use of principal component analysis for the assimilation of high-resolution infrared sounder observations for numerical weather prediction. *Q. J. R. Meteorol. Soc.* **136**, 2038–2050. URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.701. doi:10.1002/qj.701

Cordoba, M., Dance, S. L., Kelly, G. A., Nichols, N. K. and Waller, J. A. 2017. Diagnosing atmospheric motion vector observation errors for an operational high resolution data assimilation system. *Q. J. R. Meteorol. Soc.* **143**, 333–341. doi:10.1002/qj.2925

Dando, M. L., Thorpe, A. J. and Eyre, J. R. 2007. The optimal density of atmospheric sounder observations in the Met Office NWP system. *Q. J. R. Meteorol. Soc.* **133**, 1933–1943. doi:10.1002/qj.175

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

Fowler, A., Dance, S. and Waller, J. 2018. On the interaction of observation and a-priori error correlations. *Q. J. R. Meteorol. Soc.* **144**, 48–62. doi:10.1002/qj.3183

Fowler, A. M. 2017. A sampling method for quantifying the information content of IASI channels. *Mon.Wea. Rev.* **145**, 709–725. doi:10.1175/MWR-D-16-0069.1

Goldberg, M. D., Qu, Y., McMillin, L. M., Wolf, W., Zhou, L. and Divakarla, M. 2003. AIRS near-real-time products and algorithms in support of operational numerical weather prediction. *IEEE Trans. Geosci. Remote Sensing* **41**, 379–389. doi:10.1109/TGRS.2002.808307

Golub, G. H. and Van Loan, C. F. 1996. *Matrix Computations*. 3rd ed. The John Hopkins University Press, Baltimore, MD.

Gray, R. 2006. *Toeplitz and Circulant Matrices: A Review*. Foundations and Trends in Communications and Information Theory: Vol 2: No. 3, pp 155–239. doi:10.1561/0100000006

Haben, S. A., Lawless, A. S. and Nichols, N. K. 2011. Conditioning of incremental variational data assimilation, with application to the Met Office system. *Tellus A* **64**, 782–792.

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

Hilton, F., Collard, A., Guidard, V., Randriamampianina, R. and Schwaerz, M. 2009. Assimilation of IASI radiances at European NWP centres. In: *Proceedings of Workshop on the assimilation of IASI data in NWP, ECMWF, Reading, UK, 6-8 May 2009*.

Houtekamer, P. L. and Zhang, F. 2016. Review of the ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.* **144**, 4489–4532. doi:10.1175/MWR-D-15-0440.1

Huang, H.-L. and Purser, R. J. 1996. Objective measures of the information density of satellite data. *Meteorl. Atmos. Phys.* **60**, 105–117. doi:10.1007/BF01029788

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. doi:10.1016/j.physd.2006.11.008

Janjic, T., Bormann, N., Bocquet, M., Carton, J. A., Cohn, S. E. and co-authors. 2017. On the representation error in data assimilation. *Q. J. R. Meteorol. Soc.* **144**, 1257–1278. doi:10.1002/qj.3130

Kalnay, E. 2003. *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, New York,.

Kent, T., Bokhove, O. and Tobias, S. 2017. Dynamics of an idealized fluid model for investigating convective-scale data assimilation. *Tellus A* **69**, 1369332. URL doi:10.1080/16000870.2017.1369332

Khare, S. P. and Anderson, J. L. 2006. An examination of ensemble filter based adaptive observation methodologies. *Tellus* **58A**, 179–195.

Li, Z., Ballard, S. P. and Simonin, D. 2018. Comparison of 3D-Var and 4D-Var data assimilation in an NWP-based system for precipitation nowcasting at the Met Office. *Q. J. R. Meteorol. Soc.* **144**, 404–413. URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3211. doi:10.1002/qj.3211

Liu, Z.-Q. and Rabier, F. 2002. The interaction between model resolution, observation resolution and observational density in data assimilation: a one-dimensional study. *Q. J. R. Meteorol. Soc.* **128**, 1367–1386. doi:10.1256/003590002320373337

Liu, Z.-Q. and Rabier, F. 2003. The potential of high-density observations for numerical weather prediction: a study with simulated observations. *Q. J. R. Meteorol. Soc.* **129**, 3013–3035. doi:10.1256/qj.02.170

Lorenz, E. N. 1995. Predictability – a problem partly solved. In: *Seminar Proceedings I*, Reading, UK, ECMWF, pp. 1–18.

Lorenz, E. N. and Emanuel, K. A. 1998. Optimal sites for supplementary weather observations: simulation with a small model. *J. Atmos. Sci.* **55**, 399–414. doi:10.1175/1520-0469(1998)055<0399:OSFSWO>2.0.CO;2

Majumdar, S. J. 2016. A review of targeted observations. *Bull. Am. Meteor. Soc.* **97**, 2287–2303. URL doi:10.1175/BAMS-D-14-00259.1

Migliorini, S. 2013. Information-based data selection for ensemble data assimilation. *Q. J. R. Meteorol. Soc.* **139**, 2033–2054. doi:10.1002/qj.2104

Migliorini, S. 2015. Optimal ensemble-based selection of channels from advanced sounders in the presence of cloud. *Mon. Wea. Rev.* **143**, 3754–3773. URL doi:10.1175/MWR-D-14-00249.1

Miyoshi, T., Lien, G.-Y., Satoh, S., Ushio, T., Bessho, K. and co-authors. 2016. Big data assimilation" toward post-petascale severe weather prediction: An overview and progress. *Proc. IEEE* **20**, 2155–2179.

Oke, P. R., Sakov, P. and Corney, S. P. 2007. Impacts of localisation in the EnKF and EnOI: experiments with a small model. *Ocean Dyn.* **57**, 32–45. doi:10.1007/s10236-006-0088-8

Rabier, F., Jarvinen, H., Klinker, E., Mahfouf, J.-F. and Simmons, A. 2000. The ECMWF operational implementation of four-dimensional variational data assimilation. I: experimental results and simplified physics. *Q J. R. Meteorol. Soc.* **126**, 1143–1170. doi:10.1002/qj.49712656415

Rabier, F., Fourrié, N., Chafa I, D. and Prunet, P. 2002. Channel selection methods for infrared atmospheric sounding interferometer radiances. *Q J. R Meteorol. Soc.* **128**, 1011–1027. doi:10.1256/0035900021643638

Rainwater, S., Bishop, C. H. and Campbell, W. F. 2015. The benefits of correlated observation errors for small scales. *Q. J. R. Meteorol. Soc.* **141**, 3439–3445. doi:10.1002/qj.2582

Rawlins, F., Ballard, S. P., Bovis, K. J., Clayton, A. M., Li, D. and co-authors. 2007. The Met Office global four-dimensional variational data assimilation scheme. *Q. J. R. Meteorol. Soc.* **133**, 347–362. doi:10.1002/qj.32

Rodgers, C. D. 2000. *Inverse Methods for Atmospheric Sounding*. World Scientific Publishing, Singapore.

Seaman, R. S. 1977. Absolute and differential accuracy of analyses achievable with specified observational network characteristics. *Mon. Wea. Rev.* **105**, 1211–1222. doi:10.1175/1520-0493(1977)105<1211:AADAOA>2.0.CO;2

Stewart, L. M., Dance, S. L., Nichols, N. K., Eyre, J. R. and Cameron, J. 2014. Estimating interchannel observation-error correlations for IASI radiance data in the Met Office system. *Q. J. R. Meteorol. Soc.* **140**, 1236–1244. doi:10.1002/qj.2211

Tabeart, J. M., Dance, S. L., Haben, S. A., Lawless, A. S., Nichols, N. K. and co-authors. 2018. The conditioning of least squares problems in variational data assimilation. *Numer. Linear Algebra Appl.* **0**, e2165. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/nla.2165.

van Leeuwen, P. J. 2015. Representation errors and retrievals in linear and nonlinear data assimilation. *Q. J. R. Meteorol. Soc* **141**, 612–1623.

Waller, J. A., Dance, S. L., Lawless, A. S. and Nichols, N. K. 2014. Estimating correlated observation error statistics using an ensemble transform Kalman filter. *Tellus A* **66**, 23294. doi:10.3402/tellusa.v66.23294

Waller, J. A., Ballard, S. P., Dance, S. L., Kelly, G., Nichols, N. K. and co-authors. 2016a. Diagnosing horizontal and inter-channel observation error correlations for SEVIRI observations using observation-minus-background and observation-minus-analysis statistics. *Remote Sens.* **8**, 581. doi:10.3390/rs8070581

Waller, J. A., Simonin, D., Dance, S. L., Nichols, N. K. and Ballard, S. P. 2016b. Diagnosing observation error correlations for Doppler radar radial winds in the Met Office UKV model using observation-minus-background and observation-minus-analysis statistics. *Mon. Wea. Rev.* **144**, 3533. doi:10.1175/MWR-D-15-0340.1

Weston, P. P., Bell, W. and Eyre, J. R. 2014. Accounting for correlated error in the assimilation of high-resolution sounder data. *Q. J. R. Meteorol. Soc.* **140**, 2420–2429. doi:10.1002/qj.2306

Xu, Q. 2007. Measuring information content from observations for data assimilation: relative entropy versus Shannon entropy difference. *Tellus* **59A**, 198–209.

Xu, Q., Wei, L. and Healy, S. 2009. Measuring information content from observations for data assimilation: connection between different measures and application to radar scan design. *Tellus* **61A**, 144–153.