University of Reading

The impact of using reconditioned correlated observation error covariance matrices in the Met office 1D-Var system

Article

Accepted Version

Tabeart, J. M., Dance, S.L. ORCID: https://orcid.org/0000-0003-1690-3338, Lawless, A.S. ORCID: https://orcid.org/0000-0002-3016-6568, Migliorini, S., Nichols, N. K. ORCID: https://orcid.org/0000-0003-1133-5220, Smith, F. and Waller, J. A. (2020) The impact of using reconditioned correlated observation error covariance matrices in the Met office 1D-Var system. Quarterly Journal of the Royal Meteorological Society, 146 (728). pp. 1372-1390. ISSN 1477-870X doi: 10.1002/qj.3741 Available at https://centaur.reading.ac.uk/88395/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1002/qj.3741

Publisher: Royal Meteorological Society

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in

the End User Agreement.

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

The impact of using reconditioned correlated observation error covariance matrices in the Met Office 1D-Var system

Jemima M. Tabeart [*†], Sarah L. Dance [‡], Amos S. Lawless [*], Stefano Migliorini [§], Nancy K. Nichols [*], Fiona Smith [§¶], Joanne A. Waller [‡]

January 8, 2020

Abstract

Recent developments in numerical weather prediction have led to the use of correlated observation error covariance (OEC) information in data assimilation and forecasting systems. However, diagnosed OEC matrices are often ill-conditioned and may cause convergence problems for variational data assimilation procedures. Reconditioning methods are used to improve the conditioning of covariance matrices while retaining correlation information. In this paper we study the impact of using the 'ridge regression' method of reconditioning to assimilate Infrared Atmospheric Sounding Interferometer (IASI) observations in the Met Office 1D-Var system. This is the first systematic investigation of how changing target condition numbers affects convergence of a 1D-Var routine. This procedure is used for quality control, and to estimate key variables (skin temperature, cloud top pressure, cloud fraction) that are not analysed by the main 4D-Var data assimilation system. Our new results show that the current (uncorrelated) OEC matrix requires more iterations to reach convergence than any choice of correlated OEC matrix studied. This suggests that using a correlated OEC matrix in the 1D-Var routine would have computational benefits for IASI observations. Using reconditioned correlated OEC matrices also increases the number of observations that pass quality control. However, the impact on skin temperature, cloud fraction and cloud top pressure is less clear. As the reconditioning parameter is increased, differences between retrieved variables for correlated OEC matrices and the operational diagonal OEC matrix reduce. As correlated choices of OEC matrix yield faster convergence, using stricter convergence criteria along with these matrices may increase efficiency and improve quality control.

Key words: Correlated observation errors, IASI, 1D-Var, reconditioning, data assimilation

1 Introduction

In numerical weather prediction (NWP) a data assimilation procedure is used to combine observations of the atmosphere with a model description of the system in order to obtain initial conditions for forecasts. The contribution of each component is weighted by its respective error statistics. In recent years, interest in the understanding and use of correlated observation error covariance (OEC) matrices has grown (e.g. Janjić et al. [2018]). This increased interest has been motivated by results showing that neglecting correlated observation errors hinders forecasts [Rainwater et al., 2015, Stewart et al., 2008], and that even including poorly approximated correlation structures is better than using a diagonal error matrix in the presence of correlated errors [Stewart et al., 2013, Healy and White, 2005].

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

[†]Now at School of Mathematics, University of Edinburgh, UK (jemima.tabeart@ed.ac.uk)

[‡]School of Mathematical, Physical and Computational Sciences, University of Reading, UK

[§]Met Office, Exeter, UK

[¶]Now at Bureau of Meteorology, Melbourne, Australia

Previously, diagonal observation error covariance matrices were used for all observations, even when it was known that non-zero error correlations were present. Determining error statistics is a non-trivial problem, as they cannot be observed directly and must be estimated in a statistical sense. It was also thought that it would not be possible to use correlated observation error covariance matrices operationally due to the increased computational cost associated with inverting a dense matrix rather than a diagonal matrix [Stewart et al., 2013]. The development of a new method to check error consistency by Desroziers et al. [2005] was first applied to explicitly diagnose error correlations using the Met Office system [Stewart et al., 2009]. Since then, the diagnostic introduced in Desroziers et al. [2005] (henceforth referred to as DBCP) has been used widely at operational centres [Weston, 2011, Weston et al., 2014, Stewart et al., 2014, Bennitt et al., 2017, Bormann et al., 2011, 2016, Campbell et al., 2017, Gauthier et al., 2018, Wang et al., 2018], although diagonal OEC matrices are still used operationally for most instruments. Although much of the initial use of the diagnostic to estimate observation errors focussed on interchannel correlations, this has been extended to spatial correlations [Waller et al., 2014, 2016a,c, Cordoba et al., 2016] and temporal correlations [Bennitt et al., 2017]. In addition to the direct use of estimated covariance matrices in data assimilation schemes, information obtained using the DBCP diagnostic has been used to model error covariance functions and operators (e.g. Stewart et al. [2013], Michel [2018], Simonin et al. [2019], Guillet et al. [2019]). Theoretical work has also demonstrated how well the diagnostic is expected to perform depending on either the accuracy of the initial choice of background and OEC matrices for the single step [Waller et al., 2016b] and the iterative form of the diagnostic [Ménard, 2016, Bathmann, 2018]. The use of the diagnostic in data assimilation schemes using localization has also been considered [Waller et al., 2017].

The output of the diagnostic cannot be used directly in the assimilation procedure. Diagnosed matrices are asymmetric, and some are not positive definite [Stewart et al., 2014, Weston et al., 2014] and are therefore not valid covariance matrices. Typically, the matrices are symmetrised, and negative and zero eigenvalues are set to be small and positive [Weston, 2011]. Additionally, diagnosed OEC matrices are often ill-conditioned. This means that small perturbations to the observations will result in large changes to the analysis, and that iterative methods are likely to converge slowly. Indeed, the direct use of diagnosed matrices has led to problems with non-convergence of the minimisation of the data assimilation procedure [Weston, 2011, Weston et al., 2014]. Weston [2011] suggested that part of these problems were due to small minimum eigenvalues of the diagnosed OEC matrix, $\mathbf{R}$.

One way to study the effect of changes to the assimilation system on the convergence of the objective function minimisation is by using the condition number of the Hessian of the variational objective function as a proxy for convergence. This was done in Haben [2011] for the case of a linear observation operator. In Tabeart et al. [2018] the minimum eigenvalue of the OEC matrix, $\mathbf{R}$, appears in bounds on the condition number of the Hessian of the variational assimilation problem, indicating that this term will also be important for convergence of the objective function minimisation. The bounds indicate that decreasing the minimum eigenvalue of the OEC matrix will increase the condition number of the Hessian.

Increased understanding of how the eigenvalues of $\mathbf{R}$ affect the convergence of the data assimilation problem motivated investigation into 'reconditioning' methods [Weston, 2011, Weston et al., 2014, Campbell et al., 2017, Tabeart et al., 2018]. These methods increase eigenvalues of the matrix $\mathbf{R}$ to improve the conditioning of the OEC matrix, while maintaining much of the existing correlation structure of the diagnosed matrix. Two methods are commonly used by NWP centres: 'ridge regression' which increases all eigenvalues of $\mathbf{R}$ by the same amount, and the 'minimum eigenvalue' method which changes only the smallest eigenvalues. These methods were investigated theoretically in Tabeart et al. [2019] where it was found that both methods increase standard deviations, and that the ridge regression method strictly reduces all off-diagonal correlations. Both methods were compared in an operational system in Campbell et al. [2017], where the sensitivity of forecasts to the choice of method was found to be small, but the ridge regression outperformed the minimum eigenvalue method in terms of convergence. A method similar to the minimum eigenvalue method is used at the European Centre for Medium Range Weather Forecasts (ECMWF) [Bormann et al., 2016], but will not be discussed further in this paper.

The aim of this paper is to investigate the use of the ridge regression method within the Met Office system. At the Met Office, in addition to the 4D-Variational data assimilation routine (4D-Var) that

is used to produce the initial conditions for weather forecasts, a 1D-Variational data assimilation routine (1D-Var) is used for quality control and pre-processing purposes [Eyre, 1989]. The 1D-Var routine assimilates observations individually, and is used to remove observations that are likely to cause problems with convergence in the 4D-Var routine, as well as to estimate model variables that are not included in the 4D-Var state vector [Pavelin and Candy, 2014, Pavelin et al., 2008]. After the work of Weston [2011], Weston et al. [2014], correlated OEC matrices were introduced in the 4D-Var routine for IASI (Infrared Atmospheric Sounding Instrument) and other observation types. However, this was not the case for the 1D-Var routine, where a diagonal OEC matrix continues to be used. Previous work found that diagnosed observation error correlations were small for most channels for the 1D-Var routine [Weston, 2011, Stewart et al., 2014] and the proportional increase in computational cost was estimated to be large compared with using correlated OEC matrices in 4D-Var [Weston et al., 2014]. However, this was partly due to the use of an extremely ill-conditioned choice of OEC matrix.

In this paper we study how the use of reconditioning methods affects the 1D-Var routine when applied to interchannel OEC matrices for the Infrared Atmospheric Sounding Interferometer (IASI). We examine whether the ridge regression method of reconditioning allows us to include correlated observation error information more efficiently than the diagnosed OEC matrix. This method of reconditioning is used at the Met Office to recondition OEC matrices that are used in the 4D-Var routine. We compare a selection of reconditioned OEC matrices with the current diagonal operational error covariance matrix, and an inflated diagonal OEC matrix. This is the first time that multiple levels of reconditioning have been compared systematically in an operational system. We study the impact of reconditioning in terms of the computational efficiency as well as the effect on important meteorological variables.

In Section 2 the data assimilation problem is defined and the ridge regression method of reconditioning is introduced. In Section 3 we provide an overview of the experimental design. In Sections 4 and 5 we discuss the impact of changing the OEC matrix on the 1D-Var procedure, and alterations to the quality control and pre-processing for the 4D-Var routine respectively. We find that convergence is improved for any of the choices of reconditioning compared to the current operational choice of OEC matrix. Additionally, increasing the amount of reconditioning results in faster convergence - which corresponds to theoretical results for the linear variational data assimilation problem in Tabeart et al. [2018]. However, the quality control procedure is altered by changing the OEC matrix, with a larger number of observations being accepted for reconditioned correlated OEC matrices compared to the current diagonal choice of OEC matrix. We also find that for most variables, the difference between retrieved values for different choices of OEC matrix are small compared to retrieved standard deviations. However, there are a significant minority of observations for which differences are very large. Finally, in Section 6 we summarise our results and conclusions.

2 Variational data assimilation and reconditioning

2.1 Data assimilation

In data assimilation, a weighted combination of observations, $\mathbf{y} \in \mathbb{R}^p$, with a background, or 'prior', field, $\mathbf{x}_b \in \mathbb{R}^n$, is used to obtain the analysis, or posterior, $\mathbf{x}_a \in \mathbb{R}^n$, where p denotes the number of observations and n the number of state variables. The weights are the respective error statistics of the two components. The matrix $\mathbf{R} \in \mathbb{R}^{p\times p}$ is the observation error covariance (OEC) matrix and $\mathbf{B} \in \mathbb{R}^{n\times n}$ is the background error covariance matrix. In order to compare observations with the background field, the, possibly non-linear, observation operator $H : \mathbb{R}^n \to \mathbb{R}^p$ is used to map from state space to observation space. The weighted combination is written in the form of an objective function in terms of $\mathbf{x} \in \mathbb{R}^n$, the model state vector. In the case of 3D-Var the objective function is given by:

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^\top \mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) + \frac{1}{2}(\mathbf{y}-H[\mathbf{x}])^\top \mathbf{R}^{-1}(\mathbf{y}-H[\mathbf{x}]). \tag{1}$$

The value of $\mathbf{x}$ that minimises (1) is given by $\mathbf{x}_a$.

The first order Hessian, or matrix of second derivatives, of the objective function (1) is given by

$$\nabla^2 J \equiv \mathbf{S} = \mathbf{B}^{-1} + \mathbf{H}^\top \mathbf{R}^{-1} \mathbf{H}, \tag{2}$$

where $\mathbf{H} \in \mathbb{R}^{p \times n}$ is the Jacobian of the observation operator, $H[\mathbf{x}]$, linearised about the current best estimate of the optimal solution of (1).

We now define the condition number of a matrix. Although covariance matrices are symmetric positive semi-definite by definition, in practice $\mathbf{B}$ and $\mathbf{R}$ are required to be strictly positive definite in order that they can be inverted in (1). This means that $\mathbf{S}$ is symmetric positive definite, and its condition number is given by

$$\kappa(\mathbf{S}) = \frac{\lambda_{max}(\mathbf{S})}{\lambda_{min}(\mathbf{S})}. \tag{3}$$

We note that the minimum possible value of the condition number of any matrix is one. The condition number of the Hessian is of interest because it can be used to study the sensitivity of the solution to small changes in the background or observation data [Golub and Van Loan, 1996, Sec 2.7]. As (1) is non-linear, it is solved using a sequence of Gauss-Newton iterations with an inner linearised problem solved using the conjugate gradient method [Haben et al., 2011]. The rate of convergence of the minimisation of the linearised problem by a conjugate gradient function can also be bounded by $\kappa(\mathbf{S})$ [Golub and Van Loan, 1996], although this bound can be quite pessimistic. In particular, clustering of eigenvalues can result in much faster convergence than is predicted by $\kappa(\mathbf{S})$ [Nocedal, 2006].

2.2 Reconditioning: motivation and definition

In Weston [2011], observations from IASI were used at the UK Met Office for an initial study investigating the feasibility of using correlated observation error matrices in their 4D-Var system. A first guess of the OEC matrix was obtained using the DBCP diagnostic.

One problem that was encountered in Weston [2011] and Weston et al. [2014] was the ill-conditioning of the matrix resulting from the DBCP diagnostic. The use of an ill-conditioned OEC matrix can result in slower convergence of a variational scheme [Weston et al., 2014, Tabeart et al., 2018]. Similar problems were encountered at ECMWF where a degradation in the forecast was seen when the raw output of the DBCP diagnostic was tested [Lupu et al., 2015]. Weston [2011] suggested that the convergence problems were caused by very small minimum eigenvalues of the diagnosed error covariance matrix.

Tabeart et al. [2018] developed bounds for the condition number of the Hessian in terms of its constituent matrices. This provides an indication of the role of each matrix in the conditioning of $\mathbf{S}$, and therefore the convergence of the associated minimisation problem. We recall that there are well known cases where the condition number provides a very loose upper bound on convergence of a conjugate gradient method, for example in the case of clustered or repeated eigenvalues. However, we expect that qualitative conclusions will still hold: if changing a component of the data assimilation system reduces the condition number of the Hessian, it is likely to lead to improved convergence. The bound which separates the role of each matrix is given by

$$\begin{aligned} \max \left\{ \frac{1 + \frac{\lambda_{max}(\mathbf{B})}{\lambda_{min}(\mathbf{R})} \lambda_{min}(\mathbf{H}\mathbf{H}^\top)}{\kappa(\mathbf{B})}, \frac{1 + \frac{\lambda_{max}(\mathbf{B})}{\lambda_{max}(\mathbf{R})} \lambda_{max}(\mathbf{H}\mathbf{H}^\top)}{\kappa(\mathbf{B})}, \frac{\kappa(\mathbf{B})}{1 + \frac{\lambda_{max}(\mathbf{B})}{\lambda_{min}(\mathbf{R})} \lambda_{max}(\mathbf{H}\mathbf{H}^\top)} \right\} \\ \leq \kappa(\mathbf{S}) \leq \Big(1 + \frac{\lambda_{min}(\mathbf{B})}{\lambda_{min}(\mathbf{R})} \lambda_{max}(\mathbf{H}\mathbf{H}^\top)\Big) \kappa(\mathbf{B}). \end{aligned} \tag{4}$$

These bounds show that the minimum eigenvalue, $\lambda_{min}(\mathbf{R})$, of the OEC matrix is a key term in the upper bound for $\mathbf{S}$, meaning that increasing the minimum eigenvalue of $\mathbf{R}$ is a reasonable heuristic for reducing the condition number of $\mathbf{S}$ and improving the conditioning of the problem (1).

In the case that the error covariance matrices can be written as the product of a scalar variance with a correlation matrix, e.g. $\mathbf{R} = \sigma_o^2\mathbf{D}$ and $\mathbf{B} = \sigma_b^2\mathbf{C}$, and observations are restricted to model variables, we can simplify the bound (4) to

$$\max\left\{\frac{1+\frac{\sigma_b^2}{\sigma_o^2}\frac{\lambda_{max}(\mathbf{C})}{\lambda_{min}(\mathbf{D})}}{\kappa(\mathbf{C})}, \frac{\kappa(\mathbf{C})}{1+\frac{\sigma_b^2}{\sigma_o^2}\frac{\lambda_{max}(\mathbf{C})}{\lambda_{min}(\mathbf{D})}}\right\} \leq \kappa(\mathbf{S}) \leq \left(1+\frac{\sigma_b^2}{\sigma_o^2}\frac{\lambda_{min}(\mathbf{C})}{\lambda_{min}(\mathbf{D})}\right)\kappa(\mathbf{C}). \tag{5}$$

The qualitative conclusions of Tabeart et al. [2018] can be summarised as follows.

- The minimum eigenvalue of $\mathbf{R}$ was shown to be important for determining both the conditioning of the Hessian, and the speed of convergence of a minimisation procedure. This can be seen in (4) and (5).
- The ratio of the background and observation variances was also shown to be important for conditioning of the Hessian. This can be seen in (5) explicitly for the case of direct observations where variances are homogeneous for both background and observation error covariance matrices. However, we expect the conclusion to hold more broadly, for example in the case where all standard deviation values corresponding to an OEC matrix were larger than those corresponding to another OEC matrix, then the bounds would be smaller for the first choice of OEC matrix.
- Although (4) and (5) separate the contribution of each term, numerical experiments revealed that the level of interaction between observation error and background error statistics depends on the choice of observation network. Examples of observation operators which yield identical bounds for (4) but different dependence of $\kappa(\mathbf{S})$ on $\mathbf{B}$ and $\mathbf{R}$ were found experimentally in Tabeart et al. [2018].
- Numerical experiments reveal that for an idealised framework, changes to the components of the assimilation scheme that decrease the condition number of the Hessian also lead to faster convergence of a conjugate gradient method.

Therefore for the unpreconditioned data assimilation problem, reducing $\kappa(\mathbf{S})$ is likely to result in improved convergence of a minimisation method. However, this conclusion may be affected by changes to the data assimilation scheme, such as the use of preconditioners or more complicated observation operators.

The importance of the minimum eigenvalue of the OEC matrix in controlling the condition number of the Hessian motivates the use of reconditioning methods. In order to make operational implementation of correlated observation error matrices feasible, it is necessary to reduce the impact of the very small eigenvalues of the matrix $\mathbf{R}$ by increasing its condition number. To achieve this, different methods of inflation or reconditioning are used to improve conditioning of correlation matrices for a variety of applications. The ridge regression method is used to recondition OEC matrices at the Met Office [Weston, 2011, Weston et al., 2014], and hence will be the reconditioning method that is considered in the remainder of this paper. The ridge regression method adds a scalar multiple of the identity to $\mathbf{R}$ to obtain the reconditioned matrix $\mathbf{R}_{RR}$. This scalar, δ, is chosen such that $\kappa(\mathbf{R}_{RR}) = \kappa_{max}$, a user-specified condition number. The method for calculating δ for a given choice of κ_{max} was formally defined in Weston et al. [2014], Tabeart et al. [2019] as follows:

Definition 1 ***Ridge regression reconditioning constant,*** δ ***[Weston et al., 2014, Tabeart et al., 2019]***
Define $\delta = (\lambda_{max}(\boldsymbol{R}) - \lambda_{min}(\boldsymbol{R})\kappa_{max})/(\kappa_{max}-1)$.
Set $\boldsymbol{R}_{RR} = \boldsymbol{R} + \delta\boldsymbol{I}$

We note that this choice of δ yields $\kappa(\mathbf{R}_{RR}) = \kappa_{max}$. Mathematical theory describing the effect of this reconditioning method on the correlations and variances of any covariance matrix was developed in Tabeart et al. [2019]. This work proved that the ridge regression method increases variances for all variables, and decreases the magnitude of all off-diagonal correlations. Numerical experiments

illustrate the effect of using the ridge regression method of reconditioning on the variances [Tabeart et al., 2019, Figure 2] and matrix of correlations [Tabeart et al., 2019, Figure 3] corresponding to the OEC matrix for IASI estimated by the Met Office. Tabeart et al. [2019] also studied the impact of using the ridge regression method of reconditioning on the variational objective function (1). Use of the ridge regression method was found to reduce the weight on all observation information in the analysis, with a relatively larger effect on information associated with small eigenvalues.

In this paper we study the implementation of ridge regression that is used operationally at the Met Office as described in Weston et al. [2014]: the covariance matrix is reconditioned directly. We therefore do not consider alternative approaches, including the application of ridge regression to correlation matrices or the use of variance rescaling procedures that may be used at other NWP centres. We investigate whether the qualitative conclusions from Tabeart et al. [2018] hold in the case of a non-linear observation operator, and we study the impact of reconditioning methods in an operational system.

3 Experimental Overview

3.1 Met Office System

The experiments carried out in this manuscript will use observations from the IASI instrument on the EUMETSAT MetOp constellation. IASI is an infrared Fourier transform spectrometer, and measures infrared radiation emissions from the atmosphere and surface of the earth [Chalon et al., 2001]. We note that the observation operator for this instrument, a radiative transfer model, is highly non-linear so the conclusions from Tabeart et al. [2018] will not necessarily apply to this problem. The infrared spectrum is split into channels corresponding to different wavelengths; this means that an observation at a single location will provide information for up to 8641 channels. An early use of the DBCP diagnostic focused on observations from IASI implemented in the Met Office system [Stewart et al., 2009]. Much of the subsequent research on correlated observation error uses IASI observations [Weston, 2011, Weston et al., 2014, Stewart et al., 2014, Bormann et al., 2016]. In particular IASI has channels that are sensitive to water vapour which have been found to have errors with large correlations [Stewart et al., 2014, Weston et al., 2014, Bormann et al., 2016].

One attraction of IASI, and other hyperspectral instruments, is the large number of available channels, which provides high vertical resolution. However, using all of these channels is not feasible in current operational NWP systems for reasons including computational expense, and not requiring too many observations of a similar type. Additionally, when IASI was first used, there was a reluctance to include correlated observation errors so an effort was made to choose channels that are spectrally different and hence less likely to have correlated errors [Stewart et al., 2014]. This means that of the 8461 available channels, only a few hundred are used at most NWP centres [Stewart et al., 2014]. At the time of the experiments, the Met Office stored a subset of 314 channels with a maximum of 137 being used in the 4D-Var system. A list of these channels is given by Stewart [2010, Appendix A]. As there is a large degree of redundancy between channels [Collard et al., 2010], directly assimilating a larger number of channels is likely to make the conditioning of the OEC matrix worse. This has motivated alternative approaches such as principal component compression [Collard et al., 2010] and the use of transformed retrievals [Prates et al., 2016], which will not be considered in this work.

A larger number of channels is used for the 1D-Var assimilation than for the Met Office 4D-Var assimilation; standard deviation values for these channels are filled in from the current operational (diagonal) OEC matrix. We choose to focus on the channels used in the 4D-Var system in order to be consistent between both assimilation systems. We also note that not all channels are used for each assimilation; for example, some channels are not used in the presence of cloud. In this case, rows and columns corresponding to channels that are affected by cloud are deleted from the OEC matrix. As the submatrix chosen from the full OEC matrix used could change at each observation time, there may be a difference in the condition number of the OEC matrix used in practice and the OEC matrices presented in this work. However, the Cauchy interlacing theorem [Bernstein, 2009, Lemma 8.4.4] states that the condition number will not be increased by deleting rows and columns of a symmetric positive definite matrix. This means that the values given here are upper bounds for

$\kappa(\mathbf{S})$ even if the quality control procedure excludes some channels.

We test the impact of using correlated OEC matrices in the Met Office 1D-Var system and consider the effect of using the ridge regression method of reconditioning with different choices of target condition number. At the Met Office, 1D-Var is run prior to every 4D-Var assimilation procedure, meaning that computational efficiency and speed are vitally important. We note that a single IASI observation consists of brightness temperature values for each of the channels that are used in the assimilation. A 1D-Var procedure assimilates observations separately at each location to retrieve variables such as temperature and humidity over a 1D column of the atmosphere. This procedure is much cheaper and more parallelisable than a 4D-Var algorithm.

The 1D-Var routine performs two main functions:

1. Observations that require more than 10 iterations for the 1D-Var minimisation to reach convergence are not passed to the 4D-Var routine. This is because it is assumed that observations for which the retrieval procedure takes too long to converge for the 1D-Var minimisation will also result in slow convergence for a 4D-Var minimisation. A Marquardt-Levenberg minimisation algorithm is used, with convergence criteria based on the value of the cost function and normalised gradient [Pavelin et al., 2008]. Changing the OEC matrix will alter the speed of convergence of 1D-Var, and hence affect which observations are accepted.

2. Estimation of values for certain variables that are not included in the 4D-Var state vector: values for skin temperature, cloud fraction, cloud top pressure and emissivity over land are fixed by the 1D-Var procedure. Altering the OEC matrix will change retrieved values for these variables.

Changing the OEC matrix is therefore likely to have two main effects on results of the 1D-Var procedure: changing the observations that are accepted by the quality control, and changing the values of those variables not included in the 4D-Var state vector. Skin temperature (ST), cloud fraction (CF) and cloud top pressure (CTP) are retrieved as scalar values at each observation location. In contrast, surface emissivity is retrieved as a spectrum, which is represented as a set of leading principal components [Pavelin and Candy, 2014]. As we expect the interactions between the choice of $\mathbf{R}$ and the retrieved values to be complex, in this work we only consider the effect of changing $\mathbf{R}$ on the three scalar variables: skin temperature, cloud top pressure and cloud fraction.

3.2 Experimental Design

We now describe the experimental framework and key areas of interest that will be investigated in Sections 4 and 5. We use the operational Met Office 1D-Var framework at the time of the experiments (July 2016), and consider how the results change for different choices of OEC matrix. Background data are obtained from the Unified Model (UM) background files for the corresponding configuration. A number of different times and dates for the six months between December 2015 and June 2016 were considered, but as results were similar across all trials we only present results from experiments for the six-hourly data assimilation window centred on 16th June 2016 0000 UTC.

The correlated choices of $\mathbf{R}$ are calculated using the method introduced in Weston et al. [2014]; applying the ridge regression method of reconditioning to the diagnosed matrix for a variety of choices of κ_{max}. The matrices estimated by the DBCP diagnostic depend considerably on the choice of background and observation error matrices. For all OEC matrices produced, the same 4 days of IASI and background NWP data (03/12/15-06/12/15), were used as input data. We note that the estimated OEC matrix was obtained using background and OEC matrices from the 4D-Var assimilation routine rather than the 1D-Var routine. Previous studies have found stronger correlations between a larger block of water vapour sensitive channels for 4D-Var diagnosed matrices and 1D-Var diagnosed matrices [Weston, 2011, Stewart et al., 2014]. This means that using an OEC matrix estimated using 4D-Var is likely to result in under-weighting of information from these channels, and hence sub-optimal use of the IASI observations. However, the use of 4D-Var error statistics within the 1D-Var assimilation allows us to better understand the impact that our changes are likely to have on 4D-Var. We are using 1D-Var as a pre-processing step for 4D-Var to remove observations that are likely to cause convergence issues in the main assimilation algorithm.

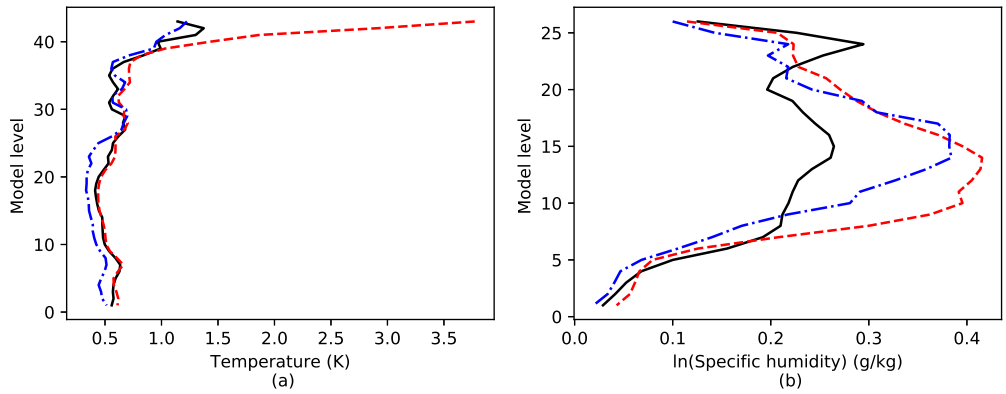

Figure 1: Standard deviation values for the operational background error covariance matrices, **B**, for the northern hemisphere (solid line), tropics (dot-dashed line) and southern hemisphere (dashed line) for temperature (a) and ln(specific humidity) (b).

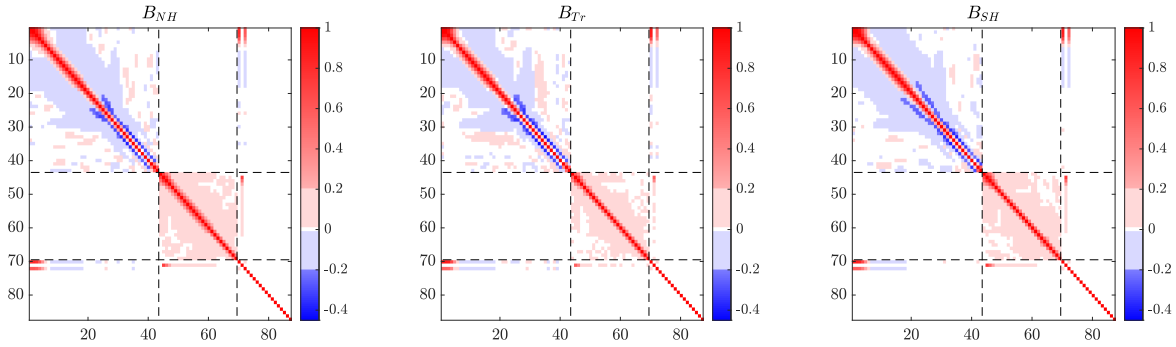

Figure 2: Correlation matrices for the operational background error covariance matrices, **B**, for the northern hemisphere (a), tropics (b) and southern hemisphere (c). Dashed vertical and horizontal lines separate inter and cross correlations between temperature, ln(specific humidity) and other variables (from left to right). Variables 1- 43 correspond to temperature, 44-70 to ln(specific humidity), 71 to surface air temperature, 72 to surface specific humidity, 73 to skin temperature, 74 to surface pressure, 75 to cloud top pressure, 76 to cloud fraction and 77-88 to the first 12 principal components of surface emissivity.

Table 1: Background standard deviation values for variables not included in the 4D-Var state vector: Cloud Fraction (CF), Cloud Top Pressure (CTP) and Skin Temperature (ST).

Variable	Standard deviation
Cloud Fraction (global)	1
Cloud Top Pressure (hPa) (global)	1000
Skin Temperature (K) (Northern Hemisphere)	2.24
Skin Temperature (K) (Tropics)	1.92
Skin Temperature (K) (Southern Hemisphere)	2.02

We use the operational background error covariance matrix, $\mathbf{B}$, at the time of the experiments. This was calculated using the so-called randomization method as described in Andersson et al. [2000], Eyre and Hilton [2013]. In this method the matrix used to transform from "control" space to model space is multiplied by normally-distributed random vectors (vertical profiles) with zero mean and unit covariance. The $\mathbf{B}$ matrix is then determined as the average of an ensemble of these outer products. We note that the ensemble of random vectors was chosen to represent different locations and times within the assimilation window. Figure 1 shows background error standard deviation (BESD) values for temperature and humidity variables, and Table 1 gives BESD values for CF, CTP and ST for each of the choices of $\mathbf{B}$. The 43 model levels are determined by the 43 evenly distributed pressure levels in the radiative transfer retrieval algorithm. Figures in this paper, are plotted with model level 1 corresponding to the atmospheric level closest to the surface, and model level 43 corresponding to the top of the atmosphere. This is the opposite ordering to that which is used by RTTOV. We note that standard deviations for cloud variables are assumed to be very large so that the background is effectively ignored for these variables [Pavelin et al., 2008]. In Sections 4 and 5 we will compare the standard deviations from the background error covariance matrix against retrieved standard deviations for the observations as well as differences between observations for different choices of $\mathbf{R}$. Figure 2 shows that correlations corresponding to the three choices of $\mathbf{B}$ are qualitatively very similar. Cross-correlations between variables are quite weak, with no correlations between temperature and specific humidity. Most correlations larger than 0.2 occur for adjacent model levels for temperature and specific humidity. Correlations greater than 0.2 also occur between surface temperature and temperature, and ST and temperature for lower model levels, and surface specific humidity and specific humidity at lower model levels. CTP and CF are uncorrelated with all other variables.

We apply the DBCP diagnostic to the subset of 137 channels that are assimilated in the 4D-Var routine. The 1D-Var routine uses additional channels [Hilton et al., 2009], with a total of 183 channels being assimilated. Observation errors for these additional channels are assumed to be uncorrelated, and filled in with values from the diagonal error covariance matrix $\mathbf{R}_{diag}$. If additional channels are included in future versions of the operational system, it would be advisable to recompute the DBCP diagnostic applied to all channels.

The seven different choices of the matrix $\mathbf{R}$ that were tested are now listed:

- $\mathbf{R}_{infl}$ which is an inflated diagonal matrix. This matrix was used prior to the introduction of correlated observation error in the 4D-Var assimilation scheme, and is the operational OEC matrix used in Weston et al. [2014] for 4D-Var. In particular variances are inflated to account for the fact that the assumption of uncorrelated errors is incorrect. The standard deviations (square root of the diagonal entries of $\mathbf{R}_{infl}$) are shown in Weston et al. [2014, Figure 1] by the black dashed line. The largest value entry of $\mathbf{R}_{infl}$ is 16, and the smallest entry is 0.25. The construction of $\mathbf{R}_{infl}$ is described in Hilton et al. [2009].

- $\mathbf{R}_{diag}$, the current operational matrix for the 1D-Var retrievals, which is diagonal. The standard deviations are calculated as instrument noise plus $0.2K$ forward-model noise [Collard, 2007]. The variances of $\mathbf{R}_{diag}$ are shown in Stewart et al. [2014, Figure 7] by the red line. The variances are much smaller than for $\mathbf{R}_{infl}$; for the first 120 channels, the diagonal elements of $\mathbf{R}_{diag}$ are all less than 0.27 and the largest value of $\mathbf{R}_{diag}$ is given by 0.49.

- $\mathbf{R}_{est}$, the symmetrised raw output of the code that produces the DBCP diagnostic. This is

Table 2: Minimum eigenvalues and condition number of $\mathbf{R}$ for each experiment.

Experiment name	E_{diag}	E_{est}	E_{1500}	E_{1000}	E_{500}	E_{67}	E_{infl}
Choice of $\mathbf{R}$	$\mathbf{R}_{diag}$	$\mathbf{R}_{est}$	$\mathbf{R}_{1500}$	$\mathbf{R}_{1000}$	$\mathbf{R}_{500}$	$\mathbf{R}_{67}$	$\mathbf{R}_{infl}$
$\lambda_{min}(\mathbf{R})$	0.025	0.00362	0.00482	0.007244	0.0145	0.1010	0.0625
$\kappa(\mathbf{R})$	9.263	2730	1500	1000	500	67	64

computed by $\mathbf{R}_{est} = 0.5(\mathbf{R}_{DBCP} + \mathbf{R}_{DBCP}^T)$, where $\mathbf{R}_{DBCP} \in \mathbb{R}^{137\times137}$ is the output of the DBCP diagnostic.

- Reconditioned versions of $\mathbf{R}_{est}$ so that the correlated submatrix has a condition number of 1500, 1000, 500 and 67, referred to respectively as $\mathbf{R}_{1500}, \mathbf{R}_{1000}, \mathbf{R}_{500}$ and $\mathbf{R}_{67}$.

We refer to the experiments using each choice of OEC matrix as E with subscript corresponding to that of the OEC matrix (i.e. $E_{diag}, E_{est}, E_{1500}E_{1000}, E_{500}, E_{67}$ and E_{infl}). The correlation matrices and standard deviations corresponding to each choice of $\mathbf{R}$ are shown in Figure S1 and S2 respectively in the Supplementary Material.

Details of the conditioning, and minimum eigenvalues of each of the choices of $\mathbf{R}$ can be found in Table 2. We see that for the non-diagonal matrices, as we decrease the target condition number, we increase the minimum eigenvalue of $\mathbf{R}$. This agrees with the theoretical results of Tabeart et al. [2019]. We also see that of the two diagonal choices of $\mathbf{R}$, $\lambda_{min}(\mathbf{R}_{infl}) > \lambda_{min}(\mathbf{R}_{diag})$, suggesting that we might expect better convergence with $\mathbf{R}_{infl}$ compared to $\mathbf{R}_{diag}$. We also notice that the largest value of $\lambda_{min}(\mathbf{R})$ occurs for $\mathbf{R}_{67}$. It will be of interest to consider whether the introduction of correlations has more effect on convergence and conditioning than the value of $\lambda_{min}(\mathbf{R})$. We note that the inclusion of 46 extra channels in the 1D-Var algorithm, in addition to the 137 channels used in the 4D-Var algorithm, could change the condition numbers presented in Table 2 by the introduction of very small or very large eigenvalues.

Our numerical experiments will be broadly split into two groups. Firstly we will consider the effect of changing the OEC matrix, $\mathbf{R}$, on the 1D-Var procedure itself in Section 4. This includes the impact on retrieved values and the convergence of the 1D-Var assimilation. Secondly, in Section 5, we will consider the impact of these changes on the 4D-Var procedure, by looking at how the number of accepted observations varies, and how the retrieved values of skin temperature, cloud top pressure, and cloud fraction retrievals are altered.

4 Impact on Met Office 1D-Var routine

In this section we consider the impact of changing the OEC matrix used in the Met Office 1D-Var system on the conditioning of the Hessian and on individual retrievals of temperature and humidity. In particular, the conditioning of the Hessian is important in terms of speed of convergence of the minimisation procedure. We recall (Section 3.1) that in 1D-Var information for each observation location is assimilated separately. Here a single observation corresponds to information from a column of IASI channels valid at one location. There are 97330 observations over the data assimilation window centred on 16th June 2016 0000 UTC with objective functions that converge in 10 or fewer iterations for all choices of E_{exp}. For much of the discussion that follows we will consider statistics of this set of 97330 observations to understand how changing the OEC matrix affects 1D-Var for IASI observations. We note that these observations are distributed across the globe, and hence the choice of background error covariance matrix (introduced in Section 3.2) will vary depending on observation location.

4.1 Influence of observation error covariance matrix on convergence and conditioning of the 1D-Var routine

We begin by investigating explicitly the effect of changing the OEC matrix, $\mathbf{R}$, on the 1D-Var routine. We consider two variables: the number of iterations required for convergence for the minimisation routine and the condition number of the Hessian of the 1D-Var cost function.

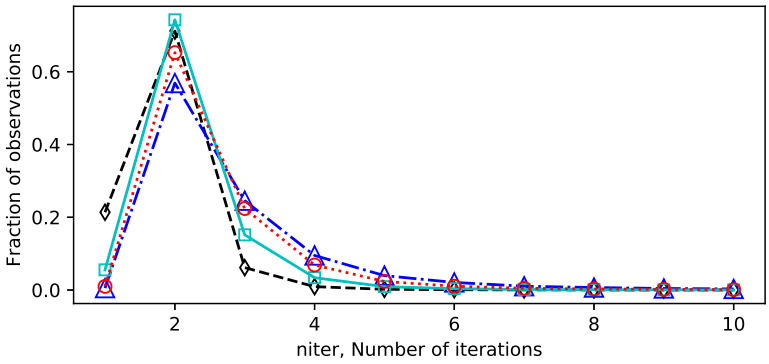

Figure 3: Number of iterations required for convergence of the minimization of the 1D-Var cost function as a fraction of the total number of observations common to all choices of $\mathbf{R}$. Symbols correspond to: $\mathbf{R}_{diag}$ ($\triangle$), $\mathbf{R}_{est}$ ($\circ$), $\mathbf{R}_{67}$ ($\square$) and $\mathbf{R}_{infl}$ ($\diamond$).

Table 3: Maximum, mean and median values of $\kappa(\mathbf{S})$ for control and experiments.

	E_{diag}	E_{est}	E_{1500}	E_{1000}	E_{500}	E_{67}	E_{infl}
max $\kappa(\mathbf{S})$	3.01×10^{12}	7.546×10^{11}	7.469×10^{11}	7.30×10^{11}	7.02×10^{11}	3.71×10^{11}	1.74×10^{11}
mean $\kappa(\mathbf{S})$	2.78×10^{10}	6.71×10^{9}	6.62×10^{9}	6.43×10^{9}	6.00×10^{9}	4.01×10^{9}	2.83×10^{9}
median $\kappa(\mathbf{S})$	2.09×10^{8}	1.31×10^{8}	1.32×10^{8}	1.33×10^{8}	1.37×10^{8}	1.78×10^{8}	2.89×10^{8}

Firstly we consider the number of iterations required for the minimisation of the 1D-Var cost function to reach convergence for each assimilated observation. For NWP centres, this is a variable of significant interest, as the extra expense of introducing correlated error predominantly comes from the increase in the number of iterations needed before convergence in the case of interchannel errors [Weston, 2011]. We note that this may not be the case for other types of error correlation such as spatial and temporal correlations (where the computation of matrix-vector products may require additional communication between processors [Simonin et al., 2019]). The minimisation is deemed to have converged when $|\mathbf{x}^i_{k+1} - \mathbf{x}^i_k| < 0.4\sigma^i_{\mathbf{B}}$ for all values of i, where $\mathbf{x}^i_{k+1}$ denotes the ith component of the estimated state vector $\mathbf{x}_{k+1}$ at iteration $k+1$ and $\sigma^i_{\mathbf{B}}$ is the ith component of the vector of background error variances. Values deemed to be unphysical, such as temperature components falling out of the range $70K - 340K$, are discarded.

For each observation, we store the number of iterations required for the corresponding 1D-Var objective function to converge, *niter*. Figure 3 shows the fraction of observations that have objective functions that converge in *niter* iterations for four choices of $\mathbf{R}$. We note that the behaviour for the other correlated experiments is similar to the behaviour for E_{est} and hence only the distributions for E_{est} and E_{67} are shown. We see that for all experiments $niter = 2$ is the modal class and contains over 50% of the observations. We begin by considering experiments corresponding to correlated choices of the matrix $\mathbf{R}$. Our results show that as the minimum eigenvalue of the matrix $\mathbf{R}$ increases, there is a decrease in the required number of iterations. This agrees with the theoretical conclusions of Tabeart et al. [2018]. However, the overall effect of reconditioning on convergence speed is less for 1D-Var than was observed in the case of 3D-Var or 4D-Var as described in Weston [2011]. It is likely that this is because the average number of iterations is greater in 3D and 4D-Var, and the maximum permitted number of iterations is much larger than the 10 allowed for the 1D-Var minimisation.

We now consider the two diagonal choices of OEC matrix, $\mathbf{R}_{infl}$ and $\mathbf{R}_{diag}$. The distribution corresponding to E_{diag} is more heavily weighted towards a higher number of iterations than any of the correlated cases. This is not what we might expect from an uncorrelated choice of OEC matrix, particularly as it is well-conditioned compared to most other choices of OEC matrix. In particular, $\lambda_{min}(\mathbf{R}_{diag})$ is greater than the minimum eigenvalue for all choices of correlated OEC matrix apart from $\mathbf{R}_{67}$ (see Table 2). In contrast, for the experiment E_{infl} convergence is faster than for any of the other experiments.

As we noted in Section 2.2, the minimum eigenvalue of the matrix $\mathbf{R}$ is not the only important property for determining the speed of convergence. The distribution of standard deviations for E_{diag} and E_{infl} is shown in Figure 1 of Weston et al. [2014]. As the standard deviations for $\mathbf{R}_{infl}$ are much larger than the standard deviations for any other choice of $\mathbf{R}$, the ratio of background variance to observation variance will be smaller for E_{infl} than other experiments, resulting in smaller condition numbers of the Hessian and hence faster convergence of the 1D-Var minimisation. We recall from Section 2.2 that the ratio of background to observation error variances appears in the bounds on the condition number of the Hessian given by (5) in Tabeart et al. [2018] and similar bounds in Haben [2011]. It is clear from these bounds that decreasing the observation error variance will increase the value of the bounds. We can therefore explain the worse convergence seen for E_{diag} by considering channels 107-121 and 128-137, where variances for $\mathbf{R}_{diag}$ are smaller than the variances for correlated choices of $\mathbf{R}$. These channels are sensitive to water vapour, and also correspond to the strongest positive correlations in $\mathbf{R}_{est}$. Typically, inflation is used when correlated errors are not accounted for; here we have the opposite effect with smaller variances for uncorrelated $\mathbf{R}_{diag}$. In terms of the minimisation of the 1D-Var objective function, this means that E_{diag} is pulling much closer to observations for those channels than any of the correlated experiments. This makes it harder to find a solution, resulting in slower convergence.

We now consider how the condition number of the Hessian of the 1D-Var cost function, $\kappa(\mathbf{S})$, changes with the experiment E. From theoretical results developed in Tabeart et al. [2018], in particular the result of Corollary 1, we expect $\kappa(\mathbf{S})$ to decrease as $\lambda_{min}(\mathbf{R})$ increases. The minimum eigenvalues for each choice of OEC matrix, $\mathbf{R}$, discussed here can be seen in Table 2. The condition number of $\mathbf{S}$ is computed separately for each objective function. We can therefore consider the maximum, mean and median value of $\kappa(\mathbf{S})$ over the 97330 observations for each experiment. This information is shown in Table 3. As discussed in Section 2.1 the condition number of any matrix is bounded below by one. We therefore do not include the minimum values of $\kappa(\mathbf{S})$ in the table. We firstly note that the maximum values of $\kappa(\mathbf{S})$ are extremely large, with the largest value occurring for the matrix $\mathbf{R}_{diag}$. For experiments with correlated OEC matrices, increasing $\lambda_{min}(\mathbf{R})$ results in a decrease in the maximum value of $\kappa(\mathbf{R})$. We note that the changes to the condition number for $\mathbf{R}_{67}$ compared to $\mathbf{R}_{500}$ are much larger than the difference in conditioning between other experiments. The maximum value of $\kappa(\mathbf{R})$ for the OEC matrix $\mathbf{R}_{infl}$ is the smallest of all choices of OEC matrix. A decrease in the maximum value of $\kappa(\mathbf{S})$ corresponds to a distribution that has increased weight at the lower end of the spectrum for the iteration count distribution shown in Figure 3.

We now consider the mean and the median of $\kappa(\mathbf{S})$. Firstly we note that the values of the mean and median differ by at least one order of magnitude. The distribution of $\kappa(\mathbf{S})$ is not symmetric: it is bounded below by 1, with very large maximum values. The mean is skewed by such outliers and we note that for a boxplot of this data (not shown) the mean does not lie within the interquartile range (IQR) of the data for all experiments other than E_{67} and E_{infl}. Both the maximum and mean of $\kappa(\mathbf{S})$ decrease with increasing $\lambda_{min}(\mathbf{R})$, for correlated OEC matrices. The largest values occur for the experiment E_{diag}, and the smallest for the experiment E_{infl}. In contrast, the median is largest for the experiment E_{infl}, and decreasing $\lambda_{min}(\mathbf{R})$ increases the median value of $\kappa(\mathbf{S})$ for experiments with correlated choices of OEC matrix. Considering the deciles indicates that the spread of $\kappa(\mathbf{S})$ across all observations reduces as more reconditioning is applied.

We have seen that introducing correlated OEC matrices improves convergence and reduces $\kappa(\mathbf{S})$ compared to the current operational choice. Additionally, reducing the target condition number results in further improvements. This behaviour agrees with the theoretical conclusions of Tabeart et al. [2018] that were summarised in Section 2.2. For a linear observation operator we expect the upper bound on the condition number of the Hessian to decrease as the minimum eigenvalue of the OEC matrix, $\mathbf{R}$, increases. This is shown in (4). This equation also shows that the ratio between background and observation variance is important for the conditioning of $\mathbf{S}$. The final column of Table 3 shows that range of $\kappa(\mathbf{S})$ for the experiment E_{infl} is less than the range for any experiment with a correlated choice of OEC matrix. The variances for $\mathbf{R}_{infl}$ are much larger than the variances for any other OEC matrix considered in this work. We therefore conclude that the qualitative conclusions of Tabeart et al. [2018], as presented in Section 2.2, hold in this framework, even in the case of a non-linear observation operator.

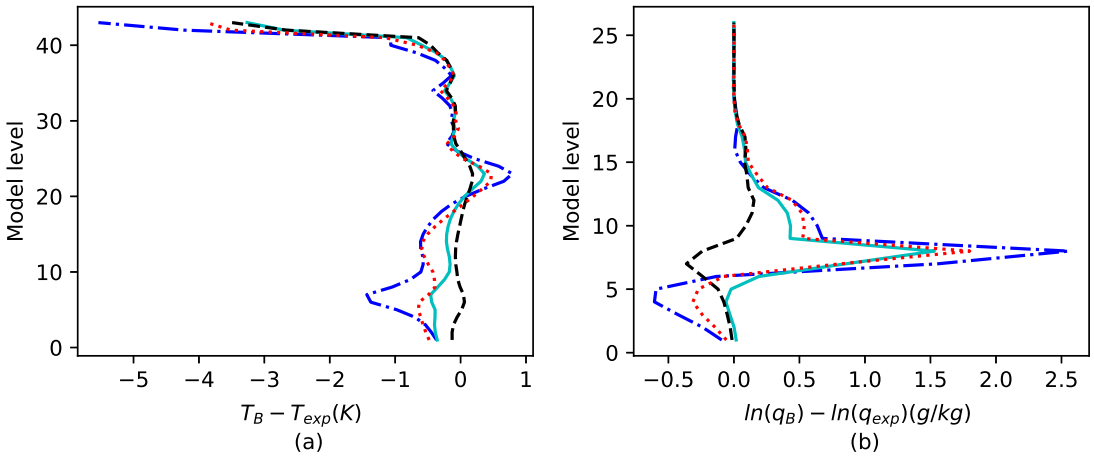

Figure 4: Background minus retrieved profiles from observation at (-33.16N,-32.70E) for 16th June 2016 0000UTC for (a) temperature (b) ln(specific humidity). Differences are shown for E_{diag} (dot-dashed line), E_{est} (dotted line), E_{67} (solid line) and E_{infl} (dashed line).

4.2 Effect of changing the observation error covariance matrix on 1D-Var Retrievals

In this section we consider how changing the OEC matrix impacts the retrieved values of physical variables. In particular we focus on temperature and specific humidity, as we obtain profiles that occur across multiple model levels rather than individual values. We note that the retrieved temperature and humidity values are not passed to the 4D-Var assimilation procedure. However, studying how these variables change for different choices of OEC matrix helps us understand the impact of changing the OEC matrix, **R**, on the 1D-Var assimilation. Additionally, as part of the 1D-Var assimilation procedure, retrieval error standard deviation (RESD) values for each of the retrieval values are derived. The RESD values are calculated as the square root of the diagonal entries of the inverse of the Hessian given by (2), i.e. the retrieved analysis error covariance in state variable space. For each 1D-Var assimilation we obtain a different value for RESD for each retrieved variable. We therefore consider the average RESD value for a given experiment and retrieved variable. For temperature and specific humidity this means that we obtain different RESD values for each model level. Comparing the range of differences between retrievals to the RESD values will allow us to determine whether the difference made when changing the OEC matrix, **R**, is of a similar order to expected variation, or much larger (and hence results in significant differences). We will also compare RESD and differences against BESD values as shown in Figure 1. We note that the Hessian only gives a good estimate of the retrieval error when the assumed and actual observation (and background) errors agree. For example, if the assumed observation errors were much smaller than the true errors, RESD values would underestimate the true analysis error standard deviations.

Figure 4 shows background profiles minus retrieved profiles for temperature and humidity for observations at the location (-33.16N,-32.70E). Retrievals are shown at pressure levels in the atmosphere. These model levels are determined by the 43 evenly distributed pressure levels in the radiative transfer retrieval algorithm. Specific humidity is only calculated for the lowest 26 model levels. We note that this is the configuration that was used at the time of the experiments (July 2016). Differences from the background are larger for specific humidity profiles than for temperature. However qualitative behaviour is similar for both variables. In both cases E_{diag} is the most different from the background, implying that the use of correlated observation errors increases the weighted importance of the background. Increasing the amount of reconditioning used decreases the norm of the difference between the retrieved profile and the background for all correlated OEC matrices. Hence, applying a larger amount of reconditioning results in a retrieved profile that is closer to the background. Finally, the retrieval corresponding to E_{infl} is closest to the background for both variables. For this case, standard deviations have been inflated, meaning that we expect the retrieved profile to fit closer to

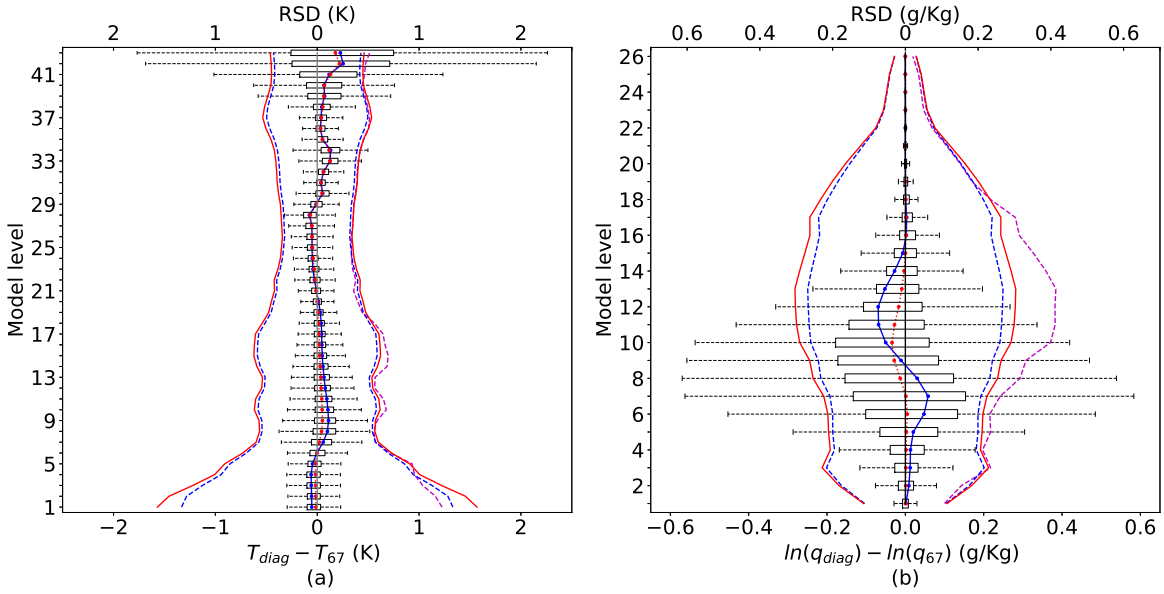

Figure 5: Differences in retrievals between E_{diag} and E_{67} for trial on 16th June 0000UTC for (a) temperature and (b) ln(specific humidity) for 97330 observations. Dashed lines and solid lines give the mean retrieval error standard deviations values for E_{diag} and E_{67} respectively. Dashed lines with dots denote the median and solid lines with dots denote the mean for each pressure level. The solid box contains the middle 50% of the data, and the whiskers (dashed horizontal lines) extend to the quartiles plus/minus 1.5 times the interquartile range - the difference between the third and first quartiles. Outliers, which lie outside the range of the whiskers, are not shown.

the background. This is particularly evident for specific humidity where there is a large difference between background and retrieved values for model level 8 (approximately 800hPa) for E_{diag}, E_{est} and E_{67}. This occurs due to large differences between background and retrieved brightness temperature for channels 128-137, which have water vapour mixing ratio Jacobians that peak at pressure level 8 [Stewart, 2010]. We recall that these channels are sensitive to water vapour, and have the strongest positive correlations in $\mathbf{R}_{est}$. As the correlations diagnosed using 1D-Var would be likely to have much weaker correlations for water vapour sensitive channels, it is probable that information from these channels is being under-weighted for correlated choices of $\mathbf{R}$, leading to over-fitting to background information. This explains why specific humidity is particularly affected by changes to the OEC matrix for this model level, although we note that noticeable changes also occur for temperature for this model level.

We now consider the differences between retrieved values for E_{diag} and E_{67} for all 97330 observations that were accepted by the 1D-Var routine for all choices of OEC matrix. Figures 5a and b are box plots showing the distribution of these differences across each model level for temperature and ln(specific humidity) profiles respectively. The qualitative behaviour for other experiments is very similar and is not shown here. Figure 5a shows that for most model levels the whiskers are contained within the average RESD values, and for model levels $1 - 41$ the central 50% of differences lie within the averaged RESD. This indicates that changing from an uncorrelated to correlated choice of OEC matrix has a generally small impact on temperatures for the majority of model levels compared to RESD.

Mean RESD values for E_{67} and E_{diag} are also very similar, with larger mean RESD values for E_{67} than E_{diag} for all model levels. This is observed for all correlated choices of OEC matrix; the mean RESD is increased for all model levels compared to E_{diag}. This suggests that using a correlated choice of OEC matrix increases the mean RESD for temperature i.e. by introducing correlations we have less confidence in the retrieved values, or 1D-Var analysis. This increase to standard deviations is expected from theoretical and idealised studies [Stewart et al., 2008, Rainwater et al., 2015, Fowler et al., 2018].

We also note that by including correlations we put less weight on the individual channels but allow more freedom to fit multivariate information arising from the combination of channels. Comparing the RESD values to the BESD values given by Figure 1 we find that across all three choices of $\mathbf{B}$, the standard deviation values are similar to RESD values for most model levels. For model levels where the BESDs are small, differences between retrievals using $\mathbf{R}_{diag}$ and correlated choices of $\mathbf{R}$ are small. As the background error is relatively small, the observations will be given low weight and hence any changes to the assumed observation errors will have a small effect on the resulting retrieval.

Figure 5b shows the differences between retrieved values of specific humidity for E_{diag} and E_{67} for 23 model levels. As was the case for temperature, the mean RESD values for all other choices of experiment are larger than those for E_{diag}. We note that differences for model levels 1 and $18-23$ are very small compared to RESD. However, for model levels $5-12$, the whiskers lie outside the values for mean RESD. This means there is a large proportion of model levels where changing the OEC matrix has a larger impact on retrieved specific humidity values than we would expect due to instrument noise and other quantified types of uncertainty. We also note that for these model levels we have non-zero and non-equal means and medians. This suggests that the distribution of differences is not symmetric. Again, BESD values are larger than RESD values for the majority of model levels for specific humidity. However, whiskers still extend past the BESD values for levels 5 - 10 for all choices of $\mathbf{B}$.

The effect of changing the OEC matrix, $\mathbf{R}$, seems to affect a larger proportion of the retrieved specific humidity values than temperature values. This coincides with the findings of Bormann et al. [2016], Weston et al. [2014]. They found large changes to humidity fields with the introduction of correlated OEC matrices in 4D-Var assimilation procedures, which resulted in improved NWP skill scores.

5 Impact on variables that influence 4D-Var routine

In Section 4 we showed that the choice of OEC matrix, $\mathbf{R}$, does make a difference to the 1D-Var routine in terms of convergence, and the individual retrieval values. We now consider variables that directly impact the main 4D-Var procedure that is used to initialise forecasts. Changes to the OEC matrix in the 1D-Var routine affect 4D-Var in two main ways: firstly by altering the observations that are accepted by the quality control procedure, and secondly via retrieved values of variables that are not analysed in the 4D-Var state vector. We will consider these two aspects in turn.

5.1 Changes to the quality control procedure

In Section 4.1 we showed that increasing $\lambda_{min}(\mathbf{R})$ increases the speed of convergence of the 1D-Var routine. We now investigate whether changing the OEC matrix, $\mathbf{R}$, alters the number of observations that pass the quality control step that was described in Section 3.1. We also consider how the number of observations accepted by each experiment (respectively E_{diag}) and rejected by E_{diag} (respectively each experiment) changes for different choices of OEC matrix. This information is presented in Table 4.

We begin by considering in more detail why changing the OEC matrix would result in changes to the number of observations that pass quality control. Observations are rejected if the minimisation of the 1D-Var procedure requires more than 10 iterations to converge. In Section 4.1 we found that introducing correlated observation error reduces the number of iterations required for convergence, and that decreasing the target condition number increases convergence speed further. This suggests that introducing correlated OEC matrices and using reconditioning will result in a larger number of observations that converge fast enough to pass this aspect of quality control. We therefore expect the use of reconditioning methods to result in a larger number of accepted observations.

The first row of Table 4 shows that the number of accepted observations increases as $\lambda_{min}(\mathbf{R})$ (see Table 2) increases and the largest number of accepted observations occurs for experiment E_{infl}. This coincides with what we would expect due to alterations in the quality control procedure. However, we note that the number of accepted observations is slightly larger for E_{diag} than E_{est} even though convergence for E_{est} was faster than for E_{diag} across the set of common observations. The second

Table 4: Number of observations accepted by the 1D-Var quality control for each experiment (E_{exp}) compared to E_{diag}. For E_{diag} the total number of accepted observations is 100686. Here T refers to the total number of distinct observations (defined in Section 3.1) accepted by E_{exp} for each experiment. The number of observations accepted by all experiments is 97330.

	E_{exp}					
Experiment	E_{est}	E_{1500}	E_{1000}	E_{500}	E_{67}	E_{infl}
No. of accepted obs (T)	100655	100795	101002	101341	102333	102859
No of obs accepted by both E_{diag} and E_{exp}	99039	99175	99352	99656	100382	100679
Accepted by E_{exp}, rejected by E_{diag}	1616	1620	1650	1685	1951	2180
Accepted by E_{diag}, rejected by E_{exp}	1647	1511	1334	1030	304	7

row of Table 4 shows that most observations are accepted by both E_{exp} and E_{diag}. We see that the number of accepted observations increases with $\lambda_{min}(\mathbf{R})$ for correlated choices of $\mathbf{R}$. The largest number of observations is accepted by E_{infl}. The third and fourth rows of Table 4 shows the number of observations that are accepted by each experiment and rejected by R_{diag}, and vice versa. However, this number is smaller than 2.2% of the total number of observations for all choices of E_{exp}. For what follows we shall consider the large majority of observations that are accepted by both E_{diag} and E_{exp}. Although observations that are accepted by only one of E_{diag} and E_{exp} are of interest, the fact that there are very few observations in either of these sets makes it hard to study their properties statistically.

5.2 Changes to retrieved values for variables that are not included in the 4D-Var control vector

In this Section we consider how altering the OEC matrix used in the 1D-Var routine alters the retrieved values of variables that are not included in the 4D-Var control vector. For all three variables, Figure 6 shows that the majority of retrievals are changed by a small amount for each choice of experiment. The largest differences occur between E_{diag} and E_{infl} for ST, CF and CTP, where the IQR and whiskers are much larger than for any correlated choice of OEC matrix. For correlated OEC matrices, we see a reduction in IQR and whisker length as $\lambda_{min}(\mathbf{R})$ increases. This indicates that as we increase the amount of reconditioning that is applied, the differences between E_{diag} and E_{exp} reduce. However, there are some differences between the variables.

Firstly, for ST all choices of OEC matrix yield whiskers that are equal to or exceed the RESD values corresponding to E_{diag} (solid line), and all except E_{67} exceed the RESD values for the corresponding experiment (dashed line). In contrast, the whiskers for correlated choices of OEC matrix are well within both RESD values for CF and CTP, as well as the BESD values given in Table 1. This shows that compared to expected observation variability, differences between CF and CTP retrievals are small for correlated choices of OEC matrix. However, we recall that BESD values for cloud variables were artificially inflated [Pavelin et al., 2008].

For ST and CTP the values of the mean and median are close for all correlated choices of E_{exp}, and the box and whiskers are fairly symmetric about 0. In contrast, for CF, differences between the mean and median occur, and the box extends further into the positive axis. Cloud errors are expected to vary greatly with the cloud state, meaning that it is difficult to interpret gross statistics [Eyre, 1989]. We include them here for completeness.

For all variables, the majority of retrievals change by a small amount, relative to RESD, when comparing the experiment to E_{diag}. However, Table 5 shows that over 15% of observations are classed as outliers for all three variables. These outliers, defined as observations with retrieval differences that are not between $Q_1 - 1.5IQR$ and $Q_3 + 1.5IQR$, where Q_1 and Q_3 denote the first and third quartiles of the data respectively, are not shown in Figure 6. Not all of these outliers represent large differences between retrieved values. Instead we consider 'large' outliers, which we define in this setting as differences larger than 25% of the maximum differences for each variable. For cloud variables the maximum difference is defined by the possible range of values: ± 1 and $\pm 913.25hPa$ for CF and CTP respectively. For ST we use the maximum difference between retrievals from the data set. These

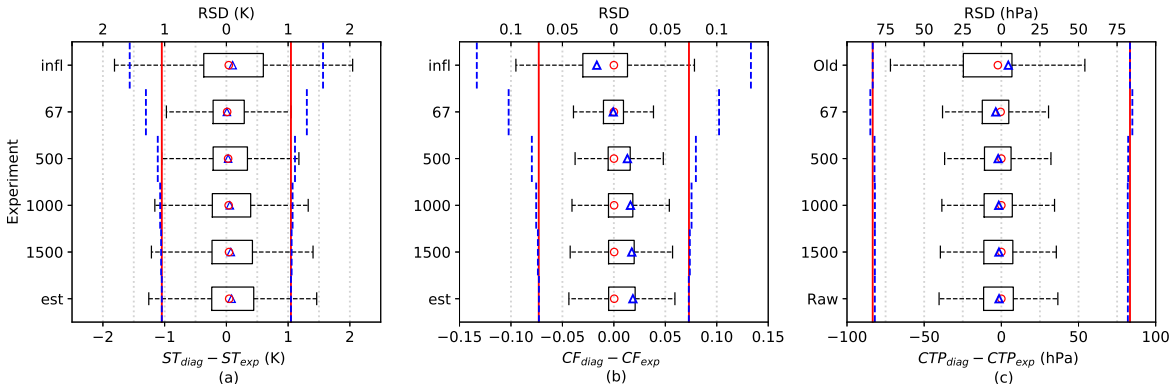

Figure 6: Box plot showing differences between retrieved variables for $E_{diag} - E_{exp}$ for (a) ST (skin temperature) (b) CF (cloud fraction) and (c) CTP (cloud top pressure). The circle shows the median, the triangle depicts the mean, the solid box contains the central 50% of data (the interquartile range), and the dashed horizontal lines show the whiskers which extend to the quartile $\pm 1.5 \times IQR$. Vertical dashed lines show the mean retrieval error standard deviation (RESD) values for the experiment, and the solid vertical lines shows the mean RESD values for E_{diag}. Outliers (not shown) lie in the range (a) $\pm 33.52K$, (b) ± 1 and (c) $\pm 913.25hPa$. The number of outliers and extreme outliers for these experiments is presented in Tables 4 and 5

Table 5: Percentage of outliers for cloud fraction (CF), cloud top pressure (CTP) and skin temperature (ST). Outliers are differences which fall outside the whiskers shown in Figure 6.Maximum and minimum differences are shown for skin temperature only; maximum differences for cloud fraction and cloud top pressure are ± 1 and $\pm 913.25hPa$ respectively, for all choices of $\mathbf{R}$.

	E_{est}	E_{1500}	E_{1000}	E_{500}	E_{67}	E_{infl}
% outliers (ST)	15.1	15.3	15.6	16.3	17.6	15.9
% of outliers (CF)	23.9	24.02	24.2	24.6	25.3	21.4
% outliers (CTP)	22.8	22.8	23.0	22.9	21.4	18.8
Maximum difference (ST (K))	21.67	21.12	21.14	22.38	21.03	26.83
Minimum difference (ST (K))	-33.52	-33.01	-32.14	-29.76	-23.82	-20.88

values are given in Table 5.

Table 6 shows the percentage of large outliers for each variable, which is much smaller than the total number of outliers for all variables and experiments. For all variables, the number of large outliers decreases with $\lambda_{min}(\mathbf{R})$ for correlated experiments. The experiment E_{infl} has a much greater number of large outliers than any experiment with a correlated choice of OEC matrix, agreeing with earlier findings that the qualitative and quantitative differences between E_{infl} and E_{diag} are much larger than for any other experiment.

As background information has almost no weight for cloud variables, due to inflated BESD values, changing the OEC matrix could result in much larger differences between retrieved values for CF and CTP than for other variables. However, this is not the case for ST, where the maximum differences given in Table 5 are extremely large compared to RESD and BESD values. The number of observations with extremely large retrieval differences is small: for correlated experiments fewer than 10 observations yield absolute differences larger than 20K. These observations can be considered as failures of the 1D-Var algorithm and should be removed by the quality control procedure. This emphasises that when altering the OEC matrix, the quality control procedure needs to be altered as well.

Previous studies by Stewart et al. [2014], Weston et al. [2014], Bormann et al. [2016], Campbell et al. [2017] have shown that the largest impacts of applying the DBCP diagnostic to IASI occur for humidity sounding channels, which will affect clouds and retrieved values associated with clouds. Skin temperature is also sensitive to cloud; although in partly overcast conditions it is possible to retrieve

Table 6: Number of large outliers for cloud fraction (CF), cloud top pressure (CTP) and skin temperature (ST) for each experiment. Large outliers are defined as observations with absolute differences greater than 0.25 for CF, $225hPa$ for CTP and $5K$ for ST. This corresponds to absolute differences greater than approximately 25% of the maximum differences presented in Table 5

	E_{est}	E_{1500}	E_{1000}	E_{500}	E_{67}	E_{infl}
% extreme outliers ($\|ST\| > 5K$)	1.6	1.5	1.5	1.4	1.4	3.6
% extreme outliers ($\|CF\| > 0.25$)	4.9	4.7	4.4	3.9	3.2	7.5
% extreme outliers ($\|CTP\| > 225hPa$)	3.3	3.3	3.3	3.3	2.7	4.4

estimates of skin temperature, errors in the modelling of cloud effects are likely to dominate the surface signal [Stewart et al., 2014, Pavelin and Candy, 2014]. In terms of impact, under cloudy conditions the 4D-Var assimilation procedure is less sensitive to skin temperature [Pavelin and Candy, 2014], so it is possible that these large changes to retrievals will not result in large impacts when passed to 4D-Var. However, further work is needed to understand the origin and consequences of these extreme differences fully.

6 Conclusions

It is widely known that many observing systems in numerical weather prediction (NWP) have errors that are correlated [Janjić et al., 2018] for reasons including scale mismatch between observation and model resolution, approximations in the observation operator or correlations introduced by preprocessing. However, diagnosed error covariance matrices have been found to be extremely ill-conditioned, and cause convergence problems when used in existing NWP computer systems [Campbell et al., 2017, Weston et al., 2014]. Tabeart et al. [2018] established that increasing the minimum eigenvalue of the OEC matrix improves bounds on the conditioning of the associated linear variational data assimilation problem. This provided insights into possible reconditioning methods which could permit the inclusion of correlation information while ensuring computational efficiency [Tabeart et al., 2019].

In this paper we have investigated the impact of changing the OEC matrix for the IASI instrument in the Met Office 1D-Var system, an operational non-linear assimilation system. In particular we have considered how reconditioning methods could permit the implementation of correlated observation error matrices. The 1D-Var system is used for quality control purposes and to retrieve values of variables that are not included in the 4D-Var state vector. As each observation is assimilated individually, it is more straightforward to understand and isolate the effects of using different choices of OEC matrix on retrieved variables and convergence compared to the more complicated 4D-Var procedure.

We found that:

- The current operational choice of observation error covariance (OEC) matrix for IASI results in the slowest convergence of the 1D-Var routine of all OEC matrices considered. Increasing the amount of reconditioning applied to correlated OEC matrices improves convergence of the 1D-Var routine, in accordance with the qualitative theoretical conclusions of Tabeart et al. [2018, 2019].

- Most experimental choices of correlated OEC matrix resulted in a larger number of IASI observations that were accepted by the 1D-Var routine than the current diagonal operational choice. Increasing the amount of reconditioning applied to correlated OEC matrices increases the number of IASI observations that converge in fewer than 10 iterations, and hence pass the quality control component of 1D-Var.

- Retrieval differences for skin temperature, cloud fraction and cloud top pressure are smaller than retrieved standard deviation values for over 75% of IASI observations for all choices of correlated OEC matrix. Up to 5% of retrievals have large differences relative to the retrieved standard deviation.

- As the minimum eigenvalue of the OEC matrix is increased, the difference between E_{diag} (using the current operational diagonal OEC matrix) and experimental retrieved values reduces.

We also find that for most variables studied RESD values are of a similar size to BESD values. We note that the BESD values for cloud variables are artificially inflated, and are hence an order of magnitude larger than the corresponding RESD values. This indicates that observation information has as large or a larger weight in the 1D-Var objective function than background profiles. However, as the OEC matrix used for these experiments was estimated using the results of a 4D-Var assimilation algorithm, artificially strong correlations between water vapour channels may lead to information from these channels being under-weighted in the analysis.

The qualitative conclusions from this work agree with the theoretical results of Tabeart et al. [2018], which prove that for a linear observation operator, increasing the minimum eigenvalue of the OEC matrix is important in terms of convergence of a variational data assimilation routine.

We emphasise that these convergence results contradict the common assumption that the use of correlated OEC matrices in a variational data assimilation scheme will always cause convergence problems. We find that decreasing variances negatively affects convergence, meaning that previous convergence issues were likely due to the large difference in variances between inflated diagonal matrices and correlated OEC matrices. However, as the current operational diagonal OEC matrix is not inflated, moving to a correlated choice of OEC matrix yields an increase in convergence speed, particularly when combined with reconditioning methods. At the Met Office the 1D-Var routine is run every 6 hours for the global model so reducing the cost of the routine would save significant computational effort. Additionally, the faster convergence that is achieved by correlated choices of OEC could permit stricter convergence criteria, e.g. reducing the maximum number of iterations from 10 to 8, which would also result in computational savings. However, care needs to be taken to consider how this will interact with other aspects of the quality control procedure and ensure that 'good' observations are not rejected. We also note that our results are presented in terms of iteration count rather than computational processing time. As we have considered a single 6 hour assimilation period, making general conclusions about changes to computational time may be misleading. A longer study would be needed to evaluate changes to wall clock time for different choices of OEC matrix. One aspect of the system that should be considered is whether the cost of inverting a full matrix will dominate the 1D-Var assimilation time. Weston [2011] found that using a correlated OEC matrix in 1D-Var resulted in a large increase to wall-clock time. However, the specific OEC matrix being used was extremely ill-conditioned. We expect that reconditioning will mitigate the impact on computation time as well as the number of iterations required for convergence of the 1D-Var assimilation, in a similar manner to the improvements seen for 4D-Var [Weston, 2011, Tabeart et al., 2019].

Changes to OEC matrices also alter the quality control aspect of the 1D-Var procedure, so care needs to be taken to ensure that these changes to the system are well understood. In particular, reducing the number of iterations required for convergence of the 1D-Var routine means that a larger number of observations were accepted by our tests and passed to the 4D-Var routine. For observations that were accepted by all experiments, we considered changes to retrieved estimates for skin temperature, cloud top pressure and cloud fraction. Although changes to the retrieved values with different OEC matrices were small for the majority of observations, for a small percentage of observations, the differences between retrieved values were very large. As ST, CTP and CF are not estimated as part of the 4D-Var procedure, such large changes may have significant effects on the analysis for 4D-Var. The most extreme of these differences (particularly for ST) are unrealistic and can be viewed as 1D-Var failures. This highlights that changes to the 1D-Var system, such as with the introduction of correlated OEC matrices, must be made in conjunction with tuning of the quality control procedures. Furthermore, we have not studied whether the use of correlated OEC matrices and reconditioning techniques are beneficial or detrimental in terms of the quality of 4D-Var analyses or forecasts. Improvements to convergence will need to be be balanced with impacts on other aspects of the assimilation system, such as changes to quality control, analysis fit and forecast skill.

acknowledgements

This work is funded in part by the EPSRC Centre for Doctoral Training in Mathematics of Planet Earth, the NERC Flooding from Intense Rainfall programme (NE/K008900/1), the EPSRC DARE project (EP/P002331/1) and the NERC National Centre for Earth Observation.

References

Erik Andersson, Mike Fisher, Rosemary Munro, and Anthony McNally. Diagnosis of background errors for radiances and other observable quantities in a variational data assimilation scheme, and the explanation of a case of poor convergence. *Quarterly Journal of the Royal Meteorological Society*, 126(565):1455–1472, 2000.

K. Bathmann. Justification for estimating observation-error covariances with the Desroziers diagnostic. *Quarterly Journal of the Royal Meteorological Society*, 144(715):1965–1974, 2018. doi: 10.1002/qj.3395. URL `https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3395`.

G. V. Bennitt, H. R. Johnson, P. P. Weston, J. Jones, and E. Pottiaux. An assessment of ground-based GNSS Zenith Total Delay observation errors and their correlations using the Met Office UKV model. *Quarterly Journal of the Royal Meteorological Society*, 2017. ISSN 1477-870X. doi: 10.1002/qj.3097. URL `http://dx.doi.org/10.1002/qj.3097`.

D. S. Bernstein. *Matrix mathematics : theory, facts, and formulas*. Princeton University Press, Princeton, N.J. ; Oxford, 2nd ed. edition, 2009. ISBN 9780691132877.

N. Bormann, A. J. Geer, and P. Bauer. Estimates of observation-error characteristics in clear and cloudy regions for microwave imager radiances from numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 137:2014–2023, 2011.

N. Bormann, M. Bonavita, R. Dragani, R. Eresmaa, M. Matricardi, and A. McNally. Enhancing the impact of IASI observations through an updated observation error covariance matrix. *Quarterly Journal of the Royal Meteorological Society*, 142(697):1767–1780, 2016.

W. F. Campbell, E. A. Satterfield, B. Ruston, and N. L. Baker. Accounting for correlated observation error in a dual-formulation 4D variational data assimilation system. *Monthly Weather Review*, 145(3):1019–1032, 2017. doi: 10.1175/MWR-D-16-0240.1. URL `https://doi.org/10.1175/MWR-D-16-0240.1`.

G. Chalon, F. Cayla, and D. Diebel. IASI: An advanced sounder for operational meteorology. In *Proceedings of IAF, Toulouse, France*, 2001.

A. D. Collard. Selection of iasi channels for use in numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 133(629):1977–1991, 2007. doi: 10.1002/qj.178. URL `https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.178`.

A. D. Collard, A. P. McNally, F. I. Hilton, S. B. Healy, and N. C. Atkinson. The use of principal component analysis for the assimilation of high-resolution infrared sounder observations for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 136(653):2038–2050, 2010. doi: 10.1002/qj.701.

M. Cordoba, S. L. Dance, G. A. Kelly, N. K. Nichols, and J. A. Waller. Diagnosing Atmospheric Motion Vector observation errors for an operational high resolution data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 2016. doi: 10.1002/qj.2925.

G. Desroziers, L. Berre, B. Chapnik, and P. Poli. Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society*, 131: 3385–3396, 2005.

J. R. Eyre. Inversion of cloudy satellite sounding radiances by nonlinear optimal estimation. I: Theory and simulation for TOVS. *Quarterly Journal of the Royal Meteorological Society*, 115(489):1001–1026, 1989. doi: 10.1002/qj.49711548902.

J. R. Eyre and F. I. Hilton. Sensitivity of analysis error covariance to the mis-specification of background error covariance. *Quarterly Journal of the Royal Meteorological Society*, 139(671):524–533, 2013.

A. M. Fowler, S. L. Dance, and J. A. Waller. On the interaction of observation and prior error correlations in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 144(710): 48–62, 2018. doi: 10.1002/qj.3183.

P. Gauthier, P. Du, S. Heilliette, and L. Garand. Convergence issues in the estimation of interchannel correlated observation errors in infrared radiance data. *Monthly Weather Review*, 146(10):3227–3239, 2018. doi: 10.1175/MWR-D-17-0273.1.

G. H. Golub and C. F. Van Loan. *Matrix Computations*. The John Hopkins University Press, third edition, 1996.

O. Guillet, A. T. Weaver, X. Vasseur, Y. Michel, S. Gratton, and S. Grol. Modelling spatially correlated observation errors in variational data assimilation using a diffusion operator on an unstructured mesh. *Quarterly Journal of the Royal Meteorological Society*, 145(722):1947–1967, 2019. doi: 10.1002/qj.3537.

S. A. Haben. *Conditioning and preconditioning of the minimisation problem in variational data assimilation*. PhD thesis, Department of Mathematics and Statistics, University of Reading, 2011.

S. A. Haben, A. S. Lawless, and N. K. Nichols. Conditioning of incremental variational data assimilation, with application to the Met Office system. *Tellus A*, 64(4):782–792, 2011.

S. B. Healy and A. A. White. Use of discrete Fourier transforms in the 1D-Var retrieval problem. *Quarterly Journal of the Royal Meteorological Society*, 131(605):63–72, 2005.

F. Hilton, N. C. Atkinson, S. J. English, and J. R. Eyre. Assimilation of IASI at the Met Office and assessment of its impact through observing system experiments. *Quarterly Journal of the Royal Meteorological Society*, 135:495–505, 2009.

T. Janjić, N. Bormann, M. Bocquet, J. A Carton, S. E. Cohn, S. L. Dance, S. N. Losa, N. K. Nichols, R. Potthast, J. A. Waller, and P. Weston. On the representation error in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 144(713):1257–1278, 2018.

C. Lupu, C. Cardinali, and A. P. McNally. Adjoint-based forecast sensitivity applied to observation-error variance tuning. *Quarterly Journal of the Royal Meteorological Society*, 141:3157–3165, 2015.

R. Ménard. Error covariance estimation methods based on analysis residuals: theoretical foundation and convergence properties derived from simplified observation networks. *Quarterly Journal of the Royal Meteorological Society*, 142:257–273, 2016.

Y. Michel. Revisiting Fisher's approach to the handling of horizontal spatial correlations of observation errors in a variational framework. *Quarterly Journal of the Royal Meteorological Society*, 144(716): 2011–2025, 2018.

J. Nocedal. *Numerical optimization*. Springer series in operations research and financial engineering. Springer, New York ; London, 2nd ed. edition, 2006. ISBN 9780387303031.

E. G. Pavelin and B. Candy. Assimilation of surface-sensitive infrared radiances over land: Estimation of land surface temperature and emissivity. *Quarterly Journal of the Royal Meteorological Society*, 140(681):1198–1208, 2014. doi: 10.1002/qj.2218.

E. G. Pavelin, S. J. English, and J. R. Eyre. The assimilation of cloud-affected infrared satellite radiances for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 134(632):737–749, 2008. doi: 10.1002/qj.243.

C. Prates, S. Migliorini, L. Stewart, and J. Eyre. Assimilation of transformed retrievals obtained from clear-sky IASI measurements. *Quarterly Journal of the Royal Meteorological Society*, 142: 1697–1712, 2016.

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

D. Simonin, J. A. Waller, S. P. Ballard, S. L. Dance, and N. K. Nichols. A pragmatic strategy for implementing spatially correlated observation errors in an operational system: an application to Doppler radar winds. *Quarterly Journal of the Royal Meteorological Society*, 2019. doi: https://doi.org/10.1002/qj.3592.

L. M. Stewart. *Correlated observation errors in data assimilation*. PhD thesis, University of Reading, 2010.

L. M. Stewart, S. L. Dance, and N. K. Nichols. Correlated observation errors in data assimilation. *Int. J. Numer. Meth.*, 56:1521–1527, 2008.

L. M. Stewart, J. Cameron, S. L. Dance, S. English, J. Eyre, and N. K. Nichols. Observation error correlations in IASI radiance data. Mathematics report series, 1/2009, University of Reading, Reading, UK, 2009. URL `https://www.reading.ac.uk/web/files/maths/obs_error_IASI_radiance.pdf`.

L. M. Stewart, S. L. Dance, and N. K. Nichols. Data assimilation with correlated observation errors: experiments with a 1-D shallow water model. *Tellus A*, 65:19546 (14pp), 2013. URL `http://dx.doi.org/10.3402/tellusa.v65i0.19546`.

L. M. Stewart, S. L. Dance, N. K. Nichols, J. R. Eyre, and J. Cameron. Estimating interchannel observation-error correlations of IASI radiance data in the Met Office system. *Quarterly Journal of the Royal Meteorological Society*, 140:1236–1244, 2014.

J. M. Tabeart, S. L. Dance, S. A. Haben, A. S. Lawless, N. K. Nichols, and J. A. Waller. The conditioning of least squares problems in variational data assimilation. *Numerical Linear Algebra with Applications*, doi:10.1002/nla.2165 (22pp), 2018. URL `http://dx.doi.org/10.1002/nla.2165`.

J. M. Tabeart, S. L. Dance, A. S. Lawless, N. K. Nichols, and J. A. Waller. Improving the conditioning of estimated covariance matrices. 2019. In press, Tellus A.

J. A. Waller, S. L. Dance, A. S. Lawless, N. K. Nichols, and J. R. Eyre. Representativity error for temperature and humidity using the Met Office high-resolution model. *Q.J.R. Meteorol. Soc*, 140: 1189–1197, 2014.

J. A. Waller, S. P. Ballard, S. L. Dance, G. Kelly, N. K. Nichols, and D. Simonin. Diagnosing Horizontal and Inter-Channel Observation Error Correlations for SEVIRI Observations Using Observation-Minus-Background and Observation-Minus-Analysis Statistics. *Remote Sensing*, 8 (7):851, 2016a. URL `10.3390/rs8070581`.

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

J. A. Waller, D. Simonin, S. L. Dance, N. K. Nichols, and S. P. Ballard. Diagnosing Observation Error Correlations for Doppler Radar Radial Winds in the Met Office UKV Model Using Observation-Minus-Background and Observation-Minus-Analysis Statistics. *Monthly Weather Review*, 144(10): 3533–3551, 2016c.

J. A. Waller, S. L. Dance, and N. K. Nichols. On diagnosing observation-error statistics with local ensemble data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 143(708):2677–2686, 2017. doi: 10.1002/qj.3117.

T. Wang, J. Fei, X. Cheng, X. Huang, and J. Zhong. Estimating the correlated observation-error characteristics of the chinese fengyun microwave temperature sounder and microwave humidity sounder. *Advances in Atmospheric Sciences*, 35(11):1428–1441, 2018. ISSN 1861-9533. doi: 10.1007/s00376-018-8014-9.

P. Weston. Progress towards the implementation of correlated observation errors in 4D-Var. Forecasting research technical report 560, Met Office, Exeter, UK, 2011.

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