On the interaction of observation and prior error correlations in data assimilation

RMetS
Royal Meteorological Society

On the interaction of observation and prior error correlations in data assimilation

A. M. Fowler,[a,b]* S. L. Dance[a] and J. A. Waller[a]

[a]*School of Mathematical, Physical and Computational Sciences, University of Reading, UK*
[b]*National Centre for Earth Observation, University of Reading, UK*

*Correspondence to: A. M. Fowler, Department Meteorology, University of Reading, RG6 6BB, UK. E-mail: a.m.fowler@reading.ac.uk

The importance of prior error correlations in data assimilation has long been known; however, observation-error correlations have typically been neglected. Recent progress has been made in estimating and accounting for observation-error correlations, allowing for the optimal use of denser observations. Given this progress, it is now timely to ask how prior and observation-error correlations interact and how this affects the value of the observations in the analysis. Addressing this question is essential to understanding the optimal design of future observation networks for high-resolution numerical weather prediction. This article presents new results, which unify and advance upon previous studies on this topic.

The interaction of the prior and observation-error correlations is illustrated with a series of two-variable experiments in which the mapping between the state and observed variables (the observation operator) is allowed to vary. In an optimal system, the reduction in the analysis-error variance and spread of information is shown to be greatest when the observation and prior errors have complementary statistics: for example, in the case of direct observations, when the correlations between the observation and prior errors have opposite signs. This can be explained in terms of the relative uncertainty of the observations and prior on different spatial scales. The results from these simple two-variable experiments are used to inform the optimal observation density for observations of a circular domain (with 32 grid points). It is found that dense observations are most beneficial when they provide a more accurate estimate of the state at smaller scales than the prior estimate. In the case of second-order auto-regressive correlation functions, this is achieved when the length-scales of the observation-error correlations are greater than those of the prior estimate and the observations are direct measurements of the state variables.

Key Words: variational data assimilation; Bayes' theorem; data thinning; network design

Received 26 June 2017; Revised 27 September 2017; Accepted 16 October 2017; Published online in Wiley Online Library

1. Introduction

Variational data assimilation (DA) combines observations and a prior estimate (background) of the state of the system, weighting them according to the inverse of their specified error covariances. The output, referred to as the analysis, should theoretically be more accurate than either individual source of information. Data assimilation has proven to be essential for accurate weather forecasting by providing the initial conditions for numerical weather prediction (NWP) (Rabier *et al.*, 2000; Rawlins *et al.*, 2007).

It has long been known that background-error correlations (BECs) are important in DA. The structure of BECs describes how corrections of the prior estimate of one variable should spread to another. As such, the modelling of the BECs has received much attention (e.g. Bannister, 2008a, 2008b).

Observations can also have significant error correlations. However, until recently, observation-error correlations (OECs) have been neglected in NWP. To account for this, the data may be thinned (Dando *et al.*, 2007) or 'super-obbed' (Berger and Forsythe, 2004; van Leeuwen, 2015) or the error variances inflated (Hilton *et al.*, 2009). Therefore, accounting for OECs correctly could allow for denser observations to be assimilated, which could be important for high-resolution weather forecasting (Browne *et al.*, 2014; Sun *et al.*, 2014).

Correlations in observation errors have a very different origin from those in the background. In general, OECs can be attributed to errors in the comparison of observations with model variables, known as representation error, rather than instrument noise (Janjic and Cohn, 2006; Waller *et al.*, 2014b; Hodyss and Nichols, 2015; Hodyss and Satterfield, 2017; Janjic *et al.*, 2017). As such, they can be state- and model-dependent (Waller *et al.*, 2014a) and

are found predominantly in observation types with complex observation operators (such as satellite radiances: Bormann and Bauer, 2010; Stewart *et al.*, 2014; Bormann *et al.*, 2016; Waller *et al.*, 2016a; Campbell *et al.*, 2017), observation types measuring features with natural length- and time-scales that are different from those resolved by the model (e.g. Doppler radial winds: Waller *et al.*, 2016c) and high-level observation products that 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).

The inclusion of OECs clearly has the potential to improve the optimality of the DA system in the case in which they were previously neglected, or only accounted for by the inflation of diagonal error variances. Evaluating the problems caused by erroneously not including OECs has been the subject of many recent articles, e.g. Liu and Rabier (2002, 2003), Stewart *et al.* (2008, 2013), Miyoshi *et al.* (2013), Jacques and Zawadzki (2014) and Rainwater *et al.* (2015).

The first attempts at operational centres to account for Infrared Atmospheric Sounding Interferometer (IASI) inter-channel error correlations (which previously relied on variance inflation) have been shown to lead to an improvement in forecast skill scores (Weston *et al.*, 2014; Bormann *et al.*, 2016; Campbell *et al.*, 2017). This has motivated estimation of OECs for a range of other observations, for example using methods based on innovation consistency diagnostics (Desroziers *et al.*, 2005). The question is: how do the background and observation-error correlations interact and how does this affect the value of the observations in the analysis? Addressing this question is essential to understanding how future observation networks should be designed in order to meet the needs of high-resolution NWP. A brief review of the literature, addressing aspects of this question, is provided below. This article aims to unify these previous works and provide new insight in order to extend these previous conclusions.

When the OECs are positive, modelling OECs correctly has been shown to allow an instrument to provide more information about small scales (Seaman, 1977; Rainwater *et al.*, 2015). As such, observations with correlated errors may provide a more accurate analysis of gradients and small-scale but intense features compared with observations with uncorrelated errors. We will show in section 3 how these results can be extended to explain how the scales at which the observations are able to constrain the analysis depend not only on OECs but also upon BECs and the mapping between state and observation variables. The latter is described by the observation operator.

Miyoshi *et al.* (2013) and Terasaki and Miyoshi (2014) showed that the interaction between OECs and the observation operator are very important. In particular, they showed that when the observations are direct measurements of the state variables, observations with correlated errors (either positive or negative correlations) can reduce the entropy of the posterior probability distribution function (PDF) compared with the case when the observation errors are uncorrelated. However, when the observation operator is expressed as a linear combination of the state variables with positive coefficients, the sign of the OECs becomes important. In this case, OECs only reduce the entropy of the posterior PDF, compared with uncorrelated errors, if the OECs are negative. Through a series of idealized experiments, we will show how these results can be explained in terms of the analysis scales that the observations can constrain (see section 4).

Liu and Rabier (2002) studied the case in which OECs originate from the observations measuring different scales from those modelled. This effect was simulated in these experiments by generating observations from a model run at a higher spectral resolution than that used in the assimilation. In this case, the OECs and the observation operator are intrinsically related. Liu and Rabier (2002) presented results on the optimal thinning of these observations with correlated errors. They found that if the OECs are correctly modelled, then increasing the observation density beyond a certain threshold does not reduce the analysis error significantly. In contrast, if the errors are uncorrelated, denser (i.e. more) observations will always lead to a reduction in analysis error. This was similarly the conclusion of Bergman and Bonner (1976), who also studied analysis errors as a function of observation density with spatially correlated errors. We are able to extend these results using metrics other than analysis root-mean-square error (RMSE), to show that this is only half the picture. In section 6, the value of dense observations is shown to depend not only on OECs but how they relate to BECs. In particular, here we also consider the case in which the OEC length-scales are longer than the BEC length-scales, which may be the case, for example, for AMVs (Cordoba *et al.*, 2017).

This article is structured as follows. In section 2, we will provide the theoretical background on three different metrics of the observation impact on the analysis. In section 3, these metrics will be considered for the case in which the background- and observation-error covariances and the observation operator can all be described by circulant matrices. In sections 4 and 5, a series of numerical experiments is presented. It is shown that the reduction in analysis-error variances and spread of information is greatest when the observation and prior errors have complementary statistics. This is explained by the fact that the different correlation structures allow the observations and prior to provide accurate estimates of the state at different scales. In section 6, the implications of these results on the optimal thinning of a variety of observations will be demonstrated. It is found that dense observations are most beneficial when they provide more small-scale information than that available from the prior estimate. In the case of second-order autoregressive correlation functions, this may be achieved when the length-scales of the OECs are greater than those of the BECs and the observations are direct measurements of the state variables.

2. Variational data assimilation

Variational data assimilation aims to find the most probable state, $\mathbf{x}^{\mathrm{a}} \in \mathbb{R}^n$, given a prior estimate (commonly known as the 'background'), $\mathbf{x}^{\mathrm{b}} \in \mathbb{R}^n$, and observations, $\mathbf{y} \in \mathbb{R}^p$. The background in general will represent the discretized model variables (known as state space). The observations are not necessarily in state space and the mapping between the state and observation space is characterized by the observation operator, $h : \mathbb{R}^n \to \mathbb{R}^p$. The observation operator can be used to relate the observations to the discretized version of the truth (in state space), $\mathbf{x}^{\mathrm{t}}$:

$$\mathbf{y} = h(\mathbf{x}^{\mathrm{t}}) + \epsilon, \tag{1}$$

where ϵ is the observation error. The observation operator may be as simple as linear interpolation from the model grid to the observation, for example in the case of radiosonde measurements of temperature and humidity or sea-surface temperature products derived from satellite instruments. Alternatively the observation operator may be more complicated, for example it may include a radiative transfer model in the case of top-of-the-atmosphere radiances measured from satellite instruments (Rodgers, 2000). In this case, the linearized radiative transfer model can be interpreted as weighting functions representing the sensitivity of the measured radiance to the state variables, with the altitude of the peak of the weighting function and its width depending on the instrument characteristics and the wavelength observed. If the satellite instrument measures at a number of close wavelengths, it can be expected that the weighting functions will overlap (Rodgers, 2000).

It is assumed that both the background and observations are unbiased estimates of the truth with Gaussian errors. The error covariances are given by $\mathbf{B} \in \mathbb{R}^{n\times n}$ and $\mathbf{R} \in \mathbb{R}^{p\times p}$, respectively. The DA problem can therefore be expressed in terms of Bayes' theorem, as finding the state that maximizes the posterior PDF, $p(\mathbf{x}|\mathbf{y})$. Bayes' theorem states:

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{x})p(\mathbf{y}|\mathbf{x}), \tag{2}$$

where $p(\mathbf{x})$ is the prior PDF and $p(\mathbf{y}|\mathbf{x})$ is the likelihood PDF.

Although the observation error, ϵ, is in observation space, in calculating (2) the likelihood may be thought of as a function of the state variables, $\mathbf{x}$. If the observation operator is (near) linear, observation errors with a Gaussian distribution will result in a (near-) Gaussian likelihood in state space. Note that in practice p is generally less than n, so that there will be directions of state space for which the likelihood has infinite uncertainty, i.e. the observations provide no information.

Variational data assimilation computes an analysis at a given time by minimizing the cost function, $J = -2\ln(p(\mathbf{x}|\mathbf{y}))$, with respect to $\mathbf{x}$. The minimum of J can be shown to correspond to the maximum *a posteriori* state (which due to the Gaussian and linear assumptions is also the minimum variance estimate) and is given by

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

where $\mathbf{K} \in \mathbb{R}^{n\times p}$ is known as the Kalman gain matrix, $\mathbf{K} = \mathbf{B}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}$, with $\mathbf{H} \in \mathbb{R}^{p\times n}$ the observation operator linearized about the best estimate of the state (see Kalnay (2003) for an introduction to variational data assimilation).

2.1. The analysis accuracy

On assimilation of the observations, it is assumed that the analysis is a more accurate estimate of the state than the background. That is, the volume of the region of high probability in the posterior is reduced compared with the prior.

Under the assumption of an optimal $\mathbf{B}$, $\mathbf{R}$ and $\mathbf{H}$, the analysis-error covariance matrix, $\mathbf{P}^{\mathrm{a}} \in \mathbb{R}^{n\times n}$, is given by

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

(Kalnay, 2003). The accuracy of the analysis (given by $(\mathbf{P}^{\mathrm{a}})^{-1}$) is therefore the sum of the accuracy of the background ($\mathbf{B}^{-1}$) and the accuracy of the observations mapped to state space ($\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}$). In the case when $n > p$, $\mathbf{H}$ will have a null space and in this part of state space the accuracy of the analysis will equal the accuracy of the background.

The success of the DA system can be judged in terms of the magnitude of $\mathbf{P}^{\mathrm{a}}$, for example in terms of the trace of this matrix, which should be related to the analysis RMSE. The analysis RMSE was used as a metric of the impact of OECs in many previous studies, e.g. Bergman and Bonner (1976), Seaman (1977), Liu and Rabier (2002), Miyoshi *et al.* (2013), Stewart *et al.* (2013) and Rainwater *et al.* (2015). However, in the results presented here, we will show that the impact of OECs on analysis-error correlations is just as important as their impact on the analysis-error variances.

The Gaussian, unbiased and linear assumptions mean that $\mathbf{P}^{\mathrm{a}}$ describes the uncertainty of the posterior fully. From (2), we see that the posterior PDF is a product of the prior and likelihood. The seemingly trivial statement that high posterior probability will be in regions of state space with high prior and likelihood probability will turn out to be crucial for understanding how the structure of $\mathbf{R}$, $\mathbf{B}$ and $\mathbf{H}$ interacts in the computation of $\mathbf{P}^{\mathrm{a}}$. This will be illustrated in section 4.1 for a simple two-variable experiment.

2.2. The sensitivity matrix

In addition to the analysis-error covariances, it is interesting to study the influence of the observations on the analysis itself. This can be quantified by the sensitivity matrix, $\mathbf{S} \in \mathbb{R}^{p\times p}$, defined as

$$\mathbf{S} = \frac{\partial h(\mathbf{x}^{\mathrm{a}})}{\partial \mathbf{y}} \approx \mathbf{H}\mathbf{K} = \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1} \tag{5}$$

(Cardinali *et al.*, 2004). In an optimal system, the sensitivity matrix can be related to the analysis-error covariances as $\mathbf{S} = \mathbf{H}\mathbf{P}^{\mathrm{a}}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}$.

The diagonal elements of the sensitivity matrix, S_{ii}, $i = 1, 2, ..., p$, known as the self-sensitivities, measure the sensitivity of $h(\mathbf{x}^{\mathrm{a}})_i$ to $\mathbf{y}_i$. The off-diagonal elements, S_{ij}, $j = 1, 2, ..., p$, known as the cross-sensitivities, measure the sensitivity of $h(\mathbf{x}^{\mathrm{a}})_i$ to $\mathbf{y}_j$. The cross-sensitivities therefore provide a measure of the spread of information from the observations.

The sensitivity matrix can be summarized to give the overall information content of the observations. One way to do this is to calculate the trace of $\mathbf{S}$ that gives the degrees of freedom for the signal (dfs: Cardinali *et al.*, 2004). The dfs, by definition, is only a function of the self-sensitivities. The dfs has previously been used for objectively choosing a subset of the most useful channels for IASI, which measures more channels than can practically be stored, transmitted and assimilated (Rabier *et al.*, 2002; Collard, 2007; Eresmaa *et al.*, 2014; Ventress and Dudhia, 2014).

2.3. Mutual information

A second way to summarize the sensitivity matrix is based on mutual information, also known as Shannon information content. Mutual information measures the reduction in entropy (uncertainty) due to assimilation of the observations and has also previously been used in the objective selection of IASI channels (Rabier *et al.*, 2002; Fowler, 2017).

Mutual information measures the change in entropy from the prior to the posterior. For a general PDF, ξ, entropy is defined as

$$E(\xi) = -\int \xi(x)\ln\xi(x)\,\mathrm{d}x. \tag{6}$$

The entropy of a Gaussian distribution with covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$ is therefore

$$E(\xi) = n\ln(2\pi e)^{1/2} + \frac{1}{2}\ln[\det(\Sigma)] \tag{7}$$

(Rodgers, 2000).

The effect of error correlations on the entropy is directly related to the effect on the determinant of the covariance matrix. For example, let a given covariance matrix, $\Sigma \in \mathbb{R}^{n\times n}$, be decomposed in terms of its standard deviations, $\mathbf{D} \in \mathbb{R}^{\mathbf{n}\times\mathbf{n}}$ ($=\sqrt{\mathrm{diag}(\Sigma)}$), and correlations, $\mathbf{C} \in \mathbb{R}^{n\times n}$, such that

$$\Sigma = \mathbf{D}\mathbf{C}\mathbf{D}. \tag{8}$$

Let λ_i^{c} be the ith eigenvalue of $\mathbf{C}$ for $i = 0, 1, ..., n-1$. The trace of $\mathbf{C}$ is a constant ($\sum_{i=0}^{n-1}\lambda_i^{\mathrm{c}} = n$). However, the determinant of $\mathbf{C}$, $\det(\mathbf{C}) = \prod_{i=0}^{n-1}\lambda_i^{\mathrm{c}}$, is sensitive to the structure of the correlations and is known to satisfy

$$0 < \det(\mathbf{C}) \leq 1. \tag{9}$$

This is a special case of Hadamard's inequality (see Pahl and Damrath, 2012, p914). The upper bound is attained when $\mathbf{C} = \mathbf{I}$, i.e. no correlations are present and $\lambda_i^{\mathrm{c}} = 1$, $\forall i$. For a perfectly correlated matrix, $\mathbf{C}$ is a matrix populated by ones. In this case, the largest eigenvalue is n and the rest are zero. Heuristically, we might therefore expect that, as the errors become more correlated, $\det(\mathbf{C})$ tends to zero and the region of uncertainty collapses along the leading eigenvector.

This is illustrated in Figure 1, which shows the region of 95% probability of a Gaussian PDF for a two-variable case when the covariance matrix is given by (i) $\mathbf{I}$ (black dashed line) and (ii) $[(1, 0.99)^{\mathrm{T}}, (0.99, 1)^{\mathrm{T}}]$ (solid line). In the first case the entropy is 2.8379 and in the second case the entropy is 0.8794. Note that the entropy effectively measures the volume of the uncertainty and not its shape or orientation, so a Gaussian PDF with covariance matrix given by $[(1, -0.99)^{\mathrm{T}}, (-0.99, 1)^{\mathrm{T}}]$ would also have an entropy of 0.8794, as would a Gaussian PDF with covariance matrix given by $0.1411\mathbf{I}$.

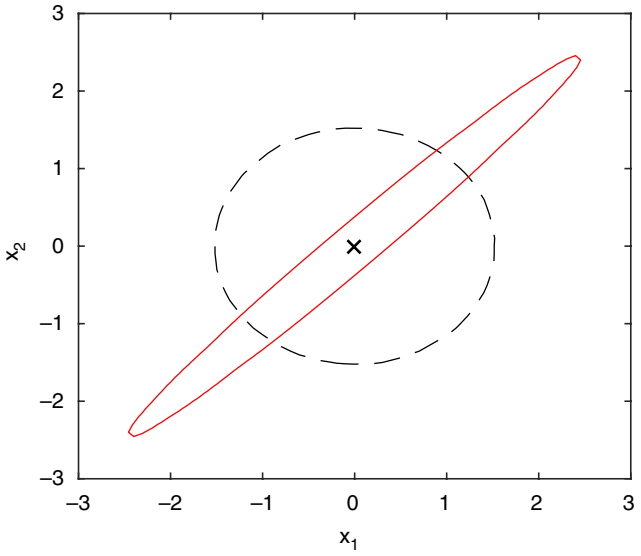

Figure 1. Illustration of the 95th percentile of the region of uncertainty given by a Gaussian distribution when the errors in x_1 and x_2 are uncorrelated (black dashed) and correlated with a coefficient of 0.99 (solid). The cross marks the mean value. [Colour figure can be viewed at wileyonlinelibrary.com].

Mutual information is given by the difference between the prior and posterior entropy and hence is a function of the background and analysis-error covariances only:

$$MI = 0.5 \ln \left(\det \left[\mathbf{B} \left(\mathbf{P}^{\mathrm{a}} \right)^{-1} \right] \right) \tag{10}$$

(Rodgers, 2000).

This can be written as a function of the sensitivity matrix, $\mathbf{S}$, using the identities $\mathbf{B}(\mathbf{P}^{\mathrm{a}})^{-1} = (\mathbf{I}_n - \mathbf{KH})^{-1}$, and $\det(\mathbf{I}_n - \mathbf{KH}) = \det(\mathbf{I}_p - \mathbf{HK})$ (Pozrikidis, 2014, p271):

$$MI = -0.5 \ln \det(\mathbf{I}_p - \mathbf{S}) = -0.5 \sum_{k=0}^{p-1} \ln(1 - \lambda_k^{\mathrm{s}}), \tag{11}$$

where λ_k^{s} is the kth eigenvalue of the sensitivity matrix and $\mathbf{I}_m$ is the identity matrix with dimension m. MI therefore differs fundamentally from the degrees of freedom for the signal, as it takes into account the cross-sensitivities and can therefore lead to a different interpretation of how the information content of the observations depends on the inputs of the DA system. From the previous discussion of entropy, we can expect MI to be greatest when the observations not only have had the effect of reducing the analysis-error variances compared with the background-error variances but also have resulted in strongly correlated analysis errors.

2.4. Interpretation of measures of observation impact when errors are uncorrelated

Both the analysis-error covariance matrix, $\mathbf{P}^{\mathrm{a}}$, and the sensitivity matrix, $\mathbf{S}$, are functions of only the error covariances, $\mathbf{B}$ and $\mathbf{R}$, and the observation operator, $\mathbf{H}$. If $\mathbf{B}$ and $\mathbf{R}$ are diagonal and $\mathbf{H} = \mathbf{I}$, then $\mathbf{P}^{\mathrm{a}}$ and $\mathbf{S}$ will also be diagonal, with diagonal elements P_{ii}^{a} and S_{ii} given by

$$P_{ii}^{\mathrm{a}} = \frac{(\sigma_i^{\mathrm{o}})^2 (\sigma_i^{\mathrm{b}})^2}{(\sigma_i^{\mathrm{o}})^2 + (\sigma_i^{\mathrm{b}})^2} \tag{12}$$

and

$$S_{ii} = \frac{(\sigma_i^{\mathrm{b}})^2}{(\sigma_i^{\mathrm{o}})^2 + (\sigma_i^{\mathrm{b}})^2}, \tag{13}$$

where σ_i^{o} and σ_i^{b} are the observation- and background-error standard deviations of the ith variable, respectively.

In this case, the analysis-error variances will decrease as either the background- or observation-error variances decrease and the analysis will become more sensitive to the observations as either the observation-error variances decrease or the background-error variances increase.

The mutual information, in this case, is given by

$$MI = 0.5 \sum_{k=0}^{p-1} \ln \frac{(\sigma_k^{\mathrm{o}})^2 + (\sigma_k^{\mathrm{b}})^2}{(\sigma_k^{\mathrm{o}})^2}. \tag{14}$$

In a similar way to the sensitivity of the analysis to the observations, MI is seen to increase as the observation-error variances decrease and the background-error variances increase.

When the errors are correlated or the observations are no longer direct, it is less straightforward to interpret how the analysis-error covariance and sensitivity matrix depend on the error characteristics. In the next section, we introduce a theoretical framework to give insight into how these measures change as a function of the error correlation length-scales of the background and observation errors and the structure of the observation operator.

3. Circulant matrices

A matrix $\mathbf{C} \in \mathbb{R}^{N \times N}$ can be described as circulant if it can be expressed in the following way:

$$\mathbf{C} = \begin{pmatrix} c_0 & c_1 & \dots & c_i & \dots & c_N \\ c_N & c_0 & \dots & c_{i-1} & \dots & c_{N-1} \\ \vdots & \vdots & & \vdots & & \vdots \\ c_1 & c_2 & \dots & c_{i+1} & \dots & c_0 \end{pmatrix}, \tag{15}$$

that is each row vector is shifted cyclically one element to the right relative to the preceding row vector. For this circulant matrix also to be a correlation matrix, it must be symmetric such that $c_i = c_{N-i}$, reducing the number of independent elements, N_c, to $N/2$ if N is even or $(N+1)/2$ if N is odd. As the matrix, $\mathbf{C}$, is fully specified by one vector $(\ c_0 \quad c_1 \quad \dots \quad c_{N_c} \quad \dots \quad c_1\ )$, the structure of the covariance matrix can be described in terms of a single correlation function.

The eigenvalues of a circulant matrix can be found using a discrete Fourier transform, with the eigenvectors being the discrete Fourier basis (Gray, 2006). The mth eigenvector is therefore given by

$$\frac{1}{\sqrt{N}} (1, \mathrm{e}^{-2\pi m/N}, \dots, \mathrm{e}^{-2\pi m(N-1)/N})^{\mathrm{T}}. \tag{16}$$

It can be seen from (16) that the eigenvalues are ordered according to wave number, with the first eigenvalue relating to the Fourier mode with the largest length-scales. In general, this ordering is not linked to the magnitude of the eigenvalues. The mth eigenvalue for any circulant matrix is given by

$$l_m = \sum_{k=0}^{N-1} c_k \mathrm{e}^{-2\pi imk/N}, \tag{17}$$

where c_k is the kth element of the vector characterizing the circulant matrix. Note that the eigenvalues of a correlation matrix are always greater than or equal to zero.

Circulant matrices of the same dimension have identical eigenvectors (seen from (16)). Therefore, assuming both the background and observations are vectors of the same length, $p = n$, their error covariances can be described by a circulant

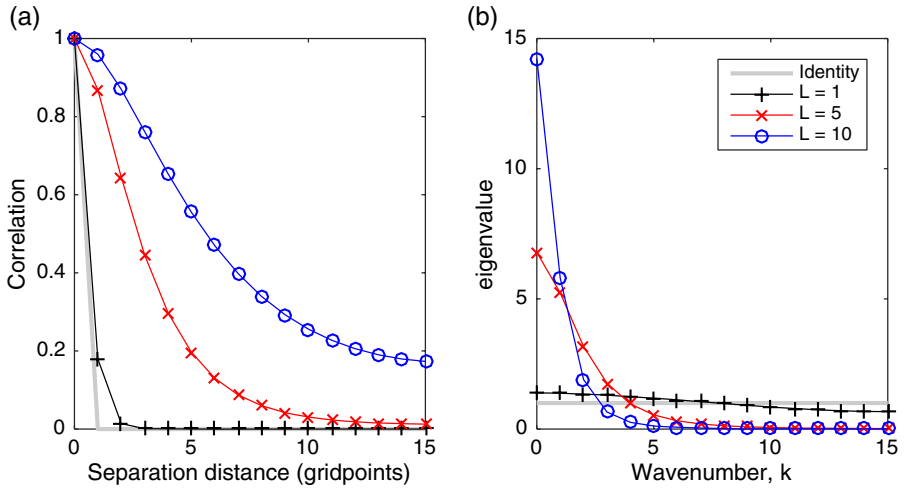

Figure 2. (a) The SOAR function and (b) its eigenspectrum for varying correlation length-scales, assuming a circulant correlation matrix of the form (15) and $\Delta x = \pi$. The expression for the SOAR function is given by (19). [Colour figure can be viewed at wileyonlinelibrary.com].

matrix and the observation operator is also of the same form, we can write them as

$$\begin{aligned} \mathbf{B} &= \beta \mathbf{F} \Gamma \mathbf{F}^{\mathrm{T}}, \\ \mathbf{R} &= \rho \mathbf{F} \Psi \mathbf{F}^{\mathrm{T}}, \\ \mathbf{H} &= \mathbf{F} \Phi \mathbf{F}^{\mathrm{T}}, \end{aligned} \tag{18}$$

where $\mathbf{F}$ is a matrix containing the eigenvectors common to each circulant matrix and β and ρ are the error variances of the background and observations, respectively, which we have assumed to be constant. The diagonal matrices Γ and Ψ contain the eigenvalues of the correlations of the background and observation errors respectively. The diagonal matrix Φ contains the eigenvalues of the observation operator.

This framework allows us to give insight into how changing the error correlation length-scales changes the uncertainty at different wave numbers. We know that, as the trace of the correlation matrix is conserved, the effect of changing the correlation structure will mean that at some scales the eigenvalues will be increased whilst at other scales the eigenvalues will decrease. When the correlation coefficients are positive and monotonically decreasing, increasing the correlation length-scale will result in an increase in the magnitude of the eigenvalues at small wave number and a decrease at large wave number (Seaman, 1977; Rainwater *et al.*, 2015; Waller *et al.*, 2016b). This is illustrated in Figure 2 for the case in which the correlations are described by a SOAR (Second-Order AutoRegressive) function, defined as

$$c_k = (1 + r_k/L)\mathrm{e}^{-r_k/L}, \tag{19}$$

where r_k is the distance between two points and L is the correlation length-scale. We are therefore, in the case of positive monotonically decreasing correlation functions, able to associate an increase in error correlation length-scale with an increase in uncertainty at large scales and a decrease in uncertainty at small scales.

3.1. *Application to measures of observation impact*

We can substitute the expressions for $\mathbf{B}$, $\mathbf{R}$ and $\mathbf{H}$ (with the additional assumption of symmetry) in (18) into (4) and (5) to show that the analysis-error covariance matrix and the sensitivity matrix in this case will also be circulant, with the same eigenvectors as $\mathbf{H}$, $\mathbf{B}$ and $\mathbf{R}$. The kth eigenvalue of $\mathbf{P}^{\mathrm{a}}$ is given by

$$\lambda_k^{\mathrm{a}} = \frac{\rho\beta\psi_k\gamma_k}{\beta\phi_k^2\gamma_k + \rho\psi_k}, \tag{20}$$

where $\psi_k = \Psi_{kk}$, $\gamma_k = \Gamma_{kk}$ and $\phi_k = \Phi_{kk}$. The kth eigenvalue of $\mathbf{S}$ is given by

$$\lambda_k^{\mathrm{s}} = \frac{\beta\phi_k^2\gamma_k}{\beta\phi_k^2\gamma_k + \rho\psi_k}. \tag{21}$$

Note that ψ_k and γ_k will be greater than or equal to zero for all k, due to the positive semi-definite constraint of correlation matrices.

Let us first examine the case when $\phi_k = 1, \forall k$ (i.e $\mathbf{H} = \mathbf{I}$ and we have direct observations of the whole state). In this case, the analysis uncertainty at scales associated with the kth wave number, λ_k^{a}, will decrease as $\beta\gamma_k$ and $\rho\psi_k$ decrease. Additionally the analysis will become more sensitive to observations at scales associated with the kth wave number, λ_k^{s}, as $\beta\gamma_k$ increases and $\rho\psi_k$ decreases.

Let L_{B} and L_{R} describe the correlation length-scales of $\mathbf{B}$ and $\mathbf{R}$ in the case in which the correlation structures are assumed to be given by the same positive function that is monotonically decreasing with separation distance (e.g. the SOAR function) as described in the introduction to this section. As L_{R} and L_{B} increase, ψ_k and γ_k will therefore increase at small wave numbers and decrease at large wave numbers (recall Figure 2). Hence, we can conclude from (20) that, when L_{R} and L_{B} increase, the analysis uncertainty will decrease at small scales and increase at large scales. Similarly, from (21) we can conclude that when $L_{\mathrm{R}} < L_{\mathrm{B}}$ the analysis will be more sensitive to observations at large scales than at small scales and vice versa when $L_{\mathrm{R}} > L_{\mathrm{B}}$. When $L_{\mathrm{R}} = L_{\mathrm{B}}$ and $\gamma_k = \psi_k, \forall k$, the analysis has the same sensitivity $(\beta/(\beta + \rho))$ at all scales. This was also noted by Daley (1996, section 4.8).

Now, consider the case in which the observation operator also has a length-scale, L_H, describing a positive weighting function. In this case, analogous to an increase in the correlation length-scales, as L_H increases ϕ_k will increase at small wave numbers and decrease at large wave numbers. We can see from (20) that this would lead to a decrease in the analysis uncertainty associated with large scales compared with direct observations and an increase in analysis uncertainty at small scales. Likewise, we can see from (21) that, as L_H increases, the analysis will become more sensitive to observations at large scales compared with direct observations and less sensitive to observations at small scales. This will be illustrated further in the numerical results of section 4.

In this circulant framework, the mutual information (10) can be expressed as

$$MI = 0.5\sum_{k=0}^{p-1} \ln\left(\frac{(\beta\gamma_k\phi_k^2 + \rho\psi_k)}{\rho\psi_k}\right). \tag{22}$$

When $\phi_k = 1, \forall k$ and $\gamma_k = \psi_k, \forall k$ (e.g. $\mathbf{H} = \mathbf{I}$ and $L_\mathrm{R} = L_\mathrm{B}$), the mutual information reduces to

$$\frac{p}{2} \ln \left(\frac{\beta + \rho}{\rho} \right).$$

Therefore in this special case *MI* is independent of the exact value of the background- and observation-error correlation length-scales. To provide some intuition as to how *MI* may depend on the structure of **H**, **B** and **R** in the more general case, in the next section we present numerical results for a series of simple experiments.

4. Two-variable problem

Let us consider a simple experimental design in which OECs, BECs and a non-identity observation operator are present. Let the circulant error covariance and observation operator matrices be given by

$$\mathbf{B} = \beta \begin{bmatrix} 1 & \chi_\mathrm{b} \\ \chi_\mathrm{b} & 1 \end{bmatrix} \tag{23}$$

where $-1 < \chi_\mathrm{b} < 1$,

$$\mathbf{R} = \rho \begin{bmatrix} 1 & \chi_\mathrm{r} \\ \chi_\mathrm{r} & 1 \end{bmatrix} \tag{24}$$

where $-1 < \chi_\mathrm{r} < 1$ and

$$\mathbf{H} = \begin{bmatrix} 1 & a \\ a & 1 \end{bmatrix}. \tag{25}$$

This experimental design allows us to consider the case in which the error correlations are negative as well as positive and the observation operator is a linear combination of the state variables with negative coefficients (e.g. $a < 0$) as well as positive coefficients (e.g. $a > 0$).

Within this system, two scales are present; a large scale associated with the first eigenvector and a small scale associated with the second eigenvector. In this case the eigenvectors (see (16)) are given by $(1/\sqrt{2})\left[[1,1]^\mathrm{T}, [1,-1]^\mathrm{T}\right]$, respectively. Therefore, the large scale is represented in eigenspace as a scaled version of the mean value in physical space and the small scale is represented in eigenspace as a scaled version of the gradient in physical space. The eigenvalues of the correlation matrices and observation operator given by (23)–(25) can be expressed as

$$\begin{aligned} \gamma_0 &= 1 + \chi_\mathrm{b}, \quad & \gamma_1 &= 1 - \chi_\mathrm{b}, \\ \psi_0 &= 1 + \chi_\mathrm{r}, \quad & \psi_1 &= 1 - \chi_\mathrm{r}, \\ \phi_0 &= 1 + a, \quad & \phi_1 &= 1 - a, \end{aligned} \tag{26}$$

where γ_i, ψ_i and ϕ_i are the corresponding eigenvalues of **B**, **R** and **H**, respectively.

In section 3.1, we discussed the effect of indirect observations on the analysis uncertainty and analysis sensitivity to observations for the case in which **H** can be described by a circulant matrix with a positive weighting function. For the simple two-variable problem described above, these conclusions also hold when a is positive and less than 1. That is, as a increases, the observations are able to constrain the analysis less at small scales but are able to provide a greater constraint at large scales. As such, the analysis becomes more sensitive to observations at large scales and less at small scales than if the observations were direct measurements of the state variable.

For this simple two-variable problem, it is also possible to show that the reverse is true when the observation-operator coefficient is negative. That is, when $-1 < a < 0$ as $|a|$ increases, the observations are able to constrain the analysis *more* at small scales but provide a *reduced* constraint at large scales compared with having direct observations. As such, the analysis becomes *less* sensitive to the observations at large scales and *more* at small scales than if the observations were direct measurements of the state variable.

Table 1. The values of the analysis-error covariances for the two-variable cases illustrated in Figure 3, for varying values of the observation-operator coefficient, a, and the observation-error correlation, χ_r. **B**, **R** and **H** are given by (23)–(25) with $\beta = 1$, $\rho = 1$ and $\chi_\mathrm{b} = 0.9$. The correlations (corr) are given by $P^\mathrm{a}_{ij}/P^\mathrm{a}_{ii}$.

	$a = 0$			$a = 0.5$		
$\lvert\chi_\mathrm{b} - \chi_\mathrm{r}\rvert$	P^a_{ii}	P^a_{ij}	(corr)	P^a_{ii}	P^a_{ij}	(corr)
0	0.500	0.450	(0.900)	0.332	0.252	(0.759)
0.9	0.370	0.282	(0.756)	0.229	0.131	(0.452)
1.8	0.095	0	(0)	0.071	−0.028	(−0.394)

4.1. *Illustration in terms of Bayes' theorem*

To provide a geometrical interpretation of the interaction of **H**, **B** and **R**, we first consider Bayes' theorem (2). Figure 3 illustrates the resulting posterior distribution for six different cases of **H**, **B** and **R**. In each case, **B** is given by (23) with $\beta = 1$ and $\chi_\mathrm{b} = 0.9$, **R** is given by (24) with $\rho = 1$ and χ_r allowed to vary and **H** is given by (25) with a allowed to vary. It is assumed that the truth is given by $\mathbf{x}^\mathrm{t} = (0, 0)$.

In Figures 3(a)–(c), the observation operator is taken to be the identity, that is $a = 0$ in (25). We can see that the region of high posterior probability (third column) coincides with the region where the *a priori* (first column) and likelihood (second column) high probabilities coincide. Therefore when the structure of the prior and likelihood PDFs are very different from each other, the region of high posterior probability is small and consequently the analysis variance is small. In this example, we therefore see that the analysis-error variance decreases as $|\chi_\mathrm{b} - \chi_\mathrm{r}|$ increases (see Table 1). The analysis-error correlations are also seen to reduce as $|\chi_\mathrm{b} - \chi_\mathrm{r}|$ increases.

In Figures 3(d)–(f), the observation operator is no longer the identity and has the form (25) with $a = 0.5$; **B** and **R** remain the same as in Figures 3(a)–(c). The first column shows the effect this choice of observation operator has on the likelihood PDF in state space.

As expected, we see that the effect of $a = 0.5$ is to rotate and deform the likelihood in state space so that, as a function of the state variables, it is more negatively correlated than as a function of the observation variables. As discussed in section 3, this is consistent with the observations providing a more accurate estimate of the state at large scales and a less accurate estimate of the state at small scales, compared with having direct observations.

The effect of **H** on the correlations explains the results given by Miyoshi *et al.* (2013) and Terasaki and Miyoshi (2014), which showed that when a and χ_r are different signs the observations have the greatest information (see e.g. Figure 3(f)). This is because under these conditions the observation-error correlations are strengthened and the entropy of the likelihood is reduced.

In this case (**H** invertible), the observation-error covariance mapped to state space is given by

$$\mathbf{H}^{-1}\mathbf{R}\mathbf{H}^{-1} = \frac{\rho}{(a^2-1)^2} \begin{bmatrix} 1 - 2a\chi_\mathrm{r} + a^2 & \chi_\mathrm{r} - 2a + a^2\chi_\mathrm{r} \\ \chi_\mathrm{r} - 2a + a^2\chi_\mathrm{r} & 1 - 2a\chi_\mathrm{r} + a^2 \end{bmatrix}. \tag{27}$$

(For practical applications, we note that **H** is rarely invertible.)

It can be shown that if

$$2a\chi_\mathrm{r} < a^2(3 - a^2), \tag{28}$$

then the observation operator increases the variances of the observation errors mapped to state space (i.e. $(\mathbf{H}^{-1}\mathbf{R}\mathbf{H}^{-1})_{ii} > \mathbf{R}_{ii}$). In the cases illustrated, $a = 0.5$, so we can expect the variances to be increased when $\chi_\mathrm{r} < 0.6875$, as is the case in Figures 3(e) and (f).

In the cases illustrated, the analysis-error variance is always smaller when $a = 0.5$ compared with when $a = 0$ (see Table 1).

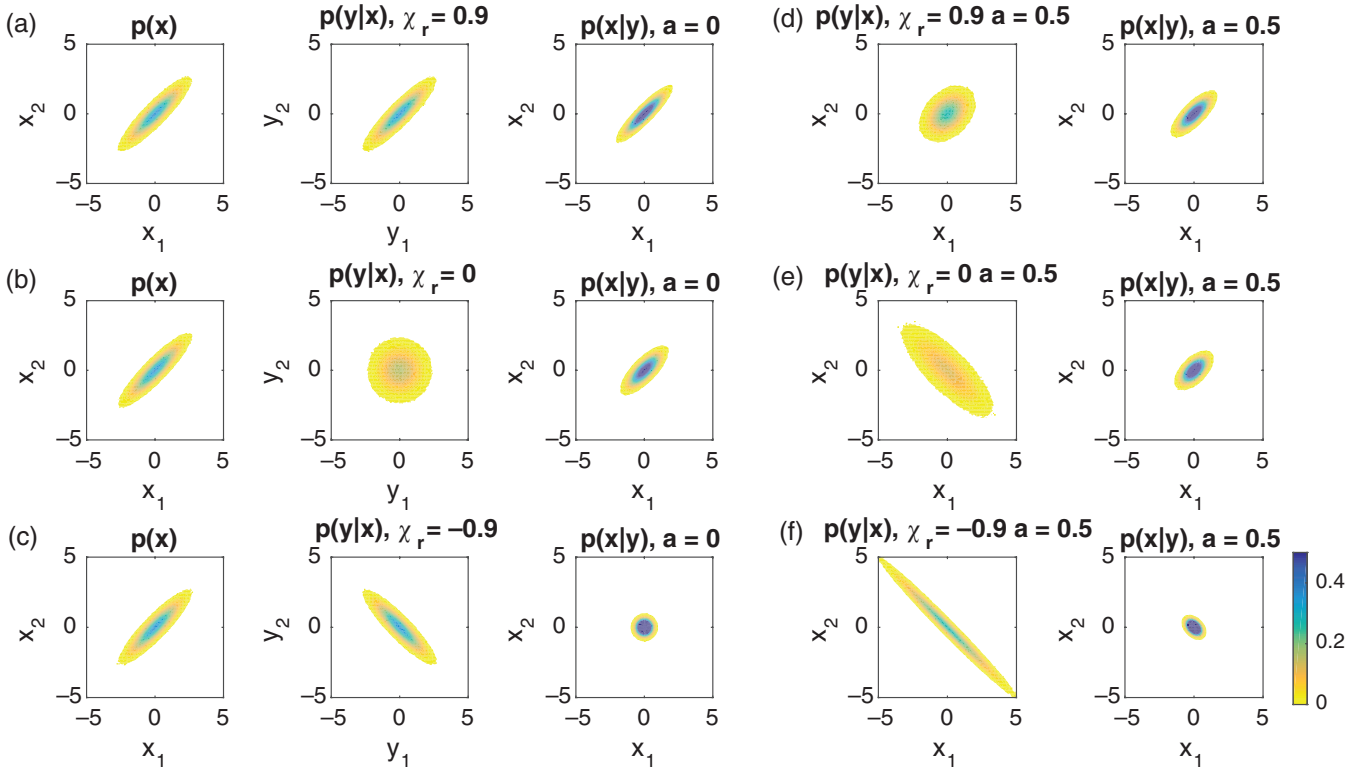

Figure 3. Illustration of the effect of observation-error correlations on analysis-error covariances. (a)–(c) Colours show (from left to right) the prior PDF, likelihood PDF in observation space and resulting posterior PDF when $\mathbf{H} = \mathbf{I}$. (d)–(f) Colours show (from left to right) the likelihood distribution in state space and resulting posterior PDF when $\mathbf{H} = [(1, 0.5)^{\mathrm{T}}, (0.5, 1)^{\mathrm{T}}]$ (the prior PDF is the same as in (a)–(c), so is not replotted). In each example, the background and observation-error variances are 1 and the background-error correlations χ_{b} are 0.9. The observation-error correlations χ_{r} are 0.9 in examples (a) and (d), 0 in examples (b) and (e) and −0.9 in examples (c) and (f).

This is because, in this case ($\chi_{\mathrm{b}} > 0$), the background is more accurate at small scales than at large scales and so the observation operator (with $a > 0$) has a beneficial effect, by increasing the influence of the observations at large scales.

These results could be expected to be sensitive to the choice of structure of the observation operator. Other observation operators for simple two-variable examples were considered, such as $\mathbf{H} = [(\sqrt{1+a^2}, a)^{\mathrm{T}}, (a, \sqrt{1+a^2})^{\mathrm{T}}]$ and $\mathbf{H} = [1/(1+a)][(1, a)^{\mathrm{T}}, (a, 1)^{\mathrm{T}}]$. It was found that the first (which conserves $\det(\mathbf{H}) = 1$) had little effect on the results. In contrast, the latter ($\mathbf{H} = [1/(1+a)][(1, a)^{\mathrm{T}}, (a, 1)^{\mathrm{T}}]$) had a large impact on the variances of the likelihood in state space, increasing them substantially.

In the remainder of this section, we will show what the results presented for the Bayes' illustration mean in terms of our three metrics of observation impact. In each case, $\mathbf{B}$ and $\mathbf{R}$ are given by (23) and (24), respectively, with $\beta = 2$ and $\rho = 1$. The choice of unequal error variances is to highlight when the metric of observation impact as a function of the error correlations depends qualitatively on β/ρ and when it does not. The observation operator is given by (25) with $a = [-0.5, 0, 0.5]$. We see that the observation operator, $\mathbf{H}$, given by (25) has the nice property that the response of the different measures of observation impact is symmetric between a positive and a negative.

4.2. *Analysis-error covariance matrix*

4.2.1. *Analysis-error variances*

For this simple experimental design, the analysis-error covariance matrix is also circulant. Thus the error variance, $\mathbf{P}^{\mathrm{a}}_{ii}$, is the same for each variable. $\mathbf{P}^{\mathrm{a}}_{ii}$ can be written in terms of the eigenvalues of the analysis-error covariance matrix as follows:

$$\mathbf{P}^{\mathrm{a}}_{ii} = (\lambda^{\mathrm{a}}_0 + \lambda^{\mathrm{a}}_1)/2, \tag{29}$$

where λ^{a}_k for $k = 0, 1$ are given by (20). $\mathbf{P}^{\mathrm{a}}_{ii}$ therefore depends upon the analysis uncertainty at all scales.

Figure 4 shows the analysis-error variance, $\mathbf{P}^{\mathrm{a}}_{ii}$, as a function of background-error correlation, χ_{b}, and observation-error correlation, χ_{r}, for three different values of the observation-operator coefficient: $a = -0.5$ (left), 0 (middle) and 0.5 (right).

We first consider the case in which $a = 0$ (middle panel). As anticipated from the illustration in section 4.1, we see that the analysis-error variance is greatest when $\chi_{\mathrm{b}} = \chi_{\mathrm{r}}$. This is independent of the values of ρ and β (see Appendix A). This is because when $\chi_{\mathrm{b}} = \chi_{\mathrm{r}}$ the background and observations are each most accurate at the same scales and so the assimilation will provide an accurate analysis only at that scale. However, if $\chi_{\mathrm{b}} \neq \chi_{\mathrm{r}}$ and the background and observations are most accurate at different scales, then the assimilation will be able to weight the data appropriately to provide an accurate analysis at each of those scales.

When $a = 0.5$ (right-hand panel), this corresponds to the cases illustrated in Figure 3(d)–(f). We see that, when $\chi_{\mathrm{b}} > 0.25$ and a is positive, for most values of χ_{r} the analysis-error variance is reduced compared with the case when $a = 0$. This is because, when χ_{b} is positive, the background estimate of the state is more accurate at small scales than at large scales. In contrast, a positive a reduces the accuracy of the observations in state space at small scales but increases the accuracy at large scales, as noted in the discussion of the eigenvalues of $\mathbf{H}^{-1}\mathbf{R}\mathbf{H}^{-1}$ in section 3. Each source of information is therefore able to reduce the analysis uncertainty effectively at different scales. However, when $\chi_{\mathrm{b}} < 0$ and a is positive, the analysis-error variance is increased compared with when $a = 0$. This is because, when χ_{b} is negative, the background estimate of the state is more accurate at large scales than at small scales. Therefore the effect of a positive a is no longer beneficial. The reverse argument can be used to understand the pattern for the analysis-error covariances when a is negative (left-hand panel).

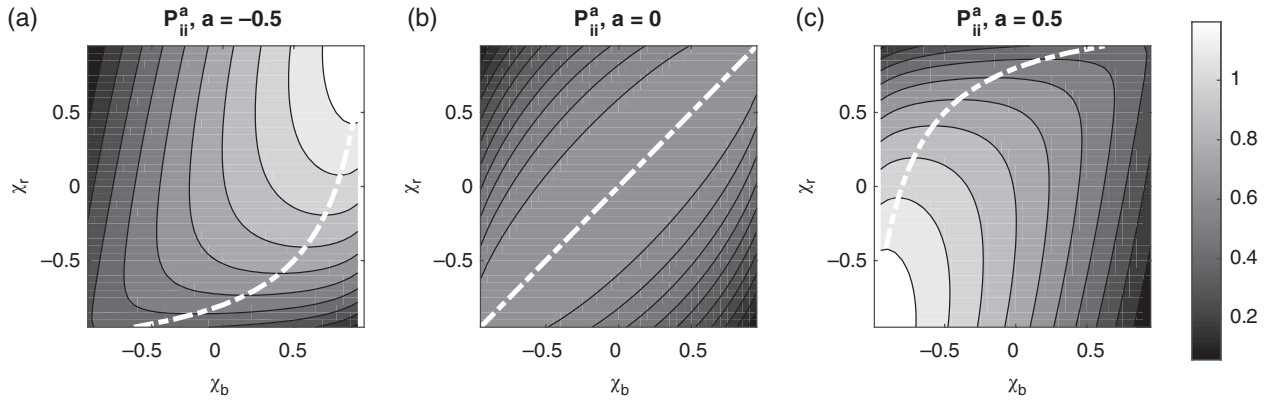

Figure 4. The analysis-error variance, $\mathbf{P}^{a}_{ii}$, as a function of background-error correlation, χ_b, and observation-error correlation, χ_r. In each case, **B** and **R** are given by (23) and (24), respectively, with $\beta = 2$ and $\rho = 1$. **H** is given by (25) with (a) $a = -0.5$, (b) $a = 0$ and (c) $a = 0.5$. The dash–dotted white line indicates the value of χ_b that results in the maximum $\mathbf{P}^{a}_{ii}$ as a function of χ_r as derived in Appendix A; this is found to be independent of the choice of error variances, ρ and β.

4.2.2. *Analysis-error covariances*

Figure 5 shows the analysis-error covariance, $\mathbf{P}^{a}_{ij}$, as a function of background-error correlation, χ_b, and observation-error correlation, χ_r, for the three different values of the observation-operator coefficient shown in Figure 4. $\mathbf{P}^{a}_{ij}$ can be written in terms of the eigenvalues of the analysis-error covariance matrix as follows:

$$\mathbf{P}^{a}_{ij} = (\lambda^{a}_{0} - \lambda^{a}_{1})/2. \tag{30}$$

The sign of $\mathbf{P}^{a}_{ij}$ can be interpreted in terms of the analysis uncertainty at different scales. Positive values (red) mean the analysis is more uncertain at large scales than at small scales ($\lambda^{a}_{0} > \lambda^{a}_{1}$), negative values (blue) mean the analysis is more uncertain at small scales than at large scales ($\lambda^{a}_{0} < \lambda^{a}_{1}$). When $\mathbf{P}^{a}_{ij}$ is zero, the uncertainty at both scales is the same.

When a is positive (right-hand panel), this has the effect of making the analysis errors much more negatively correlated, implying that there is much more uncertainty at smaller scales. This is consistent with the effect of the observation operator in reducing the accuracy of the observations in state space at small scales. The reverse is true when a is negative (left-hand panel).

Unlike the analysis-error variances, we see that the exact relationship between the analysis-error correlations and the background- and observation-error correlations is dependent on the error variances of the observations and the background.

4.3. *The sensitivity matrix*

4.3.1. *Self-sensitivities*

For this simple experimental design, the sensitivity matrix is also circulant. An implication of this is that the self-sensitivity, S_{ii}, is the same for each variable.

The self-sensitivities for the same cases discussed in section 4.2 are plotted in Figure 6 as a function of background-error correlation, χ_b, and observation-error correlation, χ_r. Again let us first consider the case when $a = 0$ (middle panel). Mostly, an increase in the magnitude of the BECs results in a decrease in the self-sensitivity (as noted by Cardinali *et al.*, 2004). This is because the BECs allow for information in the observation of variable x_j to spread to the analysis of x_i and so the observation of x_i itself has less weight. In contrast to this, we see that an increase in the magnitude of OECs results in an increase in the influence of an observation of the analyzed variable. This is because the OECs allow an observation of x_j to reinforce the information in the observation of x_i and so this has more weight in computing the analysis of x_i. This implies that OECs increase the amount of information in the observations as measured by the dfs (as noted by Stewart *et al.*, 2008). In contrast, an increase in the magnitude of the BECs generally results in a decrease in the dfs (as noted by Cardinali *et al.*, 2004). As noted by Eresmaa *et al.* (2014), the self-sensitivities plotted in Figure 6 are not symmetric about $\chi_r = 0$ and $\chi_b = 0$. This means that when χ_b is large there can be a small decrease in the value of the dfs as χ_r increases.

When $a \neq 0$, we see that this generally reduces the self-sensitivity. The maximum value of S_{ii} now occurs when χ_r and a are of the same sign. This is likely related to the fact that this combination results in the smallest observation-error variances mapped to state space (see Figure 4 and (28)). As with the analysis-error correlations, the exact relationship between the self-sensitivities and the error correlations is dependent on the error variances of the observations and background.

4.3.2. *Cross-sensitivities*

In Figure 7, the cross-sensitivities are plotted. When $a = 0$ (middle panel), we see that, when $\chi_r = 0$, as $|\chi_b|$ increases there is an increase in the magnitude of the cross-sensitivities. This supports the argument that BECs allow for information in the observation of variable x_j to spread to the analysis of x_i. We also see that when $\chi_b = 0$ an increase in the magnitude of the OECs, $|\chi_r|$, results in an increase in the magnitude of the cross-sensitivities as well as the increase in self-sensitivities shown in Figure 6. However, note that when χ_r is positive (and $\chi_b = 0$) S_{ij} is negative, whilst when χ_b is positive (and $\chi_r = 0$) S_{ij} is positive. The cross-sensitivities can be related to the difference in the eigenvalues of the sensitivity matrix such that

$$S_{ij} = (\lambda^{s}_{0} - \lambda^{s}_{1})/2. \tag{31}$$

Therefore positive (negative) values of S_{ij} mean that the analysis at larger scales is more (less) sensitive to the observations than the analysis at smaller scales. Therefore, as χ_r becomes more positive, the analysis at smaller scales becomes more sensitive to the observations (consistent with the results of Rainwater *et al.* (2015) and Seaman (1977)), whereas when χ_b becomes more positive the analysis at larger scales becomes more sensitive to the observations.

An interesting consequence of this is that, when $a = 0$ and $\chi_r = \chi_b$, an observation of a different variable, x_j, has little influence on the analyzed variable, x_i, and the sensitivity of the analysis to observations is the same at all scales, as noted in the discussion of (21). In Appendix B, it is shown that this is independent of the choice of ρ and β. In effect, the benefits of the spread of information to the analysis of x_i from an observation of x_j due to the BECs is cancelled out by the OECs.

The magnitude of the cross-sensitivities increases when jointly $\chi_b \to \pm 1$ and $\chi_r \to \mp 1$. We therefore see that the greatest spread in information also corresponds to the lowest analysis-error variance seen in Figure 4. In contrast, the pattern of the

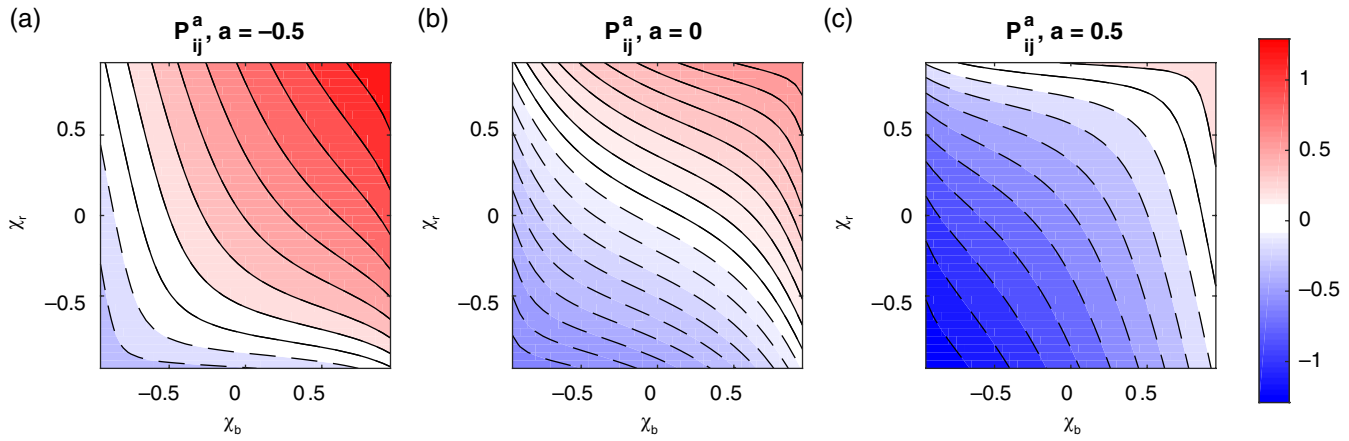

Figure 5. The analysis-error covariance, $\mathbf{P}^{a}_{ij}$, plotted as in Figure 4. Negative contours are represented by a dashed line. [Colour figure can be viewed at wileyonlinelibrary.com].

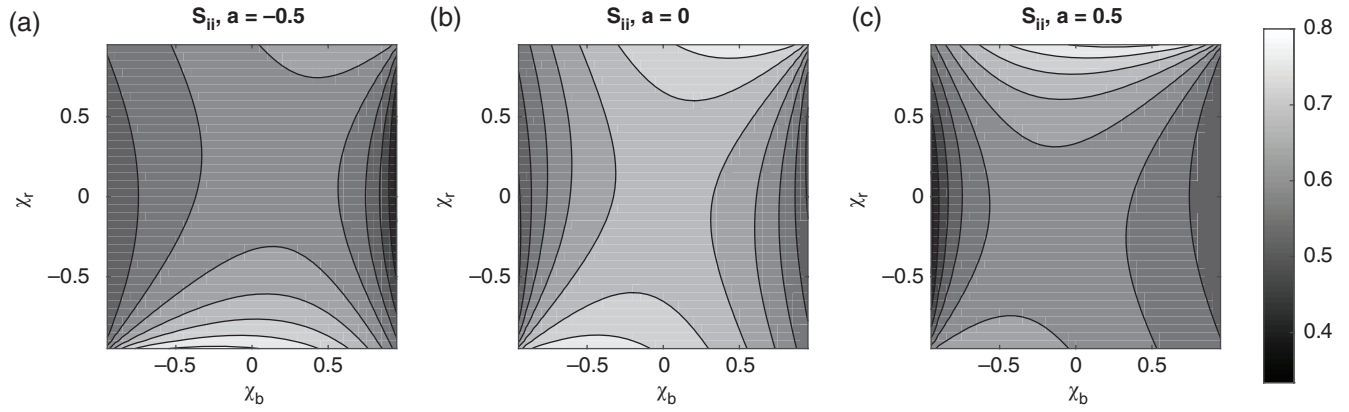

Figure 6. The self-sensitivities, $\mathbf{S}_{ii}$, plotted as in Figure 4.

greatest self-sensitivities seen in Figure 6 has little agreement with the pattern in the lowest analysis-error variances.

When a is positive, the cross-sensitivity becomes more positive. This is because the effect of a being positive is to magnify the importance of the large-scale uncertainty in $\mathbf{B}$ and so the analysis at large scales becomes more sensitive to the observations. The reverse is true when a is negative.

4.4. *Mutual information*

MI summarizes both the self- and cross-sensitivities (11), but is defined in terms of the entropy reduction of the posterior compared with the prior. As such, *MI* is also a function of both analysis-error variances and correlations (10). This is demonstrated in Figure 8, in which *MI* is plotted for the same cases discussed in sections 4.2 and 4.3 as a function of background-error correlation, χ_b, and observation-error correlation, χ_r.

When $a = 0$ (middle panel), we see that in general *MI* increases as $|\chi_r|$ increases, but it is greatest when $|\chi_r - \chi_b|$ is large. The smallest *MI* occurs when $|\chi_b|$ is large and $|\chi_r - \chi_b|$ is small. Note that these figures look quite different from the self-sensitivities in Figure 6, which are proportional to the dfs, demonstrating that the spread in information described by the cross-sensitivities is very important for *MI*.

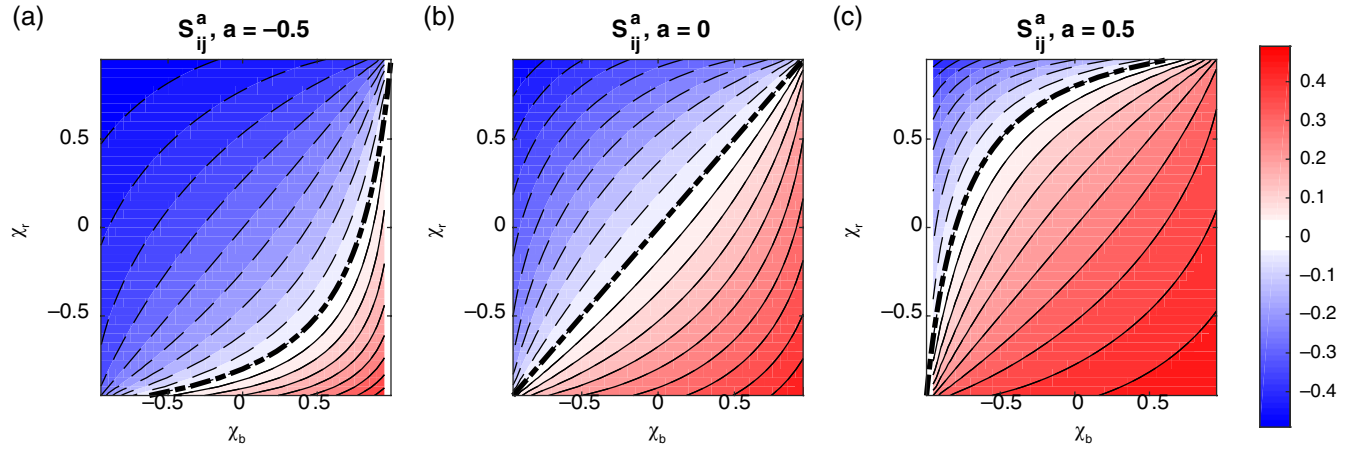

Figure 7. The cross-sensitivities, $\mathbf{S}_{ij}$, plotted as in Figure 4. Negative contours are represented by a dashed line. The bold dash–dotted black line indicates the value χ_b that results in $\mathbf{S}_{ij} = 0$ as a function of χ_r as derived in Appendix B. This is the same value of χ_b that maximizes the analysis-error variance shown by the white dash–dotted line in Figure 4. [Colour figure can be viewed at wileyonlinelibrary.com].

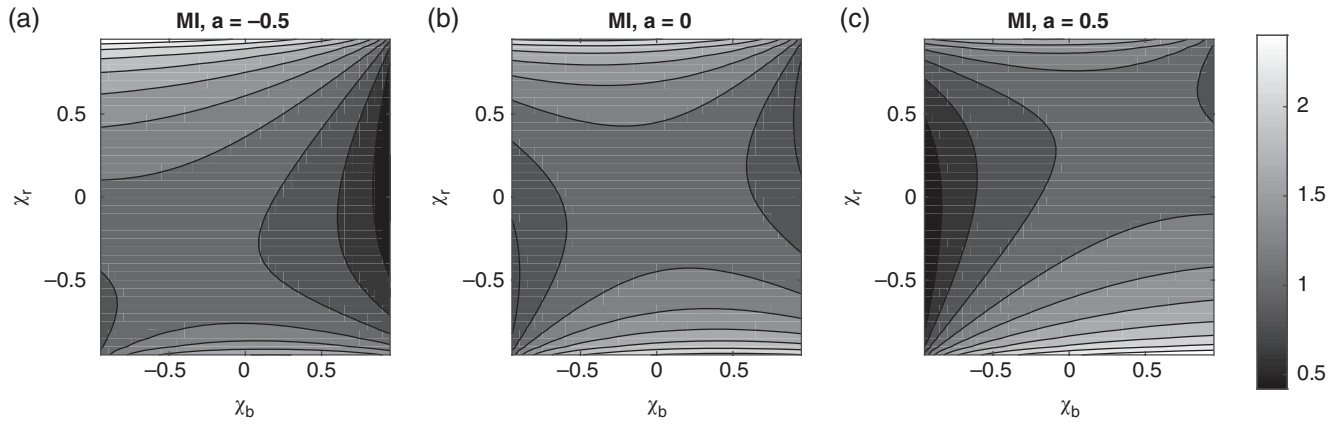

Figure 8. Mutual information, *MI*, plotted as in Figure 4.

When a is positive (right-hand panel), χ_r is positive and χ_b is negative, the value of *MI* is reduced, compared with direct observations. The opposite is true when χ_r is negative and χ_b is positive. This can only be explained by the effect of assimilating the observation on both the analysis-error variances and the analysis-error correlations. From section 2.3, we know that the entropy is smallest when both the variance is small and the correlations are large. By comparing Figure 8 with Figures 4 and 5, therefore, we see that the maximum *MI* is a compromise between the minimum analysis-error variance and maximum analysis-error correlations.

5. 32 variable problem

In this section, we show how the two-variable numerical results extend to higher dimensions. In the following numerical experiments the state is represented on a circular domain of length 32π, discretized into 32 evenly spaced points. **B** and **R** are both circulant (see section 3) with a SOAR correlation function (19). The length-scale of the background-error correlations used in these experiments is $L_B = 5$. The length-scale of the observation-error correlations, L_R, is allowed to vary. The SOAR correlation function has previously been used to model the background-error correlations in operational systems (e.g. Ingleby, 2001; Simonin *et al.*, 2014). Each state variable and observation has an error variance of 1.

The observations, made at every grid point, measure a weighted average of the surrounding state points. The observation operator is defined by a triangular weighting function, given by

$$h(x_i) = \sum_{j=i-a}^{i+a} \frac{a+1-|j-i|}{a+1} x_j, \tag{32}$$

where $a \in \mathbb{N}$ defines the weighting length-scale.

In Figure 9, the eigenvalues of **B** (black line with crosses), **R** (black line with pluses) and **H** (black line) are given as a function of wave number (computed using (17)). Also plotted are the resulting eigenvalues of the sensitivity matrix (thick red line) and the analysis-error covariance matrix (thick dashed blue line) as computed from (21) and (20), respectively. The corresponding values for the trace of $\mathbf{P}^a$ ($\mathrm{tr}(\mathbf{P}^a)$), dfs and *MI* are given in Table 2.

The top row of Figure 9 shows the case when $\mathbf{H} = \mathbf{I}$ ($a = 0$) and $L_R < L_B$ (left), $L_R = L_B$ (middle) and $L_R > L_B$ (right). It can be seen from the eigenvalues of the sensitivity matrix, λ_k^s, that, as L_R increases, the analysis at larger scales become less sensitive to observations, but the analysis at smaller scales becomes more sensitive to observations. When $L_R = L_B$ (middle), we see that the analysis has the same sensitivity at all scales, as predicted by (21). Table 2 shows that an increase in L_R also coincides with a general increase in information content of the observations as measured by the *MI* and dfs.

We also see from the eigenvalues of the analysis-error covariance matrix, λ_k^a, that, when $a = 0$, as L_R increases the analysis uncertainty becomes greater at larger scales and smaller at small scales. However, the trace of the analysis-error covariance matrix is at a maximum when $L_R = L_B$ (see Table 2), which was explained for the two-variable case in section 4.2.

Table 2. The values of the trace of the analysis-error covariance matrix, degrees of freedom for the signal and mutual information for the 32 variable cases illustrated in Figure 9.

	$a = 0$			$a = 1$		
L_R	$\mathrm{tr}(\mathbf{P}^a)$	*dfs*	*MI*	$\mathrm{tr}(\mathbf{P}^a)$	*dfs*	*MI*
1	10.2	8.2	6.4	4.9	10.2	11.1
5	16.0	16.0	11.1	7.4	13.8	11.8
10	14.3	25.6	28.0	6.5	20.2	22.5

In each case, **B** and **R** are circulant with correlations described by (19). The background-error correlation length-scale, L_B, is 5, and the observation-error correlation length-scale, L_R, is allowed to vary. The error variances are 1. **H** is circulant, with weighting function given by (32), with a allowed to vary.

The bottom row of Figure 9 shows the case when $a = 1$ and again $L_R < L_B$ (left), $L_R = L_B$ (middle) and $L_R > L_B$ (right). Unlike the case when $a = 0$, we note that the observations now lie in a different space from the state. We see, in general, that when $a = 1$ this reduces the accuracy of the analysis and its sensitivity to observations at small scales compared with the case when $a = 0$. We also see in the lower right-hand part of the figure (where $L_R > L_B$) that the eigenvalues of $\mathbf{P}^a$ are larger than those of **R** at small scales. This is consistent with the averaging effect of the observation operator, meaning that the accuracy of the observed estimate of the state, given by $\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$, is reduced at small scales (compared with direct observations). The reduction in the accuracy of the analysis at small scales coincides with an increase in the accuracy of the analysis and its sensitivity to observations at large scales; again, this is consistent with the averaging effect of the observation operator.

Table 2 shows that an increase in L_R is still seen to coincide with a general increase in information content of the observations as measured by *MI* and dfs when $a = 1$. Again, the trace of the analysis-error covariance matrix is at a maximum when $L_R = L_B$. In all cases, $\mathrm{tr}(\mathbf{P}^a)$ is reduced when $a = 1$ compared with the case when $a = 0$.

Experiments were also performed (not shown) in which the OECs were modelled by the following Gaussian function:

$$c_k = \mathrm{e}^{-r_k/2L_R^2}, \tag{33}$$

whilst the BECs were still modelled by the SOAR function given by (19). It was found that in this case the general conclusions still hold: i.e. the trace of the analysis-error covariances still peaks when $L_R \approx L_B$ and the values of the dfs and *MI* still increase as L_R increases. It was also found that, when $a = 1$, the trace of the analysis-error covariances decreased compared with the case when $a = 0$.

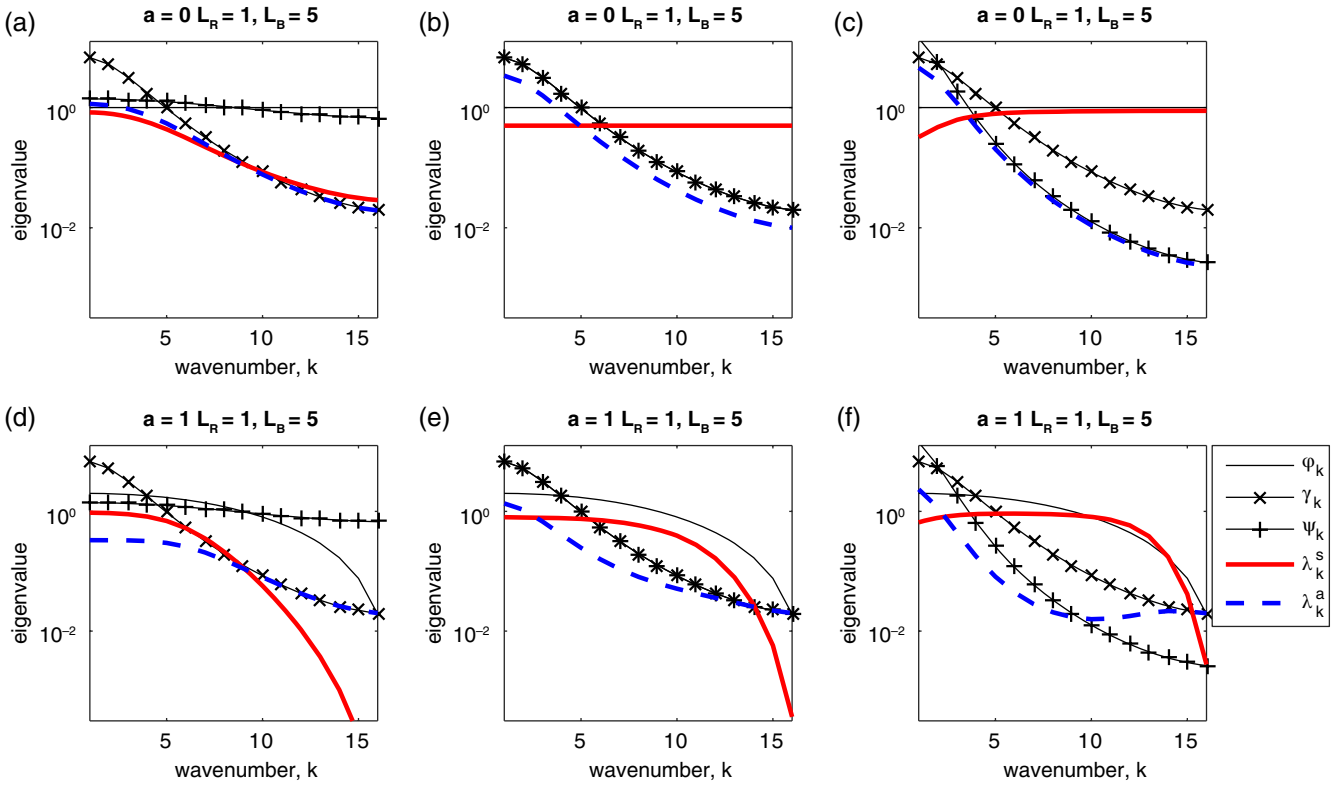

Figure 9. Example of the eigenvalues of the analysis-error covariance matrix (λ^a, thick dashed line) and sensitivity matrix (λ^s, thick solid line) plotted as a function of wave number. The error covariance matrices **B** and **R** are described in section 5. In each case, the correlation length-scale of the background errors is $L_B = 5$ and the correlation length-scale of the observation errors is (a, d) $L_R = 1$, (b, e) $L_R = 5$ and (c, f) $L_R = 10$. The observation operator is given by (32), with $a = 0$ (top row, i.e. direct observations) and $a = 1$ (bottom row). The observation- and background-error variances are 1. [Colour figure can be viewed at wileyonlinelibrary.com].

6. Data thinning

In sections 4 and 5, it was seen that the impact of observations is highly dependent upon both the observation- and background-error correlations and the observation operator. In this section, we investigate how these results could be used for the design of instruments and observation networks.

In practice, we have little control over the error correlation length-scales or the observation operator. However, we can show how the results presented could be used to influence the choice of the density of data we assimilate.

One of the motivating factors for accounting for OECs is that it will assist the use of denser observations, which could be useful for high-resolution and high-impact weather forecasting. In this section, we perform experiments using the same circular grid and experimental design presented in section 5. However, now we investigate the effect of reducing the number of the observations (keeping them evenly spaced). An important consequence of this is that the circulant matrix assumption is no longer applicable, as now $n \geq p$ and hence the linearized observation operator, **H**, is not necessarily square.

In the top row of Figure 10, $\mathrm{tr}(\mathbf{P}^a)$ (left) and MI/p (right) are plotted as a function of the thinning distance of the observations for $a = 0$ (i.e. direct observations) and three different values of $L_R = [1, 5, 10]$ (recall $L_B = 5$ in each case). The ratio MI/p quantifies the average information content of each observation assimilated. It is therefore important for optimizing the efficiency of a DA system. Conversely, $\mathrm{tr}(\mathbf{P}^a)$ gives a measure of just the variance of the analysis error and is insensitive to the analysis-error correlations. A measure of $\mathrm{tr}(\mathbf{P}^a)$ alone is therefore not a useful diagnostic for optimizing the efficiency of the DA system, as more observations (denser observations) in general will always mean a smaller analysis-error variance.

Consistent with the results shown in sections 4 and 5, the analysis-error variances are greatest when $L_R = L_B$ (crosses; for reference, the trace of **B** is plotted as a grey thick line). As such, in this case $\mathrm{tr}(\mathbf{P}^a)$ is least sensitive to an increase in the thinning distance of the observations from $1\Delta x$ to $2\Delta x$ to $4\Delta x$ when $L_R = L_B$. When the thinning distance is greater than $8\Delta x$, $\mathrm{tr}(\mathbf{P}^a)$ increases by a similar amount irrespective of the OEC length-scale.

The average information content, MI/p, behaves in a different way (as shown in sections 4 and 5). In the case of direct observation ($a = 0$), the information increases as the OEC length-scale increases (consistent with our previous results). The effect of the density of the observations on MI/p can be understood from the definition of MI in terms of the eigenvalues of **S**, given by (11). Thinning the observations will only increase MI/p if the sensitivity of the analysis to the observations at the smallest scales resolved by the observations is less than the sensitivity at larger scales (see Figure 9 for examples of how the eigenvalues of **S** can be expected to change with wave number). This means that reducing the number of observations has little effect on the average information content per observation when $L_R = L_B$, as the eigenvalues of **S** are constant (for the circulant case). In other words, the analysis sensitivity to the observations is the same at all scales and so MI is directly proportional to p. However, when $L_R < L_B$, reducing the number of observations increases the average amount of information per observation, as in this case the background is more accurate at small scales than the observations and so thinning the observations is not detrimental. Conversely, when $L_R > L_B$, reducing the number of observations decreases the average amount of information per observation, as in this case the observations are more accurate than the background at small scales and so thinning the observations loses valuable small-scale information.

An example of a case where $L_R < L_B$ is Doppler radial wind observations. Waller *et al.* (2016c) estimated Doppler radial wind observation errors to have correlation length-scales of the order of 20 km, compared with background-error correlation length-scales of the order of 100 km for the Met Office's high resolution U.K. (UKV) 1.5 km limited-area model (Lean *et al.*, 2008). At the UK Met Office, these data are currently thinned to a distance of 6 km (Simonin *et al.*, 2014). From Figure 10, we can conclude that, if the OECs were correctly modelled in the assimilation,

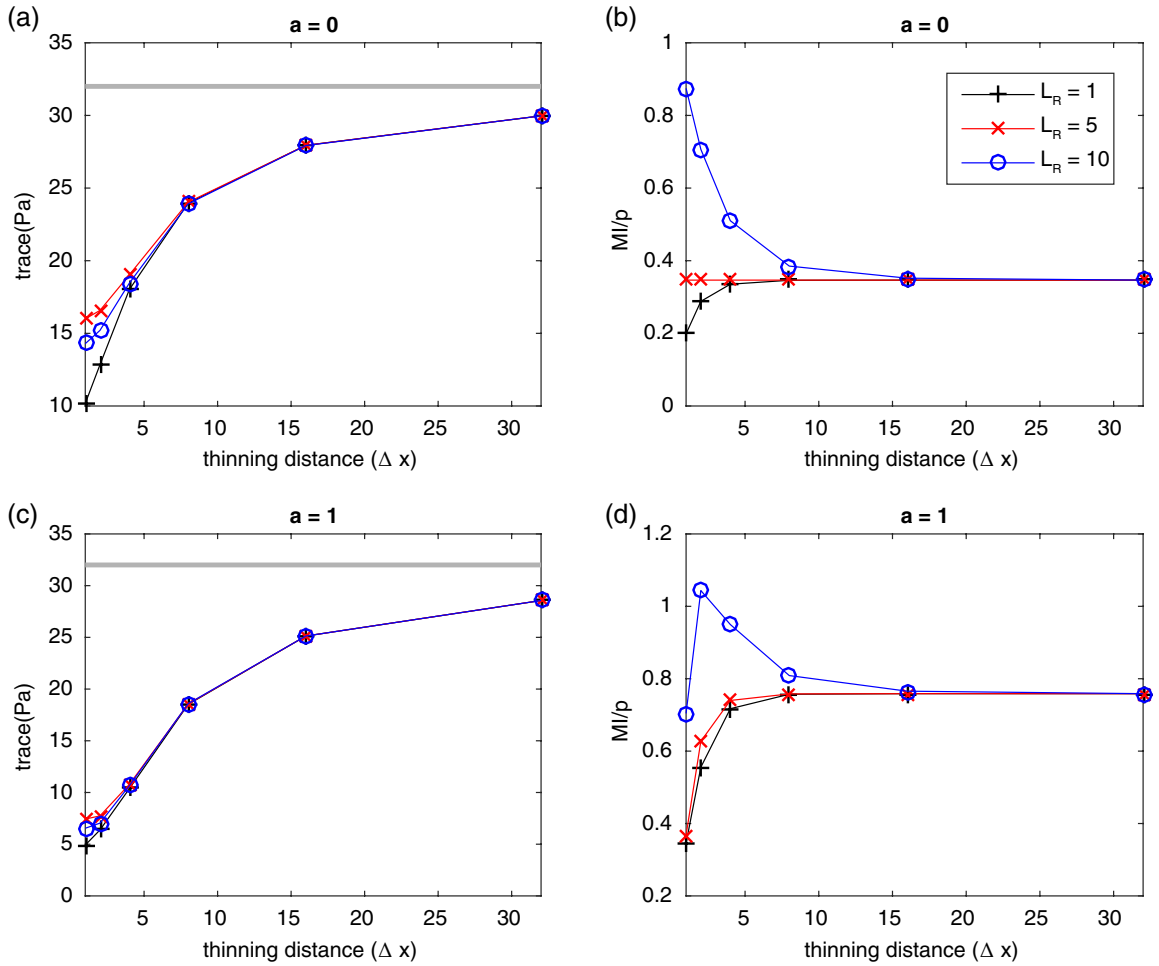

Figure 10. (a, c) The trace of the analysis-error covariance matrix and (b, d) the average mutual information per observation as a function of thinning distance, given by the number of grid points between observations. In the left-hand panels, the trace of **B** is plotted as a grey thick line. In each case, **B** is equal to a circulant matrix with a SOAR correlation function and length-scale $L_B = 5$. **R** is equal to a circulant matrix with a SOAR correlation function and length-scale $L_R = 1$ (plus symbols), 5 (crosses) and 10 (circles). The observation operator is given by (32) with (a, b) $a = 0$ (i.e. direct observations) and (c,d) $a = 1$. [Colour figure can be viewed at wileyonlinelibrary.com].

using all the data may see a small improvement in the analysis-error variances, but could actually reduce the average amount of information per observation as measured by MI/p. Therefore, it may be beneficial to thin these data to allow for resources to be used instead to assimilate a different data source, for example.

An example of a case where $L_R > L_B$ is the winds derived from atmospheric motion vectors. Cordoba *et al.* (2017) estimated AMV errors to have correlation length-scales of the order of 150 km, compared with background-error correlation length-scales of the order of 100 km for the UKV. At the UK Met Office, these data are currently thinned to a distance of 20 km. From Figure 10, we can conclude that, if the OECs were correctly modelled in the assimilation, using all the data may again see a small improvement in the analysis-error variances. However, unlike when $L_R < L_B$, using denser AMVs could actually increase the average amount of information per observation as measured by MI/p.

In the bottom row of Figure 10, we see the effect of increasing the weighting length-scale of the observation operator, a, in (32). When $a = 1$ (i.e. each observation is modelled as a weighted average of three neighbouring grid points), we see that the analysis-error variance is still greatest when $L_R = L_B$, but in all cases when we have dense observations, $\mathrm{tr}(P^a)$ is smaller than when we had direct observations. The effect of overlapping weighting functions reducing the analysis-error variance when the BECs are positive was already noted in section 4. As a is increased, we also generally see more information in the observations (as measured by MI).

In each example of L_R, we now find that a thinning distance greater than the grid length increases the amount of information per observation. This is because, as noted in section 3, the effect of overlapping weighting functions is to reduce the accuracy of the observation in state space at smaller scales, hence thinning the observations becomes less detrimental. In particular, when $L_R > L_B$ we now have a peak at a thinning distance of $2\Delta x$ and so there is no longer a benefit to having observations at every grid point. This can be explained by the fact that, at a thinning distance of $2\Delta x$, the overlapping of the weighting functions is reduced, so that each individual observation is able to provide more independent information about the state; however, the observation-error correlations are still significant and so further thinning is not beneficial.

An example of observations for which this experimental design could be relevant is satellite measurements of top-of-the-atmosphere radiances at different wavelengths, which have corresponding weighting functions peaking at different heights in the atmosphere. Instruments such as IASI use objective channel-selection methods to select a subset of the channels to be assimilated based on their information content. Ventress and Dudhia (2014) developed an efficient method for performing the channel selection, whilst accounting for spectral correlations using the dfs. From the results presented above, it is expected that the OECs will be unlikely to have a significant impact on the channel selection, unless the OECs are large in comparison with the overlapping of the weighting functions, for example comparing **R** with $\mathbf{H}\mathbf{H}^{\mathrm{T}}$. To maximize the average amount of information per observation, spectral thinning should be done in order to minimize the overlap of the weighting functions in **H** if the OECs are large compared with the BECs mapped to observation space. If the OECs are small compared with

the BECs mapped to observation space, then the efficiency of the DA system could be improved further by further spectral thinning.

In practice, the correlation structure of **B** and **R** is unlikely to coincide. For example, it has been seen that the observation-error covariances are often a combination of white noise and coloured noise, most likely due to different sources of observation uncertainty (e.g. Waller *et al.*, 2016c), whilst the background-error correlation is more likely to be a smooth function of separation distance. However, interpreting the optimum thinning distance in terms of the relative sensitivity of the analysis to the observations at small scales, as discussed above, allows the results to be extended to other correlation structures.

7. Summary and conclusions

The inclusion of OECs in NWP DA is a relatively new area of research. The first attempts to include OECs for IASI data have been shown to have a positive impact on the skill scores of the forecast (Weston *et al.*, 2014; Bormann *et al.*, 2016; Campbell *et al.*, 2017). This has led to a surge of research to estimate the error correlations for a myriad of other observation types. This article has aimed to give insight into how the expected benefits of including OECs in DA depend on the other characteristics of the DA system, namely the BECs and observation operator. To give an insight into how the OECs, BECs and observation operator interact, we have shown a series of two-variable experiments in which the observation- and background-error correlations and the correlations between the observations themselves (described by the observation operator) were allowed to vary. The impact of varying these correlations on the analysis was measured using a variety of metrics: namely analysis-error covariances, sensitivity of the analysis to the observations and mutual information. This, therefore, differs from previous studies (e.g. Liu and Rabier, 2002; Stewart *et al.*, 2008, 2013; Rainwater *et al.*, 2015) in which the effect of OECs on only one metric was studied and the dependence of the impact was largely considered in isolation from the BECs and the observation operator.

For the idealized experiments shown, it was found that, in general, as the magnitude of the OECs increases, the information content of the observations as measured by the dfs (the sum of the self-sensitivities) and MI increase; this agrees with the results of Stewart *et al.* (2008) and Eresmaa *et al.* (2014). However, the impact of the OECs on other metrics cannot be understood in isolation from the background-error statistics and the observation operator. It was found that the results could be explained in terms of the scales of the greatest observation and background uncertainties, as described by the likelihood and prior PDFs respectively. When the prior and likelihood are accurate at different scales, it was found that

- the analysis-error variances are smallest;
- there is the greatest spread in information (i.e. cross-sensitivities are largest); and
- the observations have the greatest mutual information.

If the observations measure the state variables directly, then the scales of the observation and background uncertainty can be interpreted by comparing **R** and **B** directly. However, if we have indirect observations then we need to consider the effect of **H** on the accuracy of the observations in state space (or equivalently the projection of the background uncertainty into observation space). The effect of a positive observation-operator coefficient (i.e. observations themselves are positively correlated) was shown to reduce the accuracy of the observations in state space at small scales and increase the accuracy of the observations in state space at large scales. This allows us to explain the results given by Miyoshi *et al.* (2013), who found that the observations contained more information (when the background errors are uncorrelated) when the sign of the observation-operator coefficient and the observation-error correlations was different. This is because, in the case when the observations themselves are positively correlated and the OECs are negative, they combine to enhance the accuracy of the observations in state space at large scales. Similarly, if the observations themselves are negatively correlated and the OECs are positive, then they combine to enhance the accuracy of the observations in state space at small scales. The positive observation-operator coefficient was seen to have a beneficial impact when the background errors were positively correlated, i.e. the background is more accurate at small scales than at large scales.

In section 5, it was shown that these conclusions can be extended to higher dimensional cases. In particular, it was demonstrated that there is a peak in the trace of the analysis-error covariance matrix when the OECs and BECs are the same. However, these experiments continued to assume that the background- and observation-error covariances and the observation operator could be described by a circulant matrix. A limitation of this is that the error variances are the same for each variable.

It is anticipated that these results could be used for the design of instruments and observation networks. In section 6, the results were used to inform the optimal thinning distance of observations with a variety of error correlation length-scales and observation operators. For these experiments, the assumption of a circulant observation operator was no longer applicable, although the background and observation-error covariances continued to have circulant form. We found the following.

- When the observations measured the state variables directly, it was most beneficial to have dense observations when the sensitivity of the analysis to the observations at the smallest scales was greater than the sensitivity at larger scales.
- If the observations are not direct but instead are sensitive to a range of state variables, such that the observations have overlapping weighting functions (reducing the accuracy of the observation in state space at small scales), thinning should be performed in terms of reducing the overlap of the weighting functions rather than the correlation length-scales.

These conclusions differ from those of Liu and Rabier (2002) and Bergman and Bonner (1976), who considered only OECs with length-scales less than the BECs and only the effect on the analysis RMSE.

A final point is that, in practice, the error correlations in **B** and **R** may be flow-dependent. As the variances fluctuate, the impact of the observations will fluctuate. Similarly, as the correlation length-scales fluctuate, we have shown that the relationship between the length-scales in **B** and **R** will largely determine how the impact of the observations will fluctuate.

To conclude, for all observation types it is important that the OECs are estimated and correctly allowed for. However, for some observation types it may still be preferable to thin the data to the assumed correlation length-scales rather than assimilating all data. Future efforts should focus on accurately estimating and accounting for OECs for observation types that are thought to have longer length-scales than the BECs. Increasing the density, when assimilating these observations, will show the most benefit for improving the efficiency of the assimilation system, particularly when analyzing small-scale features. However, we note that the software engineering challenges of implementing such long length-scales in high-performance computing environments are significant.

Appendix A: Maximum analysis-error variance of the two-variable problem

For the two-variable problem introduced in section 4, it can be shown that the background-error correlation, χ_b, which results in the maximum analysis-error variance is a function of the observation-error correlation, χ_r, and observation-operator

coefficient, a, only. In particular, when the observations observe the state directly, it can be shown that the analysis-error variance is largest when the background and observation-error correlations are identical, i.e. $\chi_b = \chi_r$.

From (17), it can be shown that the analysis-error variance for the two-variable problem is given by $P^a_{ii} = \frac{1}{2}(\lambda^a_0 + \lambda^a_1)$, where λ^a_i is the ith eigenvalue of the analysis-error covariance matrix. The eigenvalues of the analysis-error covariance matrix can be written explicitly in terms of the eigenvalues of the background and observation-error correlation matrices, γ_i and ψ_i, and their variances, β and ρ respectively, and the eigenvalues of the observation operator, ϕ_i (see (20)):

$$\lambda^a_i = \frac{\rho\beta\gamma_i\psi_i}{\beta\gamma_i\phi_i^2 + \rho\psi_i}. \tag{A1}$$

For the two-by-two circulant matrices described by (23) and (24), the eigenvalues are given by (26).

To find the χ_b that results in the maximum P^a_{ii}, we find the χ_b that satisfies $\mathrm{d}P^a_{ii}/\mathrm{d}\chi_b = 0$. The derivative of the analysis-error variance with respect to the background-error correlations is given by

$$\frac{\mathrm{d}P^a_{ii}}{\mathrm{d}\chi_b} = \frac{1}{2}\Big(\frac{\partial\lambda^a_0}{\partial\gamma_0}\frac{\partial\gamma_0}{\partial\chi_b} + \frac{\partial\lambda^a_1}{\partial\gamma_1}\frac{\partial\gamma_1}{\partial\chi_b} + \frac{\partial\lambda^a_0}{\partial\psi_0}\frac{\partial\psi_0}{\partial\chi_b} + \frac{\partial\lambda^a_1}{\partial\psi_1}\frac{\partial\psi_1}{\partial\chi_b} + \frac{\partial\lambda^a_0}{\partial\phi_0}\frac{\partial\phi_0}{\partial\chi_b} + \frac{\partial\lambda^a_1}{\partial\phi_1}\frac{\partial\phi_1}{\partial\chi_b}\Big), \tag{A2}$$

where

$$\frac{\partial\lambda^a_i}{\partial\gamma_i} = \frac{\beta(\rho\psi_i)^2}{(\rho\psi_i + \beta\gamma_i\phi_i^2)^2}, \tag{A3}$$

$$\frac{\partial\gamma_0}{\partial\chi_b} = -\frac{\partial\gamma_1}{\partial\chi_b} = 1 \tag{A4}$$

and

$$\frac{\partial\psi_0}{\partial\chi_b} = \frac{\partial\psi_1}{\partial\chi_b} = \frac{\partial\phi_0}{\partial\chi_b} = \frac{\partial\phi_1}{\partial\chi_b} = 0. \tag{A5}$$

Therefore

$$\frac{\mathrm{d}P^a_{ii}}{\mathrm{d}\chi_b} = \frac{1}{2}\beta\left(\frac{(\rho\psi_0)^2}{(\rho\psi_0 + \beta\gamma_0\phi_0^2)^2} - \frac{(\rho\psi_1)^2}{(\rho\psi_1 + \beta\gamma_1\phi_1^2)^2}\right). \tag{A6}$$

For this to equal zero, this implies that

$$\frac{(\rho\psi_0)^2}{(\rho\psi_0 + \beta\gamma_0\phi_0^2)^2} = \frac{(\rho\psi_1)^2}{(\rho\psi_1 + \beta\gamma_1\phi_1^2)^2}, \tag{A7}$$

which is satisfied when

$$\frac{\rho\psi_0}{\rho\psi_0 + \beta\gamma_0\phi_0^2} = \frac{\rho\psi_1}{\rho\psi_1 + \beta\gamma_1\phi_1^2}. \tag{A8}$$

Therefore $\mathrm{d}P^a_{ii}/\mathrm{d}\chi_b = 0$ when $\gamma_0\phi_0^2\psi_1 = \gamma_1\phi_1^2\psi_0$. Substituting (26) into (A8), it is seen that this is satisfied when

$$\chi_b = \frac{\chi_r(1+a^2) - 2a}{1 + a^2 - 2a\chi_r}, \tag{A9}$$

with no dependence on the values of β and ρ. Further analysis (not shown here) finds that, at this point, P^a_{ii} as a function of χ_b is at a maximum.

In the case when we have direct observations of the state and $a = 0$, we see that the maximum analysis-error variance occurs when $\chi_b = \chi_r$. Substituting $\chi_b = \chi_r$ and $a = 0$ into the expression for P^a_{ii} implies that the maximum analysis-error variance is given by

$$P^a_{ii} = \frac{\beta\rho}{\beta + \rho}.$$

Appendix B: The analysis cross-sensitivities of the two-variable problem

For the two-variable problem introduced in section 4, it can be shown that the background-error correlation, χ_b, that results in zero analysis cross-sensitivity is given by the same function of the observation-error correlation, χ_r, and observation-operator coefficient, a, that results in the maximum analysis-error variance derived in Appendix A.

From (26), it can be shown that the cross-sensitivities for the two-variable problem are given by $S_{ij} = \frac{1}{2}(\lambda^s_0 - \lambda^s_1)$, where λ^s_i is the ith eigenvalue of the sensitivity matrix.

As with the eigenvalues of the analysis-error covariance matrix, the eigenvalues of the sensitivity matrix can be written in terms of the eigenvalues of the background- and observation-error covariances, γ_i and ψ_i, their variances, β and ρ respectively, and the eigenvalues of the observation operator, ϕ_i (see (21)):

$$\lambda^s_i = \frac{\beta\gamma_i\phi_i^2}{\beta\gamma_i\phi_i^2 + \rho\psi_i}. \tag{B1}$$

For the cross-sensitivities to be equal to zero, we require that

$$\lambda^s_0 = \lambda^s_1. \tag{B2}$$

From (B1), we see that this is satisfied when $\gamma_0\phi_0^2\psi_1 = \gamma_1\phi_1^2\psi_0$. Substituting this into (26) implies that $S_{ij} = 0$ when

$$(1+\chi_b)(1+a)^2(1-\chi_r) = (1-\chi_b)(1-a)^2(1+\chi_r). \tag{B3}$$

Rearranging for χ_b we see that this holds when

$$\chi_b = \frac{\chi_r(1+a^2) - 2a}{1 + a^2 - 2a\chi_r}, \tag{B4}$$

which is the same as Eq. (A9).

Acknowledgements

AMF has been funded by the Natural Environment Research Centre's support to the National Centre for Earth Observation. SLD and JAW were supported in part by UK NERC grant NE/K008900/1 (FRANC) and UK EPSRC grant EP/P002331/1 (DARE).

References

Bannister RN. 2008a. A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics and measurements of forecast error covariances. *Q. J. R. Meteorol. Soc.* **134**: 1951–1970.

Bannister RN. 2008b. A review of forecast error covariance statistics in atmospheric variational data assimilation. II: Modelling the forecast error covariance statistics. *Q. J. R. Meteorol. Soc.* **134**: 1971–1996.

Berger H, Forsythe M. 2004. 'Satellite wind superobbing'. Met Office Forecasting Research Technical Report 451. http://research.metoffice.gov.uk/research/nwp/publications/papers/technical_reports/2004/FRTR451/FRTR451.pdf.

Bergman KH, Bonner WD. 1976. Analysis error as a function of observation density for satellite temperature soundings with spatially correlated errors. *Mon. Weather Rev.* **104**: 1308–1316.

Bormann N, 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. Meteorol. Soc.* **136**: 1036–1050.

Bormann N, Saarinen S, Kelly G, Thépaut J-N. 2003. The spatial structure of observation errors in atmospheric motion vectors from geostationary satellite data. *Mon. Weather Rev.* **131**: 706–718.

Bormann N, Bonavita M, Dragani R, Eresmaa R, Matricardi M, McNally A. 2016. Enhancing the impact of IASI observations through an updated observation-error covariance matrix. *Q. J. R. Meteorol. Soc.* **1420**: 1767–1780, https://doi.org/10.1002/qj.2774.

Browne PA, Charlton-Perez C, Dance SL. 2014. RMetS Special Interest Group Meeting: High resolution data assimilation. *Atmos. Sci. Lett.* **150**: 354–357, https://doi.org/10.1002/asl2.512.

Campbell WF, Satterfield EA, Ruston B, Baker NL. 2017. Accounting for correlated observation error in a dual formulation 4D-variational data assimilation system. *Mon. Weather Rev.* **145**: 1019–1032. https://doi.org/10.1175/MWR-D-16-0240.1.

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

Collard AD. 2007. Selection of IASI channels for use in numerical weather prediction. *Q. J. R. Meteorol. Soc.* **133**: 1977–1991.

Cordoba M, Dance SL, Kelly GA, Nichols NK, Waller JA. 2017. Diagnosing atmospheric motion vector observation errors for an operational high resolution data assimilation system. *Q. J. R. Meteorol. Soc.* **143**: 333–341, https://doi.org/10.1002/qj.2925.

Daley R. 1996. *Atmospheric Data Analysis*, *Cambridge Atmospheric and Space Science Series. Cambridge University Press: New York, NY.*

Dando ML, Thorpe AJ, Eyre JR. 2007. The optimal density of atmospheric sounder observations in the Met Office NWP system. *Q. J. R. Meteorol. Soc.* **133**: 1933–1943.

Desroziers G, Berre L, Chapnik B, Poli P. 2005. Diagnosis of observation, background and analysis-error statistics in observation space. *Q. J. R. Meteorol. Soc.* **1310**: 3385–3396.

Eresmaa R, Bormann N, McNally A. 2014. 'Implications of observation-error correlation on the assimilation of interferometric radiances'. In *Proceedings of the ITSC-XIX*, 26 March–1 April 2014. Jeju Island, South Korea.

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

Gray R. 2006. Chapter in 'Foundations and Trends in Communications and Information Theory 2', 155–239. Now publishers: Boston-Delft.

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

Hodyss D, Nichols NK. 2015. The error of representation: Basic understanding. *Tellus A* **67**: 24 822–24 839.

Hodyss D, Satterfield E. 2017. *The Treatment, Estimation, and Issues with Representation Error Modelling*. Springer International Publishing: Cham, Switzerland, http://dx.doi.org/10.1007/978-3-319-43415-5_8.

Ingleby B. 2001. The statistical structure of forecast errors and its representation in the Met. Office Global 3-D Variational data assimilation scheme. *Q. J. R. Meteorol. Soc.* **127**: 209–231.

Jacques D, Zawadzki I. 2014. The impacts of representing the correlation of errors in radar data assimilation. Part I: Experiments with simulated background and observation estimates. *Mon. Weather Rev.* **142**: 3998–4016.

Janjic T, Cohn SE. 2006. Treatment of observation error due to unresolved scales in atmospheric data assimilation. *Mon. Weather Rev.* **134**: 2900–2915.

Janjic T, Bormann N, Bocquet M, Carton JA, Cohn SE, Dance SL, Losa SN, Nichols NK, Potthast R. 2017.On the representation error. *Q. J. R. Meteorol. Soc.* In Press, https://doi.org/10.1002/qj.3130.

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

Lean HW, Clark PA, Dixon M, Roberts NM, Fitch A, Forbes R, Halliwell C. 2008. Characteristics of high-resolution versions of the met office unified model for forecasting convection over the united kingdom. *Mon. Weather Rev.* **1360**: 3408–3424, https://doi.org/10.1175/2008MWR2332.1.

Liu Z-Q, 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.

Liu Z-Q, 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, https://doi.org/10.1256/qj.02.170.

Miyoshi T, Kalnay E, Li H. 2013. Estimating and including observation-error correlations in data assimilation. *Inverse Prob. Sci. Eng.* **210**: 387–398, https://doi.org/10.1080/17415977.2012.712527.

Pahl PJ, Damrath R. 2012. *Mathematical Foundations of Computational Engineering: A Handbook*. Springer Science & Business Media: Berlin.

Pozrikidis C. 2014. *An Introduction to Grids, Graphs and Networks*. Oxford University Press: Oxford, UK.

Rabier F, Jarvinen H, Klinker E, Mahfouf J-F, 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.

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

Rainwater S, Bishop CH, Campbell WF. 2015. The benefits of correlated observation errors for small scales. *Q. J. R. Meteorol. Soc.* **141**: 3439–3445.

Rawlins F, Ballard SP, Bovis KJ, Clayton AM, Li D, Inverarity GW, Lorenc AC, Payne TJ. 2007. The Met Office global four-dimensional variational data assimilation scheme. *Q. J. R. Meteorol. Soc.* **133**: 347–362.

Rodgers CD. 2000. *Inverse Methods for Atmospheric Sounding*. World Scientific Publishing: Singapore.

Seaman RS. 1977. Absolute and differential accuracy of analyses achievable with specified observational network characteristics. *Mon. Weather Rev.* **105**: 1211–1222.

Simonin D, Ballard SP, Li Z. 2014. Doppler radar radial wind assimilation using an hourly cycling 3D-Var with a 1.5 km resolution version of the Met Office Unified Model for nowcastings. *Q. J. R. Meteorol. Soc.* **140**: 2298–2314. https://doi.org/10.1002/qj.2298.

Stewart LM, Dance SL, Nichols NK. 2008. Correlated observation errors in data assimilation. *Int. J. Numer. Methods Fluids* **560**: 1521–1527.

Stewart LM, Dance SL, Nichols NK. 2013. Data assimilation with correlated observation errors: Experiments with a 1-D shallow water model. *Tellus A* **65**: 19546.

Stewart LM, Dance SL, Nichols NK, Eyre JR, 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.

Sun J, Xue M, Wilson JW, Zawadzki I, Ballard SP, Onvlee-Hooimeyer J, Joe P, Barker DM, Li P-W, Golding B, Xu M, Pinto J. 2014. Use of NWP for nowcasting convective precipitation: Recent progress and challenges. *Bull. Am. Meteorol. Soc.* **950**: 409–426, https://doi.org/10.1175/BAMS-D-11-00263.1.

Terasaki K, Miyoshi T. 2014. Data assimilation with error-correlated and non-orthogonal observations: Experiments with the Lorenz-96 model. *SOLA* **10**: 210–213.

van Leeuwen PJ. 2015. Representation errors and retrievals in linear and nonlinear data assimilation. *Q. J. R. Meteorol. Soc.* **141**: 612–1623, https://doi.org/10.1002/qj.2464.

Ventress L, Dudhia A. 2014. Improving the selection of IASI channels for use in numerical weather prediction. *Q. J. R. Meteorol. Soc.* **140**: 2111–2118.

Waller JA, Dance SL, Lawless AS, Nichols NK. 2014a. Estimating correlated observation error statistics using an ensemble transform Kalman filter. *Tellus A* **66**: 23294.

Waller JA, Dance SL, Lawless AS, Nichols NK, Eyre JE. 2014b. Representativity error for temperature and humidity using the Met Office high-resolution model. *Q. J. Meteorol. Soc.* **140**: 1189–1197.

Waller JA, Ballard SP, Dance SL, Kelly G, Nichols NK, Simonin D. 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. https://doi.org/10.3390/rs8070581.

Waller JA, Dance SL, Nichols NK. 2016b. Theoretical insight into diagnosing observation-error correlations using observation-minus-background and observation-minus-analysis statistics. *Q. J. R. Meteorol. Soc.* **142**: 418–431.

Waller JA, Simonin D, Dance SL, Nichols NK, Ballard SP. 2016c. 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. Weather Rev.* **144**: 3533–3551. https://doi.org/10.1175/MWR-D-15-0340.1.

Weston PP, Bell W, Eyre JR. 2014. Accounting for correlated error in the assimilation of high-resolution sounder data. *Q. J. Meteorol. Soc.* **140**: 2420–2429.