University of Reading

# The conditioning of least squares problems in variational data assimilation

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Tabeart, J. M., Dance, S. L. ORCID: https://orcid.org/0000-0003-1690-3338, Haben, S. A., Lawless, A. S. ORCID: https://orcid.org/0000-0002-3016-6568, Nichols, N. K. ORCID: https://orcid.org/0000-0003-1133-5220 and Waller, J. A. (2018) The conditioning of least squares problems in variational data assimilation. Numerical Linear Algebra with Applications, 25 (5). e2165. ISSN 1099-1506 doi: 10.1002/nla.2165 Available at https://centaur.reading.ac.uk/74981/

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/nla.2165

Publisher: John Wiley and Sons

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

University of Reading

**CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

Received: 1 August 2017 | Revised: 22 December 2017 | Accepted: 15 January 2018

DOI: 10.1002/nla.2165

RESEARCH ARTICLE

WILEY

# The conditioning of least-squares problems in variational data assimilation

Jemima M. Tabeart[1,3] | Sarah L. Dance[1,2] | Stephen A. Haben[1,4] | Amos S. Lawless[1,2,3] | Nancy K. Nichols[1,2,3] | Joanne A. Waller[2]

[1]Department of Mathematics and Statistics, University of Reading, UK

[2]Department of Meteorology, University of Reading, UK

[3]NCEO, Reading, UK

[4]Mathematical Institute, University of Oxford, UK

**Correspondence**
Jemima M. Tabeart, Department of Mathematics and Statistics, University of Reading, PO Box 220, Reading RG6 6AX, UK.
Email: jemima.tabeart@pgr.reading.ac.uk

**Funding information**
UK Engineering and Physical Sciences Research Council Centre for Doctoral Training in Mathematics of Planet Earth; UK Natural Environmental Sciences Research Council Flooding from Intense Rainfall programme, Grant/Award Number: NE/K008900/1; EPSRC DARE, Grant/Award Number: EP/P002331/1 and NERC National Centre for Earth Observation

**Summary**

In variational data assimilation a least-squares objective function is minimised to obtain the most likely state of a dynamical system. This objective function combines observation and prior (or background) data weighted by their respective error statistics. In numerical weather prediction, data assimilation is used to estimate the current atmospheric state, which then serves as an initial condition for a forecast. New developments in the treatment of observation uncertainties have recently been shown to cause convergence problems for this least-squares minimisation. This is important for operational numerical weather prediction centres due to the time constraints of producing regular forecasts. The condition number of the Hessian of the objective function can be used as a proxy to investigate the speed of convergence of the least-squares minimisation. In this paper we develop novel theoretical bounds on the condition number of the Hessian. These new bounds depend on the minimum eigenvalue of the observation error covariance matrix and the ratio of background error variance to observation error variance. Numerical tests in a linear setting show that the location of observation measurements has an important effect on the condition number of the Hessian. We identify that the conditioning of the problem is related to the complex interactions between observation error covariance and background error covariance matrices. Increased understanding of the role of each constituent matrix in the conditioning of the Hessian will prove useful for informing the choice of correlated observation error covariance matrix and observation location, particularly for practical applications.

**KEYWORDS**

condition number, correlated observation errors, data assimilation, Hessian, least squares

## 1 | INTRODUCTION

Data assimilation combines the output from a numerical model of a dynamical system, the background or prior, with observations of the system to yield an accurate description of the current dynamical state (analysis).

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2018 The Authors. *Numerical Linear Algebra With Applications* Published by John Wiley & Sons Ltd.

Contributions from observations and the background are weighted according to their relative uncertainty via error covariance matrices, meaning that assessing and quantifying observation error are crucial in order to obtain an accurate analysis sufficiently quickly.[1,2] One of the most well-known applications of data assimilation is to numerical weather prediction (NWP), where observations of the atmosphere and ocean are combined with a prior model state of the atmosphere in order to produce the initial conditions for a weather forecast. Until recently, diagonal observation error covariance matrices have been used operationally at all major NWP centres,[3] a choice that is only valid in the case that observation errors are uncorrelated. It has been shown that implementing diagonal error covariance matrices inappropriately, that is, when error correlations are nonzero, may lead to suboptimal results.[4–8] However, using diagnosed full observation error covariance matrices directly in the assimilation has been shown to cause problems with the speed of convergence of the assimilation scheme.[9]

Variational assimilation, a popular data assimilation method,[10–12] finds the analysis by minimising a nonlinear least-squares objective function. This objective function, which is dependent on both observations and the background field, is minimised by an iterative method, such as the Gauss–Newton method.[13,14] This consists of an outer loop that solves the full nonlinear problem, and an inner loop that solves the linearised problem, often via a conjugate gradient method.[15] The conditioning of the Hessian matrix of the objective function provides a bound on the rate of convergence of the conjugate gradient minimisation.[16–18] Hence, it can be used as a rough estimate for the number of iterations needed to solve the inner loop problem. We note, however, that this worst-case bound on convergence can be improved significantly in the case of clustered eigenvalues.[17,19] The magnitude of the condition number also provides an indication of the sensitivity of the system to perturbations in the data.[10] Speed of convergence is critical in practise due to the need to provide timely forecasts. In this work, we investigate how introducing correlated observation errors affects the condition number of the Hessian and examine the associated speed of convergence of a conjugate gradient method.

Correlated observation error statistics have been diagnosed for certain observation types (e.g., see other works[20–27]), although there are problems associated with their use. In particular, the methods used to diagnose observation error covariance matrices are imperfect, and the quality of these estimates is unclear. Due to unknown observation error statistics and in order to reduce the computational cost of operational assimilation, in practise, the majority of observation errors are assumed uncorrelated. However, empirical evidence from simple model experiments indicate that even approximate correlation structures give significant benefit in terms of analysis accuracy.[7,28] Similar conclusions can be drawn for practical implementations.[4]

In the works of Stewart[6] and Stewart et al.,[26] it was shown that there were problems with the use of diagonal observation error covariance matrices in the variational data assimilation for certain instruments. Motivated by this work, in 2011, the UK Met Office first trialled the use of correlated observation errors in their operational system.[3] However, there were problems with the convergence of the minimisation algorithm, which necessitated the "reconditioning" of observation error covariance matrices (by altering their eigenvalues), prior to their use in the system. In the works of Weston[3] and Weston et al.,[9] it was suggested that slow convergence was caused by the very small minimum eigenvalues of the diagnosed observation error covariance matrix. This work provides motivation to investigate further the role of the minimum eigenvalue of the observation error covariance matrix on the conditioning of the variational data assimilation problem; in turn, developing this crucial understanding will permit the optimal use of correlated observation errors in data assimilation systems.

Even in the case of uncorrelated observation errors, the minimisation problem for any large system is very ill conditioned. Preconditioning, where the original problem is transformed into an equivalent but less ill-conditioned problem, is used operationally to mitigate against the slow convergence of the minimisation.[29] In data assimilation, the most common method of preconditioning is the control variable transform,[16,30] where the preconditioner is based on the background error covariance matrix. The optimal choice of preconditioning depends on the formulation of the data assimilation problem,[31] and practical constraints may require the use of a less computationally intensive preconditioner.[32] In this work, an unpreconditioned framework will be used, as it is unknown whether the introduction of correlated observation errors will alter the optimal choice of preconditioner. This framework also has practical relevance, as the UK Met Office uses an unpreconditioned 1D-Var routine, where each observation is assimilated individually, for quality control purposes. Hence, the bounds and conclusions presented here will apply directly to that case.

In this article, we develop a new theory for bounding the condition number of the Hessian of the least-squares objective function. This theory applies to both uncorrelated and correlated choices of observation error. We investigate the impact of introducing these correlations via small-scale numerical tests, which illustrate the influence of observation correlations associated with a physical length scale. We begin in Section 2 by defining a notation common to data assimilation and the condition number. We explain why the conditioning of the system and the rate of convergence of the minimisation

are linked and present results from linear algebra that will be used to construct the bounds discussed in Section 3. Three new sets of bounds will be introduced in Section 3; these will have a varying number of additional constraints on the constituent matrices. Bounds that separate the contribution of each of the constituent terms have been developed for both general matrices and matrices with additional assumptions on observation location and observation error correlations. In Section 4, we discuss our numerical framework for the experiments of Section 5. The results of these numerical tests support the theoretical conclusions presented in Section 3. In particular, we see that the minimum eigenvalue of the observation error covariance matrix and the ratio of background variance to observation variance are important terms for controlling the conditioning of the variational problem for both the bounds in Section 3 and the numerical results from Section 5. We conclude in Section 6 that even in a simple linear setting, the choice of observation operator has a significant effect on the conditioning. The theoretical conclusions indicate how correlated error statistics in the observation and background can be expected to interact, and highlight areas where reconditioning and similar techniques could be used to reduce the increased computational cost associated with using correlated observation errors operationally. Although the primary motivation for the investigation of the impact of correlated observation errors arises from their application in meteorology, the theory and conclusions presented here are very general and apply to any other application of variational data assimilation such as in neuroscience[33,34] and ecology.[35,36]

## 2 | VARIATIONAL ASSIMILATION AND CONDITION NUMBER

### 2.1 | Notation

In data assimilation, information from observations, $\mathbf{y} \in \mathbb{R}^p$, is combined with information from a background, or "prior", field, $\mathbf{x}_b \in \mathbb{R}^N$. The analysis, $\mathbf{x}_a \in \mathbb{R}^N$, or posterior, is found by weighting each of the two components using their respective error statistics. It is assumed that observation errors and background errors are unbiased and mutually uncorrelated. The background and observation error covariance matrices are denoted by the symmetric positive semidefinite matrices $\mathbf{B} \in \mathbb{R}^{N\times N}$ and $\mathbf{R} \in \mathbb{R}^{p\times p}$, respectively (although in practise, we assume $\mathbf{B}$ and $\mathbf{R}$ are positive definite matrices). Usually, there are far fewer observations than state variables, that is, $p \ll N$. Observation and background information may describe different variables or be situated at different locations in space. The observation operator $h : \mathbb{R}^N \to \mathbb{R}^p$, which may be nonlinear, is used to map from state space to observation space to allow the comparison of observations with the background; in particular, $\mathbf{y}$ will be compared with $h[\mathbf{x}]$.

For variational assimilation methods, the analysis is found by minimising an objective function. In this work, we focus on 3D-Var, a particular variational assimilation method, which assimilates variables at a single fixed time in the assimilation window over the entire spatial domain.[37] In the case of 3D-Var, the objective function is given by the following:

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

The state vector $\mathbf{x}_a$ that minimises this objective function is then used as the initial condition to produce a forecast. When $h$ is linear, this equation has an analytic solution (see equation 2.4 in the work of Haben[16]), but (1) is too expensive to be solved explicitly on an operational scale. In NWP, where observation operators can be nonlinear and high dimensional, a gradient descent algorithm, such as the Gauss–Newton method, is used to solve a sequence of linearised problems, in order to converge iteratively to the solution, $\mathbf{x}_a$.[10] We note that $\mathbf{x}_a$ corresponds to the maximum a posteriori estimate under the assumption that all probability distributions are Gaussian.[38,39]

### 2.2 | Condition number

In practice, to solve the nonlinear problem, the Gauss–Newton method is used to solve a sequence of linearised problems, often via a conjugate gradient method.[29] We will now consider the linearised problem, where the nonlinear problem given by (1) is linearised about $\mathbf{x}_a$, the optimal solution.

As the linearisation of (1) is a quadratic function,[37] finding $\mathbf{x}_a$ is equivalent to solving a linear system of the form

$$\mathbf{S}\mathbf{w} = \mathbf{b}, \tag{2}$$

where $\mathbf{w} \in \mathbb{R}^N$ and $\mathbf{b} \in \mathbb{R}^N$ are given by (3.10) in Section 3.2 in the work of Haben.[16] (This formulation will be used in numerical experiments in Section 5.) Here, $\mathbf{S} \in \mathbb{R}^{N\times N}$ is the Hessian of the linearisation of the objective function (1)

given by

$$\mathbf{S} = \mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}, \tag{3}$$

where $\mathbf{H} \in \mathbb{R}^{p\times N}$ is the Jacobian of the observation operator $h$ linearised about the optimal state. The Hessian can be used to study the sensitivity of the solution to small changes in observation or background data, by considering its condition number (see Section 2.7 in the work of Golub et al.[18]). As $\mathbf{B}$ and $\mathbf{R}$ are symmetric positive definite, $\mathbf{S}$ is also symmetric positive definite, and hence, the $L_2$ condition number of $\mathbf{S}$ can be represented in terms of its eigenvalues.

## 2.3 | Eigenvalue theory

For the remainder of the paper, the following ordering of eigenvalues of matrix $\mathbf{D}$ will be used: For a matrix $\mathbf{D} \in \mathbb{R}^{N\times N}$, let $\lambda_{\max}(\mathbf{D}) = \lambda_1(\mathbf{D}) \geq \lambda_2(\mathbf{D}) \geq \cdots \geq \lambda_N(\mathbf{D}) = \lambda_{\min}(\mathbf{D})$.

**Theorem 1.** *If $\mathbf{S} \in \mathbb{R}^{N\times N}$ is a symmetric and positive definite matrix, then we can write the condition number in the $L_2$ norm as*

$$\kappa_2(\mathbf{S}) = \frac{\lambda_1(\mathbf{S})}{\lambda_N(\mathbf{S})}, \tag{4}$$

*where $\lambda_1(\mathbf{S})$ and $\lambda_N(\mathbf{S})$ correspond to the largest and smallest eigenvalues of $\mathbf{S}$, respectively.*

*Proof.* (See Section 2.7.2 in the work of Golub et al.[18]) □

Henceforth, $\kappa_2(\mathbf{S})$ will be referred to as the condition number of $\mathbf{S}$ and will be denoted $\kappa(\mathbf{S})$.

In order to determine the bounds on the condition number of the Hessian we make use of the following result from linear algebra.

**Theorem 2.** *Consider two symmetric matrices $\mathbf{S}_1, \mathbf{S}_2 \in \mathbb{R}^{N\times N}$. The kth eigenvalue of the matrix sum $\mathbf{S}_1 + \mathbf{S}_2$ satisfies the following:*

$$\lambda_k(\mathbf{S}_1) + \lambda_N(\mathbf{S}_2) \leq \lambda_k(\mathbf{S}_1 + \mathbf{S}_2) \leq \lambda_k(\mathbf{S}_1) + \lambda_1(\mathbf{S}_2). \tag{5}$$

*Proof.* (See Chapter 2, Theorem 44 in the work of Wilkinson.[40]) □

This result allows us to separate the contributions of $\mathbf{B}^{-1}$ and $\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$ when bounding the condition number of $\mathbf{S}$ given by (3) and is discussed in Section 3.

A result bounding the eigenvalues of matrix products in terms of the eigenvalues of the constituent matrices is given by the following.

**Theorem 3.** *If $\mathbf{F}, \mathbf{G} \in \mathbb{C}^{N\times N}$ are positive semidefinite Hermitian matrices, then*

$$\prod_{i=1}^{k}\lambda_i(\mathbf{FG}) \leq \prod_{i=1}^{k}\lambda_i(\mathbf{F})\lambda_i(\mathbf{G}), \quad k = 1, \ldots, N-1. \tag{6}$$

*Proof.* (See Section 9 H.1.a. in the work of Marshall et al.[41]) □

**Theorem 4.** *If $\mathbf{F}, \mathbf{G} \in \mathbb{C}^{N\times N}$ are positive semidefinite Hermitian and $1 \leq i_1 < \cdots < i_k \leq N$, then*

$$\prod_{t=1}^{k}\lambda_t(\mathbf{FG}) \geq \prod_{t=1}^{k}\lambda_{i_t}(\mathbf{F})\lambda_{N-i_t+1}(\mathbf{G}), \tag{7}$$

*with equality for $k = N$.*

*Proof.* (See the work of Wang et al.[42]) □

## 3 | THEORETICAL RESULTS

We now present new bounds on the condition number of the Hessian given by (3). We begin in Section 3.1 by considering the general case: $\mathbf{B}$ and $\mathbf{R}$ are general covariance matrices, and $\mathbf{H}$ is any linear observation operator. In Section 3.2, we then introduce further assumptions that constrain $\mathbf{H}$ to only observe state variables. Finally, in Section 3.3, we restrict the form of $\mathbf{B}$ and $\mathbf{R}$ to have a particular structure.

### 3.1 | General bounds on the condition number

We begin by introducing bounds on the eigenvalues of $\mathbf{S}$ in terms of the eigenvalues of $\mathbf{B}$, $\mathbf{R}$, and $\mathbf{H}$.

**Lemma 1.** *For $\mathbf{S} = \mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}$, where $\mathbf{B} \in \mathbb{R}^{N\times N}$, $\mathbf{R} \in \mathbb{R}^{p\times p}$ are symmetric positive definite covariance matrices, and $\mathbf{H} \in \mathbb{R}^{p\times N}$ with $p < N$, we can bound the eigenvalues of $\mathbf{S}$ below by*

$$\lambda_k(\mathbf{S}) \geq \max\left\{\lambda_k(\mathbf{B}^{-1}), \quad \lambda_N(\mathbf{B}^{-1}) + \lambda_k(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})\right\} \tag{8}$$

*and above by*

$$\lambda_k(\mathbf{S}) \leq \min\left\{\lambda_k(\mathbf{B}^{-1}) + \lambda_1(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}), \quad \lambda_1(\mathbf{B}^{-1}) + \lambda_k(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})\right\}, \tag{9}$$

*where $\lambda_k(\mathbf{S})$ is the kth eigenvalue of $\mathbf{S}$.*

*Proof.* The bounds follow immediately from the result of Theorem 2 by exchanging the order of addition. Note that $\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$ is not full rank, meaning that $\lambda_N(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}) = 0$. □

As we wish to bound the condition number of $\mathbf{S}$, we are primarily interested in bounding $\lambda_1(\mathbf{S})$ and $\lambda_N(\mathbf{S})$. In this case, the bounds given by (8) and (9) then simplify to

$$\lambda_N(\mathbf{B}^{-1}) \leq \lambda_N(\mathbf{S}) \leq \min\left\{\lambda_N(\mathbf{B}^{-1}) + \lambda_1(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}), \quad \lambda_1(\mathbf{B}^{-1})\right\} \tag{10}$$

and

$$\max\left\{\lambda_1(\mathbf{B}^{-1}), \quad \lambda_N(\mathbf{B}^{-1}) + \lambda_1(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})\right\} \leq \lambda_1(\mathbf{S}) \leq \lambda_1(\mathbf{B}^{-1}) + \lambda_1(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}). \tag{11}$$

We note that this applies to any choice of correlation matrices $\mathbf{B}$ and $\mathbf{R}$ and for any linear choice of observation operator $\mathbf{H}$. This suggests that we expect the eigenvalues, and hence condition number, of $\mathbf{S}$ to vary based on the interactions between $\mathbf{B}$ and $\mathbf{R}$. We now introduce a new bound on the condition number of (3) for 3D-Var for the most general choice of $\mathbf{B}$, $\mathbf{R}$, and $\mathbf{H}$.

**Theorem 5.** *Let the background and observation error covariance matrices, $\mathbf{B} \in \mathbb{R}^{N\times N}$ and $\mathbf{R} \in \mathbb{R}^{p\times p}$, respectively, be symmetric positive definite covariance matrices, with $p < N$. Additionally, let $\mathbf{H} \in \mathbb{R}^{p\times N}$ be the observation operator. Then, the following bounds are satisfied by the condition number of the Hessian (given by (3)):*

$$\begin{aligned}\max\left\{\frac{1+\lambda_1(\mathbf{B})\lambda_1(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})}{\kappa(\mathbf{B})}, \quad \frac{\kappa(\mathbf{B})}{1+\lambda_1(\mathbf{B})\lambda_1(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})}\right\} &\leq \kappa(\mathbf{S}) \\ &\leq \left(1+\lambda_N(\mathbf{B})\lambda_1(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})\right)\kappa(\mathbf{B}).\end{aligned} \tag{12}$$

*(This is a slightly modified form of Theorem* 6.1.1 *in the work of Haben.[16])*

*Proof.* To obtain an upper bound for the condition number of (3), we take the upper bound for $\lambda_1(\mathbf{S})$ in (11) and the lower bound (10) for $\lambda_N(\mathbf{S})$,

$$\kappa(\mathbf{S}) \leq \left(1+\lambda_N(\mathbf{B})\lambda_1(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})\right)\kappa(\mathbf{B}), \tag{13}$$

using the fact that $(\lambda_1(\mathbf{B}^{-1}))^{-1} = \lambda_N(\mathbf{B})$. We can obtain a lower bound for the condition number similarly by taking the lower bound for $\lambda_1(\mathbf{S})$ in (11) and the upper bound for $\lambda_N(\mathbf{S})$ in (10). This gives two possible bounds for $\kappa(\mathbf{S})$, depending on which of the two terms is larger,

$$\kappa(\mathbf{S}) \geq \max\left\{\kappa(\mathbf{B})\left(1+\lambda_1(\mathbf{B})\lambda_1\left(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)\right)^{-1}, (\kappa(\mathbf{B}))^{-1}\left(1+\lambda_1(\mathbf{B})\lambda_1\left(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)\right)\right\} \tag{14}$$

using the fact that $(\lambda_1(\mathbf{B}))^{-1} = \lambda_N(\mathbf{B}^{-1})$. Combining these inequalities completes the proof. □

We note that the two terms in (14) are reciprocals. This means that the lower bound will always be greater than or equal to one. Any condition number is bounded below by one.[18]

We now extend this result to write it in a form that explicitly separates the role of the observation error covariance matrices and the observation operator. This makes it easier to investigate how changes in $\mathbf{R}$, $\mathbf{B}$, and $\mathbf{H}$ affect the condition number of the Hessian.

**Corollary 1.** *Let $\mathbf{B} \in \mathbb{R}^{N\times N}$ and $\mathbf{R} \in \mathbb{R}^{p\times p}$, with $p < N$, be the background and observation error covariance matrices, respectively. Additionally, let $\mathbf{H} \in \mathbb{R}^{p\times N}$ be the observation operator. Then, the following bounds are satisfied by the condition number of the Hessian (given by (3)):*

$$\begin{aligned}\max\left\{\frac{1+\frac{\lambda_1(\mathbf{B})}{\lambda_N(\mathbf{R})}\lambda_N(\mathbf{H}\mathbf{H}^T)}{\kappa(\mathbf{B})},\quad \frac{1+\frac{\lambda_1(\mathbf{B})}{\lambda_1(\mathbf{R})}\lambda_1(\mathbf{H}\mathbf{H}^T)}{\kappa(\mathbf{B})},\quad \frac{\kappa(\mathbf{B})}{1+\frac{\lambda_1(\mathbf{B})}{\lambda_N(\mathbf{R})}\lambda_1(\mathbf{H}\mathbf{H}^T)}\right\} &\leq \kappa(\mathbf{S})\\ &\leq \left(1+\frac{\lambda_N(\mathbf{B})}{\lambda_N(\mathbf{R})}\lambda_1(\mathbf{H}\mathbf{H}^T)\right)\kappa(\mathbf{B}).\end{aligned} \tag{15}$$

*Proof.* Using Theorem 21.10.1 in the work of Harville,[43] we see that $\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$ has precisely the same nonzero eigenvalues as $\mathbf{R}^{-1}\mathbf{H}\mathbf{H}^T$. Applying the same result, $\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$ also has the same nonzero eigenvalues as $\mathbf{H}\mathbf{H}^T\mathbf{R}^{-1}$. Therefore, $\lambda_1(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}) = \lambda_1(\mathbf{R}^{-1}\mathbf{H}\mathbf{H}^T) = \lambda_1(\mathbf{H}\mathbf{H}^T\mathbf{R}^{-1})$. Applying Theorem 3 for $k = 1$ and $i = 1$ yields the following bound:

$$\lambda_1\left(\mathbf{R}^{-1}\mathbf{H}\mathbf{H}^T\right) \leq \lambda_1(\mathbf{R}^{-1})\lambda_1(\mathbf{H}\mathbf{H}^T) = \frac{\lambda_1(\mathbf{H}\mathbf{H}^T)}{\lambda_N(\mathbf{R})}, \tag{16}$$

as $\lambda_1(\mathbf{R}^{-1}) = 1/\lambda_N(\mathbf{R})$. To bound $\lambda_1(\mathbf{R}^{-1}\mathbf{H}\mathbf{H}^T)$ below, we apply Theorem 4 for $k = 1$ and $i_1 = 1$ to obtain two lower bounds, as follows:

$$\lambda_1\left(\mathbf{R}^{-1}\mathbf{H}\mathbf{H}^T\right) \geq \max\{\lambda_1(\mathbf{R}^{-1})\lambda_N(\mathbf{H}\mathbf{H}^T), \lambda_N(\mathbf{R}^{-1})\lambda_1(\mathbf{H}\mathbf{H}^T)\} = \max\left\{\frac{\lambda_1(\mathbf{H}\mathbf{H}^T)}{\lambda_1(\mathbf{R})}, \frac{\lambda_N(\mathbf{H}\mathbf{H}^T)}{\lambda_N(\mathbf{R})}\right\}. \tag{17}$$

Substituting (16) and (17) into the upper and lower bounds of Theorem 5 gives the desired result. □

We note that the upper bound in (15) increases as $\lambda_N(\mathbf{R})$ decreases. It is not immediately clear how the lower bound will change with $\mathbf{R}$. This will be discussed in Section 4.3, which provides a summary of how the bounds given by (15) vary with $\mathbf{R}$ and $\mathbf{B}$ for the numerical framework tested in Section 5.

## 3.2 | Bounds on the condition number with additional restrictions on the choice of observation operator

We now develop a further bound, which applies in the case that additional assumptions are made regarding the choice of observation operator. In particular, we restrict the observation operator to direct observations of single state variables. We note that if observations are restricted to direct observations of single state variables, then $\mathbf{H}^T\mathbf{H}$ is diagonal with $(\mathbf{H}^T\mathbf{H})_{i,i} = 1$ if variable $i$ is observed and zero otherwise, as shown by Haben et al.[44] Under this stricter assumption, we show that the value of $\lambda_1(\mathbf{H}\mathbf{H}^T)$ is the same, irrespective of the choice of observations.

**Lemma 2.** *If $\mathbf{H}^T\mathbf{H} \in \mathbb{R}^{N\times N}$ is a diagonal matrix with $p < N$ units on the diagonal and the remaining elements zero, then $\mathbf{H}\mathbf{H}^T$ is the $p \times p$ identity matrix.*

*Proof.* As $\mathbf{H}^T\mathbf{H}$ is diagonal, we can calculate its eigenvalues directly; they are simply its diagonal elements. Hence, $\mathbf{H}^T\mathbf{H}$ has $p$ unit eigenvalues and $N - p$ zero eigenvalues. By Theorem 21.10.1 in the work of Harville,[43] $\mathbf{H}\mathbf{H}^T$ has the same nonzero eigenvalues as $\mathbf{H}^T\mathbf{H}$, that is, $p$ units. As $\mathbf{H}\mathbf{H}^T$ is symmetric, these eigenvalues correspond to $p$ linearly independent eigenvectors. We now write $\mathbf{H}\mathbf{H}^T$ in terms of its eigendecomposition. Let $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_N) \in \mathbb{R}^{p\times p}$ be the matrix of eigenvalues of $\mathbf{H}\mathbf{H}^T$, and $\mathbf{V} \in \mathbb{R}^{p\times p}$ be the corresponding matrix of eigenvectors of $\mathbf{H}\mathbf{H}^T$. As the eigenvalues of $\mathbf{H}\mathbf{H}^T$ are all units, $\mathbf{\Lambda} = \mathbf{I}_p$, the $p \times p$ identity. Then,

$$\mathbf{H}\mathbf{H}^T = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1} = \mathbf{V}\mathbf{I}_p\mathbf{V}^{-1} = \mathbf{V}\mathbf{V}^{-1} = \mathbf{I}_p. \tag{18}$$

Hence, under the assumptions on $\mathbf{H}^T\mathbf{H}$, $\mathbf{H}\mathbf{H}^T$ is the $p \times p$ identity matrix. □

Hence, if observations are restricted to single state variables, then $\mathbf{H}\mathbf{H}^T = \mathbf{I}_p$. Eliminating the $\mathbf{H}$ and $\mathbf{H}^T$ terms from the bound given by Corollary 1 reduces the number of matrix multiplications required for evaluation. This result is now used to obtain a bound for the case where observation and background error covariances are correlated and where observations are limited to model variables. We additionally assume that observation variance, $\sigma_o^2$, and background variance, $\sigma_b^2$, are uniform variances, and hence, the covariance matrices can be written as a scalar variance multiplied by a correlation matrix.

**Corollary 2.** *Let $\mathbf{B} = \sigma_b^2\mathbf{C} \in \mathbb{R}^{N\times N}$ and $\mathbf{R} = \sigma_o^2\mathbf{D} \in \mathbb{R}^{p\times p}$, where $\mathbf{C}$ and $\mathbf{D}$ are symmetric positive definite correlation matrices, and $\sigma_b^2$ and $\sigma_o^2$ are positive scalars denoting the background and observation error variances, respectively. In addition, let $\mathbf{H}^T\mathbf{H}$ be a diagonal matrix with $p < N$ units on the diagonal and the remaining elements zero. Then, the following bound on the condition number of $\boldsymbol{S}$ (given by* (3)*) holds:*

$$\max\left\{\frac{1+\frac{\sigma_b^2}{\sigma_o^2}\frac{\lambda_1(\mathbf{C})}{\lambda_N(\mathbf{D})}}{\kappa(\mathbf{C})}, \frac{\kappa(\mathbf{C})}{1+\frac{\sigma_b^2}{\sigma_o^2}\frac{\lambda_1(\mathbf{C})}{\lambda_N(\mathbf{D})}}\right\} \leq \kappa(\mathbf{S}) \leq \left(1+\frac{\sigma_b^2}{\sigma_o^2}\frac{\lambda_N(\mathbf{C})}{\lambda_N(\mathbf{D})}\right)\kappa(\mathbf{C}). \tag{19}$$

*Proof.* Using (15) with the definitions of $\mathbf{B}$ and $\mathbf{R}$ in the theorem statement along with the result of Lemma 2 yields the desired result immediately. □

The bounds given by (19) are equal to those given by (15) for the case of direct observations, so the comments concerning how the bounds change with $\mathbf{R}$ and $\mathbf{B}$ following Corollary 1 also apply here. In general, it is not possible to comment on how the lower bound given by (19) will behave with changing $\mathbf{B}$ and $\mathbf{R}$. In Section 5, we provide an overview for how the terms in (19) change for some specific choices of $\mathbf{B}$, $\mathbf{R}$, and $\mathbf{H}$.

We note that the ratio $\frac{\sigma_b^2}{\sigma_o^2}$ appears in both bounds, meaning that as the observations get more accurate, and the variance $\sigma_o^2$ decreases, we will see an increased upper bound and a decreased lower bound for $\lambda_1(\mathbf{B}) > \lambda_N(\mathbf{B}) + \lambda_1(\mathbf{R})$, with an increased lower bound for $\lambda_1(\mathbf{B}) < \lambda_N(\mathbf{B}) + \lambda_1(\mathbf{R})$. This was also observed theoretically and numerically by Haben[16] for the case that $\mathbf{R}$ is uncorrelated. Both of these results assume the same variance for all observations, which is not true in general. However, they indicate the general behaviour we would expect for an increase in accuracy across a wide range of observing systems.

### 3.3 | Bounds on the condition number for circulant error covariance matrices

In this section, we present a lower bound that is tighter than those of (15) for a given matrix framework. Improved bounds are obtained for this specific case by exploiting the eigenvalue and eigenvector properties of a particular matrix structure. It is feasible that for other matrix structures, similar properties could be used to compute tighter bounds for other classes of matrices. However, as the results from Section 3.1 are general and apply to any choice of covariance matrices, we do not consider other specialised bounds in this work.

It is often desirable for error correlations to be homogeneous and isotropic, meaning that the correlation between two points is determined solely by the distance between them.[45] This makes circulant matrices a natural choice for correlation matrices on a one-dimensional periodic domain. For the numerical tests discussed in Section 5, both $\mathbf{B}$ and $\mathbf{R}$ will be chosen to be circulant matrices, although the bounds given by Theorem 5, Corollary 1, and Corollary 2 apply for any valid choice of correlation matrix.

**Definition 1.** (See the work of Davis.[46])
A circulant matrix $\mathbf{D} \in \mathbb{R}^{N\times N}$ is a matrix of the form

$$\mathbf{D} = \begin{pmatrix} d_0 & d_1 & d_2 & \cdots & d_{N-2} & d_{N-1} \\ d_{N-1} & d_0 & d_1 & \cdots & d_{N-3} & d_{N-2} \\ d_{N-2} & d_{N-1} & d_0 & \cdots & d_{N-4} & d_{N-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ d_2 & d_3 s & d_4 & \cdots & d_0 & d_1 \\ d_1 & d_2 & d_3 & \cdots & d_{N-1} & d_0 \end{pmatrix}.$$

As described by Gray,[47] the structure of a circulant matrix of the form given by Definition 1 permits the rapid calculation of eigenvalues and eigenvectors via a discrete Fourier transform. In practise, this means that we can calculate the eigenvalues of $\mathbf{D}$ directly via the following formula.

**Theorem 6.** *The eigenvalues of a circulant matrix* $\mathbf{D}$*, as given by Definition* 1*, are given by*

$$\gamma_m = \sum_{k=0}^{N-1} d_k \omega^{mk}, \tag{20}$$

with corresponding eigenvectors

$$\mathbf{v}_m = \frac{1}{\sqrt{N}}(1, \omega^m, \ldots, \omega^{m(N-1)}), \tag{21}$$

where $\omega = e^{-2\pi i/N}$ is an $N$th root of unity.

*Proof.* (See the work of Gray [47] for full derivation.) □

To avoid confusion, the eigenvalues of a circulant matrix calculated using (20) will be denoted by $\gamma_j$ rather than $\lambda_j$, as they are ordered in terms of wavenumber rather than size. We can see from (21) that the eigenvectors only depend on $N$, the dimension of the circulant matrix. Therefore, any $N \times N$ circulant matrix will have the same set of eigenvectors.

We now use this matrix structure to consider a further restriction to the case that observation error is assumed to be uncorrelated, and the background observation error matrix is required to be circulant. In particular, in the following theorem, $\mathbf{R}$ is taken to be a scalar multiple of the identity. We note that Theorem 7 was presented by Haben et al.[45] without proof.

**Theorem 7.** *Let* $\mathbf{B} = \sigma_b^2 \mathbf{C} \in \mathbb{R}^{N\times N}$*, where* $\boldsymbol{C}$ *is a symmetric positive definite circulant matrix, and* $\mathbf{R} = \sigma_o^2 \mathbf{I}_p$*, where* $\mathbf{I}_p \in \mathbb{R}^{p\times p}$ *is the identity matrix. Both* $\sigma_b^2$ *and* $\sigma_o^2$ *are positive scalars. In addition, let* $\mathbf{H}^{\mathrm{T}}\mathbf{H}$ *be a diagonal matrix with* $p < N$ *units on the diagonal and the remaining elements zero. Then the following bounds on the condition number of* $\mathbf{S}$ *(given by (3)) hold:*

$$\left(\frac{1 + \frac{p}{N}\frac{\sigma_b^2}{\sigma_o^2}\lambda_N(\mathbf{C})}{1 + \frac{p}{N}\frac{\sigma_b^2}{\sigma_o^2}\lambda_1(\mathbf{C})}\right)\kappa(\mathbf{C}) \leq \kappa(\mathbf{S}) \leq \left(1 + \frac{\sigma_b^2}{\sigma_o^2}\lambda_N(\mathbf{C})\right)\kappa(\mathbf{C}), \tag{22}$$

*where* $\lambda_1(\mathbf{C})$ *and* $\lambda_N(\mathbf{C})$ *are the largest and smallest eigenvalues of the matrix* $\mathbf{C}$*, respectively.*

*Proof.* By the assumptions on the matrices in the theorem, we can write $\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} = \sigma_o^{-2}\mathbf{H}^T\mathbf{H}$ and therefore $\lambda_1(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}) = \sigma_o^{-2}$. Additionally, we have $\lambda_N(\mathbf{B}) = \sigma_b^2\lambda_N(\mathbf{C})$. If we substitute these into the upper bound of (12), we obtain

$$\kappa(\mathbf{S}) \leq \left(1 + \frac{\sigma_b^2}{\sigma_o^2}\lambda_N(\mathbf{C})\right)\kappa(\mathbf{C}), \tag{23}$$

which establishes the upper bound. Rather than repeat this procedure with the lower bound, we produce an improved estimate by applying the Rayleigh quotient, $R_{\mathbf{S}}(\mathbf{x}), \mathbf{x} \in \mathbb{C}^N$ (defined in Section 5.9 in the work of Süli et al.[48]). Let $\mathbf{v}_1 \in \mathbb{C}^N$ be the eigenvector corresponding to the largest eigenvalue of $\mathbf{C}^{-1}$. Because $\mathbf{C}^{-1}$ is circulant, then all the components of the eigenvectors of $\mathbf{C}^{-1}$ lie on the unit circle in $\mathbb{C}$ (see (21)). In particular, this implies that for an eigenvector, $\mathbf{v}_m$, of $\mathbf{C}^{-1}$,

$$\mathbf{v}_m^\dagger \mathbf{H}^T\mathbf{H}\mathbf{v}_m = \frac{1}{N}\sum_{k\in K} \overline{e^{-2\pi ikm/N}} e^{-2\pi ikm/N} = \frac{1}{N}\sum_{k\in K} e^{2\pi ikm/N} e^{-2\pi ikm/N} = \frac{p}{N}, \tag{24}$$

where $K$ denotes the positions of the nonzero diagonal elements of $\mathbf{H}^T\mathbf{H}$, and $\mathbf{v}^\dagger$ denotes the conjugate transpose of $\mathbf{v}$. The maximum value obtained by the Rayleigh quotient of $\mathbf{S}$ occurs at the eigenvector corresponding to the largest eigenvalue of $\mathbf{S}$ (see Section 5.9 in the work of Süli et al.[48]). Hence,

$$\lambda_1(\mathbf{S}) = \max_{\mathbf{v}\in\mathbb{C}^N}(R_{\mathbf{S}}(\mathbf{v})) \geq \mathbf{v}_1^\dagger\left(\mathbf{B}^{-1} + \sigma_o^{-2}\mathbf{H}^T\mathbf{H}\right)\mathbf{v}_1 = \sigma_b^{-2}\lambda_1(\mathbf{C}^{-1}) + \sigma_o^{-2}\frac{p}{N}. \tag{25}$$

Similarly, the minimum value of the Rayleigh quotient occurs at the eigenvector corresponding to the smallest eigenvalue of $\mathbf{S}$. Let $\mathbf{v}_N$ be the eigenvector corresponding to the smallest eigenvalue of $\mathbf{C}^{-1}$. Then, again using the Rayleigh

quotient, we find

$$\lambda_N(\mathbf{S}) = \min_{\mathbf{v}\in\mathbb{C}^N}(R_{\mathbf{S}}(\mathbf{v})) \leq \mathbf{v}_N^{\dagger}\left(\mathbf{B}^{-1} + \sigma_o^{-2}\mathbf{H}^T\mathbf{H}\right)\mathbf{v}_N = \sigma_b^{-2}\lambda_N(\mathbf{C}^{-1}) + \sigma_o^{-2}\frac{p}{N}. \tag{26}$$

Combining (25) and (26), we find

$$\kappa(\mathbf{S}) \geq \frac{\sigma_b^{-2}\lambda_1(\mathbf{C}^{-1}) + \sigma_o^{-2}\frac{p}{N}}{\sigma_b^{-2}\lambda_N(\mathbf{C}^{-1}) + \sigma_o^{-2}\frac{p}{N}} = \kappa(\mathbf{C})\left(\frac{1 + \frac{\sigma_b^2}{\sigma_o^2}\frac{p}{N}\lambda_N(\mathbf{C})}{1 + \frac{\sigma_b^2}{\sigma_o^2}\frac{p}{N}\lambda_1(\mathbf{C})}\right), \tag{27}$$

giving the lower bound on the condition number. This completes the proof. □

We note that the lower bound presented here is tighter than the others introduced in this section. This comes from the restriction on the form of $\mathbf{S}$ when additional assumptions are made on $\mathbf{R}$ and $\mathbf{H}$ and does not generalise to the other results presented in this work. We also observe that the lower bound (22) has an explicit dependence on the number of observations, $p$, meaning that $\kappa(\mathbf{S})$ increases with $p$. This was studied in detail by Haben.[16] Additionally, the ratio $\frac{\sigma_b^2}{\sigma_o^2}$ appears in both bounds, meaning that the discussion following the result of Corollary 2 also applies to the result of Theorem 7.

We now have bounds that require minimal matrix multiplications for evaluation and that separate the contributions of $\mathbf{B}$, $\mathbf{R}$, and $\mathbf{H}$. In the following sections we will test these bounds numerically and discuss the impact of changing each of the constituent matrices in turn.

## 4 | NUMERICAL FRAMEWORK

We now outline the experimental framework that will be used in Section 5 to numerically investigate the bounds presented in Section 3. In particular, in Section 4.1, we introduce specific matrix structures that will be used to generate covariance matrices. These structures have been chosen as they permit easy calculation of eigenvalues of the resulting matrices. We note that these correlation structures illustrate the case where there is a physical length scale associated with our observation and background error correlations, as in the case of horizontal correlations. Different choices of observation operator will then be presented in Section 4.2. Finally, in Section 4.3, we define the experiments that will be studied in Section 5 and discuss the choice of parameters to be used in these tests in detail.

### 4.1 | Correlation and second-order auto-regressive correlation matrices

This work will make use of the second-order auto-regressive correlation (SOAR) function, which is used by the UK Met Office as a horizontal correlation function, as detailed in the work of Simonin et al.[49] It is also commonly used to model background error correlations,[7] as its relatively long tails coincide well with estimates of correlation structure. Additionally, these longer tails ensure that SOAR matrices are better conditioned for inversion than Gaussian matrices.[10,16]

The SOAR function, defined by Daley,[50] is homogeneous and isotropic and naturally extends to a circulant form when we have equally spaced observations on a periodic domain, such as a latitude circle on the Earth. We define the SOAR error correlation matrix for a 1D model with state variables (or observations) given by equally spaced grid-points on a fixed domain (the unit circle of radius $a = 1$), following the procedure given in the works of Haben[16] and Waller et al.[51] This makes use of a substitution of a chordal distance for a "great circle distance" to ensure that we obtain a valid correlation model on the circle, as discussed by Gaspari et al.[52] and Jeong et al.[53]

**Definition 2.** The SOAR error correlation matrix on the finite domain is given by

$$\mathbf{D}(i,j) = \left(1 + \frac{\left|2a\sin\left(\frac{\theta_{i,j}}{2}\right)\right|}{L}\right)\exp\left(\frac{-\left|2a\sin\left(\frac{\theta_{i,j}}{2}\right)\right|}{L}\right), \tag{28}$$

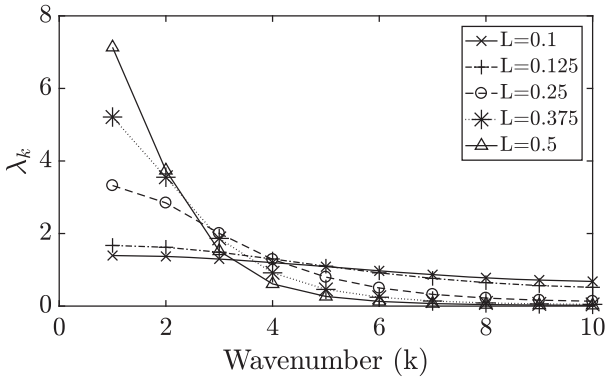


**FIGURE 1** Eigenvalues of an error correlation matrix defined by the second-order auto-regressive function given in (28) for $N = 20$ and $a = 1$

where $L > 0$ is the correlation length scale, $\theta_{i,j}$ denotes the angle between gridpoints $i$ and $j$, and $a$ is the radius of the domain. The chordal distance between adjacent gridpoints is given by

$$\Delta x = 2a \sin\left(\frac{\theta}{2}\right) = 2a \sin\left(\frac{\pi}{N}\right), \tag{29}$$

where $N$ is the number of gridpoints, and $\theta = \frac{\pi}{2N}$ is the angle between adjacent gridpoints.

As SOAR matrices are circulant by construction, we can calculate their eigenvalues directly using Equation 20. The distribution of eigenvalues is symmetric and, as shown in Figure 1, decreases monotonically towards the central value. This means that only two eigenvalues need to be calculated in order to obtain the maximum and minimum eigenvalues of any SOAR matrix: $\gamma_1$ and $\gamma_{N/2}$ (if $N$ is even) or $\gamma_{(N+1)/2}$ (if $N$ is odd). The circulant structure can hence be exploited to reduce the number of computations required for computing the bounds given by (15) and (19) for the condition number of the Hessian.

For the numerical experiments, we alter the length scales of the SOAR matrices corresponding to background and observation error. Figures will be plotted in terms of the maximum eigenvalues of $\mathbf{B}^{-1}$ and $\mathbf{R}^{-1}$ (recalling that for any matrix, $\mathbf{D} \in \mathbb{R}^{m\times m}$, $\lambda_1(\mathbf{D}^{-1}) = 1/\lambda_N(\mathbf{D})$). We note that this also means that $\lambda_1(\mathbf{D}^{-1}) = \gamma_{N/2}(\mathbf{D}^{-1})$ for $N$ even (or $\lambda_1(\mathbf{D}^{-1}) = \gamma_{(N+1)/2}(\mathbf{D}^{-1})$ for $N$ odd), using the notation established in Theorem 6. The relationship between the increasing length scale and the spectrum of a SOAR matrix is shown in Figure 1, namely that as the length scale, $L$, increases, the minimum eigenvalue of the SOAR matrix decreases and the maximum eigenvalue increases. This means that the maximum eigenvalue of the inverse of a SOAR matrix increases with length scale, and its minimum eigenvalue decreases.

Having described the choice of correlation matrices that will be used in the numerical tests in Section 5, in the next section we discuss the different choices of observation operator that will be tested in our experiments.

## 4.2 | Choice of observation operator

Most previous research into the impact of correlated observation errors on the variational assimilation problem does not investigate the impact of using different observation operators systematically. Either the operational observation operator is used (e.g., see other works[9,21]) or the experiments are carried out in a simple linear case where $\mathbf{H}$ is taken to be a variant of the identity, as in other works.[5,7,8,54] In this paper, we compare how the condition number of the Hessian is affected by different choices of linear observation operator in order to gain some theoretical insight into the role played by this operator. We define three choices of the observation operator that will be investigated in detail numerically. We are particularly interested in how important our choice of $\mathbf{H}$ is in determining both the true condition number of $\mathbf{S}$ and the value of the bounds given by (15). Firstly, we note that all bounds presented in this work require the assumption that the observation operator, $\mathbf{H}$, is linear, and the bounds given by (19) and (22) have the restriction that observations are only of single state variables. All the choices of $\mathbf{H}$ that are tested in the numerical experiments presented in this work are linear, and two correspond to direct observations of single model variables.

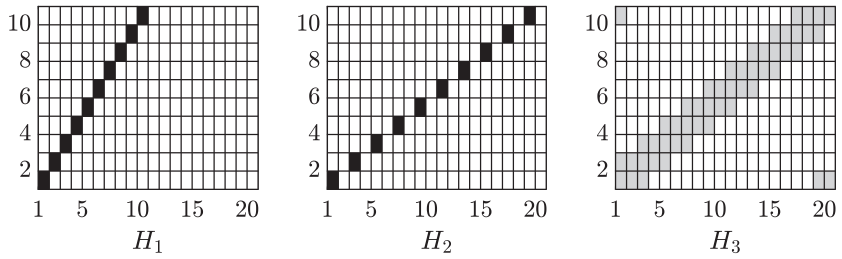


**FIGURE 2** Visualisation of the observation operators described in Definition 3 for the case $p = 10$ and $N = 20$. Shading indicates the value of the entry in the matrix; in the case of $\mathbf{H}_1$ and $\mathbf{H}_2$, all nonzero entries are 1, and for $\mathbf{H}_3$, all nonzero entries are $\frac{1}{5}$

**Definition 3.** The observation operators $\mathbf{H}_1, \mathbf{H}_2, \mathbf{H}_3 \in \mathbb{R}^{p\times N}$, for $N = 2p$, are defined as follows:

$$\mathbf{H}_1(i,j) = \begin{cases} 1, & j = i \text{ for } i = 1, \ldots, p \\ 0, & \text{otherwise.} \end{cases} \tag{30}$$

$$\mathbf{H}_2(i,j) = \begin{cases} 1, & j = 2i \text{ for } i = 1, \ldots, p \\ 0, & \text{otherwise.} \end{cases} \tag{31}$$

$$\mathbf{H}_3(i,j) = \begin{cases} \frac{1}{5}, & j \in \{2i-2, 2i-1, 2i, 2i+1, 2i+2 (\mathrm{mod} N)\} \text{ for } i = 1, \ldots, p \\ 0, & \text{otherwise.} \end{cases} \tag{32}$$

The choice of $\mathbf{H} = \mathbf{H}_1$ corresponds to observing the first $p$ state variables and to making no observations in the second half of the state space. Choosing $\mathbf{H} = \mathbf{H}_2$ corresponds to making observations at alternate state variables over the entire model domain. The observation operator $\mathbf{H} = \mathbf{H}_3$ is a smoothed version of $\mathbf{H}_2$; state variables at alternate gridpoints are smoothed over five adjacent points in state space with equal weighting. This can be thought of as a simplified version of a satellite weighting function (see Section 2.4.1 in the work of Stewart[6] and Section 2.1.3 in the work of Rodgers[39]), which measures average radiation over several model levels of the atmosphere. In Figure 2, these observation operators are depicted for a small-scale example when $p = 10$ and $N = 20$.

The choice of $\mathbf{H}_1$ was made as a check to allow the comparison of preliminary numerical tests with those from Chapter 6 in the work of Haben.[16] The bounds given by Corollary 2 in Section 3 require that $\mathbf{H}^T\mathbf{H}$ be a diagonal matrix with $p$ units on the diagonal. The observation operator $\mathbf{H}_1$ satisfies this requirement, as does $\mathbf{H}_2$, meaning that we can apply the bounds of Corollary 2 for these two cases. Additionally, by Lemma 2, $\mathbf{H}_1\mathbf{H}_1^T = \mathbf{H}_2\mathbf{H}_2^T$. This means that for fixed choices of $\mathbf{B}$ and $\mathbf{R}$, both $\mathbf{H} = \mathbf{H}_1$ and $\mathbf{H} = \mathbf{H}_2$ will yield the same upper and lower bounds. We wish to see whether there will be a significant difference in the true condition number of $\mathbf{S}$ for $\mathbf{H} = \mathbf{H}_1$ and $\mathbf{H} = \mathbf{H}_2$.

As $\mathbf{H} = \mathbf{H}_3$ does not satisfy the condition in the statement of Corollary 2, we must apply the more general bound given by (15) in Corollary 1. We would like to be able to use the same bounds to compare each of the three choices of observation operator. A short calculation reveals that we have equality of the bounds given by Corollary 1 and Corollary 2 when observations are restricted to model gridpoints for the framework described here. Hence, for what follows, we will be comparing the bounds given by (15) irrespective of the observation network chosen.

## 4.3 | Experimental design

We now discuss the experimental framework, which will be used for the numerical tests presented in Section 5. In particular, we motivate the range of parameters that will be investigated.

We fix the ratio between $p$, the number of observations, and $N$, the number of state variables, to be $N = 2p$ for all the experiments discussed below. The same ratio was used for numerical testing by Haben[16] and is not representative of what is used in practise, where observations are much less dense. Unless stated otherwise, the values $N = 200$ and $p = 100$ were used for all the plots presented here. Other choices of $p$ and $N$ were studied in detail; as qualitative results were similar for all cases considered, they will not be shown here.

Both the background error covariance matrix, $\mathbf{B} \in \mathbb{R}^{N\times N}$, and the observation error covariance matrix, $\mathbf{R} \in \mathbb{R}^{p\times p}$, are chosen to be SOAR correlation matrices (see Section 4.1) with fixed variances $\sigma_b^2 = \sigma_o^2 = 1$.

**TABLE 1** Summary of how terms that appear in (15) change with the length scales $L_B$ and $L_R$ for $\mathbf{B} \in \mathbb{R}^{200\times200}$ and $\mathbf{R} \in \mathbb{R}^{100\times100}$

| | Length scale $L_R$ or $L_B$ | | | | |
|---|---|---|---|---|---|
| | **0.1** | **0.33** | **0.66** | **0.99** | **1** |
| $\lambda_N(\mathbf{R})$ | $1.92\times10^{-2}$ | $5.74\times10^{-4}$ | $7.21\times10^{-5}$ | $2.14\times10^{-5}$ | $2.08\times10^{-5}$ |
| $\lambda_1(\mathbf{R})$ | $6.40\times10^{0}$ | $2.26\times10^{1}$ | $4.67\times10^{1}$ | $6.36\times10^{1}$ | $6.40\times10^{1}$ |
| $\lambda_N(\mathbf{B})$ | $2.54\times10^{-3}$ | $7.19\times10^{-5}$ | $8.99\times10^{-6}$ | $2.67\times10^{-6}$ | $2.59\times10^{-6}$ |
| $\lambda_1(\mathbf{B})$ | $1.28\times10^{1}$ | $4.51\times10^{1}$ | $9.35\times10^{1}$ | $1.27\times10^{2}$ | $1.28\times10^{2}$ |
| $\kappa(\mathbf{B})$ | $5.05\times10^{3}$ | $6.28\times10^{5}$ | $1.40\times10^{7}$ | $4.77\times10^{7}$ | $4.95\times10^{7}$ |

The domain for the tests is the unit circle ($a = 1$). In the experiments that follow, we will vary $L_R$, the correlation length scale of the SOAR matrix defining $\mathbf{R}$, and $L_B$, the correlation length scale of the the SOAR matrix defining $\mathbf{B}$, over a regular grid, but figures will be plotted in terms of $\lambda_1(\mathbf{B}^{-1})$ and $\lambda_1(\mathbf{R}^{-1})$. In addition to studying the impact of changing the length scale of $\mathbf{B}$ and $\mathbf{R}$ for both sets of experiments, we also consider the effect of using the different choices of $\mathbf{H}$ presented in Section 4.2.

### 4.3.1 | Condition number testing

In the numerical tests, we consider how the condition number of $\mathbf{S}$ (calculated using the Matlab 2016b function *cond*) and the bounds given by (15) change as the minimum eigenvalues of both error covariance matrices change. Of particular interest is the interaction between changes to both $\mathbf{B}$ and $\mathbf{R}$. For the results presented in this paper, the length scales of both $\mathbf{B}$ and $\mathbf{R}$ were varied between 0.1 and 1. The equivalent eigenvalues of $\mathbf{R}$ and $\mathbf{B}$ for these parameters are given in Table 1.

Table 1 presents the values of the terms that appear in (15) and depend on the background and observation error matrices for typical values of $L_B$ and $L_R$ used in the experiments. We observe the following:

- As $L_R$ increases, $\lambda_N(\mathbf{R})$ decreases; hence, the first term in the lower bound of (15) will increase with increasing $L_R$, and the third term in the lower bound of (15) will decrease with increasing $L_R$. It is therefore not possible in general to determine how the lower bound will change with increasing $L_R$.
- As $L_R$ increases, $\lambda_1(\mathbf{R})$ increases, meaning that the second term in the lower bound of (15) will decrease with increasing $L_R$.
- As $L_B$ increases, the difference between its minimum and maximum eigenvalues increases, meaning that the condition number of $\mathbf{B}$ increases with $L_B$.
- In this setting, the upper bound of (15) will increase as $L_R$ or $L_B$ increases, as $\lambda_1(\mathbf{B})$ and $\kappa(\mathbf{B})$ increase with $L_B$ and $\frac{1}{\lambda_N(\mathbf{R})}$ increases with $L_R$.
- As $L_B$ increases, the ratio $\frac{\kappa(\mathbf{B})}{\lambda_1(\mathbf{B})} = \frac{1}{\lambda_N(\mathbf{B})}$ increases, meaning that for the fixed $L_R$, the first and second terms of the lower bound of (15) will decrease, and the third term will increase.

Therefore, increasing $L_B$ for fixed $L_R$ will cause both bounds to increase. It is not possible at this stage to say whether the upper and lower bound will move closer together or further apart as $L_B$ increases. It is also not clear which term in the lower bound of (15) will be the largest for a general choice of $\mathbf{B}$, $\mathbf{R}$, and $\mathbf{H}$. This means that we cannot say how the lower bound of (15) will change with $L_R$. We will investigate how the bounds change numerically with $\mathbf{B}$ and $\mathbf{R}$ in Section 5. Although we understand the effect of changing $L_B$ and $L_R$ on the bounds of the condition number, we now want to investigate their influence on the actual value of $\kappa(\mathbf{S})$.

### 4.3.2 | Convergence of a conjugate gradient routine

In addition to studying how the condition number of the Hessian changes with $\mathbf{B}$, $\mathbf{R}$, and $\mathbf{H}$, it is of interest to determine the effect of these same changes on the rate of convergence of the minimisation of the objective function. In order to do this, we consider the convergence rate of a conjugate gradient method applied to the linear system (2) associated with the 3D-Var cost function (1).

To do this, we follow the same method that is used in Chapter 6 in the work of Haben[16]; we construct a vector $\mathbf{w}$ that has small- and large-scale features, calculate $\mathbf{b} = \mathbf{S}\mathbf{w}$, and then recover $\mathbf{w}$ by applying a linear solver, in this case, the conjugate gradient method, to $\mathbf{S}\mathbf{w} = \mathbf{b}$. Here we used the Matlab conjugate gradient routine, pcg,[55] to investigate the

change in the number of iterations to convergence. In exact arithmetic, the conjugate gradient method should converge to the true solution in exactly $n$ iterations for an $n$-dimensional problem.[17] We note that in finite precision, convergence in $n$ iterations may not occur, as the search directions lose conjugacy due to roundoff errors.[56] Operationally, however, even $n$ iterations are too many in order to obtain a solution in reasonable computational time. This problem is usually solved by preconditioning, but for this paper, we are interested in the unpreconditioned problem as discussed in Section 1. We use a tolerance of $1 \times 10^{-6}$ on the relative residual for all results presented in the next section.

We expect that the impact of changing $\mathbf{B}$ and $\mathbf{R}$ on the condition number of the Hessian will be similar for both sets of experiments (condition number and conjugate gradient convergence) due to the theoretical link between the condition number and the convergence of the conjugate gradient method.[18,19] In addition to investigating the impact of changing the length scale on the convergence of 3D-Var, we are interested in how the choice of observation operators introduced in Section 4.2 influences 3D-Var in terms of both the condition number and the convergence of the conjugate gradient method.

## 5 | NUMERICAL TESTING

Our experiments focus on how $\kappa(\mathbf{S})$ changes with both $\lambda_1(\mathbf{R}^{-1})$, for $\mathbf{R}$ correlated, and $\lambda_1(\mathbf{B}^{-1})$ for each of the choices of observation operator introduced in Section 4.2 (recalling that for any matrix $\mathbf{D} \in \mathbb{R}^{N\times N}$, $\lambda_1(\mathbf{D}^{-1}) = 1/\lambda_N(\mathbf{D})$). This extends the experiments of Haben[16] where the effect of the length scale of $\mathbf{B}$ on the conditioning of the Hessian was considered for uncorrelated $\mathbf{R}$. As the terms $\lambda_1(\mathbf{B}^{-1}) = 1/\lambda_N(\mathbf{B})$ and $\lambda_1(\mathbf{R}^{-1}) = 1/\lambda_N(\mathbf{R})$ appear in both upper and lower bounds of (15), we investigate the relationship between changing the eigenvalues of $\mathbf{B}$ and $\mathbf{R}$ and the condition number of $\mathbf{S}$. We also investigate how correlations in $\mathbf{B}$ and in $\mathbf{R}$ interact in terms of both the bounds and the true conditioning of the Hessian. We then test our conclusions in terms of a minimisation problem, to assess the impact of changing correlation length scales on the number of iterations required for the convergence of a conjugate gradient routine. We present and discuss the results for $\mathbf{H} = \mathbf{H}_1$, $\mathbf{H}_2$, and $\mathbf{H}_3$ separately before comparing the different cases.

### 5.1 | Investigating changing length scales: observing the first $p$ variables ($\mathbf{H} = \mathbf{H}_1$)

In Figure 3a, we plot the condition number of $\mathbf{S}$ (colour) with the maximum eigenvalue of $\mathbf{B}^{-1}$, shown along the $x$-axis, and the maximum eigenvalue of $\mathbf{R}^{-1}$, shown on the $y$-axis, for the case $\mathbf{H} = \mathbf{H}_1$. Both axes and the colour values are shown with a logarithmic scale. We recall that as length scale increases, $\lambda_1(\mathbf{B}^{-1})$ and $\lambda_1(\mathbf{R}^{-1})$ both increase.

We observe the following:

- For a fixed value of $\lambda_1(\mathbf{R}^{-1})$, increasing $\lambda_1(\mathbf{B}^{-1})$ results in an increased value of $\kappa(\mathbf{S})$. This behaviour is also seen in the work of Haben[16] for an uncorrelated choice of $\mathbf{R}$. The effect of this increase depends on the size of $\lambda_1(\mathbf{R}^{-1})$; larger values of $\lambda_1(\mathbf{R}^{-1})$ lead to smaller gradients in the contours of $\kappa(\mathbf{S})$. The inclusion of correlated observation errors therefore results in a more complex dependence of $\kappa(\mathbf{S})$ on $\mathbf{B}$.
- For a small fixed value of $\lambda_1(\mathbf{B}^{-1})$, increasing $\lambda_1(\mathbf{R}^{-1})$ results in an increased value of $\kappa(\mathbf{S})$, whereas for a large fixed value of $\lambda_1(\mathbf{B}^{-1})$, increasing $\lambda_1(\mathbf{R}^{-1})$ has minimal impact on the value of $\kappa(\mathbf{S})$.
- In general, the impact of changing $\lambda_1(\mathbf{B}^{-1})$ on $\kappa(\mathbf{S})$ is larger than when changing $\lambda_1(\mathbf{R}^{-1})$.

We hence note that interactions of $\lambda_1(\mathbf{B}^{-1})$ and $\lambda_1(\mathbf{R}^{-1})$ have an important effect on the condition number of $\mathbf{S}$. This agrees with the results of Corollary 1, which showed that depending on the relationship between the largest eigenvalues of $\mathbf{B}^{-1}$ and $\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$, there are two distinct bounds on the eigenvalues of $\mathbf{S}$, one in terms of $\lambda_1(\mathbf{B}^{-1})$ and the other in terms of $\lambda_1(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})$.

In Figure 3b, we see the number of iterations required for the conjugate gradient method to solve the problem described in Section 4.3. The values of $\lambda_1(\mathbf{R}^{-1})$ plotted in Figure 3b are shown in Figure 3a as horizontal lines for 80 values of $\lambda_1(\mathbf{B}^{-1})$. Firstly, for $\lambda_1(\mathbf{B}^{-1}) < 4$, increasing $\lambda_1(\mathbf{B}^{-1})$ for fixed $\lambda_1(\mathbf{R}^{-1})$ results in an increase in the number of iterations required for convergence. Additionally, for fixed $\lambda_1(\mathbf{B}^{-1})$, increasing $\lambda_1(\mathbf{R}^{-1})$ results in a clear increase in the number of iterations. This behaviour agrees well with the qualitative conclusions from the condition number experiment in Figure 3a. For $\lambda_1(\mathbf{B}^{-1}) > 4$, we see a decrease in the number of iterations as $\lambda_1(\mathbf{B}^{-1})$ increases. In this range, the value of $\kappa(\mathbf{S})$ is similar across each of the horizontal lines shown in Figure 3a, so we could expect the number of iterations to convergence to be similar. Additionally, the Hessian is extremely ill conditioned, which, combined with a small tolerance in the conjugate gradient routine, could explain the noisy values for large $\lambda_1(\mathbf{B}^{-1})$.

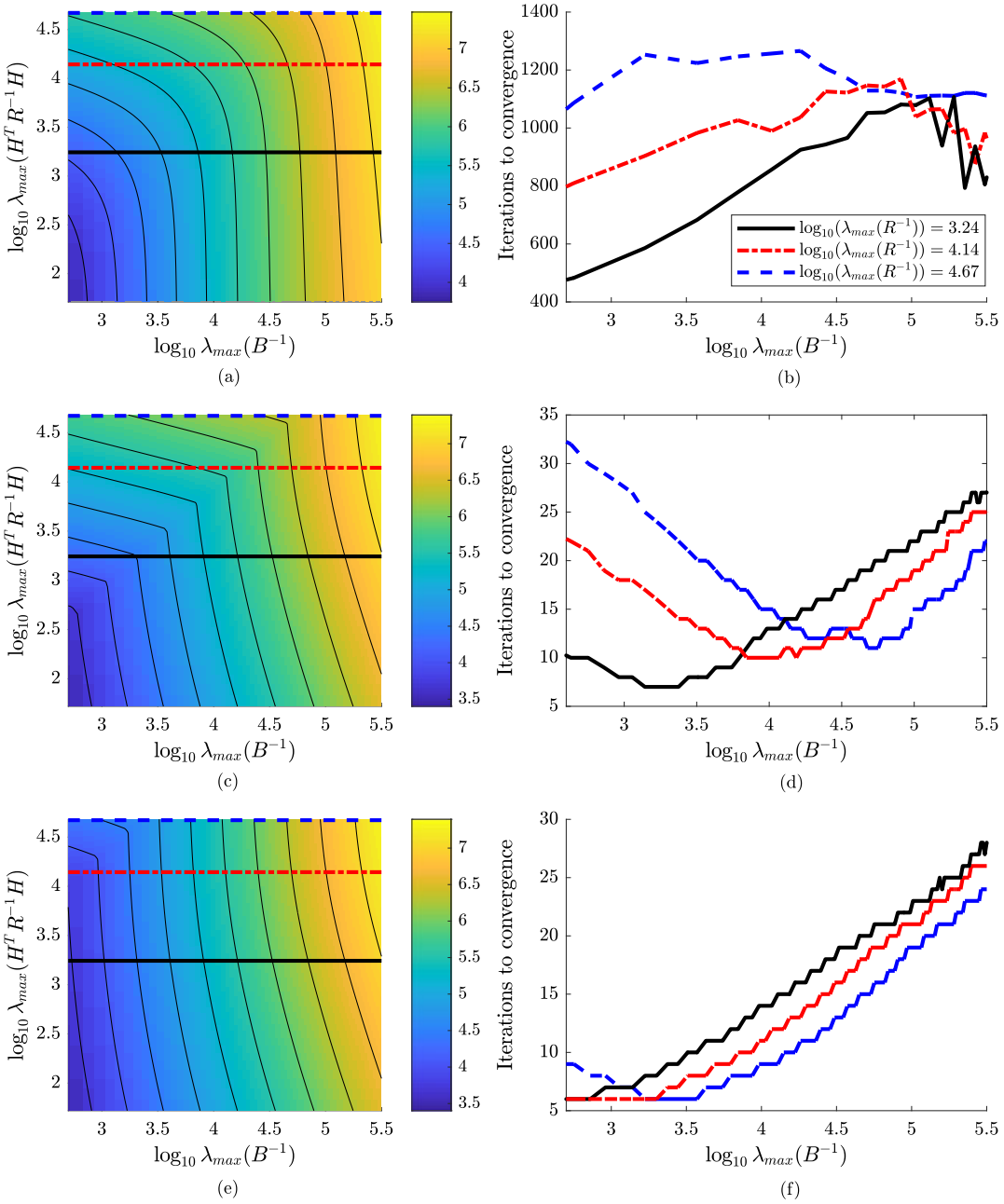


**FIGURE 3** Impact of different choices of observation operator $\mathbf{H}$ on $\kappa(\mathbf{S})$ (a, c, e) and the convergence of the conjugate gradient algorithm (b, d, f) for: (a, b) $\mathbf{H} = \mathbf{H}_1$, (c, d) $\mathbf{H} = \mathbf{H}_2$, and (e, f) $\mathbf{H} = \mathbf{H}_3$. The matrices $\mathbf{B}$ and $\mathbf{R}$ are second-order auto-regressive correlation matrices (28) for $N = 200$ and $p = 100$. The $x$-axis denotes $\log_{10}(\lambda_1(\mathbf{B}^{-1}))$. (a, c, e) The $y$-axis shows $\log_{10}(\lambda_1(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}))$, and the colour denotes $\log_{10}(\kappa(\mathbf{S}))$. We also show 10 equally spaced contours (solid lines) and horizontal lines (corresponding to the lines plotted in b, d, and f). The solid, dotted, and dash-dotted lines represent $\log_{10}(\lambda_1(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})) = 3.24$, 4.14, and 4.67, respectively

## 5.2 | Investigating changing length scales: observing $p$ alternate state variables ($\mathbf{H} = \mathbf{H}_2$)

In Figure 3c, we see how changing $\mathbf{B}$ and $\mathbf{R}$ affects the condition number of $\mathbf{S}$ for the case $\mathbf{H} = \mathbf{H}_2$. The changes in $\kappa(\mathbf{S})$ with $\lambda_1(\mathbf{R}^{-1})$ and $\lambda_1(\mathbf{B}^{-1})$ are qualitatively similar to the case $\mathbf{H}_1$ described in Section 5.1. Again, we see that the interaction between $\mathbf{B}$ and $\mathbf{R}$ has an important effect on $\kappa(\mathbf{S})$, in agreement with the results of Corollary 1. However, for $\mathbf{H} = \mathbf{H}_2$, the change of behaviour of $\kappa(\mathbf{S})$ does not occur smoothly; we observe a discontinuity in the gradient of the contours. As $\lambda_1(\mathbf{R}^{-1})$ increases, the value of $\lambda_1(\mathbf{B}^{-1})$ at which this "kink" occurs also increases linearly. We will investigate this kink further in Section 5.6 and show that it is caused by a change in regime.

In Figure 3d, we see the number of iterations required for the conjugate gradient method to converge for the case $\mathbf{H} = \mathbf{H}_2$.

- For fixed values of $\lambda_1(\mathbf{R}^{-1})$ we observe a change in behaviour as $\lambda_1(\mathbf{B}^{-1})$ increases; for smaller values of $\lambda_1(\mathbf{B}^{-1})$ we see a decrease in the number of iterations as $\lambda_1(\mathbf{B}^{-1})$ increases, and for larger values of $\lambda_1(\mathbf{B}^{-1})$ the number of iterations increases with $\lambda_1(\mathbf{B}^{-1})$. This does not agree with the results for the condition number of $\mathbf{S}$ in Figure 3c, where an increase in $\lambda_1(\mathbf{B}^{-1})$ causes an increase in $\kappa(\mathbf{S})$ for all values of $\lambda_1(\mathbf{B}^{-1})$.
- For smaller values of $\lambda_1(\mathbf{B}^{-1})$, increasing $\lambda_1(\mathbf{R}^{-1})$ leads to an increase in the number of iterations required for convergence. For larger values of $\lambda_1(\mathbf{B}^{-1})$, which occur to the right of the kink, increasing $\lambda_1(\mathbf{R}^{-1})$ decreases the number of iterations. Again, this is unlike the results seen for the condition number, where increasing $\lambda_1(\mathbf{R}^{-1})$ leads to an increase in both the actual value and the upper bound of $\kappa(\mathbf{S})$ for all values of $\lambda_1(\mathbf{B}^{-1})$.

We note that the value of $\lambda_1(\mathbf{B}^{-1})$, where this change in behaviour occurs, is the same as the value of $\lambda_1(\mathbf{B}^{-1})$, where the change in gradient of the contours occurs in Figure 3c, indicating that the kink is caused by an underlying change in regime. If we consider the eigenvalues of $\mathbf{S}$ (not shown here), the clustering of eigenvalues increases as the kink is approached. The clustering of eigenvalues is important for the convergence of a conjugate gradient method[19] and is not detected by the condition number. This explains the difference in behaviour between Figure 3c and Figure 3d with increasing $\lambda_1(\mathbf{B}^{-1})$.

## 5.3 | Investigating changing length scales: observing $p$ alternate variables smoothed over five state variables ($\mathbf{H} = \mathbf{H}_3$)

In Figure 3e, we see how changing $\mathbf{B}$ and $\mathbf{R}$ affects the condition number of $\mathbf{S}$ for the case $\mathbf{H} = \mathbf{H}_3$. The behaviour of $\kappa(\mathbf{S})$ with changing $\lambda_1(\mathbf{B}^{-1})$ and $\lambda_1(\mathbf{R}^{-1})$ is qualitatively similar to the case $\mathbf{H} = \mathbf{H}_2$. However, for $\mathbf{H} = \mathbf{H}_3$ and fixed $\lambda_1(\mathbf{B}^{-1})$, only changes to very large values of $\lambda_1(\mathbf{R}^{-1})$ result in a significant change to $\kappa(\mathbf{S})$, and this is true for only the smallest values of $\lambda_1(\mathbf{B}^{-1})$. Again, interaction between $\lambda_1(\mathbf{B}^{-1})$ and $\lambda_1(\mathbf{R}^{-1})$ has an important impact on $\kappa(\mathbf{S})$ but to much less of an extent than in the previous two cases. This agrees with the results of Corollary 1, as the value of $\lambda_1(\mathbf{H}_3^T\mathbf{R}^{-1}\mathbf{H}_3)$ is much smaller than $\lambda_1(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})$ for $\mathbf{H} = \mathbf{H}_1$ or $\mathbf{H}_2$, and hence, $L_R$ will need to take a much larger value in order that $\lambda_1(\mathbf{H}_3^T\mathbf{R}^{-1}\mathbf{H}_3) + \lambda_N(\mathbf{B}^{-1}) > \lambda_1(\mathbf{B}^{-1})$ . A discontinuity in gradient similar to the one observed for the case $\mathbf{H} = \mathbf{H}_2$ is seen here but for much larger values of $\lambda_1(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})$ than for Figure 3c.

In Figure 3f, we see the number of iterations required for the conjugate gradient to converge for the problem described in Section 4.3 when $\mathbf{H} = \mathbf{H}_3$. Similar to Figure 3d, we see an initial decrease in the number of iterations required for convergence, before a turning point where the number of iterations increases with $\lambda_1(\mathbf{B}^{-1})$. This turning point occurs for the same values of $\lambda_1(\mathbf{B}^{-1})$ as the discontinuity in gradient that was seen in Figure 3e. As the value of $\lambda_1(\mathbf{B}^{-1})$ at which this kink occurs is much smaller than for the case $\mathbf{H} = \mathbf{H}_2$, for most values of $\lambda_1(\mathbf{B}^{-1})$, increasing $\lambda_1(\mathbf{R}^{-1})$ decreases the number of iterations. As in the case $\mathbf{H} = \mathbf{H}_2$, the clustering of the eigenvalues of $\mathbf{S}$ increases as we approach the kink. The structure of the eigenvalues is more important in determining the convergence of a conjugate gradient method than the condition number in this case.

## 5.4 | Investigating bounds and actual value of $\kappa(\mathbf{S})$ for different choices of observation operator

We now compare the effect of changing the observation operator on both the condition number of $\mathbf{S}$ and the bounds of $\mathbf{S}$ introduced in Section 3. Of particular interest is how tight the bounds are for different values of $\lambda_1(\mathbf{B}^{-1})$. For clarity, the Hessian for the cases $\mathbf{H} = \mathbf{H}_1$, $\mathbf{H} = \mathbf{H}_2$, and $\mathbf{H} = \mathbf{H}_3$ will be referred to as $\mathbf{S}_1$, $\mathbf{S}_2$, and $\mathbf{S}_3$, respectively. Figure 4 displays the actual value of the condition number and the bounds from (15) for a fixed choice of $\mathbf{R}$ with $L_R = 0.33$ for all three choices of $\mathbf{H}$. We recall (Section 4.2) that the bounds for the cases $\mathbf{H} = \mathbf{H}_1$ and $\mathbf{H} = \mathbf{H}_2$ are equal, with tighter bounds for the case $\mathbf{H} = \mathbf{H}_3$. This is because the maximum eigenvalue of $\mathbf{H}_3\mathbf{H}_3^T$, which appears in both upper and lower bounds, is 0.52 rather than 1.

- Figure 4 shows cases where both the upper and lower bounds given by (15) are tight. The upper bound is close to the actual value of $\kappa(\mathbf{S})$ for $\mathbf{H}_1$, particularly when $\lambda_1(\mathbf{B}^{-1})$ is small. For small values of $\lambda_1(\mathbf{B}^{-1})$, the actual value of $\kappa(\mathbf{S})$ for $\mathbf{H}_3$ is much closer to the lower bound than the upper bound.

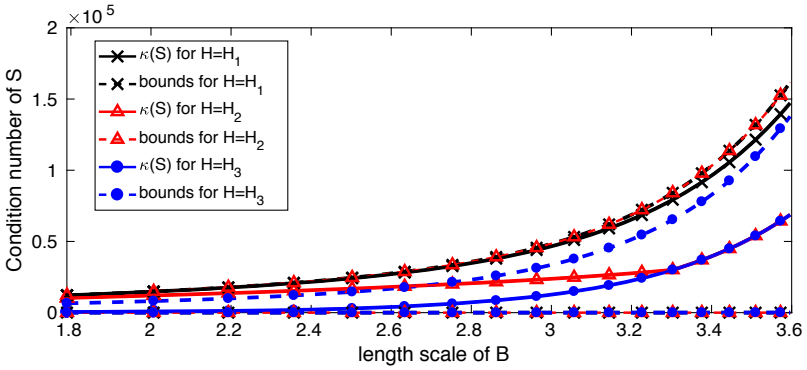


**FIGURE 4** Bounds (dashed lines) and condition number (solid lines) of $\mathbf{S}$ for $\mathbf{H}_1$ (cross), $\mathbf{H}_2$ (triangle), and $\mathbf{H}_3$ (circle) for $L_R = 0.33$. The bounds are calculated using (15) for all choices of $\mathbf{H}$. We note that the bounds for the cases $\mathbf{H}_1$ and $\mathbf{H}_2$ are the same

- The kink that was observed in Figure 3c for $\mathbf{H} = \mathbf{H}_2$ can also be seen in Figure 4. The kink occurs at the location where $\kappa(\mathbf{S}_2)$ coincides with $\kappa(\mathbf{S}_3)$. For values of $\lambda_1(\mathbf{B}^{-1})$ greater than the kink, $\kappa(\mathbf{S}_2)$ and $\kappa(\mathbf{S}_3)$ are very close to each other.
- For all choices of $\mathbf{H}$ shown in Figure 4, increasing $\lambda_1(\mathbf{B}^{-1})$ leads to the upper bound moving away from both the lower bound and the actual value of $\kappa(\mathbf{S})$.

We note that we have found different choices of $\mathbf{B}$, $\mathbf{R}$, and $\mathbf{H}$, where the actual values of $\mathbf{S}$ are close to both the upper and lower bounds given by (15). We now discuss the implications of changing $\mathbf{B}$, $\mathbf{R}$, and $\mathbf{H}$ in terms of the condition number of $\mathbf{S}$ and the number of iterations required for the conjugate gradient to converge.

## 5.5 | Comparison of results

In this section, we compare the results of the previous sections for different choices of observation operator $\mathbf{H}$, as well as different choices of $\mathbf{B}$ and $\mathbf{R}$. We recall that $\lambda_1(\mathbf{B}^{-1}) = 1/\lambda_N(\mathbf{B})$ and $\lambda_1(\mathbf{R}^{-1}) = 1/\lambda_N(\mathbf{R})$.

We begin by considering how the lower bounds given by Lemma 1 for $\lambda_1(\mathbf{S})$ change depending on whether $\lambda_1(\mathbf{B}^{-1})$ or $\lambda_1(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}) + \lambda_N(\mathbf{B}^{-1})$ is the larger term.

- For a fixed value of $L_R$ and changing $L_B$: For small values of $\lambda_1(\mathbf{B}^{-1})$, the lower bound of $\lambda_1(\mathbf{S})$ from (8) is given by $\lambda_1(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}) + \lambda_N(\mathbf{B}^{-1})$, meaning that the maximum eigenvalue of $\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$ is most important for determining $\lambda_1(\mathbf{S})$.
- As $L_B$ increases, at some point, $\lambda_1(\mathbf{B}^{-1})$ will be larger than $\lambda_1(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}) + \lambda_N(\mathbf{B}^{-1})$, meaning that $\lambda_1(\mathbf{B}^{-1})$ will be the most important term for determining $\lambda_1(\mathbf{S})$.
- Alternatively, fixing $L_B$ and changing $L_R$, we observe similar behaviour: For smaller values of $L_R$, we see less impact on $\kappa(\mathbf{S})$ when changing $L_R$ than for larger values of $L_R$, where a change in $\lambda_1(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})$ has a significant effect on the value of $\kappa(\mathbf{S})$.

This behaviour is seen for all choices of $\mathbf{H}$ in Figure 3. This bound also provides justification for the variation with $\lambda_1(\mathbf{B}^{-1})$ and $\lambda_1(\mathbf{R}^{-1})$ in the gradient of the contours seen in Figure 3a,c,e.

We now consider the similarities between different choices of observation operator for the two experiments, as follows:

- For a fixed choice of $\mathbf{H}$, there are strong similarities between the effect of increasing $\lambda_1(\mathbf{B}^{-1})$ on the convergence of the conjugate gradient method and the effect on the condition number of the Hessian. In particular, the kink in the condition number (Figure 3c,e) and the change in gradient for convergence (Figure 3d,f) occur at the same values of $\lambda_1(\mathbf{B}^{-1})$ and $\lambda_1(\mathbf{R}^{-1})$ for both $\mathbf{H} = \mathbf{H}_2$ and $\mathbf{H} = \mathbf{H}_3$. This indicates that the kink is due to a change in the underlying structure of $\mathbf{S}$.
- The effect of varying $\lambda_1(\mathbf{R}^{-1})$ and $\lambda_1(\mathbf{B}^{-1})$ for $\mathbf{H}_1$, $\mathbf{H}_2$, and $\mathbf{H}_3$ was broadly similar in terms of $\kappa(\mathbf{S})$, with the main difference being the discontinuity in the contours of $\kappa(\mathbf{S})$ seen for $\mathbf{H}_2$ and $\mathbf{H}_3$ but not for $\mathbf{H}_1$.

We also see some large differences between the two experiments. The main dissimilarity between the graphs for condition number (Figure 3a,c,e) and for convergence (Figure 3b,d,e) is that increasing $\lambda_1(\mathbf{B}^{-1})$ uniformly results in an increase in the condition number of $\mathbf{S}$, but it is not always linked to an increase in the number of iterations required for convergence. This difference was explained in Sections 5.1– 5.3 by the clustering of eigenvalues near the kink for $\mathbf{H}_2$ and $\mathbf{H}_3$.

For the conjugate gradient experiments, conclusions for the cases $\mathbf{H} = \mathbf{H}_2$ and $\mathbf{H} = \mathbf{H}_3$ were very different from those of the case $\mathbf{H} = \mathbf{H}_1$. Both $\mathbf{H} = \mathbf{H}_2$ and $\mathbf{H} = \mathbf{H}_3$ have block-circulant structures, meaning that in these cases $\mathbf{S}$ will

have a block-circulant structure. We suggest that this is the reason for the difference in eigenvalue clustering behaviour compared with the case $\mathbf{H} = \mathbf{H}_1$. This was tested through the use of an additional noncirculant observation operator made by observing 100 random state variables. The behaviour in this case is very similar to that observed for $\mathbf{H} = \mathbf{H}_1$. The fact that qualitative behaviour for the case $\mathbf{H} = \mathbf{H}_1$ is the same as for the randomly selected observation operator supports the conjecture that the rapid convergence of the conjugate gradient seen for $\mathbf{H} = \mathbf{H}_2$ and $\mathbf{H} = \mathbf{H}_3$ is caused by the inherent block-circulant structure of $\mathbf{S}_2$ and $\mathbf{S}_3$.

## 5.6 | Understanding the discontinuity in the gradient for $\mathbf{H} = \mathbf{H}_2$ and $\mathbf{H} = \mathbf{H}_3$

We now return to discuss the discontinuity in the gradient, or kink, that was observed for $\mathbf{H} = \mathbf{H}_2$ and $\mathbf{H} = \mathbf{H}_3$ for both the condition number of $\mathbf{S}$ (Figure 3c,e) and the convergence of the conjugate gradient method (Figure 3d,f). We explain this theoretically and discuss why the discontinuity in gradient is observed for $\mathbf{H}_2$ and $\mathbf{H}_3$ but not for $\mathbf{H}_1$. We begin by considering the bounds for the eigenvalues of $\mathbf{S}$ in terms of the eigenvalues of $\mathbf{B}^{-1}$ and $\mathbf{R}^{-1}$, using the bounds given by Corollary 1 and the discussion that follows in Section 3.1.

Equations (8)–(11) explain the variation with $\lambda_1(\mathbf{B}^{-1})$ and $\lambda_1(\mathbf{R}^{-1})$ that was observed in Figure 3. However, as the bounds in (10) and (11) apply to all choices of $\mathbf{H}$, they do not explain the difference between the choices of $\mathbf{H}$ for which the kink is observed ($\mathbf{H}_2$ and $\mathbf{H}_3$) and the choices of $\mathbf{H}$ that have smoothly varying values of $\kappa(\mathbf{S})$ (namely $\mathbf{H}_1$).

In order to illustrate why kink occurs for some choices of $\mathbf{H}$ but not for others, we present a tighter upper bound for the specific framework used in the numerical experiments for two cases, beginning with $\mathbf{H}_1$. By expressing $\mathbf{S}$ in terms of the difference between a circulant matrix and a low-rank update, we use (20) to directly compute the eigenvalues of the circulant component via a direct Fourier decomposition. This allows us to show that the kink occurs when there is a significant change in the wavenumber corresponding to the largest eigenvalue of $\mathbf{S}$.

**Lemma 3.** *We define $\mathbf{C}_1$ as in the Appendix. For $\mathbf{H} = \mathbf{H}_1$, we can bound the eigenvalues of $\mathbf{S}$ above by the following:*

$$\lambda_k(\mathbf{S}) \leq \lambda_k(\mathbf{C}_1), \tag{33}$$

*where the eigenvalues of $\mathbf{C}_1$ are given by the following:*

$$\gamma_m(\mathbf{C}_1) = \gamma_m(\mathbf{B}^{-1}) + \sum_{k=0}^{p-1} \omega^{mk} \mathbf{R}^{-1}_{1,k}, \quad m = 0, \ldots, N-1, \tag{34}$$

*where $\omega = e^{-2\pi i/N}$. Recall (using the notation introduced in Section 3.3) that the $\gamma_j$s are ordered in terms of wavenumber rather than by decreasing eigenvalue.*

*Proof.* See the Appendix. □

Lemma 3 yields an expression that is a sum of an eigenvalue of $\mathbf{B}^{-1}$, plus a term depending on the coefficients of $\mathbf{R}^{-1}$ and the structure of $\mathbf{H}_1$. The choice of $\mathbf{H} = \mathbf{H}_1$ is important in determining the wavenumber at which the maximum value of the second term of (34) is achieved. From Section 4.1, we recall that the largest eigenvalue of $\mathbf{B}^{-1}$ occurs for the $p$th wavenumber, $\gamma_{N/2}(\mathbf{B}^{-1})$, for $N = 2p$, or $\gamma_{(N\pm1)/2}(\mathbf{B}^{-1})$, for $N = 2p+1$. The eigenvalues of $\mathbf{B}^{-1}$ ordered by the wavenumber are shown by circles in Figure 5. The crosses in Figure 5 show the second term of (34) ordered by the wavenumber. For $\mathbf{H}_1$, the largest value of the second term of (34) occurs for the same wavenumber as the largest eigenvalue of $\mathbf{B}^{-1}$. The maximum value of this term is equal to $\lambda_1(\mathbf{R}^{-1})$. This means that as $\lambda_1(\mathbf{S}_1)$ changes from being controlled by $\lambda_1(\mathbf{R}^{-1})$ to $\lambda_1(\mathbf{B}^{-1})$, the change appears smooth, as the wavenumber associated with the frequency of the largest eigenvalue remains constant. It is clear that increasing $L_B$ will have a significant effect on the value of this bound, as changing $L_B$ increases $\lambda_1(\mathbf{B}^{-1})$ significantly, and hence the upper bound given by (33). Therefore, for both regimes, changing $L_B$ has a large impact on both bounds for $\lambda_1(\mathbf{S})$.

We now present a similar bound for $\mathbf{H} = \mathbf{H}_2$.

**Lemma 4.** *For $\mathbf{H} = \mathbf{H}_2$, the eigenvalues of $\mathbf{S}$ are bounded above by the following:*

$$\lambda_k(\mathbf{S}) \leq \lambda_k(\mathbf{C}_2), \tag{35}$$

*where the eigenvalues of $\mathbf{C}_2$ are given by the following:*

$$\gamma_m(\mathbf{C}_2) = \gamma_m(\mathbf{B}^{-1}) + \sum_{k=0}^{p-1} \omega^{2mk} \mathbf{R}^{-1}_{1,k}, \quad m = 1, 2, \ldots, p-1. \tag{36}$$

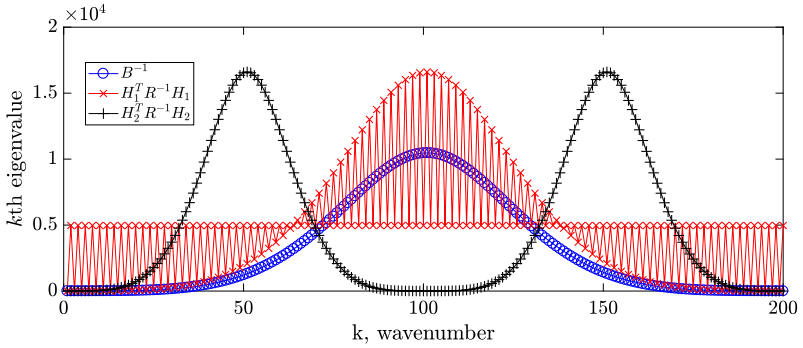


**FIGURE 5** Plots of the contribution of the background and observation terms to the eigenvalues of the circulant matrix made up of the first row of $\mathbf{S}_1$ and $\mathbf{S}_2$ for $L_R = 0.7$ and $L_B = 0.3$. Circles denote the eigenvalues of $\mathbf{B}^{-1}$ (which is a term in both (34) and (36)), crosses denote the contribution of $\mathbf{R}^{-1}$ in the second term of (34) (i.e., for $\mathbf{H} = \mathbf{H}_1$), and pluses denote the contribution of $\mathbf{R}^{-1}$ in the second term of (36) (i.e., for $\mathbf{H} = \mathbf{H}_2$)

*Recall (Section* 3.3*) that $\gamma_j$s are ordered in terms of the wavenumber rather than by the maximum eigenvalue.*

*Proof.* See the Appendix. □

Lemma 4 also yields an upper bound that is the sum of an eigenvalue of $\mathbf{B}^{-1}$ and a term depending on $\mathbf{R}^{-1}$ and the choice of $\mathbf{H}_2$. We note that the values of the second term of (36) take the same values as the second term of (34) but in a different order. These are shown by the pluses in Figure 5, where we see that in the order of wavenumber $j$, the second term of (36) yields the spectrum of $\mathbf{R}^{-1}$ twice. The second term of (36) is maximised for $j = p/2$ and $j = 3p/2$. These are different wavenumbers to the value of $j = p$, which maximises the first term.

Hence, we can bound $\lambda_1(\mathbf{S})$ above by $\lambda_1(\mathbf{R}^{-1})+\lambda_{N/4}(\mathbf{B}^{-1})$ when $\lambda_1(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})+\lambda_N(\mathbf{B}^{-1}) > \lambda_1(\mathbf{B}^{-1})$. In this case, increasing $L_B$ has a very small effect on the upper bound for $\lambda_1(\mathbf{S})$, as $\lambda_{N/4}(\mathbf{B}^{-1})$ does not change significantly with $L_B$. However, when $\lambda_1(\mathbf{B}^{-1}) > \lambda_1(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}) + \lambda_N(\mathbf{B}^{-1})$, small changes to $L_B$ will have a larger impact on $\lambda_1(\mathbf{B}^{-1})$ for all choices of $\mathbf{H}$. Similar behaviour is observed for fixed $L_B$ and changing $L_R$. This change in the wavenumber of the largest eigenvalue explains why the kink occurs in the case of $\mathbf{H}_2$.

Finally, we discuss why the kink occurs for different values of $L_B$ and $L_R$ for $\mathbf{H}_2$ and $\mathbf{H}_3$. We have shown that the kink occurs when $\lambda_1(\mathbf{B}^{-1})$ becomes larger than $\lambda_1(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})+\lambda_N(\mathbf{B}^{-1})$. For all values of $L_R$, $\lambda_1(\mathbf{H}_2^T\mathbf{R}^{-1}\mathbf{H}_2) \gg \lambda_1(\mathbf{H}_3^T\mathbf{R}^{-1}\mathbf{H}_3)$. As the contribution of $\mathbf{B}^{-1}$ is not affected by the choice of observation operator, changing from $\mathbf{H}_2$ to $\mathbf{H}_3$ increases the value of $L_R$ necessary for $\lambda_1(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})$ to be greater than $\lambda_1(\mathbf{B}^{-1})$. Hence, the kink is only visible (see Figure 3e) for $L_R \gg L_B$ for the choice $\mathbf{H} = \mathbf{H}_3$.

## 6 | CONCLUSIONS

Data assimilation is an important technique for combining information from observations with model data for the purpose of state estimation. One application of this is in NWP, where data assimilation is used to combine observations of the atmosphere with a numerical model, in order to obtain an accurate description of the current state of the atmosphere. In this case, correct specification of the uncertainty of each term is needed to produce the best forecast. The introduction of correlated observation error terms at operational NWP centres motivates investigation into the influence of observation error covariance on the convergence of the data assimilation procedure. We emphasise that the results presented here are general and are relevant for any application of variational data assimilation. Improved knowledge of the role of correlated observation error covariances will be of use in the context of engineering,[33] neuroscience,[33,34] and ecology.[35,36]

In this work, we developed theoretical bounds on the condition number of the Hessian of the 3D-Var objective function, which can be studied as a bound on the speed of convergence of the minimisation. These bounds were then tested in a simple numerical framework. We found the following:

- The bounds separate the contributions of the (correlated) observation error, background error, and observation operator, allowing us to better understand the role played by each term.We note that Theorem 5 and Corollary 1 in particular are general bounds applying to any valid covariance matrices and any choice of observation network.

- Numerical experiments for simple linear choices of observation network revealed interaction between observation error and background error terms. This interaction was also demonstrated theoretically for any choice of observation network and error covariance matrices.
- The structure of the observation network was seen to be crucial for determining how the observation and background errors interact.
- Both bounds and experiments revealed that the minimum eigenvalue of the two error covariance matrices is important for determining the conditioning of the Hessian, as well as the number of iterations required for the convergence of a minimisation procedure. This agrees with the findings of Weston et al.,[9] where small minimum eigenvalues of the observation error covariance matrix caused convergence problems in a practical setting.
- The ratio of the variances was also shown to be influential, although this was not investigated in detail in this work. This was also seen in the work of Haben.[16]

We emphasise that many of the theoretical results and conclusions presented in this work are general and apply to any valid choice of background and observation error covariance matrices and any linear observation operator. In particular, although the theoretical results presented in this paper focus on the 3D-Var problem, a natural extension to 4D-Var is obtained by replacing the observation operator $\mathbf{H}$ with the generalised 4D observation operator that incorporates dynamical model information.[16] It is therefore expected that the eigenvalues of this model will also be important for the conditioning of the Hessian in this framework.

The importance of the choice of observation operator was revealed by the numerical tests, both for the condition number of the Hessian and in terms of interaction between observation and background error covariances. Even for two observation networks with identical theoretical bounds, very different behaviour was observed numerically. This was explained by the existence of underlying structures in the Hessian, induced by the structures of the constituent error covariance and observation operator matrices. Better understanding of these interactions will be important for predicting the response of operational systems to the introduction of correlated observation errors. This is particularly applicable in practical applications where diagnosed correlated observation error covariance matrices must be adapted prior to their use in order to ensure the convergence of the minimisation of the objective function.

In the numerical experiments presented in this paper in Section 5, the observation and background error covariance matrices were altered by changing the length scales of the underlying correlation functions. This approach is mainly applicable for spatial correlations, where correlation length scales have a physical interpretation. There is significant research investigating spatial correlations,[20,23,27,51] but much current work concerns the practical implementation of interchannel correlations for satellite observations.[3,6,9,21,22,26] Although the theory presented in Section 3 applies directly to the case of interchannel correlations, it would be of interest to extend our numerical testing to the interchannel covariance case. In particular, practical experiments have revealed that the minimum eigenvalue of the observation error correlation matrix is important for the conditioning of the Hessian in the case of interchannel correlations,[3,9] which coincides with the theoretical and experimental results presented in this work. This is of particular interest as the correlation structure used by Weston et al.[9] is not circulant and demonstrates that, even beyond the numerical framework presented in this paper, our qualitative conclusions provide useful insight.

An additional area of future interest is investigation into how the best choice of preconditioning changes with the introduction of correlated observation error. Bounds on conditioning for the preconditioned case could be found by extending the results presented here, using similar theoretical techniques to those used in this work. The numerical and theoretical results discussed in this paper suggest that interactions between observation and background correlations are also likely in that framework. It is expected that understanding how the introduction of correlated observation error covariance affects the unpreconditioned 3D-Var problem will provide insight for suitable preconditioning methods in the correlated setting. One question of particular interest is whether the use of the background error covariance term as a preconditioner, as is done for the control variable transform,[30] remains optimal. One example of an operational problem that is not preconditioned is the 1D-Var used at the UK Met Office for quality control[5]; the conclusions from this paper apply directly to that implementation. The application of these results to the UK Met Office system will be discussed in a future paper.

## ACKNOWLEDGEMENTS

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

## ORCID

*Jemima M. Tabeart* http://orcid.org/0000-0001-6806-8608
*Sarah L. Dance* http://orcid.org/0000-0003-1690-3338
*Amos S. Lawless* http://orcid.org/0000-0002-3016-6568
*Joanne A. Waller* http://orcid.org/0000-0002-7783-6434

## REFERENCES

1. Buehner M. Error statistics in data assimilation: Estimation and modelling. In: Lahoz W, Khattotov B, Menard R, editors. Data assimilation: Making sense of observations. Heidelberg: Springer-Verlag, 2010. p. 93–112.
2. Janjić T, Bormann N, Bocquet M, et al. On the representation error in data assimilation. QJR Meteorol Soc. 2017. https://doi.org/10.1002/qj.3130
3. Weston P. Progress towards the implementation of correlated observation errors in 4D-Var. Forecasting research technical report 560. UK: Met Office, Exeter; 2011.
4. Rainwater S, Bishop CH, Campbell WF. The benefits of correlated observation errors for small scales. QJR Meteorol Soc. 2015;141:3439–3445.
5. Stewart LM, Dance SL, Nichols NK. Correlated observation errors in data assimilation. Int J Numer Methods. 2008;56:1521–1527.
6. Stewart LM. Correlated observation errors in data assimilation [PhD Thesis]. UK: University of Reading; 2010.
7. Stewart LM, Dance SL, Nichols NK. Data assimilation with correlated observation errors: experiments with a 1-D shallow water model. Tellus A. 2013;65(1):19546. https://doi.org/10.3402/tellusa.v65i0.19546
8. Waller JA, Dance SL, Lawless AS, Nichols NK. Estimating correlated observation error statistics using an ensemble transform Kalman filter. Tellus A. 2014;66(1):23294. https://doi.org/10.3402/tellusa.v66.23294
9. Weston PP, Bell W, Eyre JR. Accounting for correlated error in the assimilation of high-resolution sounder data. QJR Meteorol Soc. 2014;140:2420–2429.
10. Haben SA, Lawless AS, Nichols NK. Conditioning of incremental variational data assimilation, with application to the Met Office system. Tellus A. 2011;64(4):782–792.
11. Rawlins F, Ballard SP, Bovis KJ, et al. The Met Office global four-dimensional variational data assimilation scheme. QJR Meteorol Soc. 2007;133:347–362.
12. Clayton AM, Lorenc AC, Barker DM. Operational implementation of a hybrid ensemble/4DVar global data assimilation system at the Met Office. QJR Meteorol Soc. 2013;139:1445–1461.
13. Lawless AS, Gratton S, Nichols NK. Approximate iterative methods for variational data assimilation. Int J Numer Methods Fluids. 2005;47(10–11):1129–1135.
14. Gratton S, Lawless AS, Nichols NK. Approximate Gauss-Newton methods for nonlinear least squares problems. SIAM J Optim. 2007;18(1):106–132.
15. Lawless AS, Gratton S, Nichols NK. An investigation of incremental 4D-Var using non-tangent linear models. QJR Meteorol Soc. 2005;131:459–476.
16. Haben SA. Conditioning and preconditioning of the minimisation problem in variational data assimilation [PhD Thesis]. UK: University of Reading; 2011.
17. Gill PE, Murray W, Wright MH. Practical optimization. Amsterdam/London: Academic Press; 1986.
18. Golub GH, Van Loan CF. Matrix Computations. 3rd ed. Baltimore, MD: The John Hopkins University Press; 1996.
19. Nocedal J. Numerical optimization. 2nd ed. Series in operations research and financial engineering. New York/London: Springer; 2006.
20. Waller JA, Simonin D, Dance SL, Nichols NK, Ballard SP. Diagnosing observation error correlations for Doppler radar radial winds in the Met Office UKV model using observation-minus-background and observation-minus-analysis statistics. Mon Weather Rev. 2016;144(10):3533–3551.
21. Bormann N, Bonavita M, Dragani R, Eresmaa R, Matricardi M, McNally A. Enhancing the impact of IASI observations through an updated observation error covariance matrix. QJR Meteorol Soc. 2016;142(697):1767–1780.
22. Campbell WF, Satterfield EA, Ruston B, Baker NL. Accounting for correlated observation error in a dual formulation 4D-variational data assimilation system. Mon Weather Rev. 2016. https://doi.org/10.1175/MWR-D-16-0240.1.
23. Waller JA, Ballard SP, Dance SL, Kelly G, Nichols NK, Simonin D. Diagnosing horizontal and inter-channel observation error correlations for SEVIRI observations using observation-minus-background and observation-minus-analysis statistics. Remote Sens. 2016;8(7):851.
24. Bormann N, Saarinen S, Kelly G, Thépaut JN. The spatial structure of observation errors in atmospheric motion vectors from geostationary satellite data. Mon Weather Rev. 2003;131:706–718.
25. Bormann N, Geer AJ, Bauer P. Estimates of observation-error characteristics in clear and cloudy regions for microwave imager radiances from numerical weather prediction. QJR Meteorol Soc. 2011;137:2014–2023.
26. Stewart LM, Dance SL, Nichols NK, Eyre JR, Cameron J. Estimating interchannel observation-error correlations of IASI radiance data in the Met Office system. QJR Meteorol Soc. 2014;140:1236–1244.
27. Cordoba M, Dance SL, Kelly GA, Nichols NK, Waller JA. Diagnosing atmospheric motion vector observation errors for an operational high resolution data assimilation system. QJR Meteorol Soc. 2016. https://doi.org/10.1002/qj.2925

28. Healy SB, White AA. Use of discrete Fourier transforms in the 1D-Var retrieval problem. QJR Meteorol Soc. 2005;131(605):63–72.
29. Brown KL, Gejadze I, Ramage A. A multilevel approach for computing the limited-memory Hessian and its inverse in variational data assimilation. SIAM J Sci Comput. 2016;38(5):2934–2963.
30. Bannister RN. Review: A review of forecast error covariance statistics in atmospheric variational data assimilation. II: Modelling the forecast error covariance statistics. QJR Meteorol Soc. 2008;134:1971–1996.
31. Dollar HS, Gould NIM, Stoll M, Wathen AJ. Preconditioning saddle-point systems with applications in optimization. SIAM J Sci Comput. 2010;32(1):249–270. https://doi.org/10.1137/080727129
32. Pestana J, Wathen AJ. Natural preconditioning and iterative methods for saddle point systems. SIAM Rev. 2015;57(1):71–91. https://doi.org/10.1137/130934921
33. Nakamura G, Potthast R. Inverse modeling. UK: IOP Publishing; 2015;2053–2563. https://doi.org/10.1088/978-0-7503-1218-9
34. Schiff SJ. Neural control engineering: The emerging intersection between control theory and neuroscience. Cambridge, MA: MIT Press; 2011.
35. Pinnington EM, Casella E, Dance SL, et al. Investigating the role of prior and observation error correlations in improving a model forecast of forest carbon balance using four-dimensional variational data assimilation. Agr Forest Meteorol. 2016;228-229:299–314.
36. Pinnington EM, Casella E, Dance SL, et al. Understanding the effect of disturbance from selective felling on the carbon dynamics of a managed woodland by combining observations with model predictions. J Geophys Res: Biogeosci. 2017;122(4):886–902.
37. Apte A, Jones CKRT, Stuart AM, Voss J. Data assimilation: Mathematical and statistical perspectives. Int J Numer Methods Fluids. 2008;56(8):1033–1046. https://doi.org/10.1002/fld.1698
38. Cotter SL, Dashti M, Stuart AM. Variational data assimilation using targetted random walks. Int J Numer Methods Fluids. 2012;68(4):403–421. https://doi.org/10.1002/fld.2510
39. Rodgers CD. Inverse methods for atmospheric sounding: Theory and practice. Singapore: World Scientific Publishing Co.; 2000.
40. Wilkinson JH. The algebraic eigenvalue problem. Oxford: Clarendon Press; 1965.
41. Marshall AW, Olkin I, Arnold BC. Inequalities: Theory of majorization and its applications. 2nd ed. New York: Springer-Verlag; 2011.
42. Wang B, Zhang F. Some inequalities for the eigenvalues of the product of positive semidefinite Hermitian matrices. Linear Algebra Appl. 1992;160:113–118.
43. Harville DA. Matrix algebra from a statistician's point of view. New York: Springer-Verlag; 1997.
44. Haben SA, Lawless AS, Nichols NK. Conditioning of the 3DVAR Data Assimilation Problem, Dept. of Mathematics and Statistics, University of Reading, Mathematics Report, 3/2009; 2009. http://www.reading.ac.uk/web/files/maths/New_3DVarCondition_final.pdf
45. Haben SA, Lawless AS, Nichols NK. Conditioning and preconditioning of the variational data assimilation problem. Comp Fluids. 2011;46(1):252–256.
46. Davis PJ. Circulant matrices. New York: Wiley; 1979.
47. Gray RM. Toeplitz and circulant matrices: A review. Found Trends Commun Inform Theory. 2006;2(3):155–239.
48. Süli E, Mayer DF. An introduction to numerical analysis. Cambridge, UK: Cambridge University Press; 2003.
49. Simonin D, Ballard SP, Li Z. Doppler radar radial wind assimilation using an hourly cycling 3D-Var with a 1.5 km resolution version of the Met Office Unified Model for nowcasting. QJR Meteorol Soc. 2014;140:2298–2314.
50. Daley R. Atmospheric data analysis. New York: Cambridge University Press; 1991.
51. Waller JA, Dance SL, Nichols NK. Theoretical insight into diagnosing observation error correlations using observation-minus-background and observation-minus-analysis statistics. QJR Meteorol Soc. 2016;142:418–431.
52. Gaspari G, Cohn SE. Construction of correlation functions in two and three dimensions. QJR Meteorol Soc. 1999;125:723–757.
53. Jeong J, Jun M. Covariance models on the surface of a sphere: When does it matter. Stat. 2015;4:167–182.
54. Ménard R. Error covariance estimation methods based on analysis residuals: Theoretical foundation and convergence properties derived from simplified observation networks. QJR Meteorol Soc. 2016;142:257–273.
55. MATLAB (R2016b). Natick, MA: MathWorks Inc.; 2016.
56. Bardesley JM, Parker A, Solonen A, Howard M. Krylov space approximate Kalman filtering. Numer Linear Algebra Appl. 2013;20:171–184.


**How to cite this article:** Tabeart JM, Dance SL, Haben SA, Lawless AS, Nichols NK, Waller JA. The conditioning of least-squares problems in variational data assimilation. *Numer Linear Algebra Appl*. 2018;e2165. https://doi.org/10.1002/nla.2165

## APPENDIX: PROOFS

In this section, we present the proofs for Lemmas 3 and 4 (Section 5.6), in which we express **S** as the difference between a circulant matrix and a singular matrix in order to bound the eigenvalues of **S** above.

*Proof of Lemma 3.* We exploit the structure of $\mathbf{S}$ that arises from the choice of $\mathbf{H}$; entries from $\mathbf{R}^{-1}$ are only added to the top left $p \times p$ block of $\mathbf{B}^{-1}$. Let $\mathbf{C}_1$ be the circulant matrix generated by the first row of $\mathbf{S}_1$. Then, for $i = 1, \dots, N$,

$$\mathbf{C}_1(1,i) = \mathbf{B}^{-1}(1,i) + \left(\mathbf{H}_1^T \mathbf{R}^{-1} \mathbf{H}_1\right)(1,i) = \begin{cases} \mathbf{B}^{-1}(1,i) + \mathbf{R}^{-1}(1,i) & \text{for } i = 1, \dots, p \\ \mathbf{B}^{-1}(1,i) & \text{for } i = p+1, \dots, N. \end{cases} \tag{A1}$$

Let $\widetilde{\mathbf{H}}_1$ be given by

$$\widetilde{\mathbf{H}}_1(i,j) = \begin{cases} 1 & \text{for } j = i, \quad i = p+1, \dots, N \\ 0 & \text{otherwise.} \end{cases} \tag{A2}$$

Then, we can write $\mathbf{S}_1 = \mathbf{C}_1 - \widetilde{\mathbf{H}}_1^T \mathbf{R}^{-1} \widetilde{\mathbf{H}}_1$. Applying (5), we obtain

$$\lambda_k(\mathbf{S}) \leq \lambda_k(\mathbf{C}_1) + \lambda_1(-\widetilde{\mathbf{H}}_1 \mathbf{R}^{-1} \widetilde{\mathbf{H}}_1). \tag{A3}$$

As $\widetilde{\mathbf{H}}_1^T \mathbf{R}^{-1} \widetilde{\mathbf{H}}_1$ is not full rank and is positive semidefinite, its smallest eigenvalue is 0. Hence, $\lambda_1(-\widetilde{\mathbf{H}}_1^T \mathbf{R}^{-1} \widetilde{\mathbf{H}}_1) = -\lambda_N(\widetilde{\mathbf{H}}_1^T \mathbf{R}^{-1} \widetilde{\mathbf{H}}_1) = 0$, and we have that

$$\lambda_k(\mathbf{S}_1) \leq \lambda_k(\mathbf{C}_1). \tag{A4}$$

As $\mathbf{C}_1$ is circulant, we calculate its eigenvalues via a direct Fourier transform (20). In the order of wavenumber, the eigenvalues of $\mathbf{C}_1$ are given by

$$\gamma_m(\mathbf{C}_1) = \sum_{k=0}^{p-1} \omega^{mk} \left(\mathbf{B}_{1,k}^{-1} + \mathbf{R}_{1,k}^{-1}\right) + \sum_{k=p}^{N-1} \omega^{km} \mathbf{B}_{1,k}^{-1} \quad m = 0, \dots, N, \tag{A5}$$

where $\omega = e^{2\pi i/N}$ is an $N$th root of unity. Separating the contributions of $\mathbf{B}^{-1}$ and $\mathbf{R}^{-1}$ yields

$$\gamma_m(\mathbf{C}_1) = \sum_{k=0}^{N-1} \omega^{mk} \mathbf{B}_{1,k}^{-1} + \sum_{k=0}^{p-1} \omega^{mk} \mathbf{R}_{1,k}^{-1}. \tag{A6}$$

□

*Proof of Lemma 4.* We follow the same arguments as the proof for Lemma 3 above. □