New bounds on the condition number of the Hessian of the preconditioned variational data assimilation problem

Received: 19 October 2020 | Revised: 21 April 2021 | Accepted: 8 July 2021

DOI: 10.1002/nla.2405

RESEARCH ARTICLE

WILEY

New bounds on the condition number of the Hessian of the preconditioned variational data assimilation problem

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

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

[2]National Centre for Earth Observation, Reading, UK

[3]School of Mathematics, University of Edinburgh, Edinburgh, UK

[4]MetOffice@Reading, Reading, UK

Correspondence
Jemima M. Tabeart, School of Mathematics, University of Edinburgh, James Clerk Maxwell Building, The King's Buildings, Peter Guthrie Tait Road, Edinburgh EH9 3FD, UK.
Email: jemima.tabeart@ed.ac.uk

Funding information
Engineering and Physical Sciences Research Council, Grant/Award Numbers: EP/P002331/1, EP/S027785/1; Natural Environment Research Council, Grant/Award Number: NE/K008900/1; National Centre for Earth Observation

Abstract

Data assimilation algorithms combine prior and observational information, weighted by their respective uncertainties, to obtain the most likely posterior of a dynamical system. In variational data assimilation the posterior is computed by solving a nonlinear least squares problem. Many numerical weather prediction (NWP) centers use full observation error covariance (OEC) weighting matrices, which can slow convergence of the data assimilation procedure. Previous work revealed the importance of the minimum eigenvalue of the OEC matrix for conditioning and convergence of the unpreconditioned data assimilation problem. In this article we examine the use of correlated OEC matrices in the preconditioned data assimilation problem for the first time. We consider the case where there are more state variables than observations, which is typical for applications with sparse measurements, for example, NWP and remote sensing. We find that similarly to the unpreconditioned problem, the minimum eigenvalue of the OEC matrix appears in new bounds on the condition number of the Hessian of the preconditioned objective function. Numerical experiments reveal that the condition number of the Hessian is minimized when the background and observation lengthscales are equal. This contrasts with the unpreconditioned case, where decreasing the observation error lengthscale always improves conditioning. Conjugate gradient experiments show that in this framework the condition number of the Hessian is a good proxy for convergence. Eigenvalue clustering explains cases where convergence is faster than expected.

KEYWORDS

condition number, correlated observation error covariance, data assimilation, Hessian, least squares, preconditioning

1 | INTRODUCTION

Data assimilation algorithms combine observations of a dynamical system, $\mathbf{y}_i \in \mathbb{R}^{p_i}$ at times t_i, with prior information from a model, $\mathbf{x}^b \in \mathbb{R}^N$ to find $\mathbf{x}_i \in \mathbb{R}^N$, the most likely state of the system at time t_i. In variational data assimilation the posterior is computed by solving a nonlinear least squares problem. In this article we examine the effect of using correlated observation error covariance (OEC) matrices on the convergence of the preconditioned variational data assimilation problem. We develop new bounds on the condition number of the Hessian of the linearized preconditioned objective function. Numerical experiments allow us to compare the bounds to the computed condition number. We also investigate the relationship between conditioning, the full spectrum of the Hessian and convergence of a linear data assimilation test problem to assess the suitability of using the condition number of the Hessian as a proxy for convergence in this setting.

We now define the variational data assimilation objective function of interest for this article. For a time window $[t_0, t_n]$, we let $\mathbf{x}_i^t \in \mathbb{R}^N$ be the true state of the dynamical system of interest at time t_i, where N is the number of state variables. The prior, or background state, is valid at the initial time t_0 and can be written as an approximation to the true state as $\mathbf{x}^b = \mathbf{x}_0^t + \epsilon^b$. We assume that the background errors $\epsilon^b \sim \mathcal{N}(0, \mathbf{B})$, where $\mathbf{B} \in \mathbb{R}^{N\times N}$ is the background error covariance matrix. As observations can be made at different locations, or of different variables to those in the state vector $\mathbf{x}_i$, we define an observation operator $h_i : \mathbb{R}^N \to \mathbb{R}^{p_i}$ which maps from state variable space to observation space at time t_i. Observations $\mathbf{y}_i \in \mathbb{R}^{p_i}$ at time t_i are similarly expressed as $\mathbf{y}_i = h_i[\mathbf{x}_i^t] + \epsilon_i$ for $i = 0, \dots, n$: the sum of the model equivalent $h_i[\mathbf{x}_i^t]$ and an observation error $\epsilon_i \sim \mathcal{N}(0, \mathbf{R}_i)$, where $\mathbf{R}_i \in \mathbb{R}^{p_i \times p_i}$ are the OEC matrices. We additionally assume that the observation and background errors are mutually uncorrelated. The total number of observations across the whole time window is given by $p = \sum_{i=0}^n p_i$. The state $\mathbf{x}_{i-1}$ at time t_{i-1} is propagated to the next observation time t_i using a nonlinear forecast model operator, $\mathcal{M}$, to obtain

$$\mathbf{x}_i = \mathcal{M}(t_{i-1}, t_i; \mathbf{x}_{i-1}). \tag{1}$$

In variational data assimilation the analysis, $\mathbf{x}_0$, or most likely state at the initial time t_0, minimizes the full 4D-Var objective function, given by

$$J(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^b) + \frac{1}{2}\sum_{i=0}^{n} (\mathbf{y}_i - h_i[\mathbf{x}_i])^T \mathbf{R}_i^{-1}(\mathbf{y}_i - h_i[\mathbf{x}_i]). \tag{2}$$

In applications such as numerical weather prediction (NWP), the nonlinear objective function (2) is typically minimized using an iterative method. The most common implementation is the incremental formulation, which solves the variational data assimilation problem via a small number of nonlinear outer loops, and a larger number of inner loop iterations which minimize a linearized least squares problem.[1] This procedure is equivalent to a Gauss–Newton method[2-4] and will be presented in Section 2.

For many systems in the geosciences and neurosciences[5,6] the number of state variables, N, can be of the order of 10^9. In this article we consider the case where the number of state variables is greater than the number of observations, that is, $N > p$, an assumption which holds for applications with sparse measurement data. The large dimension of the state space motivates the use of a control variable transform (CVT) to model the background error covariance matrix, $\mathbf{B}$, implicitly.[7] The CVT uses the square root of $\mathbf{B}$ as a variable transform to obtain a modified objective function [8, sec 9.1], and can be interpreted as a form of preconditioning. The transformation diagonalizes the weighting on the first term of (2), making the transformed state variables uncorrelated. We refer to the incremental variational problem with the CVT as the preconditioned data assimilation problem for the remainder of this article.

As the inner iterations of the incremental 4D-Var algorithm solve a linear least squares problem, the conjugate gradient method can be used for the minimization of the linearized objective function.[9-11] Convergence of a conjugate gradient method can be bounded by the condition number of the Hessian of the objective function;[12-14] therefore the condition number of the linearized Hessian can be considered as a proxy to study how changes to a data assimilation method are likely to affect convergence of the inner loop. The Hessian of the linearized preconditioned objective function is given by

$$\hat{\mathbf{S}} = \mathbf{I} + \sum_{i=0}^{n} \mathbf{B}^{-1/2}\mathbf{M}_i^T\mathbf{H}_i^T\mathbf{R}_i^{-1}\mathbf{H}_i\mathbf{M}_i\mathbf{B}^{-1/2}, \tag{3}$$

where $\mathbf{H}_i \in \mathbb{R}^{p_i\times N}$ is the linearized observation operator, and $\mathbf{M}_i \in \mathbb{R}^{N\times N}$ is the linearization of the model operator (1). The Hessian (3) is a low rank update of the identity matrix, and hence is typically better conditioned than the Hessian corresponding to the unpreconditioned problem (as its minimum eigenvalue is one). However, the distribution of the full spectrum, and not just the extreme eigenvalues, is important for the conjugate gradient method (see [12, theorem 38.4; 15, theorems 38.3, 38.5]). In this article we will therefore consider how the condition number of the Hessian relates to convergence of the conjugate gradient method in an idealized numerical framework, and examine the distribution of the full spectrum of (3).

In recent years there has been a rise in the introduction of correlated OEC matrices ($\mathbf{R}_i$ in (2)) at NWP centers (e.g., [16-18]). The use of correlated OEC matrices brings benefit to applications by allowing users to include more observations[19,20] at higher resolutions. Correlated OEC matrices also lead to greater information content of observations, particularly on smaller scales.[19,21-23] However, the move from uncorrelated (diagonal) to correlated (full) covariance matrices has caused problems with the convergence of the data assimilation procedure in experiments at NWP centers.[17,24,25] Previous studies of the conditioning of the preconditioned Hessian have focused on the case of uncorrelated OEC matrices.[14,26] In this article we extend this theory to the case of correlated OEC matrices.

Tabeart et al.[27] considered the effect of using correlated (full) OEC matrices within the unpreconditioned data assimilation problem. The minimum eigenvalue of the correlated OEC matrix was found to be important in determining the conditioning of the Hessian of the objective function both theoretically and numerically. The condition number of the Hessian was found to be a good proxy for convergence in this framework. Haben et al.[14,26] developed bounds on the condition number of the Hessian for both the unpreconditioned and preconditioned problems in the case of uncorrelated (diagonal) OEC matrices. In the preconditioned case, reducing the observation error variance increases both the bounds on and numerical value of the condition number of the Hessian. The choice of observation network was also shown to be important for determining the conditioning and convergence of the preconditioned problem.

In this article we consider the conditioning of the preconditioned variational data assimilation problem in the case of correlated OEC matrices. We extend the analysis of Tabeart et al.[27] to the preconditioned case where there are fewer observations than state variables, that is, $p < N$. We begin in Section 2 by defining the problem and introducing existing mathematical results relating to conditioning. In Section 3 we present new theoretical bounds on the condition number of the preconditioned Hessian in terms of its constituent matrices. In Section 4 we introduce the numerical framework that will be used for our experiments. We present the results of these experiments and related discussion in Section 5. These experiments reveal the ratio between background and observation error correlation lengthscales strongly influences the conditioning of the Hessian, with minimum condition numbers occurring when the two lengthscales are equal. This contrasts with the unpreconditioned case, where the condition number of the Hessian could always be reduced by decreasing the lengthscale of the observation error covariances. We find cases where the new bounds represent the qualitative behavior of the conditioning well, as well as cases where bounds from Haben[14] are tighter. For many cases the condition number of the Hessian is a good proxy for convergence of a conjugate gradient method. Cases where convergence is much faster than expected can be explained by a single large eigenvalue with the remainder clustering around unity. Our conclusions are presented in Section 6.

2 | THE PRECONDITIONED VARIATIONAL DATA ASSIMILATION PROBLEM

2.1 | The CVT formulation of the data assimilation problem

In this section we define the preconditioned 4D-Var data assimilation problem and introduce further notation that will be used in this article. We recall[28] that covariance matrices can be decomposed as $\mathbf{B} = \Sigma_B\widetilde{\mathbf{B}}\Sigma_B$, and $\mathbf{R}_i = \Sigma_{R_i}\widetilde{\mathbf{R}}_i\Sigma_{R_i}$ where Σ_B, Σ_{R_i} are diagonal matrices containing standard deviations, and $\widetilde{\mathbf{B}}, \widetilde{\mathbf{R}}_i$ are correlation matrices with unit entries on the diagonal. By definition covariance matrices are symmetric positive semidefinite. However, we will assume in what follows that $\widetilde{\mathbf{B}}, \widetilde{\mathbf{R}}_i, \mathbf{B}$, and $\mathbf{R}_i$ are strictly positive definite, and therefore their inverses are well defined.

We now derive the linearized incremental objective function. In this formulation instead of finding the state which minimizes the objective function (2) directly, subject to the model constraint (1), we minimize a sequence of linearizations of the objective function to obtain a sequence of increments to the background, $\mathbf{x}^b$. Typically this is done via a series of outer loops, where the forecast model and observation operators are linearized about the current best estimate of $\mathbf{x}_0$.

For the lth outer loop we define $\mathbf{x}_0^{(l+1)} = \mathbf{x}_0^{(l)} + \delta\mathbf{x}_0^{(l)}$. We then consider the Taylor expansion of $\mathcal{M}(t_{i-1}, t_i; \mathbf{x}_{i-1}^{(l)})$ and obtain the linearization $\delta\mathbf{x}_i^{(l)} = \mathbf{M}_i\delta\mathbf{x}_{i-1}^{(l)}$ where $\mathbf{M}_i \in \mathbb{R}^{N\times N}$ is the linearized model operator at time t_i, linearized about the model forecast initialized at $\mathbf{x}_0^{(l)}$. Finally we denote $\delta\mathbf{x}_b^{(l)} = \mathbf{x}^b - \mathbf{x}_0^{(l)}$, with $\mathbf{x}_0^{(0)} = \mathbf{x}^b$ and $\delta\mathbf{x}_0^{(0)} = 0$.

Similarly, expanding $h_i[\mathbf{x}_i]$ about $\mathbf{x}_i^{(l)}$ we obtain the linearization $h_i[\mathbf{x}_i^{(l)} + \delta\mathbf{x}_i^{(l)}] \approx h_i[\mathbf{x}_i^{(l)}] + \mathbf{H}_i\delta\mathbf{x}_i^{(l)}$ where $\mathbf{H}_i \in \mathbb{R}^{p_i\times N}$ is the linearized observation operator at time t_i linearized about $\mathbf{x}_i^{(l)}$.

We then write the linearized objective function in terms of $\delta\mathbf{x}_0^{(l)}$,

$$\tilde{J}(\delta\mathbf{x}_0^{(l)}) = \frac{1}{2}(\delta\mathbf{x}_0^{(l)} - \delta\mathbf{x}_b^{(l)})^T\mathbf{B}^{-1}(\delta\mathbf{x}_0^{(l)} - \delta\mathbf{x}_b^{(l)}) + \frac{1}{2}\sum_{i=0}^{n}(\mathbf{d}_i^{(l)} - \mathbf{H}_i\delta\mathbf{x}_i^{(l)})^T\mathbf{R}_i^{-1}(\mathbf{d}_i^{(l)} - \mathbf{H}_i\delta\mathbf{x}_i^{(l)}), \tag{4}$$

where $\mathbf{d}_i^{(l)} = \mathbf{y}_i - h_i[\mathbf{x}_i^{(l)}]$ are the innovation vectors. These measure the misfit between the observations and the linearized state, using the full nonlinear observation operator.

In order to simplify the notation in what follows we can group the linearized forecast model and observation operator terms together as a single linear operator. We define the generalized observation operator as

$$\widehat{\mathbf{H}} = \left[\mathbf{H}_0^T, (\mathbf{H}_1\widehat{\mathbf{M}}_1)^T, \dots , (\mathbf{H}_n\widehat{\mathbf{M}}_n)^T\right]^T \in \mathbb{R}^{N(n+1)\times p(n+1)}, \tag{5}$$

where the linearized forward model from time t_0 to time t_i is given by

$$\widehat{\mathbf{M}}_i\delta\mathbf{x}_0^{(l)} = \mathbf{M}_i \dots \mathbf{M}_1\delta\mathbf{x}_0^{(l)}. \tag{6}$$

Finally we let $\widehat{\mathbf{R}} \in \mathbb{R}^{p\times p}$ denote the block diagonal matrix with the ith block consisting of $\mathbf{R}_i$. This allows us to write the Hessian of the linearized objective function, (4), in the simplified form

$$\mathbf{S} = \mathbf{B}^{-1} + \widehat{\mathbf{H}}^T\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{H}}. \tag{7}$$

The formulation of the objective function given by (4) is too expensive to be used in practice both in terms of computation, but also storage. The number of state variables, N, is very large and typically $\mathbf{B}$ cannot be stored explicitly.

The CVT formulates the objective function in terms of alternative “control variables,” which means that the background matrix $\mathbf{B}$ does not need to be stored explicitly. The CVT is described in detail by Bannister,[7,29] and is often used in NWP applications.

The CVT may be applied to the incremental form of the variational problem (4), via the change of variable $\delta\mathbf{z}_0^{(l)} = \mathbf{B}^{-1/2}\delta\mathbf{x}_0^{(l)}$. This yields the objective function

$$\mathcal{J}(\delta\mathbf{z}_0^{(l)}) = \frac{1}{2}(\delta\mathbf{z}_0^{(l)} - \delta\mathbf{z}_b^{(l)})^T(\delta\mathbf{z}_0^{(l)} - \delta\mathbf{z}_b^{(l)}) + \frac{1}{2}\left(\widehat{\mathbf{d}}^{(l)} - \widehat{\mathbf{H}}\mathbf{B}^{1/2}\delta\mathbf{z}_0^{(l)}\right)^T\widehat{\mathbf{R}}^{-1}\left(\widehat{\mathbf{d}}^{(l)} - \widehat{\mathbf{H}}\mathbf{B}^{1/2}\delta\mathbf{z}_0^{(l)}\right), \tag{8}$$

where $\delta\mathbf{z}_b^{(l)} = \mathbf{B}^{-1/2}\delta\mathbf{x}_b^{(l)}$, and

$$\widehat{\mathbf{d}}^{(l)^T} = \left[\mathbf{d}_o^{(l)^T}, \mathbf{d}_1^{(l)^T}, \dots , \mathbf{d}_n^{(l)^T}\right] \tag{9}$$

is a vector made up of the innovation vectors.

This yields a Hessian for the incremental 4D-Var problem with the CVT given by

$$\widehat{\mathbf{S}} = \mathbf{I}_N + \mathbf{B}^{1/2}\widehat{\mathbf{H}}^T\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{H}}\mathbf{B}^{1/2}. \tag{10}$$

Therefore using the CVT is equivalent to pre- and postmultiplying the Hessian of the incremental data assimilation problem (7) by $\mathbf{B}^{1/2}$ (the uniquely defined, symmetric square root of $\mathbf{B}$). The exact value of $\mathbf{B}^{-1/2}$ is not computed, but rather an approximation is constructed using physical and statistical knowledge of the system of interest.[7] The CVT can be interpreted as preconditioning the Hessian by $\mathbf{B}^{1/2}$. The data assimilation formulation described in (8) is often referred to as the preconditioned data assimilation problem, and this naming convention will be used throughout the remainder of the article. We note that as we assume $\mathbf{B}$ and $\widehat{\mathbf{R}}$ are strictly positive definite, $\widehat{\mathbf{S}}$ is also symmetric positive definite.

The preconditioned Hessian (10) highlights the computational benefit of using the CVT. For most NWP applications there are fewer observations than state variables (typically a difference of two orders of magnitude[5]), meaning that the second term in (10) is rank deficient. Therefore the preconditioned Hessian is a low-rank update to the identity, and hence its minimum eigenvalue is unity. This guarantees that the preconditioned Hessian will not suffer from small minimum eigenvalues that often result in ill-conditioning for the unpreconditioned problem. This improved conditioning is expected to lead to faster convergence of the associated data assimilation algorithm.

In this article we study the conditioning of the Hessian of the CVT objective function (8) as a proxy for convergence of the preconditioned data assimilation problem. We develop bounds on the condition number of (10) in terms of its constituent matrices. Separating the contribution of each matrix in the bounds allows us to investigate the effect of changes to each component of the data assimilation system on conditioning and convergence. In particular, we focus on the introduction of correlated OEC matrices within the preconditioned framework.

2.2 | Some inequalities for the eigenvalues of the product of positive semidefinite Hermitian matrices

For the remainder of this article, we use the following order of eigenvalues: For a matrix $\mathbf{A} \in \mathbb{R}^{k\times k}$ let the eigenvalues λ_i be such that $\lambda_{\max}(\mathbf{A}) = \lambda_1(\mathbf{A}) \geq \lambda_2(\mathbf{A}) \geq \cdots \geq \lambda_k(\mathbf{A}) = \lambda_{\min}(\mathbf{A})$.

In this section we introduce theoretical results from linear algebra. These will be used in Section 3 to develop new bounds on the condition number of the preconditioned Hessian in terms of its constituent matrices. We also present existing bounds on the Hessian of the preconditioned 3D-Var problem. The 3D-Var problem is obtained by setting $n = 0$ in (2). In the numerical experiments of Section 5 we will compare these existing bounds with the new bounds developed in Section 3.

We begin by formally defining the condition number.

Definition 1 (13, sec. 2.7.2). For $\mathbf{A} \in \mathbb{R}^{k\times k}$ symmetric positive definite we define the condition number $\kappa(\mathbf{A})$ by

$$\kappa(\mathbf{A}) = \|\mathbf{A}\|\|\mathbf{A}^{-1}\|. \tag{11}$$

We then characterize the condition number in the 2-norm of $\mathbf{A}$ as

$$\kappa_2(\mathbf{A}) = \|\mathbf{A}\|_2\|\mathbf{A}^{-1}\|_2 = \frac{\lambda_1(\mathbf{A})}{\lambda_k(\mathbf{A})}, \tag{12}$$

where λ_i are the eigenvalues of $\mathbf{A}$. The condition number in the 2-norm shall be referred to as the condition number and denoted $\kappa(\mathbf{A})$ for the remainder of this work.

We recall our additional assumption that both $\mathbf{B}$ and $\widehat{\mathbf{R}}$ are strictly positive definite. Since $\widehat{\mathbf{S}}$ is symmetric positive definite we apply the characterization of the condition number given by (12) throughout this article.

We present two results which bound the eigenvalues of a matrix product in terms of the product of the eigenvalues of the individual matrices. These will be used in Section 3 to separate the contribution of the background and OEC matrices to $\kappa(\widehat{\mathbf{S}})$.

Theorem 1. *Let $\mathbf{F}, \mathbf{G} \in \mathbb{C}^{d\times d}$ be positive semidefinite Hermitian matrices and let $i_1, \dots i_k$ denote an ordered subset of the integers $\{1, \dots, d\}$. Then*

$$\sum_{t=1}^{k} \lambda_{i_t}(\mathbf{FG}) \leq \sum_{t=1}^{k} \lambda_{i_t}(\mathbf{F})\lambda_t(\mathbf{G}), \quad k = 1, \dots, d-1. \tag{13}$$

Proof. The proof is given by Wang and Zhang,[30] theorem 3. ■

Theorem 2. *Let $\mathbf{F}, \mathbf{G} \in \mathbb{C}^{d\times d}$ be positive semidefinite Hermitian and let $i_1, \dots i_k$ denote an ordered subset of the integers $\{1, \dots, d\}$. Then*

$$\sum_{t=1}^{k} \lambda_{i_t}(\mathbf{FG}) \geq \sum_{t=1}^{k} \lambda_{i_t}(\mathbf{F})\lambda_{d-t+1}(\mathbf{G}). \tag{14}$$

Proof. The proof is given by Wang and Zhang,[30] theorem 4. ■

These results will be used to develop bounds on the condition number of the Hessian (10).

We now present an existing bound on the condition number of the 3D-Var preconditioned Hessian, $\kappa(\widehat{\mathbf{S}})$, from Haben.[14]

Theorem 3. *Let* $\mathbf{B} \in \mathbb{R}^{N\times N}$ *be the background error covariance matrix and* $\mathbf{R} \in \mathbb{R}^{p\times p}$ *be the OEC matrix with* $p < N$. *Then the following bounds are satisfied by the condition number of the preconditioned 3D-Var Hessian* $\widehat{\mathbf{S}} = \mathbf{I}_N + \mathbf{B}^{1/2}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B}^{1/2}$

$$1 + \frac{1}{p}\sum_{i,j=1}^{p} \left(\mathbf{R}^{-1/2}\mathbf{H}\mathbf{B}\mathbf{H}^T\mathbf{R}^{-1/2}\right)_{i,j} \leq \kappa(\widehat{\mathbf{S}}) \leq 1 + \left\|\mathbf{R}^{-1/2}\mathbf{H}\mathbf{B}\mathbf{H}^T\mathbf{R}^{-1/2}\right\|_\infty. \tag{15}$$

Proof. The proof is given by Haben,[14] theorem 6.2.1. ■

We note that the result of Theorem 3 extends naturally to the 4D-Var problem by replacing $\mathbf{R}$ with $\widehat{\mathbf{R}}$ and $\mathbf{H}$ with $\widehat{\mathbf{H}}$. As discussed at the end of Section 2.1, we want to develop bounds that separate the contribution of each constituent matrix. This will allow us to study how altering a single term, particularly the OEC matrix, is likely to affect the conditioning and convergence of the preconditioned data assimilation system. As well as being interesting from a theoretical perspective, improved understanding of the influence of individual terms will be useful for practical applications. For example, when introducing new observation operators or OEC matrices into operational systems, bounds which separate the role of each matrix will provide insight into how the conditioning of the preconditioned 4D-Var problem is likely to change. However, as the bounds given by (15) do not separate out each term, they are likely to be tighter than the new bounds which are presented in Section 3. In Section 5 we will numerically compare the bounds given by (15) with those developed in Section 3.

3 | THEORETICAL BOUNDS ON THE HESSIAN OF THE PRECONDITIONED PROBLEM

In this section we develop new theoretical bounds on the condition number of the Hessian of the preconditioned variational data assimilation problem, following similar methods to the unpreconditioned case in Tabeart et al..[27] These bounds will all be presented in terms of the Hessian of the preconditioned 4D-Var problem (10). For the case $n = 0$, $\widehat{\mathbf{R}} \equiv \mathbf{R}_0 \in \mathbb{R}^{p_0\times p_0}$ and $\widehat{\mathbf{H}} \equiv \mathbf{H}_0 \in \mathbb{R}^{p_0\times N}$, meaning that the bounds in this section will also apply directly to the preconditioned 3D-Var Hessian. This relation will be used in the numerical experiments presented in Section 5.

Key Assumption. The total number of observations across the time window, p, is smaller than the number of state variables, that is, $p < N$.

The first result shows that the condition number can be calculated via the eigenvalues of the rank-p update $\mathbf{B}^{1/2}\widehat{\mathbf{H}}^T\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{H}}\mathbf{B}^{1/2}$.

Lemma 1. *Following the Key Assumption we can express the condition number of* $\widehat{\mathbf{S}}$ *as*

$$\kappa(\widehat{\mathbf{S}}) = 1 + \lambda_1(\mathbf{B}\widehat{\mathbf{H}}^T\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{H}}) \tag{16}$$

$$= 1 + \lambda_1(\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{H}}\mathbf{B}\widehat{\mathbf{H}}^T). \tag{17}$$

Proof. We begin by showing that $\kappa(\widehat{\mathbf{S}}) = 1 + \lambda_1(\mathbf{B}^{1/2}\widehat{\mathbf{H}}^T\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{H}}\mathbf{B}^{1/2})$, as was presented in Haben,[14] equation (4.2). We define $\mathbf{B}^{1/2}\widehat{\mathbf{H}}^T\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{H}}\mathbf{B}^{1/2} = \mathbf{C}$ and write $\widehat{\mathbf{S}} = \mathbf{I} + \mathbf{C}$. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$ be the eigenvalues of $\mathbf{C}$, with corresponding eigenvectors v_i. As $p < N$, $\mathbf{C}$ is rank deficient and therefore $\lambda_N = 0$.

Therefore $\lambda_N(\widehat{\mathbf{S}}) = 1$, and $\kappa(\widehat{\mathbf{S}}) = \lambda_1(\widehat{\mathbf{S}}) = 1 + \lambda_1(\mathbf{C})$. Matrices $\mathbf{AB}$ and $\mathbf{BA}$ have the same nonzero eigenvalues,[31] and therefore we can write

$$\lambda_1(\mathbf{C}) = \lambda_1(\mathbf{B}^{1/2}\widehat{\mathbf{H}}^T\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{H}}\mathbf{B}^{1/2}) = \lambda_1(\mathbf{B}\widehat{\mathbf{H}}^T\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{H}}) = \lambda_1(\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{H}}\mathbf{B}\widehat{\mathbf{H}}^T). \tag{18}$$

Hence, we obtain the result

$$\kappa(\widehat{\mathbf{S}}) = 1 + \lambda_1(\mathbf{B}\widehat{\mathbf{H}}^T\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{H}}) = 1 + \lambda_1(\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{H}}\mathbf{B}\widehat{\mathbf{H}}^T). \tag{19}$$

■

The result of Lemma 1 shows that computing $\kappa(\widehat{\mathbf{S}})$ only requires the computation of the maximum eigenvalue of a single matrix product. We also note that the matrix products that appear in (16) and (17) are of different dimensions: $\mathbf{B}\widehat{\mathbf{H}}^T\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{H}} \in \mathbb{R}^{N\times N}$ and $\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{H}}\mathbf{B}\widehat{\mathbf{H}}^T \in \mathbb{R}^{p\times p}$. Additionally, by the Key Assumption (as $p < N$) the first matrix product is always rank deficient, whereas for the case that $\widehat{\mathbf{H}}^T\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{H}}$ is rank p, the second matrix product is full rank.

Previous studies[14,27] have considered the effect of separately changing the variances and correlations associated with the background and observation covariances. In our numerical experiments in Section 5, we will focus on the role of the correlations in $\mathbf{B}$ and $\mathbf{R}$ in the conditioning of the preconditioned assimilation problem and assume the variances are constant, that is, $\mathbf{B} = \sigma_B^2\widetilde{\mathbf{B}}$, $\mathbf{R} = \sigma_{R_i}^2\widetilde{\mathbf{R}}_i$, where $\sigma_{R_i}, \sigma_B \in \mathbb{R}$. In that case it is known[26] that $\kappa(\widehat{\mathbf{S}})$ increases and decreases with the ratio of the background variance to the observation variance. We therefore assume in the experiments that the variances all take unit values and examine how changes to the background and observation correlations affect the conditioning.

3.1 | General bounds on the condition number

We now develop bounds on the condition number of $\widehat{\mathbf{S}}$ in terms of its constituent matrices. We assume that $\mathbf{B}$ and $\widehat{\mathbf{R}}$ are strictly positive definite, and that the Key Assumption holds. Otherwise we make no further restrictions on the structure of the constituent matrices in this section.

Theorem 4. *Given the Key Assumption we can bound* $\kappa(\widehat{\mathbf{S}}) = \kappa(\mathbf{I}_N + \mathbf{B}^{1/2}\widehat{\mathbf{H}}^T\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{H}}\mathbf{B}^{1/2})$ *by*

$$\begin{aligned} 1 + \max\left\{ \lambda_1(\widehat{\mathbf{H}}^T\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{H}})\lambda_N(\mathbf{B}), \quad \frac{\lambda_1(\widehat{\mathbf{H}}\mathbf{B}\widehat{\mathbf{H}}^T)}{\lambda_1(\widehat{\mathbf{R}})}, \quad \frac{\lambda_p(\widehat{\mathbf{H}}\mathbf{B}\widehat{\mathbf{H}}^T)}{\lambda_p(\widehat{\mathbf{R}})} \right\} \\ \leq \kappa(\widehat{\mathbf{S}}) \leq 1 + \min\left\{ \lambda_1(\mathbf{B})\lambda_1(\widehat{\mathbf{H}}^T\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{H}}), \quad \frac{\lambda_1(\widehat{\mathbf{H}}\mathbf{B}\widehat{\mathbf{H}}^T)}{\lambda_p(\widehat{\mathbf{R}})} \right\}. \end{aligned} \tag{20}$$

Proof. We write $\kappa(\widehat{\mathbf{S}})$ as in the statement of Lemma 1. To obtain the upper bound of (20), we use the result of Theorem 1 to separate the contribution of the background and observation term

$$\kappa(\widehat{\mathbf{S}}) = 1 + \lambda_1(\mathbf{B}\widehat{\mathbf{H}}^T\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{H}}) \leq 1 + \lambda_1(\mathbf{B})\lambda_1(\widehat{\mathbf{H}}^T\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{H}}). \tag{21}$$

Similarly the alternative formulation from Lemma 1 yields

$$\begin{aligned} \kappa(\widehat{\mathbf{S}}) &= 1 + \lambda_1(\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{H}}\mathbf{B}\widehat{\mathbf{H}}^T) \\ &\leq 1 + \frac{1}{\lambda_p(\widehat{\mathbf{R}})}\lambda_1(\widehat{\mathbf{H}}\mathbf{B}\widehat{\mathbf{H}}^T). \end{aligned} \tag{22}$$

Combining these two expressions yields the upper bound in the theorem statement.

To compute the lower bound of (20), we apply the result of Theorem 2 to (16) with $k = 1, i_1 = 1, d = N$. This yields

$$\begin{aligned} \lambda_1(\mathbf{B}\widehat{\mathbf{H}}^T\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{H}}) &\geq \max\left\{ \lambda_1(\widehat{\mathbf{H}}^T\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{H}})\lambda_N(\mathbf{B}), \quad \lambda_N(\widehat{\mathbf{H}}^T\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{H}})\lambda_1(\mathbf{B}) \right\} \\ &\geq \lambda_1(\widehat{\mathbf{H}}^T\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{H}})\lambda_N(\mathbf{B}). \end{aligned} \tag{23}$$

This last inequality is due to the fact that $\widehat{\mathbf{H}}^T\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{H}}$ is rank deficient. It follows from fact 5.11.14 of Bernstein[32] that $\lambda_i(\widehat{\mathbf{R}}^{-1}) = \frac{1}{\lambda_{p-i+1}(\widehat{\mathbf{R}})}$. Applying the result of Theorem 2 to (17) with $k = 1, i_1 = 1, d = N$ we obtain

$$\lambda_1(\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{H}}\mathbf{B}\widehat{\mathbf{H}}^T) \geq \max\left\{ \frac{\lambda_1(\widehat{\mathbf{H}}\mathbf{B}\widehat{\mathbf{H}}^T)}{\lambda_1(\widehat{\mathbf{R}})}, \quad \frac{\lambda_p(\widehat{\mathbf{H}}\mathbf{B}\widehat{\mathbf{H}}^T)}{\lambda_p(\widehat{\mathbf{R}})} \right\}. \tag{24}$$

Combining the results of (21)–(24) yields (20) as required. ■

We can separate the contribution of the OEC matrix from the observation operator to give the following bound.

Corollary 1. *Under the same conditions as in Theorem 4, we can bound $\kappa(\widehat{\mathbf{S}})$ by*

$$1+\max\left\{\frac{\lambda_p(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^T)\lambda_N(\mathbf{B})}{\lambda_p(\widehat{\mathbf{R}})},\frac{\lambda_1(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^T)\lambda_N(\mathbf{B})}{\lambda_1(\widehat{\mathbf{R}})}\right\}\le\kappa(\widehat{\mathbf{S}})\le 1+\frac{\lambda_1(\mathbf{B})}{\lambda_p(\widehat{\mathbf{R}})}\lambda_1(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^T). \tag{25}$$

Proof. We begin by considering the upper bound of (20). By theorem 21.10 of Harville,[31] $\widehat{\mathbf{H}}^T\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{H}}$ has precisely the same nonzero eigenvalues as $\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{H}}\widehat{\mathbf{H}}^T$. It follows from fact 5.11.14 of Bernstein[32] that $\lambda_i(\widehat{\mathbf{R}}^{-1})=\frac{1}{\lambda_{p-i+1}(\widehat{\mathbf{R}})}$. Applying Theorem 1 for $k=1, i_1=1, d=p$ to $\lambda_1(\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{H}}\widehat{\mathbf{H}}^T)$ yields:

$$\lambda_1(\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{H}}\widehat{\mathbf{H}}^T)\le\frac{\lambda_1(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^T)}{\lambda_p(\widehat{\mathbf{R}})}. \tag{26}$$

By theorem 21.10 of Harville,[31] $\widehat{\mathbf{H}}\mathbf{B}\widehat{\mathbf{H}}^T$ has precisely the same nonzero eigenvalues as $\mathbf{B}\widehat{\mathbf{H}}^T\widehat{\mathbf{H}}$. Applying Theorem 1 for $k=1, i_1=1, d=N$ yields:

$$\lambda_1(\mathbf{B}\widehat{\mathbf{H}}^T\widehat{\mathbf{H}})\le\lambda_1(\mathbf{B})\lambda_1(\widehat{\mathbf{H}}^T\widehat{\mathbf{H}})=\lambda_1(\mathbf{B})\lambda_1(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^T). \tag{27}$$

The final equality arises as the nonzero eigenvalues of $\widehat{\mathbf{H}}\widehat{\mathbf{H}}^T$ are equal to those of $\widehat{\mathbf{H}}^T\widehat{\mathbf{H}}$. Therefore the two cases from Theorem 4 yield the same "factorized" upper bound, and gives the upper bound in (25).

We now consider the first term in the lower bound of (20) and bound $\lambda_1(\widehat{\mathbf{H}}^T\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{H}})$ below. We separate the contribution of $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{H}}\widehat{\mathbf{H}}^T$ using Theorem 2 for $k=1, i_1=1, d=p$. This yields

$$\lambda_1(\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{H}}\widehat{\mathbf{H}}^T)\ge\max\left\{\frac{\lambda_1(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^T)}{\lambda_1(\widehat{\mathbf{R}})},\frac{\lambda_p(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^T)}{\lambda_p(\widehat{\mathbf{R}})}\right\}. \tag{28}$$

Multiplying this by $\lambda_N(\mathbf{B})$ gives the two terms that appear in the lower bound of (25).

We now consider the second term of (20) and bound $\lambda_1(\widehat{\mathbf{H}}\mathbf{B}\widehat{\mathbf{H}}^T)$ below. We separate the contribution of $\mathbf{B}$ and $\widehat{\mathbf{H}}^T\widehat{\mathbf{H}}$ using Theorem 2 for $k=1, i_1=1, d=N$. This yields

$$\begin{aligned}\lambda_1(\mathbf{B}\widehat{\mathbf{H}}^T\widehat{\mathbf{H}})&\ge\max\left\{\lambda_1(\mathbf{B})\lambda_N(\widehat{\mathbf{H}}^T\widehat{\mathbf{H}}),\lambda_N(\mathbf{B})\lambda_1(\widehat{\mathbf{H}}^T\widehat{\mathbf{H}})\right\}\\&\ge\lambda_N(\mathbf{B})\lambda_1(\widehat{\mathbf{H}}^T\widehat{\mathbf{H}}).\end{aligned} \tag{29}$$

The last inequality follows as $\widehat{\mathbf{H}}^T\widehat{\mathbf{H}}$ is not full rank and therefore $\lambda_N(\widehat{\mathbf{H}}^T\widehat{\mathbf{H}})=0$. Multiplying this result by $1/\lambda_1(\widehat{\mathbf{R}})$ gives the same value as the second term in (25).

Finally, we bound the third term of the lower bound in (20). By theorem 21.10 of Harville,[31] $\lambda_p(\widehat{\mathbf{H}}\mathbf{B}\widehat{\mathbf{H}}^T)=\lambda_p(\mathbf{B}\widehat{\mathbf{H}}^T\widehat{\mathbf{H}})$. Applying Theorem 2 for $k=1, i_1=p, d=N$ yields

$$\begin{aligned}\lambda_p(\mathbf{B}\widehat{\mathbf{H}}^T\widehat{\mathbf{H}})&\ge\max\{\lambda_p(\mathbf{B})\lambda_N(\widehat{\mathbf{H}}^T\widehat{\mathbf{H}}),\lambda_N(\mathbf{B})\lambda_p(\widehat{\mathbf{H}}^T\widehat{\mathbf{H}})\}\\&\ge\lambda_N(\mathbf{B})\lambda_p(\widehat{\mathbf{H}}^T\widehat{\mathbf{H}}).\end{aligned} \tag{30}$$

Multiplying the second term of (30) by $1/\lambda_p(\widehat{\mathbf{R}})$ gives the first term in (25), as $\lambda_p(\widehat{\mathbf{H}}^T\widehat{\mathbf{H}})=\lambda_p(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^T)$. ∎

In general it is not possible to determine which term in the lower bound of (25) is larger, as this will depend on the choice of $\mathbf{B}, \widehat{\mathbf{H}}$, and $\widehat{\mathbf{R}}$. However, we are able to comment on how the bounds are likely to be altered by changes to individual matrices. As we increase $\lambda_p(\widehat{\mathbf{R}})$ both the upper bound and first term in the lower bound decrease. Increasing $\lambda_1(\widehat{\mathbf{R}})$ will lead to a decrease in the second term of the lower bound. As $\lambda_N(\mathbf{B})$ increases, the lower bound will increase but the upper bound will remain unchanged. Increasing $\lambda_1(\mathbf{B})$ will yield a larger upper bound and has no effect on the lower bound. Larger values of $\lambda_p(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^T)$ will lead to increases to the first term in the lower bound and larger values of $\lambda_1(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^T)$ will lead to increases of the upper bound and second term of the lower bound. In the experiments in Section 5 we will study how each of these terms change with interacting parameters, and assess which lower bound is tighter for a variety of situations.

3.2 | Bounds on the condition number in the case of circulant error covariance matrices

The theoretical bounds presented in Section 3.1 apply for any choice of observation and background error covariance matrices. However, for a given numerical framework, general bounds can typically be improved by exploiting specific structure of the matrices being used.[14] In this section we will show that under additional assumptions on the structure of the error covariance matrices and observation operator, the bounds given by (15) yield the exact value of $\kappa(\widehat{\mathbf{S}})$.

We begin by defining circulant matrices. Circulant matrices are a natural choice for spatial correlation matrices on a one-dimensional periodic domain, as they yield correlation matrices that are homogeneous and isotropic.[14] We will make use of this structure in the numerical experiments presented in Section 5.

Definition 2 (33). A circulant matrix $\mathbf{C} \in \mathbb{R}^{N \times N}$ is a matrix of the form

$$\mathbf{C} = \begin{pmatrix} c_0 & c_1 & c_2 & \dots & c_{N-2} & c_{N-1} \\ c_{N-1} & c_0 & c_1 & \dots & c_{N-3} & c_{N-2} \\ c_{N-2} & c_{N-1} & c_0 & \dots & c_{N-4} & c_{N-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ c_2 & c_3 & c_4 & \dots & c_0 & c_1 \\ c_1 & c_2 & c_3 & \dots & c_{N-1} & c_0 \end{pmatrix}.$$

One computationally beneficial property of circulant matrices is that their eigenvalues can be calculated directly via a discrete Fourier transform. As we shall see in Theorem 5, any circulant matrix of dimension N admits the same eigenvectors.

Theorem 5. *The eigenvalues of a circulant matrix* $\mathbf{C} \in \mathbb{R}^{N \times N}$*, as given by Definition 2, are given by*

$$\gamma_m = \sum_{k=0}^{N-1} c_k \omega^{mk}, \tag{31}$$

with corresponding eigenvectors

$$\mathbf{v}_m = \frac{1}{\sqrt{N}}(1, \omega^m, \dots, \omega^{m(N-1)}), \tag{32}$$

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

Proof. See Reference 34 for full derivation. ■

Our numerical experiments in Section 5 will use circulant background and OEC matrices. When both $\widehat{\mathbf{H}}\mathbf{B}\widehat{\mathbf{H}}^T$ and $\widehat{\mathbf{R}}$ are circulant, with some additional assumptions on the entries of matrix products, we can prove that the upper and lower bounds given by Theorem 3 are equal and yield the exact value of $\kappa(\widehat{\mathbf{S}})$.

Corollary 2. *If* $\widehat{\mathbf{H}}\mathbf{B}\widehat{\mathbf{H}}^T \in \mathbb{R}^{p \times p}$ *and* $\widehat{\mathbf{R}} \in \mathbb{R}^{p \times p}$ *are circulant matrices, and all of the entries of* $\widehat{\mathbf{R}}^{-1/2}\widehat{\mathbf{H}}\mathbf{B}\widehat{\mathbf{H}}^T\widehat{\mathbf{R}}^{-1/2}$ *are positive, then the upper and lower bounds in Theorem 3 are equal, and the bound on* $\kappa(\widehat{\mathbf{S}})$ *is exact.*

Proof. The product of circulant matrices is a circulant matrix, the inverse of a circulant matrix is circulant,[34] and the square root of a circulant matrix is also circulant.[35] Therefore if the product $\widehat{\mathbf{H}}\mathbf{B}\widehat{\mathbf{H}}^T$ is circulant then the product $\widehat{\mathbf{R}}^{-1/2}\widehat{\mathbf{H}}\mathbf{B}\widehat{\mathbf{H}}^T\widehat{\mathbf{R}}^{-1/2}$ is circulant, as $\widehat{\mathbf{R}}$ is circulant by assumption of the corollary.

The lower bound of (15) computes the average row sum of the product $\widehat{\mathbf{R}}^{-1/2}\widehat{\mathbf{H}}\mathbf{B}\widehat{\mathbf{H}}^T\widehat{\mathbf{R}}^{-1/2}$. As the product is circulant, each row has the same sum, given by $\sum_{k=0}^{p-1} c_k$, where c_i is the ith entry of the first row of the circulant matrix (as introduced in Definition 2).

The upper bound of (15) returns the maximum absolute row sum of the product. As the product is circulant with only positive entries, all absolute row sums are identically equal to $\sum_{k=0}^{p-1}|c_k| = \sum_{k=0}^{p-1}c_k$. Hence, we have equality of lower and upper bounds and hence the exact value for $\kappa(\widehat{\mathbf{S}})$. ■

This result shows that, if the additional assumptions are satisfied, we can compute $\kappa(\widehat{\mathbf{S}})$ directly using (15). If $\mathbf{B}$ and $\widehat{\mathbf{R}}$ are both circulant, and the observed state variables are regularly spaced then the first assumption of Corollary 2 is satisfied. The requirement that all entries of $\widehat{\mathbf{R}}^{-1/2}\widehat{\mathbf{H}}\mathbf{B}\widehat{\mathbf{H}}^T\widehat{\mathbf{R}}^{-1/2}$ are positive is less straightforward to guarantee a priori, and depends on the specific structure of the three matrices being considered. In particular, even if all of the entries of $\widehat{\mathbf{R}}$, $\widehat{\mathbf{H}}$, and $\mathbf{B}$ are positive, entries of the product $\widehat{\mathbf{R}}^{-1/2}\widehat{\mathbf{H}}\mathbf{B}\widehat{\mathbf{H}}^T\widehat{\mathbf{R}}^{-1/2}$ can still be negative. The result of Corollary 2 will be used in the numerical experiments in the next section to compare the performance of the new bounds given by (25) and the existing bounds given by Theorem 3.

4 | NUMERICAL FRAMEWORK

In this section we describe the numerical framework that will be used to study how the bounds on the preconditioned Hessian (10) compare with the actual value of $\kappa(\widehat{\mathbf{S}})$. Although the bounds that were developed in Section 3 were developed for the 4D-Var problem, the numerical experiments presented in Section 5 will be conducted for a 3D-Var problem. This allows us to use the framework that was introduced in Tabeart et al.,[27] and directly compare the preconditioned and unpreconditioned formulations in the same numerical setting. We note that in the case of 3D-Var, $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{H}}$ simplify to the standard OEC matrix $\mathbf{R}$ and observation operator $\mathbf{H}$, respectively in (8), (10) and all the bounds in Section 3. The Hessian that is used for the experiments in this section is therefore given by $\widehat{\mathbf{S}} = \mathbf{I}_N + \mathbf{B}^{1/2}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B}^{1/2}$.

We now define the different components of the numerical framework. Our domain is the unit circle, and we fix the ratio of the number of state variables to observations as $N = 2p$, that is, twice as many state variables as observations. Similarly to Tabeart et al.[27] we define both the observation and background error covariance matrices to have a circulant structure with unit variances. Circulant matrices are a natural choice for correlations on a periodic domain with evenly distributed state variables. They also admit useful theoretical properties as was discussed in Section 3.2. The use of circulant error covariance matrices allow us to better understand the interaction between different terms in the Hessian, and to isolate the impact of parameter changes.

The experiments presented in this article will use circulant matrices arising from the second order autoregressive (SOAR) correlation function.[36,37] SOAR matrices are used in NWP applications as a horizontal correlation function[20] and are fully defined by a correlation lengthscale for a given domain. We remark that we substitute the great circle distance in the SOAR correlation function with the chordal distance[38,39] to ensure that the properties of positive definiteness are satisfied and that we obtain a valid correlation matrix.

Definition 3. The SOAR error correlation matrix on the unit circle is given by

$$\mathbf{D}(i,j) = \left(1 + \frac{\left|2\sin\left(\frac{\theta_{i,j}}{2}\right)\right|}{L}\right)\exp\left(\frac{-\left|2\sin\left(\frac{\theta_{i,j}}{2}\right)\right|}{L}\right), \tag{33}$$

where $L > 0$ is the correlation lengthscale and $\theta_{i,j}$ denotes the angle between grid points i and j. The chordal distance between adjacent grid points is given by

$$\Delta x = 2\sin\left(\frac{\theta}{2}\right) = 2\sin\left(\frac{\pi}{N}\right), \tag{34}$$

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

Both the background and OEC matrices for the experiments presented in Section 5 will be SOAR with constant unit variance. We will denote their respective lengthscales by L_B and L_R.

We now introduce the observation operators that will be used for the 3D-Var experiments. Three of our observation operators are the same as those used in Tabeart et al.[27] which we state again for clarity.

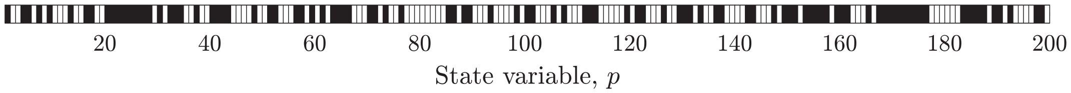

FIGURE 1 Representation of state variables that are observed for $\mathbf{H}_4$. Black denotes state variables that are observed directly and white denotes state variables that are not observed

Definition 4. 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:

$$\mathbf{H}_1(i,j) = \begin{cases} 1, & j = i \text{ for } i = 1, \dots, p \\ 0, & \text{otherwise.} \end{cases} \tag{35}$$

$$\mathbf{H}_2(i,j) = \begin{cases} 1, & j = 2i \text{ for } i = 1, \dots, p \\ 0, & \text{otherwise.} \end{cases} \tag{36}$$

$$\mathbf{H}_3(i,j) = \begin{cases} \frac{1}{5}, & j \in \{2i-2, 2i-1, 2i, 2i+1, 2i+2 \ (mod\ N)\} \text{ for } i = 1, \dots, p \\ 0, & \text{otherwise.} \end{cases} \tag{37}$$

The first choice of observation operator, $\mathbf{H}_1$, corresponds to direct observations of the first half of the domain. The second observation operator, $\mathbf{H}_2$, corresponds to direct observations of alternate state variables. The third observation operator, $\mathbf{H}_3$, is a smoothed version of $\mathbf{H}_2$. Observations of alternate state variables are smoothed equally over five adjacent state variables. The fourth choice of observation operator, $\mathbf{H}_4$, selects p random direct observations. We considered a number of choices of random observation operator, and all choices yielded similar numerical results. In order to ensure a fair comparison, we fix the same choice of $\mathbf{H}_4$ for all of the results presented in Section 5. This choice of observation operator is shown in Figure 1. Observations are spread over the whole domain, but are clustered rather than evenly distributed. Figure 2 of Tabeart et al.[27] shows a representation of a low dimensional version of the observation operator structure for $\mathbf{H}_1$, $\mathbf{H}_2$, and $\mathbf{H}_3$. For the numerical experiments in Section 5, we will use $p = 100$ observations and $N = 200$ state variables. In the unpreconditioned case, structure in the observation operator, such as regularly spaced observations, was important for the tightness of bounds and convergence of a conjugate gradient method.[27] We therefore consider $\mathbf{H}_4$ as an operator without strict structure. This will allow us to see how structure (or the lack of it) affects the preconditioned problem.

4.1 | Changes to the condition number of the Hessian

Our first set of experiments consider how different combinations of parameters will alter the value of $\kappa(\widehat{\mathbf{S}})$ and the bounds given by (25). We compute the condition number of the Hessian (10) using the Matlab 2018b function *cond*[40] and compare against the values given by our bounds. Table 1 (reproduced from Tabeart et al.[27]) shows that increasing the lengthscale of a SOAR correlation matrix will reduce its smallest eigenvalue and increase its largest eigenvalue. The maximum and minimum eigenvalues of both error covariance matrices appear in (25). We can therefore predict how the bounds will change with varying parameter values.

- As L_R (the lengthscale of the correlation function used to construct $\mathbf{R}$) increases, $\lambda_p(\mathbf{R})$ decreases. This means that both the upper bound and the first term in the lower bound of (25) will increase. However, $\lambda_1(\mathbf{R})$ increases with L_R meaning that the second term in the lower bound will decrease. It is therefore not possible to determine whether the lower bound will increase or decrease with increasing L_R in general.
- For the case of direct observations, all eigenvalues of $\mathbf{H}\mathbf{H}^T$ are equal to unity.[14] Therefore the first term in the lower bound of (25) will always be greater than the second term. Both $\mathbf{H}_1$ and $\mathbf{H}_2$ correspond to direct observations; hence for these choices of observation operator the first term in the lower bound of (25) is the lower bound of $\kappa(\widehat{\mathbf{S}})$.
- As L_B (the lengthscale of the correlation function used to construct $\mathbf{B}$) increases, $\lambda_1(\mathbf{B})$ increases. This means that the upper bound of (25) will increase with L_B. As L_B increases, $\lambda_N(\mathbf{B})$ decreases. This means that both terms in the lower bound of (25) will decrease with increasing L_B. Hence, the bounds (25) will diverge as L_B increases.

TABLE 1 Reproduction of Table 1 from Tabeart et al. [27]

	Lengthscale L_R or L_B				
	0.1	**0.33**	**0.66**	**0.99**	**1**
$\lambda_N(\mathbf{R})$	1.92×10^{-2}	5.74×10^{-4}	7.21×10^{-5}	2.14×10^{-5}	2.08×10^{-5}
$\lambda_1(\mathbf{R})$	6.40×10^{0}	2.26×10^{1}	4.67×10^{1}	6.36×10^{1}	6.40×10^{1}
$\lambda_N(\mathbf{B})$	2.54×10^{-3}	7.19×10^{-5}	8.99×10^{-6}	2.67×10^{-6}	2.59×10^{-6}
$\lambda_1(\mathbf{B})$	1.28×10^{1}	4.51×10^{1}	9.35×10^{1}	1.27×10^{2}	1.28×10^{2}

Note: Summary of changes to the eigenvalues of $\mathbf{B} \in \mathbb{R}^{200\times200}$ and $\mathbf{R} \in \mathbb{R}^{100\times100}$ with the lengthscales L_B and L_R for **B** and **R** both SOAR matrices

We wish to assess whether the qualitative behavior of $\kappa(\widehat{\mathbf{S}})$ agrees with the qualitative behavior of the bounds for our experimental framework. Additionally, we are interested in determining which term in the lower bound of (25) is largest, and whether this depends on the choice of **B**, **R**, and **H**.

In Section 5 we compare the bounds given by (25) with those of (15). As discussed at the end of Section 2, although we expect the bounds given by (15) to be tighter in many cases, separating the contribution of constituent matrices by using (25) will be qualitatively informative.

4.2 | Convergence of a conjugate gradient algorithm

Although conditioning of a problem is often used as a proxy to study convergence, there are well-known situations where the condition number provides a pessimistic indication of convergence speed, notably in the case of repeated or clustered eigenvalues (e.g., [12, theorem 38.4; 15, theorems 38.3, 38.5]). We therefore wish to investigate how well the condition number of the preconditioned system reflects the convergence of a conjugate gradient method for our experimental framework. Following a similar method to section 5.3.2. of Tabeart et al.[27] we study how the speed of convergence of a conjugate gradient method applied to the linear system $\widehat{\mathbf{S}}\mathbf{x} = \mathbf{b}$ changes with the parameters of the system. We define **x** as a vector with features at a variety of scales, and then calculate $\mathbf{b} = \widehat{\mathbf{S}}\mathbf{x}$ before recovering **x**. We use the Matlab 2018b routine *pcg.m* to recover **x** using the conjugate gradient method. As we are studying a preconditioned system, convergence is fast. In order to make the differences between parameter choices more evident we use a tolerance of 1×10^{-10} on the relative residual. In Section 5 we show results for one particular realization of **x** to enable a fair comparison between different choices of **R**, **B**, and **H**. A number of other values of **x** were tested, with similar results.

We consider how changes to lengthscale and observation operator alter the convergence of the conjugate gradient method. For cases where convergence behaves differently to conditioning, we study the spectrum of $\widehat{\mathbf{S}}$ to understand why these differences occur.

5 | 3D-VAR EXPERIMENTS

In this section we present the results of our numerical experiments. Figures will be plotted as a function of changes to correlation lengthscales for both **B** and **R**. We recall that increasing the lengthscale of a SOAR correlation matrix will reduce its smallest eigenvalue and increase its largest eigenvalue.[27,41]

Figure 2 shows how the condition number of the preconditioned Hessian (10) changes with the lengthscales of **B** and **R** for different choices of **H**. For $\mathbf{H}_1$, increasing L_R increases the value of $\kappa(\widehat{\mathbf{S}})$. Changes with L_B are much smaller, but increases to L_B lead to a slight decrease in $\kappa(\widehat{\mathbf{S}})$. For both $\mathbf{H}_2$ and $\mathbf{H}_3$, large values of $\kappa(\widehat{\mathbf{S}})$ occur for very large values of L_R and small values of L_B. For a fixed value of L_R, increasing L_B results in a rapid decrease in the value of $\kappa(\widehat{\mathbf{S}})$. For small fixed values of L_R ($L_R < 0.1$), this decrease is followed by a slow increase to $\kappa(\widehat{\mathbf{S}})$ with increasing L_B. The minimum value of $\kappa(\widehat{\mathbf{S}})$ occurs when $L_R = L_B$; in this case $\mathbf{HBH}^T = \mathbf{R}$ to machine precision for both $\mathbf{H}_2$ and $\mathbf{H}_3$. The qualitative behavior for $\mathbf{H}_2$ and $\mathbf{H}_3$ is very similar, with smaller values of $\kappa(\widehat{\mathbf{S}})$ for $\mathbf{H}_3$ than $\mathbf{H}_2$. This is also the case in the unpreconditioned setting,[27] and occurs as $\mathbf{H}_3$ can be considered as a smoothed version of $\mathbf{H}_2$. Qualitatively the behavior for $\mathbf{H}_4$ is a compromise between $\mathbf{H}_1$ and $\mathbf{H}_2$; we can reduce $\kappa(\widehat{\mathbf{S}})$ by increasing L_B or decreasing L_R. In the unpreconditioned case decreasing either

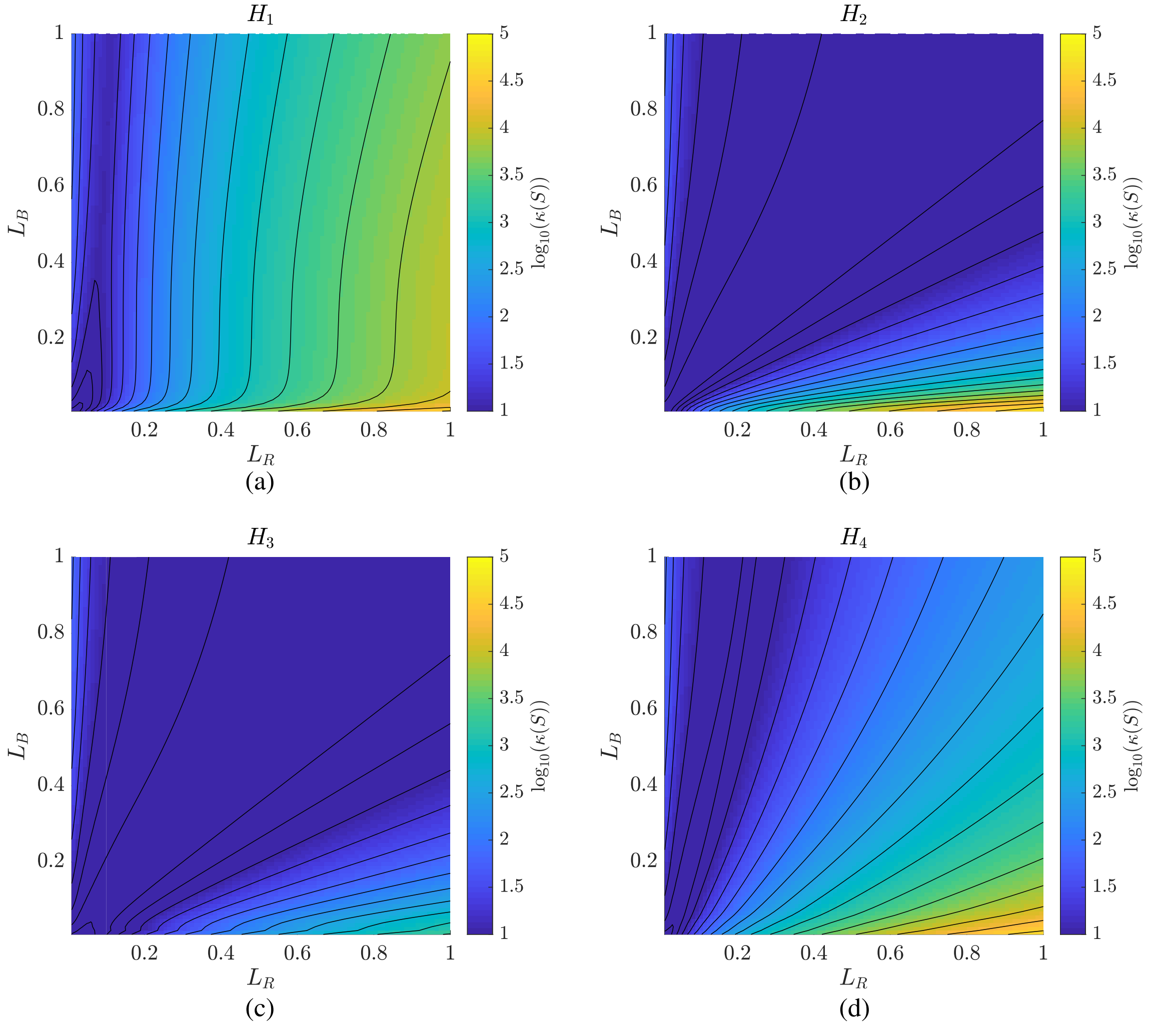

FIGURE 2 Change to $\kappa(\widehat{\mathbf{S}})$ with changes in L_R, L_B for (a) $\mathbf{H}_1$, (b) $\mathbf{H}_2$, (c) $\mathbf{H}_3$, and (d) $\mathbf{H}_4$. The color map is shown on a logarithmic scale which is standardized for all figures. Contours range from $\log_{10}(\kappa(\widehat{\mathbf{S}})) = 0.25$ to $\log_{10}(\kappa(\widehat{\mathbf{S}})) = 5$ with a contour interval of 0.25

lengthscale always reduces $\kappa(\widehat{\mathbf{S}})$. However, in the preconditioned setting the ratio between background and observation lengthscales is important, meaning that for some cases increasing L_B or L_R will reduce $\kappa(\widehat{\mathbf{S}})$.

Figure 3 shows the value of $\kappa(\widehat{\mathbf{S}})$, terms in the bounds (25), and the bounds (15) for various combinations of **H**, **R**, and **B**. The second term in the lower bound (25), given by $1 + \lambda_1(\mathbf{HH}^T)\lambda_N(\mathbf{B})(\lambda_1(\mathbf{R}))^{-1}$, is not shown, as it performs worse than the first term of (25), given by $1 + \lambda_p(\mathbf{HH}^T)\lambda_N(\mathbf{B})(\lambda_p(\mathbf{R}))^{-1}$, for all parameter combinations studied. Both the upper and lower bounds of (25) increase with L_R. They represent the increase in $\kappa(\widehat{\mathbf{S}})$ which occurs for $L_R \geq L_B$ for $\mathbf{H}_2$ and $\mathbf{H}_3$ and for larger values of L_R for $\mathbf{H}_1$ and $\mathbf{H}_4$. The initial decrease of $\kappa(\widehat{\mathbf{S}})$ with increasing L_R is not represented by the bounds of (25). Although some of the qualitative behavior is well represented, the bounds are very wide. Notably for larger values of L_B the lower bound given by (25) is very close to 1 for all values of L_R. By contrast, the bounds given by (15) represent the initial decrease in $\kappa(\widehat{\mathbf{S}})$ for small values of L_R well, both qualitatively and quantitatively. The upper bound of (15) then increases with increasing L_R and remains tight for all parameter combinations. The lower bound of (15) is monotonically decreasing, and hence does not represent the behavior of $\kappa(\widehat{\mathbf{S}})$ well for larger values of L_B and L_R. We note that for $\mathbf{H}_2$ and $\mathbf{H}_3$ the upper and lower bounds of (15) are equal for $L_B > L_R$. This results from Corollary 2 as $\mathbf{HBH}^T$ is circulant when $\mathbf{H} = \mathbf{H}_2$ or $\mathbf{H} = \mathbf{H}_3$ and all entries in the product $\mathbf{R}^{-1/2}\mathbf{HBH}^T\mathbf{R}^{-1/2}$ are positive for $L_B \geq L_R$. For panels (j) and (k) this means that the bounds given by (15) are equal to $\kappa(\widehat{\mathbf{S}})$ for all plotted values of L_R

Comparing the bounds given by (25) and (15), we find that the upper bound of (15) performs better for all parameters studied. The best lower bound depends on the choice of L_B and L_R: for lower values of L_B and larger values of L_R the first term of (25) is the tightest. Otherwise the bound given by (15) yields the tightest bound in this setting. Although the

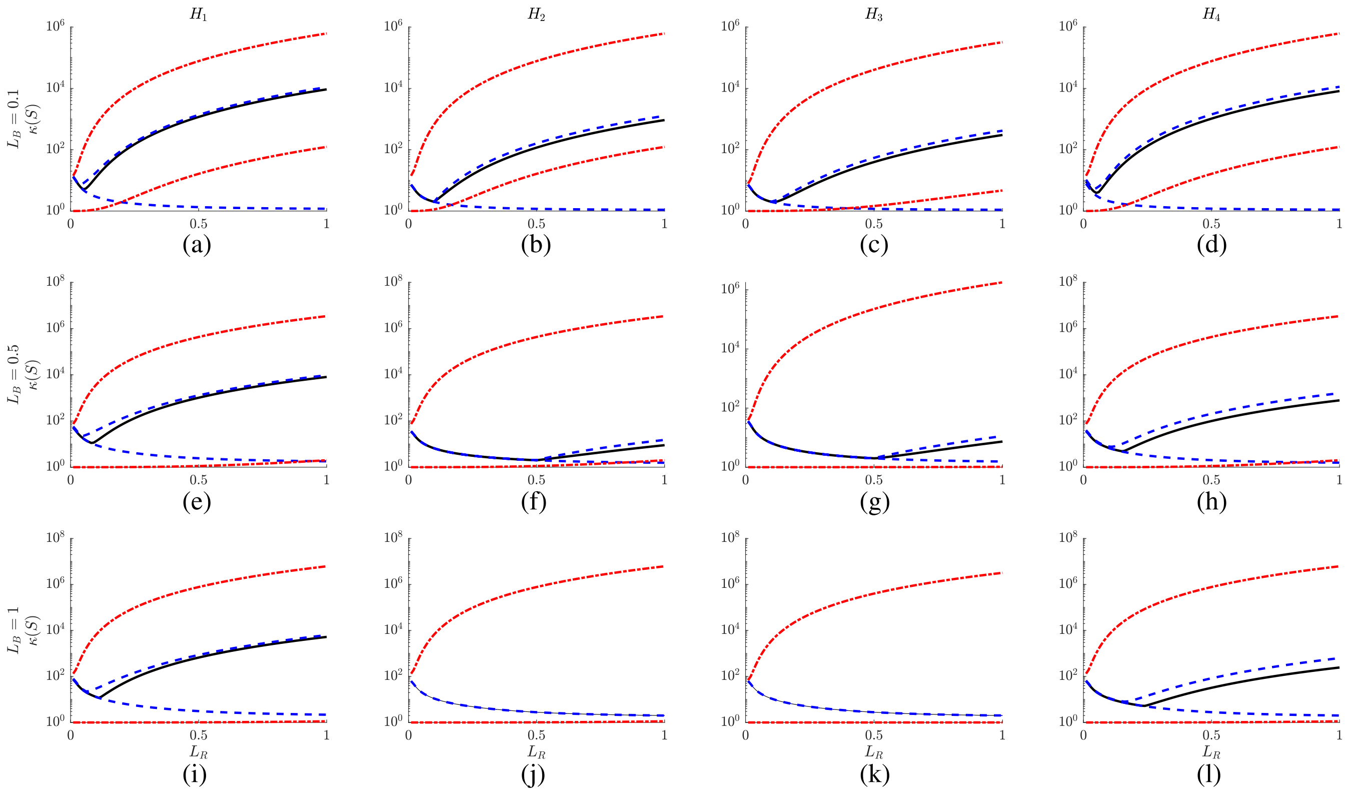

FIGURE 3 Bounds and value of $\kappa(\widehat{\mathbf{S}})$ for (a, e, i) $\mathbf{H}_1$, (b, f, j) $\mathbf{H}_2$, (c, g, k) $\mathbf{H}_3$, and (d, h, l) $\mathbf{H}_4$ as a function of L_R. Blue dashed lines denote the bounds given by (15), red dot-dashed lines denote the upper bound and first term in the lower bound of (25). The solid black line denotes the value of $\kappa(\widehat{\mathbf{S}})$ calculated using the *cond* command in Matlab 2018b.[40] The different rows correspond to different values of L_B. For (j) and (k) the upper and lower bounds of (15) are equal to $\kappa(\mathbf{S})$ for all values of L_R by the result of Corollary 4, and hence appear as a single line

bounds given by (15) represent the behavior of $\kappa(\widehat{\mathbf{S}})$ well, we note that the numerical framework considered here has a very specific structure that is unlikely to occur in practice. Observation operators are likely to be much less smooth and have less regular structure, for example, observations may not occur at the location of state variables, observation and state variables may not be evenly spaced, data may be missing, leading to different observation networks at different times or time windows. This may make a difference to the performance of both sets of bounds.

We now consider how altering the data assimilation system affects the convergence of a conjugate gradient method for the problem introduced in Section 4.2. Figure 4 shows how convergence of the conjugate gradient problem changes with L_B, L_R, and $\mathbf{H}$. For all choices of $\mathbf{H}$ the largest number of iterations occurs when L_R is large and L_B is small. Similarly to the unpreconditioned case[27], we see that for many cases $\kappa(\widehat{\mathbf{S}})$ is a good proxy for convergence: for $\mathbf{H}_2$, $\mathbf{H}_3$, and $\mathbf{H}_4$ reductions in $\kappa(\widehat{\mathbf{S}})$ and the number of iterations required for convergence occur for the same changes to L_R and L_B. The main difference in behavior is seen for $\mathbf{H}_1$, where increasing L_R increases $\kappa(\widehat{\mathbf{S}})$ for all choices of L_B, but makes no difference to the number of iterations required for convergence for $L_B \geq 0.4$.

This difference can be explained by considering the full distribution of the eigenvalues of $\widehat{\mathbf{S}}$ rather than just the condition number. Convergence of the conjugate gradient method depends on the distribution of the entire spectrum, and we expect faster convergence to occur where eigenvalues are clustered (see [12, theorem 38.4; 15, theorems 38.3, 38.5]). The eigenvalues of the full Hessian are given by $1 + \lambda(\mathbf{B}^{1/2}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B}^{1/2})$, and $N - p$ further unit eigenvalues. Figure 5 shows the nonzero eigenvalues of the low-rank update to the identity, $\mathbf{B}^{1/2}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B}^{1/2}$, for $L_B = 0.1$ and $L_R = 0.1, 0.4, 0.7, 1$. For all choices of $\mathbf{H}$ increasing L_R leads to an increase in the maximum eigenvalue of the product. Additionally, the spectrum is distributed smoothly with few clusters, meaning that the condition number is a good indicator for convergence of a conjugate gradient method. This explains why increasing L_R for $L_B = 0.1$ leads to an increase in the number of iterations required for convergence for all choices of $\mathbf{H}$.

Figure 6 shows the nonzero eigenvalues of the low-rank update to the identity, $\mathbf{B}^{1/2}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B}^{1/2}$, for $L_B = 0.5$ and $L_R = 0.1, 0.4, 0.7, 1$. Although the maximum eigenvalue of the product gets larger with increasing L_R for all choices of $\mathbf{H}$,

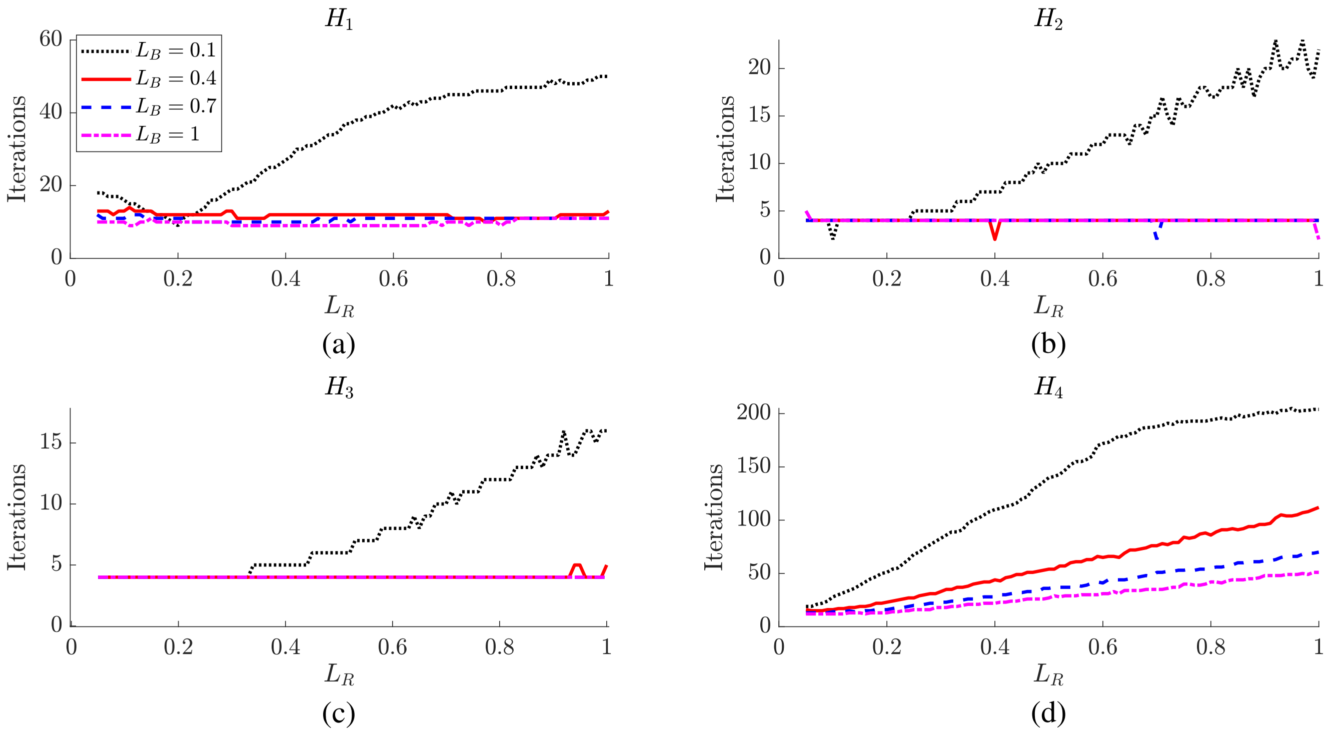

FIGURE 4 Number of iterations required for a conjugate gradient method to converge for changing values of L_R and L_B for (a) $\mathbf{H}_1$, (b) $\mathbf{H}_2$, (c) $\mathbf{H}_3$, and (d) $\mathbf{H}_4$. Note the difference in the y-axis values for each of the subplots

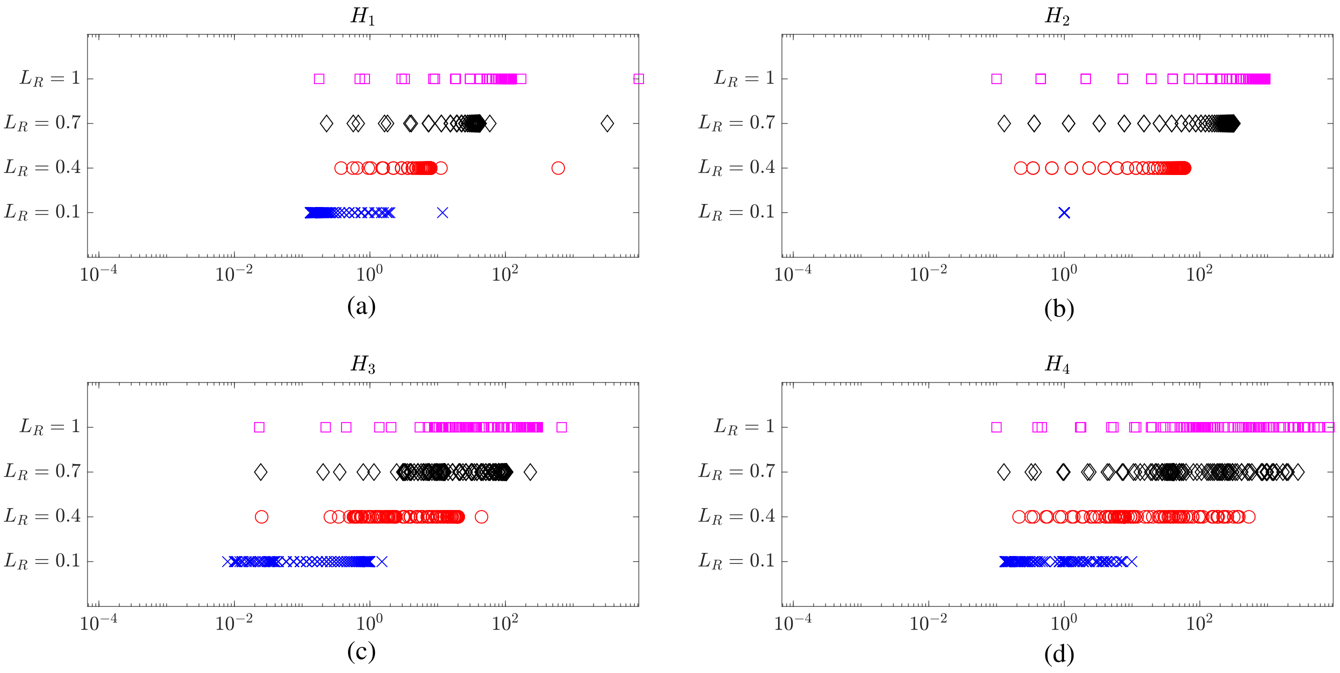

FIGURE 5 Nonzero eigenvalues of $\mathbf{B}^{1/2}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B}^{1/2}$ for $L_B = 0.1$ and $L_R = 0.1$ (crosses), $L_R = 0.4$ (circles), $L_R = 0.7$ (diamonds), and $L_R = 1$ (squares) for (a) $\mathbf{H}_1$, (b) $\mathbf{H}_2$, (c) $\mathbf{H}_3$, and (d) $\mathbf{H}_4$. Note the x-axis is plotted with a logarithmic scale

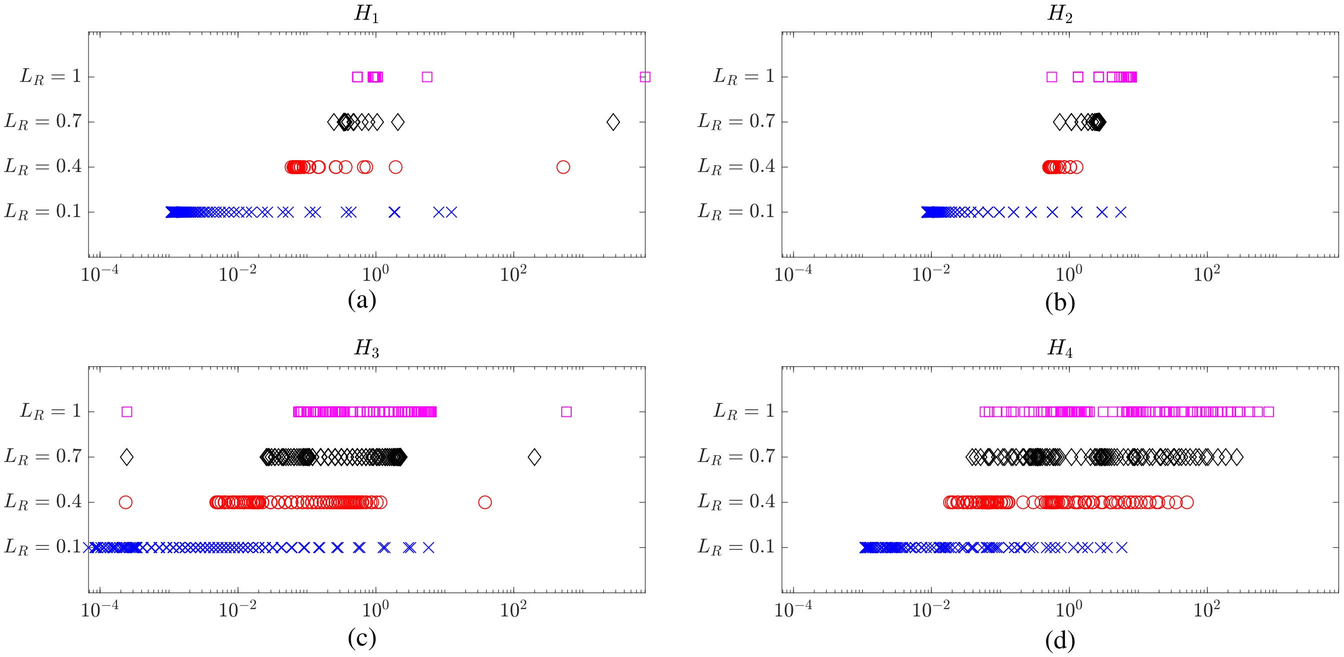

FIGURE 6 Nonzero eigenvalues of $\mathbf{B}^{1/2}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B}^{1/2}$ for $L_B = 0.5$ and $L_R = 0.1$ (crosses), $L_R = 0.4$ (circles), $L_R = 0.7$ (diamonds), and $L_R = 1$ (squares) for (a) $\mathbf{H}_1$, (b) $\mathbf{H}_2$, (c) $\mathbf{H}_3$, and (d) $\mathbf{H}_4$. Note the *x*-axis is plotted with a logarithmic scale

we see increased clustering of the remaining eigenvalues about 1 for $\mathbf{H}_1, \mathbf{H}_2$, and $\mathbf{H}_3$. Therefore, for these parameter choices the condition number of the Hessian does not represent convergence of a conjugate gradient method well. For $\mathbf{H}_4$ no such clustering is observed, which explains why increasing L_R for all values of L_B leads to slower convergence of the conjugate gradient method for this observation operator. The clustering occurs due to the very regular structures of $\mathbf{H}_1, \mathbf{H}_2$, and $\mathbf{H}_3$. Therefore if the observation operator is less regular, as may be expected in realistic observing networks,[42] the condition number is more likely to be a good proxy for convergence of a conjugate gradient method in this setting.

We conclude that in this framework changing L_B has a larger effect on convergence of a conjugate gradient method than changing L_R. This contrasts with the unpreconditioned case, where changes to both L_B and L_R had a large impact on convergence.[27] For all choices of observation operator, small values of L_B lead to poor convergence. Although changing L_R impacts $\kappa(\widehat{\mathbf{S}})$, due to an increase in clustered eigenvalues these changes do not always affect the convergence of a conjugate gradient method. Overall the condition number is a good proxy for convergence in this framework.

6 | CONCLUSIONS

The inclusion of correlated observation errors in data assimilation is important for high resolution forecasts,[22,23] and to ensure we make the best use of existing data.[19,20,43] However, multiple studies have found issues with convergence of data assimilation routines when introducing correlated OEC matrices.[16,24,44] Earlier work considers the preconditioned data assimilation problem in the case of uncorrelated OEC matrices.[14] In this article we study the effect of introducing correlated OEC matrices on the conditioning and convergence of the preconditioned variational data assimilation problem. This extends the theoretical and numerical results of a previous study by Tabeart et al.[27] that considered the use of correlated OEC matrices in the unpreconditioned variational data assimilation framework.

In this article, we developed bounds on the condition number of the Hessian of the preconditioned variational data assimilation problem, for the case that there are fewer observations than state variables. We then studied these bounds numerically in an idealized framework. We found that:

- As in the unpreconditioned case, decreasing the observation error variance or increasing the background error variance increases the condition number of the Hessian.
- The minimum eigenvalue of the OEC matrix appears in both the upper and lower bounds. This was also true for the unpreconditioned case.

- For a fixed lengthscale of the observation (background) error covariance matrix, L, the condition number of the Hessian is smallest when the lengthscale of the background (observation) error covariance matrix is also equal to L. This is in contrast to the unpreconditioned case, where for a fixed lengthscale of the observation (background) error covariance matrix, the condition number of the Hessian is smallest when the lengthscale of the background (observation) error covariance is minimized.
- Our new lower bound represented the qualitative behavior better than an existing bound for some cases. The upper bound from Haben[14] was shown to be tight for all parameter choices. We proved that under additional assumptions the upper and lower bounds from Haben[14] are equal.
- For most cases the conditioning of the Hessian performed well as a proxy for the convergence of a conjugate gradient method. However in some cases, clustered eigenvalues (induced by the specific structure of the numerical framework) meant that convergence was much faster than predicted by the conditioning.

We remark that our findings about clustered eigenvalues occur as our numerical framework has very specific structures. In particular, the eigenvectors of the background and OEC matrices are strongly related. Other experiments not presented in this article considered the use of the Laplacian correlation function for either or both of the observation and background error covariance matrices.[14] Qualitative conclusions were very similar to those shown in Section 5, even though the negative entries of the Laplacian correlation function do not satisfy the additional assumptions required for the bounds to be equal. In applications, we are likely to have more complicated observation operators, and the background and OEC matrices are less likely to have complementary structures. Satellite observations for NWP often have interchannel correlation structures that are different from the typical spatial correlations of background error covariance matrices.[17,19] We also note that our state variables were evenly distributed and homogeneous, which will not be the case for nonuniform grids.

In the unpreconditioned case using a similar numerical framework Tabeart et al.[27] found that improving the conditioning of the background or OEC matrix separately would always decrease $\kappa(\widehat{\mathbf{S}})$. The preconditioned system is more complicated; in some cases decreasing the condition number $\kappa(\mathbf{B})$ or $\kappa(\widehat{\mathbf{R}})$ increases the condition number $\kappa(\widehat{\mathbf{S}})$. We expect the relationship between each of the constituent matrices to be complicated for more general problems. This is relevant for practical applications, as estimated OEC matrices typically need to be treated via reconditioning methods before they can be used.[16,24] Currently the use of reconditioning methods is heuristic,[28] meaning that there may be flexibility to select a treated matrix that will result in faster convergence in some cases. However, popular reconditioning techniques work by increasing small eigenvalues of the OEC matrix. In the preconditioned setting, such techniques will not automatically reduce the value of $\kappa(\widehat{\mathbf{S}})$, due to the multiplication of background and observation error covariances. This means that reconditioning techniques may perform differently for the preconditioned data assimilation problem than in the unpreconditioned setting.

Although the numerical experiments in this article consider a limited choice of matrices and parameters, we note that the theory and bounds presented in this work are general and apply to any choice of covariance matrices **B** and **R**, and any linear observation operator (or generalized observation operator in the case of 4D-Var). We could consider the numerical results presented here as a "best case" due to the circulant structure of both covariance matrices. For more general choices of **B** and **R** any eigenvalue clustering is likely to be less extreme, and hence conditioning may be more influential for the convergence of a conjugate gradient method. Increased eigenvalue clustering occurred for observation operators with regular structure, whereas in practice the "randomly observed" experiment is more realistic. For the 4D-Var problem the generalized observation operator $\widehat{\mathbf{H}}$ also accounts for model evolution, and hence the structure of the linearized model is also expected to be important when considering clustering and convergence of a conjugate gradient problem. Previous work has also shown that for the unpreconditioned problem, the qualitative behavior of an operational system[25] largely followed the linear theory.[27] Similarly, for the case of uncorrelated OEC matrices, the behavior of preconditioned 4D-Var experiments broadly coincided with theory from the linear setting.[14,26] This indicates that conclusions arising from the study of linear data assimilation problems can often provide insight for a wider range of practical implementations, even if theoretical results are not directly applicable.

ACKNOWLEDGMENT

This work is funded in part by the EPSRC Centre for Doctoral Training in Mathematics of Planet Earth, the NERC Flooding from Intense Rainfall programme (NE/K008900/1), the EPSRC DARE project (EP/P002331/1), and the NERC National Centre for Earth Observation and EPSRC grant (EP/S027785/1).

CONFLICT OF INTEREST

This study does not have any conflicts to disclose.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

ORCID

Jemima M. Tabeart https://orcid.org/0000-0001-6806-8608
Sarah L. Dance https://orcid.org/0000-0003-1690-3338
Amos S. Lawless https://orcid.org/0000-0002-3016-6568
Nancy K. Nichols https://orcid.org/0000-0003-1133-5220
Joanne A. Waller https://orcid.org/0000-0002-7783-6434

REFERENCES

1. Courtier P, Thépaut J-N, Hollingsworth A. A strategy for operational implementation of 4D-Var, using an incremental approach. Q J R Meteorol Soc. 1994;120(519):1367–87.
2. Gratton S, Lawless AS, Nichols NK. Approximate Gauss-Newton methods for nonlinear least squares problems. SIAM J Optim. 2007;18(1):106–32.
3. Lawless AS, Gratton S, Nichols NK. Approximate iterative methods for variational data assimilation. Int J Numer Methods Fluids. 2005;47(10-11):1129–35.
4. Lawless AS, Gratton S, Nichols NK. An investigation of incremental 4D-Var using non-tangent linear models. Q J R Meteorol Soc. 2005;131:459–76.
5. Carrassi A, Bocquet M, Bertino L, Evensen G. Data assimilation in the geosciences: an overview of methods, issues, and perspectives. Wiley Interdiscip Rev Clim Chang. 2018;9(5):e535.
6. Schiff SJ. Neural control engineering the emerging intersection between control theory and neuroscience. Cambridge, MA: MIT Press; 2011.
7. Bannister R, Review N. A review of forecast error covariance statistics in atmospheric variational data assimilation. II: modelling the forecast error covariance statistics. Q J R Meteorol Soc. 2008;134:1971–96.
8. Lewis JM, Lakshmivarahan S, Dhall SK. Dynamic data assimilation: a least squares approach. Cambridge, MA: Cambridge University Press; 2006.
9. Fisher M. Minimization algorithms for variational data assimilation. Proceedings of the Seminar on Recent Developments in Numerical Methods for Atmospheric Modelling European Centre for Medium Range Weather Forecasts; 1998. p. 364–85; Reading, UK.
10. Liu Y, Zhang L, Lian Z. Conjugate gradient algorithm in the four-dimensional variational data assimilation system in GRAPES. J Meteorol Res. 2018;32(6):974–84.
11. Trémolet Y. Incremental 4D-Var convergence study. Tellus A Dyn Meteorol Oceanogr. 2007;59(5):706–18.
12. Gill PE, Murray W, Wright MH. Practical optimization. Amsterdam, Netherlands; London, UK: Academic Press; 1986.
13. Golub GH, Van Loan CF. Matrix computations. 3rd ed. Baltimore: The John Hopkins University Press; 1996.
14. Haben SA. Conditioning and preconditioning of the minimisation problem in variational data assimilation [PhD thesis]. Department of Mathematics and Statistics, University of Reading; 2011
15. Trefethen LN, Bau D. Numerical linear algebra. Philadelphia, PA: Society for Industrial and Applied Mathematics; 1997.
16. 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. Q J R Meteorol Soc. 2016;142(697):1767–80.
17. Weston PP, Bell W, Eyre JR. Accounting for correlated error in the assimilation of high-resolution sounder data. Q J R Meteorol Soc. 2014;140:240–2429.
18. Janjić T, Bormann N, Bocquet M, Carton JA, Cohn SE, Dance SL, et al. On the representation error in data assimilation. Q J R Meteorol Soc. 2018;144(713):1257–78.
19. Stewart LM, Dance SL, Nichols NK. Data assimilation with correlated observation errors: experiments with a 1-D shallow water model. Tellus A Dyn Meteorol Oceanogr. 2013;65:19546 (14pp).
20. Simonin D, Waller JA, Ballard SP, Dance SL, Nichols NK. A pragmatic strategy for implementing spatially correlated observation errors in an operational system: an application to Doppler radar winds. Q J R Meteorol Soc. 2019;145(723):2772-2790. https://doi.org/10.1002/qj.3592
21. Stewart LM, Dance SL, Nichols NK. Correlated observation errors in data assimilation. Int J Numer Methods Fluids. 2008;56(8):1521–7.
22. Fowler AM, Dance SL, Waller JA. On the interaction of observation and prior error correlations in data assimilation. Q J R Meteorol Soc. 2018;144(710):48–62.
23. Rainwater S, Bishop CH, Campbell WF. The benefits of correlated observation errors for small scales. Q J R Meteorol Soc. 2015;141:3439–45.

24. Weston P. Progress towards the implementation of correlated observation errors in 4D-Var met office forecasting research technical report, 560; 2011.
25. Tabeart JM, Dance SL, Hilton F, Lawless AS, Migliorini S, Nichols NK, et al. The impact of using reconditioned correlated observation error covariance matrices in the Met Office 1D-Var system. Q J R Meteorol Soc. 2020;146(728):1372-1390. https://doi.org/10.1002/qj.3741
26. Haben SA, Lawless AS, Nichols NK. Conditioning of incremental variational data assimilation, with application to the met office system. Tellus A Dyn Meteorol Oceanogr. 2011;64(4):782–92.
27. 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;25(5):e2165.
28. Tabeart JM, Dance SL, Lawless AS, Nichols NK, Waller JA. Improving the conditioning of estimated covariance matrices. Tellus A Dyn Meteorol Oceanogr. 2020;72(1):1–19.
29. Bannister RN. A review of operational methods of variational and ensemble-variational data assimilation. Q J R Meteorol Soc. 2017;143(703):607–33.
30. Wang B, Zhang F. Some inequalities for the eigenvalues of the product of positive semidefinite Hermitian matrices. Linear Algebra Appl. 1992;160:113–8.
31. Harville DA. Matrix algebra from a statistician's point of view. New York, NY: Springer-Verlag; 1997.
32. Bernstein DS. Matrix mathematics: theory, facts, and formulas. 2nd ed. Princeton, NJ: Princeton University Press; 2009.
33. Davis PJ. Circulant matrices. New York, NY: Wiley; 1979.
34. Gray RM. Toeplitz and circulant matrices: a review. Found Trends Commun Inf Theory. 2006;2(3):155–239.
35. Mei Y. Computing the square roots of a class of circulant matrices. J Appl Math. 2012;2012:647623. https://doi.org/10.1155/2012/647623
36. Daley R. Atmospheric data analysis. Cambridge, MA: Cambridge University Press; 1991.
37. Johnson C. Information content of observations in variational data assimilation [PhD thesis]. Department of Mathematics and Statistics, University of Reading; 2003
38. Gaspari G, Cohn SE. Construction of correlation functions in two and three dimensions. Q J R Meteorol Soc. 1999;125:723–57.
39. Jeong J, Jun M. Covariance models on the surface of a sphere: when does it matter. Stat. 2015;4:167–82.
40. MATLAB (R2018b) The MathWorks Inc; 2018. https://www.mathworks.com/help/matlab/
41. Waller JA, Dance SL, Nichols NK. Theoretical insight into diagnosing observation error correlations using observation-minus-background and observation-minus-analysis statistics. Q J R Meteorol Soc. 2016;142:418–31.
42. Guillet O, Weaver AT, Vasseur X, Michel Y, Gratton S, Gürol S. Modelling spatially correlated observation errors in variational data assimilation using a diffusion operator on an unstructured mesh. Q J R Meteorol Soc. 2019;145(722):1947–67.
43. Michel Y. Revisiting Fisher's approach to the handling of horizontal spatial correlations of the observation errors in a variational framework. Q J R Meteorol Soc. 2018;144(716):2011–25.
44. Campbell WF, Satterfield EA, Ruston B, Baker NL. Accounting for correlated observation error in a dual-formulation 4D variational data assimilation system. Monthly Weather Rev. 2017;145(3):1019–32.

How to cite this article: Tabeart JM, Dance SL, Lawless AS, Nichols NK, Waller JA. New bounds on the condition number of the Hessian of the preconditioned variational data assimilation problem. Numer Linear Algebra Appl. 2021;e2405. https://doi.org/10.1002/nla.2405