University of Reading

A novel localized fast multipole method for computations with spatially correlated observation error statistics in data assimilation

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Hu, G. ORCID: https://orcid.org/0000-0003-4305-3658 and Dance, S. L. ORCID: https://orcid.org/0000-0003-1690-3338 (2024) A novel localized fast multipole method for computations with spatially correlated observation error statistics in data assimilation. Journal of Advances in Modeling Earth Systems, 16 (6). e2023MS003871. ISSN 1942-2466 doi: 10.1029/2023MS003871 Available at https://centaur.reading.ac.uk/116418/

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/10.1029/2023MS003871

Publisher: American Geophysical Union

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

the End User Agreement.

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Check for updates

AGU ADVANCING EARTH AND SPACE SCIENCES

JAMES | Journal of Advances in Modeling Earth Systems®

RESEARCH ARTICLE
10.1029/2023MS003871

Key Points:

- The new method has the potential to be used in practice to assimilate a large number of observations with mutual spatial error correlations
- The new method allows fast computation of matrix-vector products involving full observation precision matrices in data assimilation
- The new method can preserve the analysis accuracy at all spatial scales

Correspondence to:
G. Hu,
guannan.hu@reading.ac.uk

Citation:
Hu, G., & Dance, S. L. (2024). A novel localized fast multipole method for computations with spatially correlated observation error statistics in data assimilation. *Journal of Advances in Modeling Earth Systems*, *16*, e2023MS003871. https://doi.org/10.1029/2023MS003871

Received 14 JUN 2023
Accepted 8 MAY 2024

Author Contributions:
Conceptualization: Guannan Hu, Sarah L. Dance
Formal analysis: Guannan Hu, Sarah L. Dance
Funding acquisition: Sarah L. Dance
Investigation: Guannan Hu, Sarah L. Dance
Methodology: Guannan Hu, Sarah L. Dance
Project administration: Sarah L. Dance
Resources: Sarah L. Dance
Software: Guannan Hu
Supervision: Sarah L. Dance
Validation: Guannan Hu, Sarah L. Dance
Visualization: Guannan Hu, Sarah L. Dance
Writing – original draft: Guannan Hu, Sarah L. Dance
Writing – review & editing: Guannan Hu, Sarah L. Dance

© 2024 The Authors. Journal of Advances in Modeling Earth Systems published by Wiley Periodicals LLC on behalf of American Geophysical Union.
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.

A Novel Localized Fast Multipole Method for Computations With Spatially Correlated Observation Error Statistics in Data Assimilation

Guannan Hu[1,2] **and Sarah L. Dance**[1,2,3]

[1]Department of Meteorology, University of Reading, Reading, UK, [2]National Centre for Earth Observation (NCEO), University of Reading, Reading, UK, [3]Department of Mathematics and Statistics, University of Reading, Reading, UK

Abstract Several observation types (e.g., geostationary satellite and Doppler radar observations) have recently been found to exhibit strong spatial error correlations. Including these error statistics in data assimilation for numerical weather prediction can improve analysis quality and forecast skill. Moreover, it allows for increases in the spatial density of observations assimilated, which is needed for the provision of information on appropriate scales for high-resolution forecasting. However, introducing correlated error statistics may increase the computational complexity and parallel communication costs of matrix-vector products involving observation precision matrices (inverse observation error covariance matrices). Without new approaches, we cannot take full advantage of new observation uncertainty estimates. We develop a new numerical approximation method based on a particular type of fast multipole method and a domain localization approach. The basic idea is to divide the observation domain into boxes and then separate calculations of matrix-vector products according to the partition. These calculations can be done in parallel with very low communication overheads. The new method is easy to implement and parallelize, and it is applicable to a wide variety of observation precision matrices. We applied the new method to a simple variational data assimilation problem and found that the computational cost of the variational minimization was dramatically reduced while preserving analysis accuracy across a range of scales. The new method has the potential to be used as an efficient technique for practical applications where a large number of observations with mutual error correlations need to be assimilated quickly.

Plain Language Summary Weather forecasts are generated by running a computational model from a starting point that describes the current weather conditions. These conditions are estimated using a data assimilation process that combines observations of the weather with computational predictions from a previous forecast. In the data assimilation process, the statistical weighting given to the observations is determined by our knowledge of the observation error statistics. Spatial relationships between errors, or spatial error correlations, are often found in remotely sensed observations, such as those from satellites. The inclusion of spatially correlated observation error statistics in the data assimilation process can substantially increase the time required to complete the process. We develop a novel method that can significantly reduce the time taken to assimilate observations with mutual spatial error correlations. Moreover, the accuracy of data assimilation can be maintained with this new method.

1. Introduction

Data assimilation is essential for numerical weather prediction (NWP) as it provides the initial conditions for computer models that simulate the evolution of the atmosphere (e.g., Lorenc et al., 2000; Nichols, 2010). In addition, data assimilation is used to create climate reanalyzes (e.g., Bollmeyer et al., 2015; Hersbach et al., 2020). Data assimilation blends observations and model forecasts by taking account of their uncertainties. Observation uncertainties are described by an observation error covariance matrix, $\mathbf{R} \in \mathbb{R}^{m \times m}$, where m is the number of observations (Hodyss & Satterfield, 2017; Janjić et al., 2018). In practical data assimilation applications, the matrix $\mathbf{R}$ is often treated as block diagonal with each block containing error statistics of one type of observation (for a given time). This is because observation errors between observations of different types are usually assumed to be uncorrelated. Moreover, observation errors between observations of the same type may be mutually uncorrelated, resulting in diagonal blocks in the matrix $\mathbf{R}$, or mutually correlated, leading to non-diagonal blocks. A widely adopted approach to estimate the matrix $\mathbf{R}$ is to use the diagnosis approach of Desroziers et al. (2005). This

19422466, 2024, 6, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023MS003871 by Test, Wiley Online Library on [13/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

approach is an indirect sampling approach that uses a large sample of assimilation residuals. For a given sample size, the sampling error of this approach is affected by the accuracy of the observation and background error covariance matrices specified in the data assimilation algorithm (Hu & Dance, 2024).

Previous studies have shown that observations such as Doppler radar radial winds (DRWs), geostationary satellite data and atmospheric motion vectors (AMVs) can exhibit strong *spatial* error correlations, leading to non-diagonal blocks of the matrix $\mathbf{R}$. For example, errors in the DRWs used by the Met Office and Deutscher Wetterdienst (DWD) NWP systems were found to have a horizontal correlation length of 12–50 km (Waller, Simonin, et al., 2016; Waller et al., 2019; Zeng et al., 2021). The spatial error correlation lengthscale estimated for observations from the Spinning Enhanced Visible and Infrared Imager (SEVIRI) used in the Met Office and Météo-France varies between 30 and 250 km (Michel, 2018; Waller, Ballard, et al., 2016). The horizontal error correlation lengthscale for the AMVs assimilated within the Met Office UKV model 3D-Var FGAT system is found to be around 140–210 km (Cordoba et al., 2017). Moreover, Bormann et al. (2003) found that some AMVs can have a correlation lengthscale up to 800 km or even broader for tropical regions in the ECMWF global data assimilation system at that time. These findings have important implications for better use of the observations with strong spatial error correlations. By directly including correlated error statistics in the assimilation scheme, we are able to use a shorter observation-thinning distance or make full use of the information from high-resolution observations (Bell et al., 2020; Fowler et al., 2018; Rainwater et al., 2015; Waller, Simonin, et al., 2016). This has been shown to improve both analysis accuracy and forecast skill in idealized data assimilation systems (e.g., Healy & White, 2005; Stewart et al., 2008, 2013) as well as operational systems (e.g., Fujita et al., 2020; Simonin et al., 2019; Yeh et al., 2022). In addition, assimilation of dense observations (e.g., Doppler radar observations) without thinning is important for convection-permitting NWPs as they require information on small scales (Waller et al., 2021). Convection-permitting NWP is key for forecasting extreme weather events such as heavy rainfall (e.g., Clark et al., 2016; Dance et al., 2019). The timing of these extreme events is often difficult to capture in a model that does not resolve small-scale weather processes (e.g., Hu et al., 2019; Hu & Franzke, 2020).

A potential problem with including spatially correlated observation error statistics in data assimilation procedures is that matrix-vector products involving the inverse of the matrix $\mathbf{R}$ can be computationally expensive. These matrix-vector products arise in the observation penalty term of the variational data assimilation cost function and its gradient with the form (see more details in Section 2)

$$\mathbf{q} = \mathbf{A}\mathbf{d}, \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{m \times m}$ denotes the observation precision matrix (the inverse of the matrix $\mathbf{R} \in \mathbb{R}^{m \times m}$) and $\mathbf{d} \in \mathbb{R}^m$ denotes the observation-minus-model departure vector,

$$\mathbf{d} = \mathbf{y} - H(\mathbf{x}), \tag{2}$$

where $\mathbf{x} \in \mathbb{R}^n$ is the model state vector, $\mathbf{y} \in \mathbb{R}^m$ is the observation vector and H is the observation operator that maps the model state vector to the observation vector (i.e., $H : \mathbb{R}^n \rightarrow \mathbb{R}^m$). The vector $\mathbf{d}$ is called the innovation vector if $\mathbf{x} = \mathbf{x}^b$. In this study, we do not address the problem of inverting the matrix $\mathbf{R}$. Instead, we assume that the matrix $\mathbf{A}$ is known and focus on fast computation of the matrix-vector products involving the matrix $\mathbf{A}$. If the matrix $\mathbf{R}$ is diagonal, then the matrix $\mathbf{A}$ is also diagonal. In this case the computation of Equation 1 requires only m floating point operations, and it can be perfectly parallelized without incurring communication costs. However, if the matrix $\mathbf{R}$ is non-diagonal, then the matrix $\mathbf{A}$ is usually a full matrix. In this case, the work required to calculate Equation 1 increases to $\mathcal{O}(m^2)$ floating point operations. Moreover, the bigger problem is that the parallelization of matrix-vector multiplications with full matrices requires a large number of communications if matrix and vector elements are distributed across multiple processors (Deng, 2012; Grama et al., 2003).

An efficient way to calculate Equation 1 for a non-diagonal matrix is to decompose the matrix $\mathbf{R}$ into a lower triangular matrix and an upper triangular matrix (e.g., using the Cholesky decomposition; Golub & Van Loan, 1996). This approach avoids the need to explicitly invert the matrix $\mathbf{R}$ and allows the linear system to be solved by forward and backward substitution. For correlated *inter-channel* error covariance matrices, the Cholesky decomposition approach is compatible with the parallelization strategy used for diagonal $\mathbf{R}$ matrices because the observations with mutually correlated errors are typically distributed in one processing element (PE;

19422466, 2024, 6, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023MS003871 by Test, Wiley Online Library on [13/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Simonin et al., 2019; Weston et al., 2014). However, for *spatially* correlated error covariance matrices, the Cholesky decomposition approach may be incompatible with this parallelization scheme because observations may be allocated to different PEs according to their geophysical locations, and hence excessive communications between PEs are required. To deal with this, Simonin et al. (2019) proposed a pragmatic strategy for operational applications where observations with mutually correlated errors were assigned to a single PE. With this new allocation of observations, the Cholesky decomposition approach can be easily applied locally in each PE without the need for communication. However, the strategy is only applicable to moderate numbers of mutually correlated observations.

Several other studies also addressed correlated observation errors in data assimilation. Guillet et al. (2019) modeled spatially correlated observation errors using diffusion operators, which facilitates a convenient operator for the inverse observation error covariance matrix. However, it is not clear how this approach can be combined with a treatment of satellite inter-channel error correlations. In another approach, eigendecomposition has been used to approximate the matrix **R** (e.g., Anderson, 2003; Fisher, 2005; Fowler, 2019; Michel, 2018; Stewart et al., 2013). However, using too few eigenpairs can cause spurious error correlations, while too many eigenpairs can lead to great computational expense (Fisher, 2005; Stewart, 2010). There have also been a few studies that account for correlated observation errors by transforming the observations into spectral space (Chabot et al., 2015, 2020; Ying, 2020). In addition, Bédard and Buehner (2020) proposed the use of the difference between adjacent observations to indirectly extract the information from observations having spatial error correlations. However, it is unclear how to use this approach to deal with non-uniformly distributed data.

The overall aim of this work is to make progress toward development of a new, approximate approach for fast computations of matrix-vector products that can be used for the treatment of spatially correlated errors in data assimilation. As this is a complicated problem, we choose to make progress in small steps. Thus, the goal of this paper is to examine a novel, hierarchical approximation approach and establish some initial evidence with an idealized model that the method has the potential to provide the key features that it will need. Furthermore, we make a pragmatic assumption that the precision matrix, **A**, is known, rather than working directly with the error covariance matrix, **R**. These assumptions allow us to establish that the approximation approach preserves (rather than discards) small scale information in the analysis as well as being fast and accurate, before we invest significant further time in mathematical derivation and numerical implementation of the approach for operational applications.

In our previous work (Hu & Dance, 2021a), we explored the use of a singular value decomposition approach to the fast multipole method (SVD-FMM; Gimbutas & Rokhlin, 2003) for fast calculation of Equation 1 with a full matrix **A**. The SVD-FMM is a hierarchical decomposition where truncated SVDs are used to compress and decompress portions of the matrix-vector product. The approach is general in that it can be applied to any covariance matrix. In Hu and Dance (2021a), we proposed a possible parallelization scheme for the SVD-FMM in data assimilation applications. We compared its computational cost (including the communication cost) with three distinct standard parallel formulations for computing large matrix-vector products (see Grama et al., 2003, Section 8.1). We carried out a series of numerical experiments to test the performance of the SVD-FMM when using different observation error correlation functions and lengthscales, and different matrix reconditioning techniques. We also tested its performance in the presence of missing observations. We showed that the SVD-FMM can accurately and rapidly compute Equation 1 under various circumstances. However, the complexity of the algorithm of the SVD-FMM may be a disadvantage for its operational use. In this work, we develop a new method based on the SVD-FMM and a domain localization approach. We call this new method the *local SVD-FMM*. The local SVD-FMM simplifies the multi-level algorithm of the SVD-FMM into a single-level one. The single-level algorithm is easier to implement and parallelize. The domain localization we use is similar to, but not identical to, that used for ensemble Kalman filter methods (e.g., Greybush et al., 2011; Janjić et al., 2011; Nino-Ruiz et al., 2019; Periáñez et al., 2014). With the ensemble filter, localization is mainly used to get rid of spurious correlations (Hamill et al., 2001; Houtekamer & Mitchell, 1998), but it also brings some computational side benefits, including the increase in effective ensemble size (Oke et al., 2007) and the reduction of computational costs (Lorenc, 2003; Petrie & Dance, 2010). With the local SVD-FMM, localization is used to exclude interactions between observations that are beyond a certain distance (given by a non-dimensional localization length parameter; see Section 3.1) in the calculation of the matrix-vector products involving the matrix **A**. This means only subsets of the matrix **A** and the vector **d** are used to calculate the vector **q** (Equation 1). Mathematically, this is equivalent to setting some of the elements of **A** to be zero. Since the spatial correlation between

19422466, 2024, 6, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023MS003871 by Test, Wiley Online Library on [13/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

two observation errors is expected to decrease as the distance between the observations increases, we anticipate that the numerical error caused by excluding the distant observations in the computation will be small (Eijkhout & Polman, 1988).

A further new contribution of this work is the application of the new method to a variational data assimilation experiment. We apply the local SVD-FMM to an idealized data assimilation problem and demonstrate its accuracy and efficiency across a range of scales in the analysis.

We organize the rest of this paper as follows: In Section 2, we introduce the three-dimensional variational data assimilation algorithm and explain where matrix-vector products involving the matrix $\mathbf{A}$ arise in the solution of the variational data assimilation problem. In Section 3, we present the algorithm for the novel local SVD-FMM along with a discussion of accuracy, efficiency and a potential parallelization scheme. In Section 4, we show the accuracy of the local SVD-FMM in computing Equation 1 for different observation error correlation lengthscales, different localization lengths and different numbers of singular vectors. In Section 5, we demonstrate the accuracy and efficiency of the local SVD-FMM in calculating the analysis in a simple variational data assimilation system. In Section 6, we discuss our results and conclude that the local SVD-FMM has potential as a numerical approximation method that can accelerate matrix-vector multiplication in data assimilation applications where a large volume of observational data with spatially correlated errors needs to be assimilated.

2. Variational Data Assimilation Algorithm

In this section, we briefly introduce three-dimensional variational data assimilation (3D-Var). We use it as an example to show how the local SVD-FMM can be applied in variational data assimilation. The objective of 3D-Var is to find the value of $\mathbf{x}$ that minimizes the cost function

$$J(\mathbf{x}) = \frac{1}{2}\left(\mathbf{x}-\mathbf{x}^b\right)^{\top}\mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{x}^b\right) + \frac{1}{2}(H(\mathbf{x})-\mathbf{y})^{\top}\mathbf{R}^{-1}(H(\mathbf{x})-\mathbf{y}), \tag{3}$$

where $\mathbf{x}^b \in \mathbb{R}^n$ denotes the background state vector, $\mathbf{B} \in \mathbb{R}^{n\times n}$ denotes the background error covariance matrix and other variables were introduced in Section 1. The minimizer of the cost function is the optimal estimate of the system state that we seek, which is called the analysis and denoted by $\mathbf{x}^a$.

In operational data assimilation, the incremental formulation of the cost function is widely used, as it reduces the computational cost of variational data assimilation (Courtier et al., 1994). The increment is defined as $\delta\mathbf{x} = \mathbf{x} - \mathbf{x}^b$. By substituting $\mathbf{x} = \mathbf{x}^b + \delta\mathbf{x}$ into Equation 3, we obtain the incremental cost function

$$J(\delta\mathbf{x}) = \frac{1}{2}\delta\mathbf{x}^{\top}\mathbf{B}^{-1}\delta\mathbf{x} + \frac{1}{2}(\mathbf{H}\delta\mathbf{x}-\mathbf{d})^{\top}\mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{x}-\mathbf{d}), \tag{4}$$

where $\mathbf{H} \in \mathbb{R}^{m\times n}$ is the observation operator linearized at $\mathbf{x} = \mathbf{x}^b$. Once we find the value of $\delta\mathbf{x}$ that minimizes Equation 4, we obtain the value of $\mathbf{x}$ that minimizes Equation 3 by

$$\mathbf{x}^a = \mathbf{x}^b + \delta\mathbf{x}^a, \tag{5}$$

where $\delta\mathbf{x}^a$ is the minimizer of the incremental cost function.

To further reduce the computational cost, the control variable transform (CVT; e.g., Courtier et al., 1998) is applied in operational variational data assimilation, where a new variable is defined as $\mathbf{z} = \mathbf{L}^{-1}\mathbf{x}$. The transformation matrix $\mathbf{L}$ satisfies $\mathbf{B} = \mathbf{L}\mathbf{L}^{\mathrm{T}}$ but is not necessarily the symmetric square root of the matrix $\mathbf{B}$, and it can be a rectangular matrix. Using $\mathbf{L} = \mathbf{B}^{1/2}$ (the symmetric square root of $\mathbf{B}$) Equation 4 reads

$$J(\delta\mathbf{z}) = \frac{1}{2}\delta\mathbf{z}^{\top}\delta\mathbf{z} + \frac{1}{2}\left(\mathbf{H}\mathbf{B}^{1/2}\delta\mathbf{z}-\mathbf{d}\right)^{\top}\mathbf{R}^{-1}\left(\mathbf{H}\mathbf{B}^{1/2}\delta\mathbf{z}-\mathbf{d}\right), \tag{6}$$

where $\delta\mathbf{z} = \mathbf{B}^{-1/2}\delta\mathbf{x}$. It can be seen from Equation 6 that the use of CVT avoids the need to invert the matrix $\mathbf{B}$.

19422466, 2024, 6, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023MS003871 by Test, Wiley Online Library on [13/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The cost function is minimized when its gradient vanishes. Differentiating Equation 6 with respect to $\delta\mathbf{z}$ gives its gradient

$$\nabla J = \left(\mathbf{I} + \mathbf{B}^{1/2}\mathbf{H}^{\top}\mathbf{R}^{-1}\mathbf{H}\mathbf{B}^{1/2}\right)\delta\mathbf{z} - \mathbf{B}^{1/2}\mathbf{H}^{\top}\mathbf{R}^{-1}\mathbf{d}. \tag{7}$$

Solving $\nabla J = 0$ is equivalent to solving the linear system

$$\mathbf{S}\delta\mathbf{z} = \mathbf{w}, \tag{8}$$

where

$$\mathbf{S} = \mathbf{I} + \mathbf{B}^{1/2}\mathbf{H}^{\top}\mathbf{R}^{-1}\mathbf{H}\mathbf{B}^{1/2} \tag{9}$$

and

$$\mathbf{w} = \mathbf{B}^{1/2}\mathbf{H}^{\top}\mathbf{R}^{-1}\mathbf{d}. \tag{10}$$

The matrix $\mathbf{S}$ is called the (preconditioned) Hessian, which must be positive definite to ensure a unique solution to the minimization of the cost function (Golub & Van Loan, 1996, Chapter 5). Since the matrix $\mathbf{S}$ is a low-rank update to the identity matrix, it has a theoretical minimum eigenvalue of one (e.g., Tabeart et al., 2022).

To solve Equation 8 for $\delta\mathbf{z}$, we may use the conjugate gradient (CG) method (e.g., Barrett et al., 1994; Tabeart, 2019; Trefethen & Bau, 1997). This iterative method has been widely used in NWP to minimize the cost function. In Section 5, we will use the local SVD-FMM in solving Equation 8, along with the CG method.

3. The Local SVD-FMM

In this section, we describe the new local SVD-FMM algorithm and propose a simple parallelization scheme for the method. We first introduce a partition of the observation domain and the domain localization approach (Section 3.1). We then separate the matrix-vector product given by Equation 1 into three terms, making use of the localized observation partition (Section 3.2). We finally describe the algorithm for calculating each term and discuss possible parallelization formulations and algorithmic complexity (Sections 3.3–3.5). The reader can refer to Gimbutas and Rokhlin (2003) and Hu and Dance (2021a) for further theoretical explanations and interpretation of the SVD-FMM, which is the basis for the local SVD-FMM.

3.1. Partition of the Observation Domain

The local SVD-FMM starts by dividing the observation domain into (approximately) equally-sized boxes. The observations could be regularly distributed or irregularly distributed. However, if observation locations are not evenly distributed, we may consider using boxes of different sizes. This ensures that the number of observations in each box is similar. If the observation domain is three-dimensional, then it can be divided into smaller cubes. In practical applications, the number of boxes can be determined such that the average number of observations in each box is smaller than a prescribed value. The optimal configuration of boxes may depend on the individual application and computer architecture. Figure 1 shows an example of the observation domain partition, where the grid spacing of observations is 6/14° in the East-West direction (equates to about 30 km at 50.5°N) and 6/11° in the North-South direction (equates to about 30 km), and the observation domain is divided into 8×8 boxes.

After observation partitioning, we define four fields for each box. The *near field* of a box b, denoted by $\mathcal{N}_b$, is made of itself and all the boxes that connect to it (see the left panel of Figure 2). The *far field* of a box b, denoted by $\mathcal{F}_b$, consists of all the other boxes. The *interaction field* of box b, denoted by $\mathcal{L}_b$, consists of a subset of boxes in $\mathcal{F}_b$, which are within a certain distance of box b. The distance is given by a non-dimensional localization length parameter (h), denoting the number of boxes. For example, the right panel of Figure 2 shows the interaction field of box b with $h = 3$, indicating that the interaction field contains boxes within a distance of three boxes from b, excluding the near-field boxes. The non-dimensional localization length parameter can also be expressed in units of length by simply multiplying it by the size of the box. The *non-interaction* boxes of b, denoted by $\mathcal{D}_b$, are given

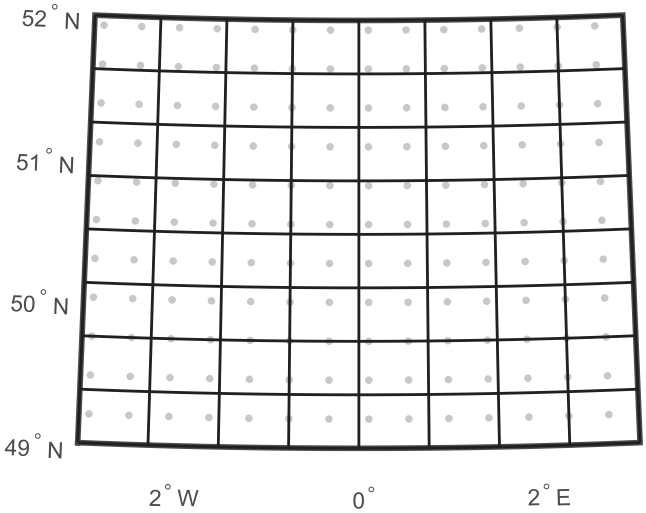

Figure 1. An example of the observation domain partition. Observations (gray dots) are approximately regularly distributed over a region from 49°N to 52°N and 3°W to 3°E with a grid spacing of about 30 km, resulting in 154 observations. The observation domain is divided into 8 × 8 boxes.

by the boxes that are in $\mathcal{F}_b$ but not in $\mathcal{L}_b$. Let m_b, $m_{\mathcal{N}_b}$, $m_{\mathcal{L}_b}$ and $m_{\mathcal{D}_b}$ denote the number of observations in box b, $\mathcal{N}_b$, $\mathcal{L}_b$ and $\mathcal{D}_b$, respectively, then we have $m = \sum_{b=1}^{N_{box}} m_b$, where N_{box} is the total number of boxes, and $m = m_{\mathcal{N}_b} + m_{\mathcal{F}_b} = m_{\mathcal{N}_b} + m_{\mathcal{L}_b} + m_{\mathcal{D}_b}$ for any b.

3.2. Separation of the Matrix-Vector Product

We first separate the vector $\mathbf{q}$ into N_{box} sub-vectors, where N_{box} is the number of boxes used to divide the observation domain. The sub-vector of the vector $\mathbf{q}$ corresponding to box b is denoted by $\mathbf{q}(\mathbf{I}_b) = \{\mathbf{q}(i)|i \in \mathbf{I}_b\}$, which contains elements of the vector $\mathbf{q}$ that correspond to the observations in box b. The vector $\mathbf{I}_b \in \mathbb{R}^{m_b}$ contains the indices of observations in box b. We then separate the calculation of the vector $\mathbf{q}(\mathbf{I}_b)$ into three matrix-vector products,

$$\mathbf{q}(\mathbf{I}_b) = \mathbf{A}(\mathbf{I}_b, \mathbf{I}_{\mathcal{N}_b})\,\mathbf{d}(\mathbf{I}_{\mathcal{N}_b}) + \mathbf{A}(\mathbf{I}_b, \mathbf{I}_{\mathcal{L}_b})\,\mathbf{d}(\mathbf{I}_{\mathcal{L}_b}) + \mathbf{A}(\mathbf{I}_b, \mathbf{I}_{\mathcal{D}_b})\,\mathbf{d}(\mathbf{I}_{\mathcal{D}_b}), \tag{11}$$

where vectors $\mathbf{I}_{\mathcal{N}_b} \in \mathbb{R}^{m_{\mathcal{N}_b}}$, $\mathbf{I}_{\mathcal{L}_b} \in \mathbb{R}^{m_{\mathcal{L}_b}}$ and $\mathbf{I}_{\mathcal{D}_b} \in \mathbb{R}^{m_{\mathcal{D}_b}}$ denote the sets of observation indices in $\mathcal{N}_b$, $\mathcal{L}_b$, and $\mathcal{D}_b$, respectively, the matrices $\mathbf{A}(\mathbf{I}_b, \mathbf{I}_{\mathcal{N}_b}) = \{\mathbf{A}(i,j)|i \in \mathbf{I}_b, j \in \mathbf{I}_{\mathcal{N}_b}\}$, $\mathbf{A}(\mathbf{I}_b, \mathbf{I}_{\mathcal{L}_b}) = \{\mathbf{A}(i,j)|i \in \mathbf{I}_b, j \in \mathbf{I}_{\mathcal{L}_b}\}$, and $\mathbf{A}(\mathbf{I}_b, \mathbf{I}_{\mathcal{D}_b}) = \{\mathbf{A}(i,j)|i \in \mathbf{I}_b, j \in \mathbf{I}_{\mathcal{D}_b}\}$ denote sub-matrices of the matrix $\mathbf{A}$ that are comprised of specific rows and columns of the matrix $\mathbf{A}$ given by the corresponding observation indices, and the vectors $\mathbf{d}(\mathbf{I}_{\mathcal{N}_b}) = \{\mathbf{d}(j)|j \in \mathbf{I}_{\mathcal{N}_b}\}$, $\mathbf{d}(\mathbf{I}_{\mathcal{L}_b}) = \{\mathbf{d}(j)|j \in \mathbf{I}_{\mathcal{L}_b}\}$, and $\mathbf{d}(\mathbf{I}_{\mathcal{D}_b}) = \{\mathbf{d}(j)|j \in \mathbf{I}_{\mathcal{D}_b}\}$ denote sub-vectors of the vector $\mathbf{d}$.

The first matrix-vector product on the right-hand side of Equation 11 is the near-field matrix-vector product, which is calculated exactly (see Section 3.3). The second matrix-vector product is the interaction-field matrix-vector product, which is approximated using the SVD (see Section 3.4). Thus, the accuracy depends on the number of singular values/vectors used in the approximation (p). The last matrix-vector product is the non-interaction-field matrix-vector product, which is completely discarded (see Section 3.5). Figure 3 shows which elements of the matrix $\mathbf{A}$ are involved in the near-field, interaction-field and non-interaction-field matrix-vector products in the calculation of $\mathbf{q}(\mathbf{I}_b)$ for all boxes. For example, the first panel of Figure 3 shows which elements of the matrix $\mathbf{A}$ are used to form $\mathbf{A}(\mathbf{I}_b, \mathbf{I}_{\mathcal{N}_b})$ for $b = 1, \ldots, N_{box}$.

To provide the reader with further intuition about the approximation, we also plot a row of the matrix $\mathbf{A}$ and some example approximated forms in physical space (since matrix elements correspond to observation locations). To do this, we first create the matrix $\mathbf{R}$ using the second-order auto-regressive (SOAR) correlation function (e.g., Daley, 1994; Tabeart et al., 2018),

$$\mathbf{C}_{SOAR}(i,j) = \sigma^2\left(1 + \frac{|\Delta_{i,j}|}{l}\right)\exp\left(\frac{-|\Delta_{i,j}|}{l}\right), \tag{12}$$

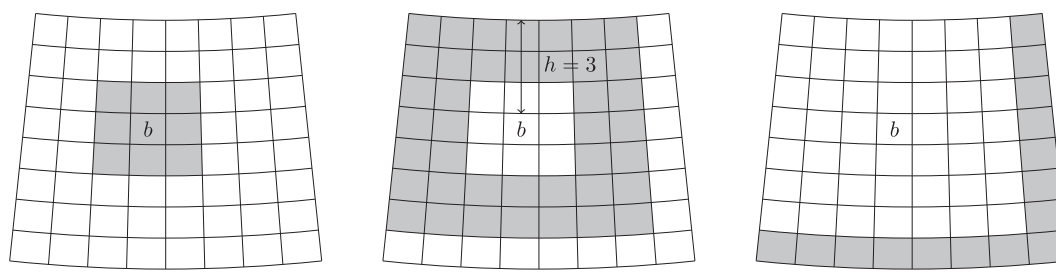

Figure 2. Illustration of the near, interaction and non-interaction fields of the box b. (left panel) The shaded boxes are in the near field of the box b. (middle panel) The shaded boxes are in the interaction field of the box b with a non-dimensional localization length parameter $h = 3$. (right panel) The shaded boxes are in the non-interaction field of the box b.

19422466, 2024, 6, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023MS003871 by Test, Wiley Online Library on [13/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

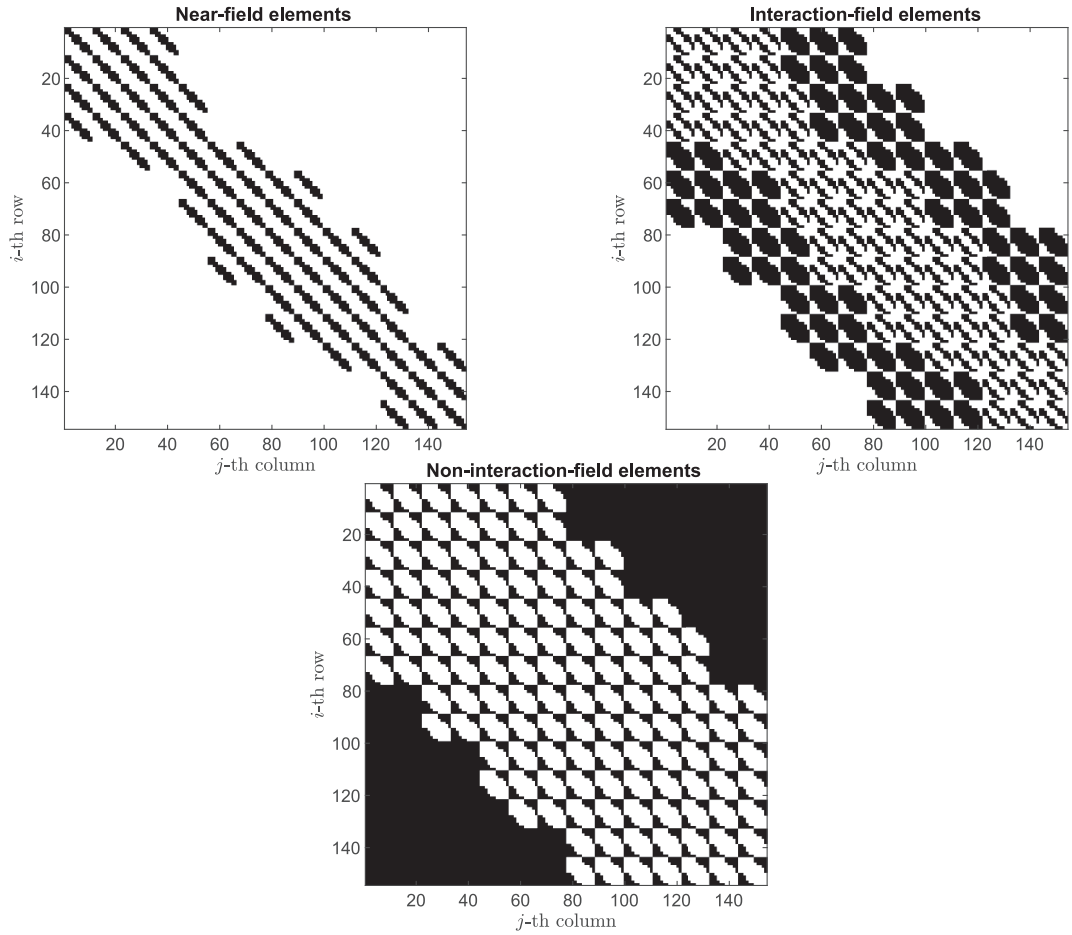

Figure 3. Plots illustrating the position (marked by black squares) of the elements of the matrix $\mathbf{A} \in \mathbb{R}^{154\times154}$ used for the near-field, interaction-field, and non-interaction-field matrix-vector products in Equation 11. Observations are distributed and partitioned as in Figure 1. The localization length is $h = 3$.

where σ is the error standard deviation, l is the error correlation lengthscale and Δ_{ij} denotes the great-circle distance between two points. We then use the MATLAB function *inv.m* (MATLAB INV, 2022) to invert the matrix **R**. The SOAR covariance matrix is symmetric positive definite and thus invertible. After obtaining the matrix **A**, we select the row corresponding to the observation located in the center of the domain and plot this row in physical space in the left panel of Figure 4. We find that the magnitude of the elements in this row decreases rapidly from the center and the sign of the elements oscillates. Similar results have been observed in other precision matrices (e.g., Morrison et al., 2022). The middle and right panels of Figure 4 show the same row of the matrix **A** that is approximated by the local SVD-FMM. The middle panel shows a good approximation of the matrix **A** where more elements are included in the interaction field and these elements are well approximated by the retained singular values/vectors. In contrast, the right panel shows a bad approximation where more elements are discarded and the interaction-field elements are not well approximated.

Although only the near-field elements of the matrix **A** are used without approximation, we still need the inverse of the full matrix **R**. This is because, for a full matrix, the inverse of a matrix sub-block typically depends on all of the other sub-blocks of the matrix (Bernstein, 2009, Chapter 2). In practice, if correlations between observation errors for different observation types are excluded, the matrix **R** can be considered block diagonal with zero blocks on the off-diagonals and each diagonal block corresponding to a different observation type (e.g., Stewart et al., 2013). In this case, the inverse of each block can be calculated independently, and the local SVD-FMM can be applied to each block separately.

19422466, 2024, 6, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023MS003871 by Test, Wiley Online Library on [13/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

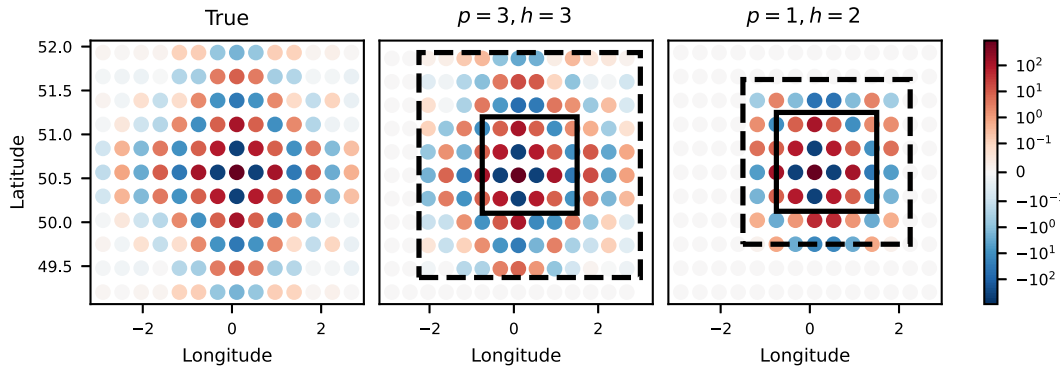

Figure 4. One row of the true and approximated observation error precision matrix, **A**, in physical space. Observations are distributed and partitioned as in Figure 1. The observation corresponding to the chosen row is located at the darkest red circle in the center of the domain. The true matrix is the inverse of the second-order auto-regressive covariance matrix with a standard deviation of 1 and a correlation lengthscale of 160 km (Equation 12). The approximated matrices are obtained by approximating the true matrix using the local singular value decomposition approach to the fast multipole method with different numbers of singular values (p) and localization lengths (h). The corresponding physical locations of near-field elements are given by circles inside the solid line square or at its edge. The corresponding physical locations of interaction-field elements are given by circles outside the solid line square, but inside or at the edge of the dashed line square. The corresponding physical locations of the non-interaction-field elements are given by circles outside the dashed line square.

3.3. The Near-Field Matrix-Vector Product

The near-field matrix-vector product is expected to contribute most to the results of Equation 11 as the elements of the matrix **A** are expected to decay with distance from the diagonal (Eijkhout & Polman, 1988). In addition, from a physical point of view, since spatial observation error correlation is expected to decrease with distance, the observations in the near field provide more information than those in the interaction and non-interaction fields. Therefore, the near-field matrix-vector product should be computed without any approximation,

$$\mathbf{A}(\mathbf{I}_b, \mathbf{I}_{\mathcal{N}_b})\,\mathbf{d}(\mathbf{I}_{\mathcal{N}_b}) = \sum_{i \in \mathbf{I}_b} \sum_{j \in \mathbf{I}_{\mathcal{N}_b}} \mathbf{A}(i,j)\mathbf{d}(j). \tag{13}$$

Equation 13 should only have a very small numerical roundoff error, which is negligible in determining the accuracy of the local SVD-FMM. Since the average size of $\mathbf{I}_b$ is m/N_{box} and the average size of $\mathbf{I}_{\mathcal{N}_b}$ is $9m/N_{box}$, the amount of work for computing Equation 13 for a box is approximately $9(m/N_{box})^2$ additions and multiplications. In total, $9m^2/N_{box}$ additions and multiplications are required for the computations for all boxes.

Near-field matrix-vector products can be calculated in parallel using a simple strategy. Let PE-b denote a PE that has been assigned to carry out the computation task for box b and let it store the vector $\mathbf{d}(\mathbf{I}_b)$, and the matrices $\mathbf{A}(\mathbf{I}_b, \mathbf{I}_{\mathcal{N}_b})$ and $\mathbf{A}(\mathbf{I}_b, \mathbf{I}_{\mathcal{L}_b})$. To calculate Equation 13, each PE needs to receive elements of **d** from the PEs holding the data for the neighboring boxes. This means that each PE needs to communicate with at most 8 PEs, as each box can have up to 8 neighbors. We note that the observation domain must be partitioned to allow each PE to have enough memory to store its portion of the vector **d** and the matrix **A**. In addition, the load balance between PEs and the balance of communication overheads per PE need to be taken into account.

3.4. The Interaction-Field Matrix-Vector Product

The interaction-field matrix-vector product is approximated by a computationally efficient algorithm. This algorithm relies on the SVD for data compression. The SVD is mathematically defined by (e.g., Golub & Van Loan, 1996)

$$\mathbf{M} = \sum_{k=1}^{\min(n_1,n_2)} \mathbf{u}_k s_k (\mathbf{v}_k)^{\top}, \tag{14}$$

19422466, 2024, 6, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023MS003871 by Test, Wiley Online Library on [13/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

19422466, 2024, 6, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023MS003871 by Test, Wiley Online Library on [13/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

where $\mathbf{M}$ is an $n_1 \times n_2$ matrix, $\mathbf{u}_k \in \mathbb{R}^{n_1}$ is the kth left singular vector, s_k is the kth singular value, and $\mathbf{v}_k \in \mathbb{R}^{n_2}$ is the right singular vector. For a square, symmetric positive-definite matrix, the SVD is equivalent to eigenvalue decomposition, and we have $\mathbf{u}_k = \mathbf{v}_k$.

There are four main steps to calculate the interaction-field matrix-vector product: (a) Obtain the numerical operators, which are used to compress, translate and decompress the information in the vector $\mathbf{d}$; (b) Compress the part of $\mathbf{d}$ that is allocated to each box; (c) For each box b, translate the compressed information from boxes in b's interaction field ($\mathcal{L}_b$) to b; and (d) Decompress the translated information for each box.

The mathematical detail, algorithmic complexity, parallelization scheme, and communication costs for each step of the interaction-field calculation are presented below. We assume that the data has already been allocated to each PE as described in Section 3.3.

1. Compute p-term truncated SVDs of the interaction-field sub-matrix for each box,

$$\mathbf{A}(\mathbf{I}_b, \mathbf{I}_{\mathcal{L}_b}) \approx \sum_{k=1}^{p} \mathbf{u}_{b,k} s_{b,k} (\mathbf{v}_{b,k})^{\top}, \tag{15}$$

where $b = 1, \ldots, N_{box}$, $\mathbf{u}_{b,k} \in \mathbb{R}^{m_b}$ denotes the kth left singular vector for box b, $\mathbf{v}_{b,k} \in \mathbb{R}^{m_{\mathcal{L}_b}}$ denotes the kth right singular vector for box b and $s_{b,k}$ denotes the kth singular value for box b, and we should have $p \ll m_b$. The computation of the truncated SVDs is perfectly parallel. Each PE can perform a matrix decomposition independently. After the SVDs, the memory allocated in PE-b for the sub-matrix $\mathbf{A}(\mathbf{I}_b, \mathbf{I}_{\mathcal{L}_b})$ can be released. The singular vectors and singular values obtained can be used to compute Equation 1 with any vector $\mathbf{d}$ as long as the matrix $\mathbf{A}$ remains the same. In practical data assimilation applications, the matrix $\mathbf{A}$ may vary each assimilation cycle due to factors such as quality control but typically does not change between iterations of the minimization scheme in a given cycle.

2. Project the sub-vector $\mathbf{d}(\mathbf{I}_b)$ for each box b onto the basis given by p left singular vectors and obtain the coefficient of projection,

$$\Phi_{b,k} = \mathbf{u}_{b,k} \cdot \mathbf{d}(\mathbf{I}_b), \tag{16}$$

where $k = 1, \ldots, p$ and the symbol $\cdot$ denotes a dot product. For a fixed b and a fixed k, Equation 16 requires $2m_b$ operations. Summing over each value of b and each value of k, the total number of operations is $2mp$. The computation of Equation 16 is perfectly parallel.

3. Transform the projection on the basis given by left singular vectors into the basis given by right singular vectors and obtain the coefficient of projection,

$$\Psi_{b,k} = \sum_{b' \in \mathcal{L}_b} \sum_{k'=1}^{p} \mathbf{T}_{b,b'}(k,k') \Phi_{b',k'}, \tag{17}$$

where $b' \in \mathcal{L}_b$ is a box in the interaction field of box b and $\mathbf{T}_{b,b'} \in \mathbb{R}^{p \times p}$ is the *translation operator* between b and b', which is given by

$$\mathbf{T}_{b,b'}(k,k') = \sum_{i=1}^{m_{b'}} \mathbf{v}_{b,k}(i) \mathbf{u}_{b',k'}(i), \tag{18}$$

where $k, k' = 1, \ldots, p$. It should be noted that $\mathbf{v}_{b,k}$ has a larger size than $\mathbf{u}_{b',k'}$. Thus, we need to multiply each element of $\mathbf{u}_{b',k'}$ by the element of $\mathbf{v}_{b,k}$ corresponding to the same observation as the element of $\mathbf{u}_{b',k'}$. Equation 17 requires less than $2(N_{box}p)^2$ operations because a box's interaction field always contains less than N_{box} entries. The parallelization of this step is the most expensive compared to the parallelization of the other steps because it requires each PE-b to communicate with all the PEs holding data for the boxes in b's interaction field. The computational cost depends on the number of entries in the interaction fields.

19422466, 2024, 6, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023MS003871 by Test, Wiley Online Library on [13/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

4. Obtain the final result for the interaction-field computation,

$$\mathbf{A}(\mathbf{I}_b, \mathbf{I}_{\mathcal{L}_b})\,\mathbf{d}(\mathbf{I}_{\mathcal{L}_b}) \approx \sum_{k=1}^{p} \mathbf{u}_{b,k} s_{b,k} \Psi_{b,k}. \tag{19}$$

This step requires $2mp$ operations and is perfectly parallel.

The computational costs for the truncated SVDs (Equation 15) and the translation operator (Equation 18) have not been discussed above because they can be pre-computed and used for any vector $\mathbf{d}$ if the matrix $\mathbf{A}$ is unchanged. The interaction-field computation could become expensive if the truncated SVDs and the translation operator need to be computed frequently. In practice, since the observations vary in space and time, the SVDs need to be recomputed for each assimilation cycle. However, since the observation precision matrix typically remains constant between iterations in a given cycle, the computed singular values and singular vectors can be reused at each iteration step.

We refer to the error arising from the interaction-field computation as the approximation error, which consists of the truncation error of the SVDs and the translation error of the translation operator. The truncation error and the translation error using p singular vectors are both dependent on the $(p+1)$th singular values of the submatrices of $\mathbf{A}$ (Bernstein, 2009; Gimbutas & Rokhlin, 2003, Fact 9.14.28). We showed in our previous work that the error of the SVD-FMM (without localization) using p singular vectors is smaller if the mean $(p+1)$th singular value of the submatrices of the matrix $\mathbf{A}$ is smaller (Hu & Dance, 2021a; Figure 9). Thus, the optimal value of p for an application should depend on how fast the singular value spectra decrease. In Section 4, we show numerically how the average of the $(p+1)$th singular values of the submatrices of $\mathbf{A}$, as well as the accuracy of the local SVD-FMM, vary with increasing p.

3.5. The Non-Interaction-Field Matrix-Vector Product

The non-interaction-field matrix-vector product in Equation 11 is completely discarded, as we expect that this term should have the smallest contribution to the result of $\mathbf{q}$ compared to the other two terms. This is because for spatially correlated observation errors (e.g., errors from geostationary satellite and radar observations), the observation error covariance matrix is expected to be banded, with the bandwidth depending on the observation error correlation lengthscale, and the inverses of positive definite band matrices are known to exhibit an exponential decay of their elements away from the main diagonal (Eijkhout & Polman, 1988). A band matrix is a matrix that contains non-zero elements near the diagonal and zero elements far away from the diagonal (Golub & Van Loan, 1996, Chapter 4). We refer to the error caused by discarding the non-interaction-field matrix-vector product as the localization error, which should depend on the observation error correlation lengthscale and the localization length. We give some example numerical results on how the localization error varies with observation error correlation lengthscale. We measure the localization error by the relative difference,

$$\mathrm{RE} = \frac{\|\mathbf{A} - \hat{\mathbf{A}}\|_F}{\|\mathbf{A}\|_F}, \tag{20}$$

where the matrix $\mathbf{A}$ is created in the same way as described in Section 3.2, the matrix $\hat{\mathbf{A}}$ is given by setting the non-interaction-field elements of the matrix $\mathbf{A}$ to be zero and $\|\cdot\|_F$ denotes the Frobenius norm. Figure 5 shows that the localization error (RE) increases as the observation error correlation lengthscale increases. Nevertheless, for all the correlation lengthscales considered, the localization error is smaller than 0.24%.

In ensemble filtering, domain localization may cause abrupt transitions between neighboring sub-domains, as the analyses are calculated independently for each sub-domain using different subsets of observations (e.g., Janjić et al., 2011). However, this should not be an issue for the local SVD-FMM because the analysis is calculated for the entire domain.

3.6. The Non-Dimensional Localization Length Parameter

In Figure 1, boxes are approximately rectangular with longer sides along the lines of latitudes. As a consequence, they may encompass more observations in the East-West direction than in the North-South direction. If we use the

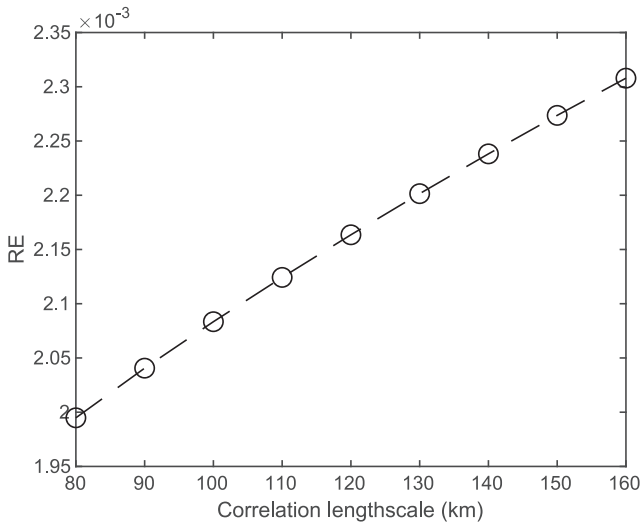

Figure 5. The relative localization error with the localization length of $h = 3$ as a function of observation error correlation lengthscale. The observations are distributed and partitioned as in Figure 1.

same localization length for both directions, we may include more observations in the East-West direction. To include a similar number of observations in the two directions, we can use different localization lengths for the two directions, as shown in Figure 6.

Moreover, we may also use non-uniform localization lengths to suit the geometry of spatial error correlations. In practical applications, observation error correlations may not be isotropic. For instance, observation error correlation lengthscales for the DRWs vary with the height of the observation and with the distance of the observation away from the radar (Waller, Simonin, et al., 2016). For such spatial error correlations, we may need to set the localization lengths for each box differently.

In general, a longer localization length means fewer matrix elements are discarded, thus improving the accuracy of the local SVD-FMM. However, at the same time, longer localization lengths increase the size of the interaction-field submatrices, thereby increasing the computational cost of the SVD (see Sections 4 and 5).

3.7. Computational Costs

The local SVD-FMM reduces the algorithmic complexity of calculating Equation 1. The standard matrix-vector multiplication requires $2m^2$ operations. The local SVD-FMM requires less than $9m^2/N_{box} + 4mp + 2(N_{box}p)^2$ operations. The operations for computing Equations 15 and 18 have not been counted, as they only need to be calculated once for the same matrix $\mathbf{A}$.

We further compare the communication costs of the parallelization scheme of the local SVD-FMM with two distinct standard parallel formulations of matrix-vector multiplication. The two formulations start with two different partitions of the matrix. The row-wise (column-wise) partitioning means an equal division of rows (columns) of the matrix between PEs. Elements of the vector are equally allocated among PEs for both cases. With row-wise partitioning, each PE needs to collect the portion of the vector stored in every other PE, which requires an all-to-all broadcast operation. The size of the message to be transferred is m/N_{PE}, where N_{PE} is the number of PEs. With column-wise partitioning, an all-to-one reduction operation is first used to add up the partial results calculated locally by each PE, and a scatter operation is then used to spread the final result to all PEs. The size of the message transmitted in the all-to-one reduction and scatter operations is m and m/N_{PE} respectively. More details on the communication operations can be found in Grama et al. (2003). In the parallelization scheme of the local SVD-FMM, communication is required in the near-field computation and the third step of the interaction-field computation. In the near-field computation, each PE should communicate with up to 8 PEs to collect the elements of the vector $\mathbf{d}$. In the third step of the interaction-field computation, each PE should communicate with a number of PEs to collect the coefficient of projection, Φ_k^b. The exact number of PEs involved depends on the choice of localization length. These communication tasks can be completed together by an all-to-all broadcast operation, where the size of the message transmitted between a box and its near field is m/N_{PE}, and the size of the message transmitted between a box and its interaction field is p. To compare the communication costs, we can compare the communication operations required and the size of messages to be transferred. An all-to-all broadcast is required in both the parallelization scheme of the local SVD-FMM and the row-wise partitioning formulation. The column-wise partitioning formulation requires two communication operations, in which the scatter operation requires a similar communication time as the all-to-all broadcast (Grama et al., 2003). Due to the smaller size of messages transferred, the

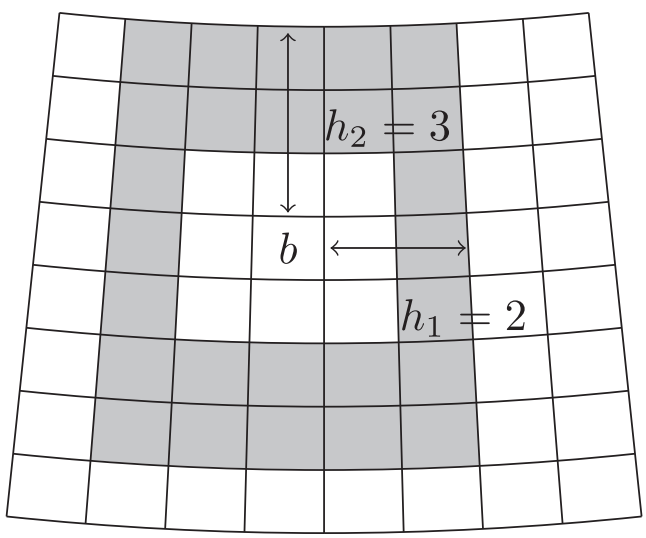

Figure 6. Illustration of using different localization lengths for the latitudinal and longitudinal directions.

19422466, 2024, 6, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023MS003871 by Test, Wiley Online Library on [13/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

19422466, 2024, 6, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023MS003871 by Test, Wiley Online Library on [13/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

parallelization scheme of the local SVD-FMM requires fewer communication costs than the two standard parallel formulations of matrix-vector multiplication.

4. The Local SVD-FMM Accuracy Experiments

The accuracy of the local SVD-FMM is determined by the number of singular vectors (p), the localization length (h) and the structure of the matrix $\mathbf{A}$. The structure of the matrix $\mathbf{A}$ is affected by the observation error correlation lengthscale (l). In this section, we carry out numerical experiments to investigate the accuracy of the local SVD-FMM in calculating Equation 1 under different values of p, h, and l. We focus on the accuracy of the matrix-vector products, and the effect of the local SVD-FMM on variational data assimilation will be examined in Section 5.

4.1. Design of Accuracy Experiments

Our observations are distributed over a region from 54°N to 60°N and 6°W to 6°E with a grid length of approximately 12 km, which results in $m = 3{,}456$ observations. We divide the observation domain into 8×8 boxes as demonstrated in Figure 1. The side length of the boxes is about 83–98 km. The distribution of our observations is similar to that of the geostationary satellite data over the UK (Waller, Ballard, et al., 2016). A moderate observation grid spacing is chosen to allow our experiments to be completed in a relatively short time. The matrix $\mathbf{R}$ is given by the SOAR covariance matrix (Equation 12). The observation error standard deviation is assumed to be 1, and the correlation lengthscale varies between 80 and 160 km with an interval of 10 km. These values are selected based on horizontal error correlations estimated for real observational data (e.g., Cordoba et al., 2017; Waller, Ballard, et al., 2016). The matrix $\mathbf{A}$ is obtained by inverting the matrix $\mathbf{R}$ using the INV function in MATLAB (MATLAB INV, 2022). The INV function uses an LDL decomposition for Hermitian matrices (Golub & Van Loan, 1996). The vector $\mathbf{d}$ is formed using random values that follow a multivariate normal distribution with zero mean and the covariance given by $\mathbf{R} + \mathbf{HBH}^{\mathrm{T}}$. The matrix $\mathbf{H}$ is considered an identity matrix, and the matrix $\mathbf{B}$ is also given by the SOAR covariance matrix. The background error standard deviation and correlation lengthscale are 0.6 and 20 km, respectively. These values are chosen based on that estimated for wind observations in km-scale NWP (Ballard et al., 2016).

We measure the accuracy of the local SVD-FMM in calculating Equation 1 using the root mean square error (RMSE),

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{m}\left(\mathbf{q}_{std}(i) - \mathbf{q}_{fmm}(i)\right)^2}{m}}, \tag{21}$$

where $\mathbf{q}_{std}(i)$ and $\mathbf{q}_{fmm}(i)$ denote the ith element of the vector $\mathbf{q}$ calculated using the standard matrix-vector multiplication and the local SVD-FMM, respectively.

4.2. Experiments With Fixed Localization Length

In these experiments, we investigate the accuracy of the local SVD-FMM in calculating the matrix-vector product when the localization length is fixed to $h = 3$, while the number of singular vectors and the observation error correlation lengthscale vary. As expected, we found that the RMSE of the local SVD-FMM decreases as more singular vectors are used (see Figure 7). We also note that the RMSE decreases quickly for the first several values of p while decreasing more slowly as its value increases further. This is related to the spectra of the singular values of the sub-matrices of the matrix $\mathbf{A}$. We explained in the last paragraph of Section 3.4 that the approximation error of the local SVD-FMM depends on the $(p + 1)$th singular values of the submatrices of the matrix $\mathbf{A}$, and in our experiments, the singular values of the submatrices of the matrix $\mathbf{A}$ drop rapidly (Figure 8, discussed further below). Since the approximation error approaches a minimum with the use of more singular vectors, the RMSE at $p = 10$ gives us a rough idea of the size of the localization error in the local SVD-FMM.

Figure 7 also shows that the RMSE of the local SVD-FMM increases with the observation error correlation lengthscale. This is attributed to an increase in the approximation error and the localization error. We have shown in Figure 5 that a larger observation error correlation lengthscale will lead to a larger localization error. The observation error correlation lengthscale also affects the approximation error because it changes the singular values of the sub-matrices of the matrix $\mathbf{A}$. We find in our experiments that if the correlation lengthscale is longer,

19422466, 2024, 6, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023MS003871 by Test, Wiley Online Library on [13/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

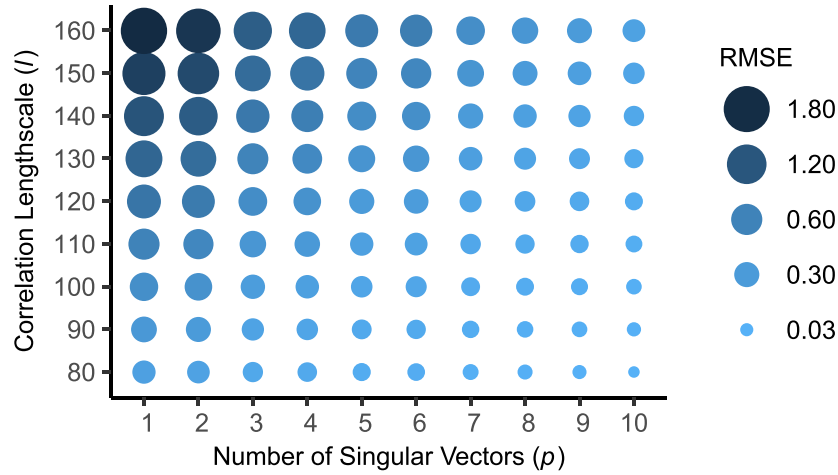

Figure 7. The root mean square error (RMSE) in using the local singular value decomposition approach to the fast multipole method to calculate the matrix-vector product given in Equation 1 under different numbers of singular vectors and correlation lengthscales. The localization length parameter is 3. The RMSE is averaged over 100 realizations of the vector. The size and the color of the circles linearly increase with the RMSE.

then the mean singular value of the submatrices of the matrix **A** is larger for a given k (Figure 8). For a given truncation, p, a larger mean $(p+1)$th singular value means a larger approximation error. Therefore, if the correlation lengthscale is longer, the approximation error will be larger. Our previous work has shown that the error of the SVD-FMM (without localization) also increases with the observation error correlation lengthscale (Hu & Dance, 2021a; Figure 10). A similar result should be expected here because the source of the error is the same.

While the singular values of the *submatrices* of **A** can be shown mathematically to control the approximation error, we have found that the singular values (or equivalently the eigenvalues) of the *full* matrix can also provide a rough guide (Hu & Dance, 2021a). Various studies have investigated the behavior of the eigenspectra of correlation matrices with varying correlation lengthscale (e.g., Fowler et al., 2018; Haben et al., 2011).

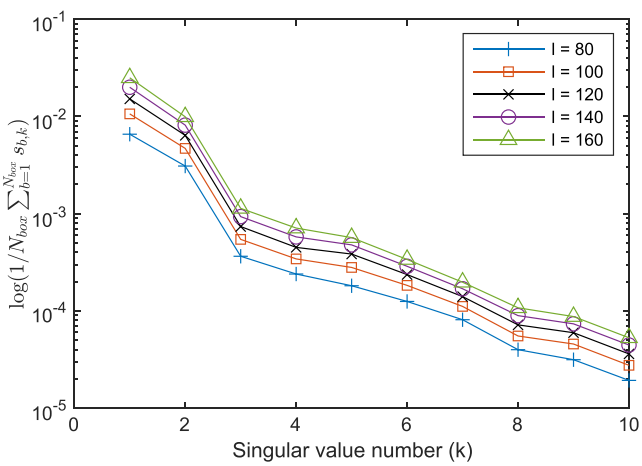

Figure 8. Average of the kth singular values ($s_{b,k}$) of N_{box} submatrices of the matrix **A**. Only the results for the first 10 singular values are plotted. The $s_{b,k}$ are calculated as shown in Equation 15. Different symbols represent results for different observation error correlation lengthscales (l). A logarithmic scale is used on the axis of ordinates to allow a clear comparison between smaller singular values.

4.3. Experiments With a Fixed Number of Singular Vectors

We now investigate the accuracy of the local SVD-FMM when using 10 singular vectors and varying the localization length and the observation error correlation lengthscale. The localization length is given by a non-dimensional parameter (h) that describes the number of boxes. In these experiments, an approximate physical localization length can be obtained by multiplying h by 90 km (the length of the box). Figure 9 shows that there is a relatively big drop in the RMSE as h increases from 2 to 3, or from 180 to 270 km in metric units. However, when increasing h further, changes in the RMSE are too small to be observed visually. The optimal value of h should depend on the configuration of boxes and the structure of the matrix **A**.

As with Figures 7 and 9 shows that the RMSE of the local SVD-FMM increases with the observation error correlation lengthscale. We discussed previously that an increase in the observation error correlation lengthscale increases the approximation error and the localization error. An additional piece of information provided by Figure 9 is that the RMSE at $h = 7$ gives the size of the approximation error because the localization error is zero when $h = 7$.

4.4. Experiments With a Fixed Correlation Lengthscale

We finally carry out experiments of the local SVD-FMM using an observation error correlation lengthscale of 80 km, and different localization lengths

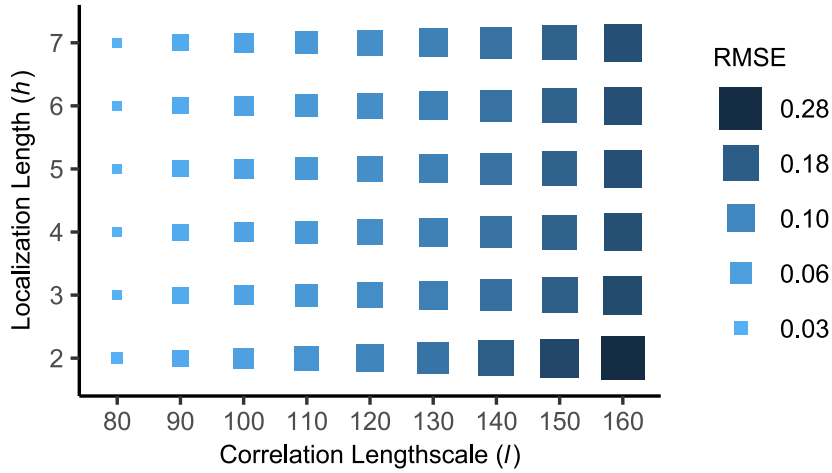

Figure 9. As Figure 7, but under different localization lengths and observation error correlation lengthscales. Ten singular vectors are used. Note that the scale used here is different to the one used in Figure 7, but the size and colors of the squares nevertheless increase linearly with the root mean square error.

and numbers of singular vectors. Figure 9 shows that increasing the number of singular vectors has a much greater effect on the RMSE than increasing the localization length. This indicates that the approximation error dominates the RMSE of the local SVD-FMM. While Figure 9 shows results using a particular observation error correlation lengthscale, we expect similar results when the observation error correlation lengthscale is increased to a certain value (see Figure 9).

5. Data Assimilation Experiments

In this section, we conduct variational data assimilation experiments using the local SVD-FMM. We investigate (a) whether the use of the local SVD-FMM affects the number of iterations to solve the cost function, (b) how much wall-clock time is saved by using the local SVD-FMM, and (c) how accurate the analyses obtained using the local SVD-FMM are. Analysis accuracy at different spatial scales is examined because we expect that including correlated observation error statistics will modulate the scales that observation information projects onto in the analysis and thus improve the analysis accuracy at these scales (Fowler et al., 2018; Stewart et al., 2008, 2013). We wish to check whether this benefit is maintained with the local SVD-FMM approximation. We do not use

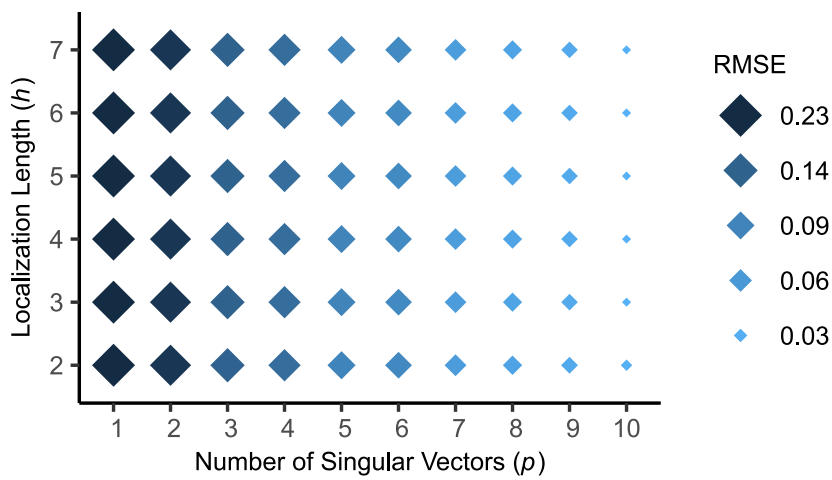

Figure 10. As Figure 7, but under different localization lengths and different numbers of singular vectors. The observation error correlation lengthscale is 80 km. Note that the scale used here is different to those used in Figures 7 and 9, but the size and colors of the diamonds nevertheless increase linearly with the root mean square error.

19422466, 2024, 6, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023MS003871 by Test, Wiley Online Library on [13/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

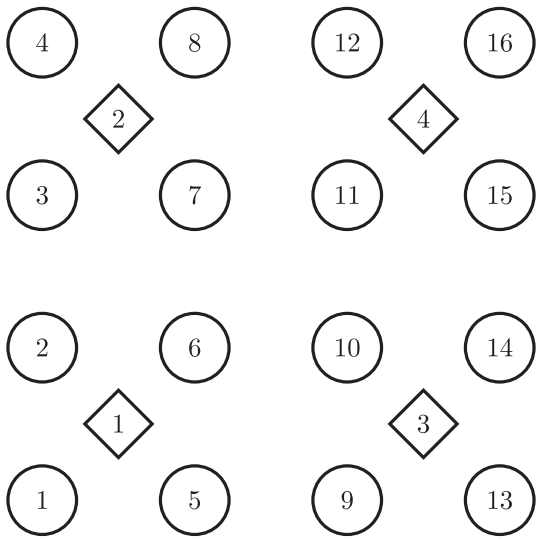

Figure 11. Location of observations (diamonds) with respect to model grid points (circles). The numbers give the order of elements when converting 2-D data into a column vector.

parallel computing because the size of the problem in this experiment is moderate and sequential computing is sufficient to provide the results we need for this initial study.

5.1. Design of Data Assimilation Experiments

In our variational data assimilation experiment, we solve Equation 8 for $\delta\mathbf{z}$ using the CG method. This is equivalent to solving the two-side preconditioned system (Haben, 2011, Chapter 4),

$$\mathbf{B}^{1/2}\widetilde{\mathbf{S}}\mathbf{B}^{1/2}\left(\mathbf{B}^{-1/2}\delta\mathbf{x}\right) = \mathbf{B}^{1/2}\widetilde{\mathbf{w}}, \tag{22}$$

where

$$\widetilde{\mathbf{w}} = \mathbf{H}^{\top}\mathbf{R}^{-1}\mathbf{d} \tag{23}$$

and

$$\widetilde{\mathbf{S}} = \mathbf{B}^{-1} + \mathbf{H}^{\top}\mathbf{R}^{-1}\mathbf{H} \tag{24}$$

is the Hessian of the cost function given by Equation 4.

To solve Equation 8, we need a background error covariance matrix ($\mathbf{B}$), an observation error covariance matrix ($\mathbf{R}$), an observation operator ($\mathbf{H}$) and an innovation vector ($\mathbf{d}$). We assume that our model grid points are distributed over the region shown in Figure 1 with a grid length of approximately 3 km. This leads to a model state vector of size $n = 15,904$. The matrix $\mathbf{B}$ is then created using the SOAR correlation function given by Equation 12 with an error standard deviation of 1 and an error correlation lengthscale of 80 km. We further assume our observations to be distributed with a grid length of about 6 km, which is similar to the footprint of SEVIRI observations in the mid-latitudes. This leads to $m = 3,976$ observations. The matrix $\mathbf{R}$ is created using the first-order auto-regressive (FOAR) correlation function (e.g., Stewart et al., 2013), which is also known as the Markov correlation function and is used for Doppler radar wind observations in operational NWP (Simonin et al., 2019). The FOAR covariance matrix is

$$\mathbf{C}_{FOAR}(i,j) = \sigma^2 \exp\left(\frac{-|\Delta_{i,j}|}{l}\right), \tag{25}$$

where σ, l and $\Delta_{i,j}$ are defined in Equation 12. We set the observation error standard deviation to be the same as the background error standard deviation and the observation error correlation lengthscale to 160 km. Since the observation error correlation lengthscale is larger than the background error correlation lengthscale, the observation information will project onto small scales in the analysis (Fowler et al., 2018). The relative position between observations and model grid points is shown in Figure 11. For each observation, our observation operator averages the values at the four nearest model grid points to that observation. This linear observation operator leads to

$$\mathbf{d} = \boldsymbol{\epsilon}^{o} - \mathbf{H}\boldsymbol{\epsilon}^{b}, \tag{26}$$

where $\boldsymbol{\epsilon}^{o} \in \mathbb{R}^{m}$ contains observation errors drawn randomly from a Gaussian distribution with zero mean and covariance given by $\mathbf{R}$ and $\boldsymbol{\epsilon}^{b} \in \mathbb{R}^{n}$ contains background errors drawn randomly from a Gaussian distribution with zero mean and covariance given by $\mathbf{B}$. Gaussian random numbers are produced using the MATLAB routine *mvnrnd.m* (MATLAB MVNRND, 2022). Figure 12 shows a sample of background and observation errors in physical space. We note that in this figure the amplitude of the observation error is larger than that of the background error. This is due to the luck of the draw. When using different random number seeds, we observe the opposite relative amplitudes.

19422466, 2024, 6, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023MS003871 by Test, Wiley Online Library on [13/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

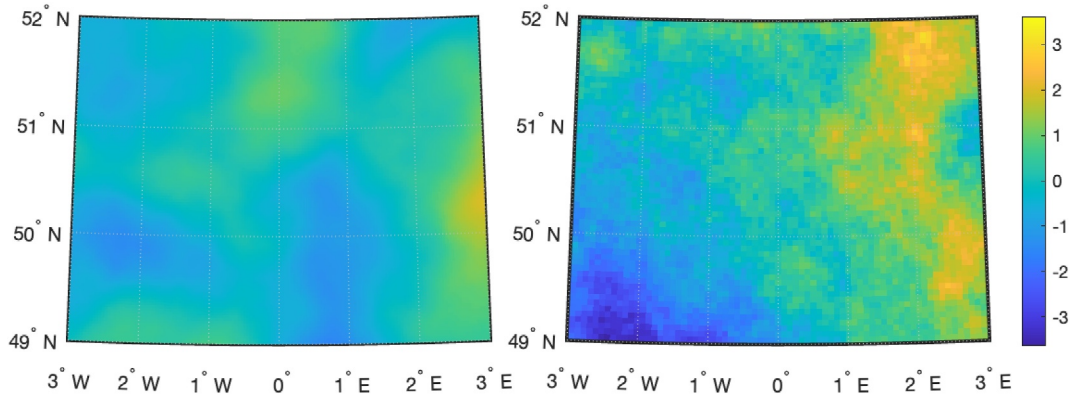

Figure 12. A realization of (left panel) the background error and (right panel) the observation error in physical space. The grid lengths of the background and observations are 3 and 6 km respectively. Background and observation error correlation lengthscales are 80 and 160 km respectively.

After obtaining the matrices **R**, **B**, and **H**, we invert the matrix **R** using the MATLAB routine *inv.m* (MATLAB INV, 2022), calculate the square root of the matrix **B** using the MATLAB routine *sqrtm.m* (MATLAB SQRTM, 2022) and compute the product of the matrices **H** and $\mathbf{B}^{1/2}$. The matrices $\mathbf{R}^{-1}$ and $\mathbf{HB}^{1/2}$ are then used in the CG method described in the first column of Table 1. The second column of Table 1 identifies where the local SVD-FMM is used in the CG method. We choose a relative tolerance of $\tau = 10^{-6}$ for the CG method, which determines when the iteration terminates. We implement the CG method using the MATLAB routine *pcg.m* (MATLAB PCG, 2022). In the control experiment, we use the full matrix **R** and standard matrix-vector multiplication. The standard matrix-vector multiplication is implemented in a naive way (e.g., no function handles are used, and calculations are performed from left to right). To use the local SVD-FMM, we divide the observation domain into 8×8 boxes (see Figure 1). We choose the localization length as $h = 2$ and $h = 3$ and the number of singular vectors as $p = 4$ and $p = 10$. We have written our own MATLAB code to implement the local SVD-FMM, which can be used together with *pcg.m*. The code can also be used to calculate matrix-vector products in other applications, provided the matrix is symmetric and the magnitude of its elements decays away from the diagonal. The MATLAB code for the local SVD-FMM is available in Hu and Dance (2021b).

Table 1
An Algorithm (the Conjugate Gradient Method; e.g., Barrett et al., 1994; Tabeart, 2019; Trefethen & Bau, 1997) for Solving Equation 8 for δz

Algorithm The conjugate gradient (CG) method	Comment Matrix-vector products involving the matrix $\mathbf{R}^{-1}$
1: Initialization $k = 0$, $(\delta\mathbf{z})_0 = \mathbf{0}$	
2: $\mathbf{r}_0 = \mathbf{w} - \mathbf{S}(\delta\mathbf{z})_0 = \mathbf{w}$	2: The calculation of **w** given by Equation 10 requires the calculation of $\mathbf{R}^{-1}\mathbf{d}$
3: $\mathbf{p}_0 = \mathbf{r}_0$	
4: **while** $\frac{\|\mathbf{r}_k\|_2}{\|\mathbf{w}\|_2} > \tau$ **do**	
5: $\alpha_{k+1} = \frac{\|\mathbf{r}_k\|_2^2}{\mathbf{p}_k^T \mathbf{S}\mathbf{p}_k}$	5: The calculation of the product of **S** (Equation 9) and $\mathbf{p}_k$ requires the calculation of $\mathbf{R}^{-1}\left(\mathbf{HB}^{1/2}\mathbf{p}_k\right)$
6: $(\delta\mathbf{z})_{k+1} = (\delta\mathbf{z})_k + \alpha_{k+1}\mathbf{p}_k$	
7: $\mathbf{r}_{k+1} = \mathbf{r}_k - \alpha_{k+1}\mathbf{S}\mathbf{p}_k$	7: The matrix-vector product, $\mathbf{S}\mathbf{p}_k$, has already been calculated in line 5
8: $\beta_{k+1} = \frac{\|\mathbf{r}_{k+1}\|_2^2}{\|\mathbf{r}_k\|_2^2}$	
9: $\mathbf{p}_{k+1} = \beta_{k+1}\mathbf{p}_k + \mathbf{r}_{k+1}$	
10: $k = k + 1$	
11: **end while**	

Note. The initial guess of $\delta\mathbf{z}$ is a zero vector. The symbols $\mathbf{r}_k = \mathbf{w} - \mathbf{S}(\delta\mathbf{z})_k$ and $\mathbf{p}_k$ are the residual and search direction at iteration step k, respectively. The scalars α_{k+1} and β_{k+1} are used to calculate the residual and search direction for iteration step $k + 1$, respectively. The symbol $\|\cdot\|_2$ denotes Euclidean norm and τ is the relative tolerance (compared to the size of $\|\mathbf{w}\|_2$) that decides when the iteration stops.

19422466, 2024, 6, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023MS003871 by Test, Wiley Online Library on [13/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

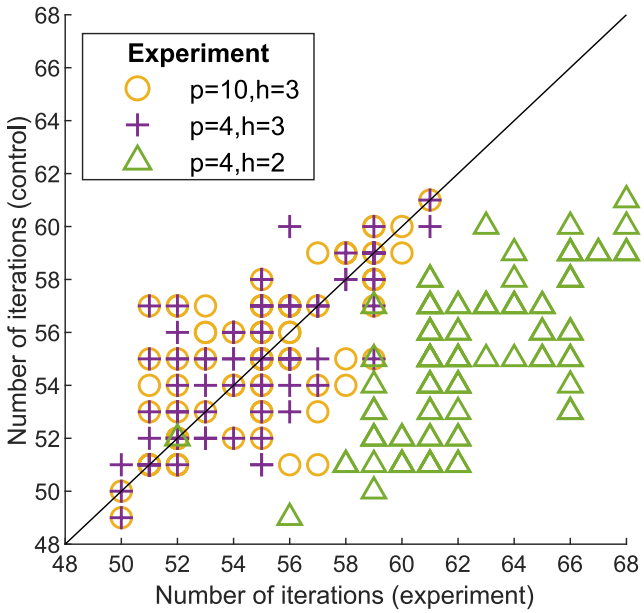

Figure 13. Comparison of the number of iterations required by the conjugate gradient method to minimize the variational data assimilation cost function. The abscissa is the number of iterations in experiments that use the local singular value decomposition approach to the fast multipole method with different numbers of singular vectors (p) and localization lengths (h). The ordinate is the number of iterations in the control run that uses the standard matrix-vector product. Each marker represents the result obtained from one realization of background and observation errors.

In addition to the local SVD-FMM experiments, we perform experiments where the off-diagonal elements of the matrix $\mathbf{R}$ are completely ignored. To account for the effect of neglecting observation error correlations, the diagonal elements are inflated by a factor of 1.7. Observation error variance inflation avoids overfitting the largest scale (e.g., Stewart et al., 2008). The chosen inflation factor has a similar magnitude to inflation factors used in operational systems (e.g., Campbell et al., 2017). We also conduct experiments to combine the local SVD-FMM ($p = 4$ and $h = 2$) with observation error variance inflation.

After obtaining the vector $\delta\mathbf{z}^a$ using the algorithm described in Table 1, we calculate the analysis increment vector as

$$\delta\mathbf{x}^a = \mathbf{B}^{1/2}\delta\mathbf{z}^a. \tag{27}$$

The analysis error vector can then be calculated as

$$\boldsymbol{\epsilon}^a = \boldsymbol{\epsilon}^b + \delta\mathbf{x}^a. \tag{28}$$

To investigate analysis accuracy at different spatial scales, we use the discrete Fourier transform (DFT),

$$\mathbf{F}(k,l) = \frac{1}{n_x}\frac{1}{n_y}\sum_{m=0}^{n_x-1}\sum_{n=0}^{n_y-1}\boldsymbol{\epsilon}_{2d}^a(m,n)e^{-i2\pi\left(k\frac{m}{n_x}+l\frac{n}{n_y}\right)}, \tag{29}$$

where $k = 0, \ldots, n_x - 1$, $l = 0, \ldots, n_y - 1$, i denotes the imaginary unit, $n_x = 112$ and $n_y = 142$ are numbers of grid points in North-South and East-West directions, $\mathbf{F} \in \mathbb{R}^{n_x \times n_y}$ denotes a complex matrix containing Fourier coefficients, and $\boldsymbol{\epsilon}_{2d}^a \in \mathbb{R}^{n_x \times n_y}$ is a matrix containing the analysis error in two-dimensional space. The matrix $\boldsymbol{\epsilon}_{2d}^a$ can be obtained by converting the vector $\boldsymbol{\epsilon}^a$ into a matrix. Equation 29 can also be applied to the background error. We carry out the DFT using the MATLAB routine *fft2.m* (MATLAB FFT2, 2022). The magnitude of Fourier coefficients is

$$\boldsymbol{\gamma}(k,l) = \sqrt{\mathrm{Re}^2(\mathbf{F}(k,l)) + \mathrm{Im}^2(\mathbf{F}(k,l))}, \tag{30}$$

where $\mathrm{Re}(\cdot)$ and $\mathrm{Im}(\cdot)$ denote the real and imaginary parts, respectively. The magnitude $\boldsymbol{\gamma}(k,l)$ measures the size of the error in the scale represented by the wavenumber of k and l. A larger wavenumber in spectral space corresponds to a smaller scale in physical space.

5.2. Experimental Results

We perform the CG method (described in Table 1) 100 times using different samples of random background and observation errors. We first look at whether the use of the local SVD-FMM affects the number of iterations needed by the CG method. We find that how much the number of iterations varies depends on how many singular vectors are used and which localization length is chosen (Figure 13). When using too few singular vectors and a too short localization length, the matrix-vector product will be poorly approximated by the local SVD-FMM (see Section 4), and we would expect the number of iterations to vary considerably (see green triangles in Figure 13). The convergence rate of the CG method is known to be affected by the eigenstructure of the preconditioned Hessian $\mathbf{S}$ (e.g., Tabeart et al., 2022; Trefethen & Bau, 1997, Theorem 38.5). When using the local SVD-FMM, we ignore some elements in the matrix $\mathbf{A}$ and approximate its sub-matrices as low-rank matrices. This changes the eigenstructure of the matrix $\mathbf{A}$ and that of the Hessian $\mathbf{S}$. In our experiments, the matrix $\mathbf{S}$ is positive definite for all the parameters we considered (number of singular vectors, p and localization length, h). This ensures that there is a unique solution to the minimization of the cost function. However, when the approximation of the matrix $\mathbf{A}$ is poor (e.g., when $h = 2$ and $p = 4$), the minimum eigenvalue of $\mathbf{S}$ is found to be less than its theoretical value of one.

19422466, 2024, 6, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023MS003871 by Test, Wiley Online Library on [13/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

19422466, 2024, 6, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023MS003871 by Test, Wiley Online Library on [13/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 2
CPU Time in Seconds Used by MATLAB R2023b on MacBook Air (M1, 2020) to Minimize the Variational Data Assimilation Cost Function

	CPU times (s)		
Matrix-vector multiplication	Minimum	Mean	Maximum
Standard	168.32	179.81	186.05
Diagonal matrix **R**	6.99	7.81	8.64
The local SVD-FMM ($p = 4$, $h = 2$)	0.26 + 11.39	0.55 + 13.97	1.08 + 15.39
The local SVD-FMM ($p = 4$, $h = 3$)	1.18 + 10.37	1.57 + 12.40	2.18 + 13.86
The local SVD-FMM ($p = 10$, $h = 3$)	1.73 + 11.07	2.19 + 12.50	2.82 + 13.85

Note. The minimum, mean, and maximum values are taken from 100 runs with different samples of background and observation errors. For experiments where the local singular value decomposition approach to the fast multipole method (SVD-FMM) is used, the computational time is given as the initialization time plus the iteration time (see the main text for more details).

A smaller minimum eigenvalue can cause a larger condition number, and a larger condition number typically means a slower convergence (e.g., Tabeart et al., 2020). Figure 13 also shows that, for a given choice of the number of singular vectors and localization length, the number of iterations also varies with different innovation vectors (variation of points marked with the same symbol). This is because the innovation vector and the local SVD-FMM are used to calculate the vector **w** (right-hand side of Equation 8), which also affects the convergence rate (Trefethen & Bau, 1997, p. 300).

We now look at the computational time for implementing the CG method (see Table. 2). As expected, using the diagonal matrix **R** gives the shortest computation time and the standard approach the longest. Using the local SVD-FMM can substantially reduce the computational time relative to the standard method. The computational cost of the local SVD-FMM can be separated into two parts. The first part includes the time taken to partition the observation domain (Section 3.1) and to calculate SVDs (Equation 15) and translation operators (Equation 18). We refer to this part as the initialization time of the local SVD-FMM. The second part is the time taken to perform the CG method, which is referred to as the iteration time and is determined by the number of iterations and the computation cost for each iteration. When using a larger number of singular vectors or a longer localization length, the initialization time increases due to the increased time used to calculate the SVD. The longer iteration time for $p = 4$ and $h = 2$ is the result of a slower convergence rate (see Figure 13). In addition, we observe a slight increase in the mean iteration time from $p = 4$ and $h = 3$ to $p = 10$ and $h = 3$. This is because the number of the iterations is comparable in both cases (Figure 13), while the computation cost per iteration increases if p is larger. To conclude, our numerical experiments show that the use of the local SVD-FMM greatly improves the speed of data assimilation, even considering the additional computational cost of the SVDs.

We finally investigate analysis accuracy at different spatial scales. The magnitude of Fourier coefficients ($\boldsymbol{\gamma}$; Equation 30) is used as a measure of analysis and background errors. Figure 14 shows $\boldsymbol{\gamma}$ as a function of the wavenumber given by $\sqrt{k^2 + l^2}$. Since we only have $n_x = 112$ grid points in the North-South direction, we cut off the plot at the wavenumber $n_x/2 - 1 = 55$. The upper panel of Figure 14 compares the mean analysis error and the mean background error in the control run, where the analysis is calculated using the standard matrix-vector product. The analysis and background errors are averaged over 100 realizations of the background and observation error vectors. We find that the mean analysis error is smaller than the mean background error at all scales, with the largest errors at large scales. This is consistent with the findings of Fowler et al. (2018). The lower panel of Figure 14 shows the absolute differences between the mean analysis error in the control run and the mean analysis errors in the experiments. The smallest difference is found for the local SVD-FMM experiment with $p = 10$ and $h = 3$. In particular, we find that the differences at the smaller scales are on the order of 10^{-8}. This indicates that the use of the local SVD-FMM can preserve the benefit of including correlated observation error statistics in data assimilation (see the beginning of Section 5). In comparison, a larger difference is found for the experiments with diagonal **R** at all scales. These results comparing analysis accuracy with diagonal and non-diagonal observation error variance matrices are consistent with the findings of previous studies (e.g., Rainwater et al., 2015; Simonin et al., 2019; Stewart et al., 2008, 2013). When smaller h and p are used for the local

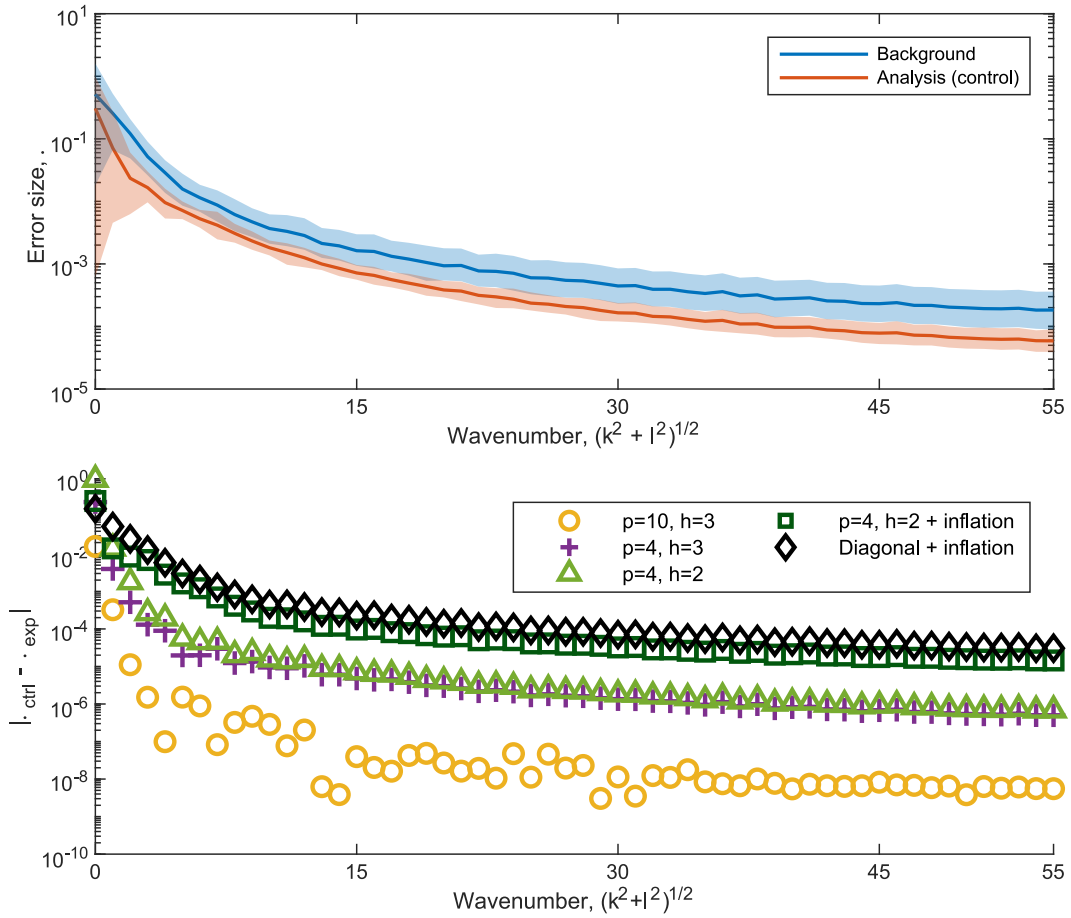

Figure 14. (upper panel) Background and analysis errors at different physical scales (represented by wavenumbers in spectral space) in the control run. Solid lines are the mean error size from 100 runs using different samples of background and observation errors. Shaded areas are bounded by the minimum and maximum errors from these runs. (lower panel) Absolute differences in the mean analysis error at different physical scales between the control run and experiments. The control run and experiments are described in Section 5.1. A logarithmic scale is used on the axis of ordinates to allow a clear comparison between results at all wavenumbers.

SVD-FMM, we observe an increase in the difference compared to the case of $p = 10$ and $h = 3$. At the largest scale, the differences for the local SVD-FMM experiments with $p = 4$ and $h = 2$ or 3 are even larger than that for the diagonal **R** experiment. It is worth noting that increasing h from 2 to 3 brings little benefits when $p = 4$. This is consistent with the results in Section 4.3, where we showed that for a fixed p, increasing h may have little effect on the accuracy of the local SVD-FMM once h reaches a certain value. In this case, it is reasonable to retain more singular values/vectors due to the increased matrix size. The lower panel of Figure 14 also shows that after inflating the observation error variance, the difference for the local SVD-FMM experiment with $p = 4$ and $h = 2$ at the largest scale is reduced. However, at the same time, the difference at the smaller scales increases.

In addition to the comparison of mean analysis and background errors, Figure 15 compares the analysis and background errors at the largest scale in each realization. We find that in some realizations the analysis error is larger, while in other realizations the background error is larger. When using the standard matrix-vector product, or when using the local SVD-FMM with a large number of singular vectors and a long localization length, the markers are mostly on the upper left of the diagonal in the scatterplot (the upper panels of Figure 15). This means that the analysis error is usually smaller than the background error at the largest scale. However, when using fewer singular vectors or a shorter localization length, more markers are found on the lower right of the diagonal (the middle panels of Figure 15), meaning that the analysis is usually less accurate than the background at the largest scale. In this case, we find that inflating the observation error variance improves the analysis accuracy at the

19422466, 2024, 6, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023MS003871 by Test, Wiley Online Library on [13/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

19422466, 2024, 6, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023MS003871 by Test, Wiley Online Library on [13/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

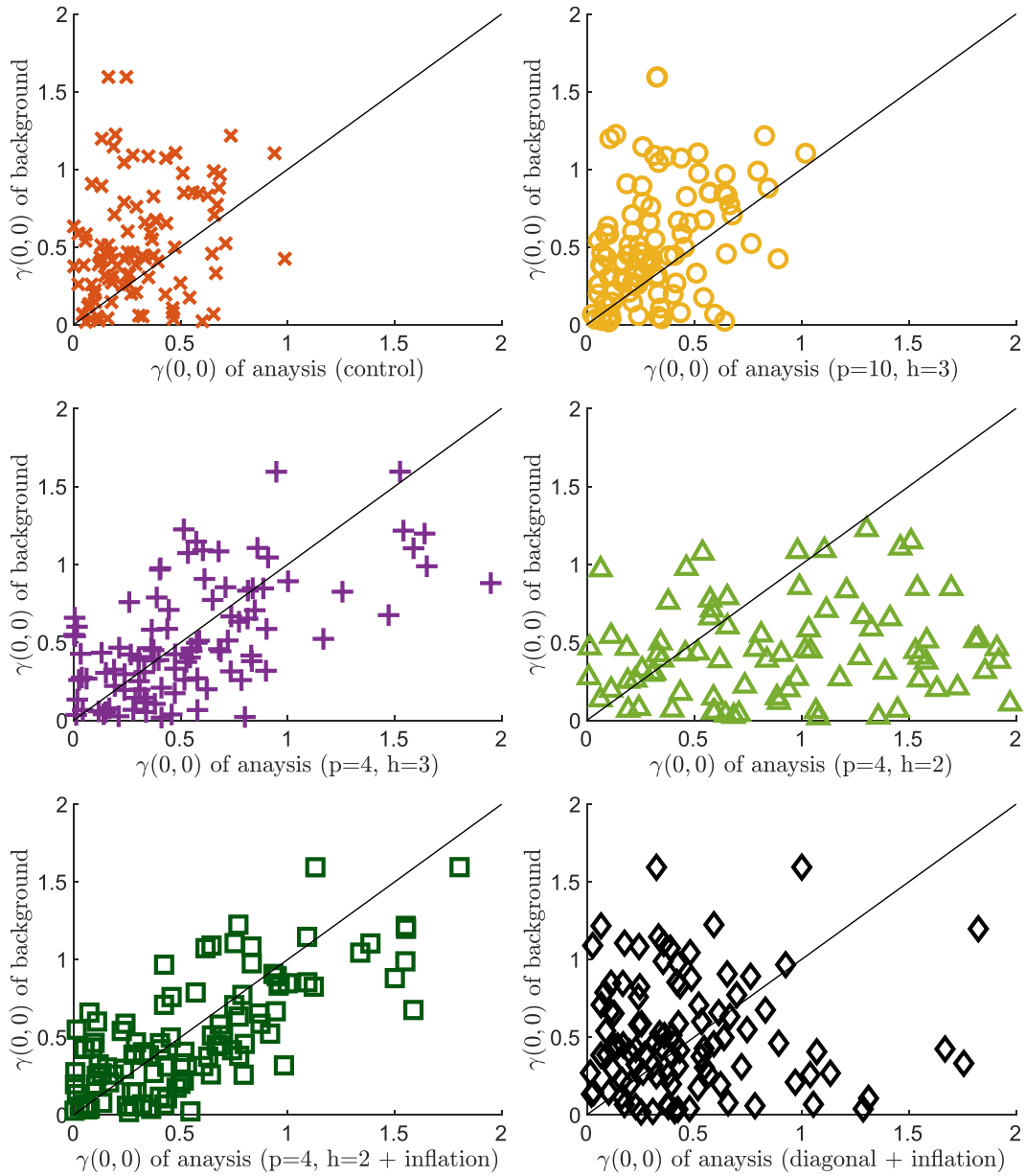

Figure 15. Background errors against analysis errors at the largest scale. Analyses are calculated differently for each panel (see Section 5.1 for details). The markers and colors used for each analysis are consistent with Figure 14. Each marker in a panel represents the result obtained from one realization of background and observation errors.

largest scale (the bottom-left panel of Figure 15). This is similar to what happens when the observation error variance is inflated using a diagonal matrix, to compensate for ignoring observation error correlations (not shown).

In summary, as long as the number of singular vectors and the localization length are chosen so that the local SVD-FMM is a good approximation of the full matrix-vector product, it preserves the information in the analysis at all scales.

6. Conclusion

Accounting for spatially correlated observation error statistics in data assimilation systems has been shown to improve both the accuracy of data assimilation and forecasts (e.g., Fujita et al., 2020; Healy & White, 2005;

19422466, 2024, 6, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023MS003871 by Test, Wiley Online Library on [13/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Simonin et al., 2019; Stewart et al., 2008, 2013). The inclusion of correlated error statistics allows for the assimilation of dense observational data sets, which can provide information on appropriate scales for high-resolution forecasting. Such data sets are currently often thinned to get rid of the spatial error correlations. A potential downside of including correlated error statistics is that it may increase the computational cost of the variational minimization problem. This is because including these error statistics makes the observation precision matrix a full matrix. Computing a matrix-vector product involving a full matrix requires more floating-point operations and higher parallel computing communication costs than computing a matrix-vector product involving a diagonal matrix.

To reduce the computational cost, we proposed a new numerical approximation method, called the local SVD-FMM, which is easy to implement in both sequential and parallel computations and is suitable for matrices whose elements decay rapidly away from the main diagonal. The local SVD-FMM was developed based on the SVD-FMM (without localization), which was originally proposed by Gimbutas and Rokhlin (2003) and explored in Hu and Dance (2021a) for its feasibility in data assimilation. There are two new aspects in the local SVD-FMM. First, it is a single-level algorithm that requires a single-level partitioning of the observation domain. In contrast, the SVD-FMM is a multi-level algorithm that uses a hierarchical structure of boxes, which leads to a more complicated parallelization scheme. Second, a new non-dimensional localization length parameter is introduced in the local SVD-FMM. The use of localization compensates for the increased cost by replacing the multi-level algorithm with a single-level algorithm. Although the localization introduces errors by ignoring error correlations between distant observations, we showed that these errors can be very small and do not affect the accuracy of the analysis too much. The main drawback of the local SVD-FMM for data assimilation is the need to precompute the observation precision matrix before the method can be applied. In practice we believe that there are three main ways we could deal with this issue, which may each be appropriate for different observation types and situations, but would require further research to evaluate: (a) Starting from a statistical estimate of the precision matrix rather than the observation error covariance matrix. For example, Le and Zhong (2022) and Yuan (2010) each provide numerically efficient approaches to estimate a precision matrix starting from the sample covariance (that could be obtained in data assimilation from a Desroziers et al. (2005)-type method). (b) Rather than inverting a new matrix each cycle (in response to quality control changes in observation distribution) compute low-rank downdates and updates, with only a small number of operations relative to the dimension of the matrix (e.g., Brand, 2006; Golub & Van Loan, 1996). (c) Combine this type of hierarchical approximation approach with a matrix factorization such as the Cholesky decomposition (Hackbusch, 2016).

The accuracy of the local SVD-FMM depends on many factors, including the localization length, the number of singular vectors used, and the observation error correlation lengthscale. Generally, the accuracy increases with the localization length and the number of singular vectors but decreases with the correlation lengthscale. Moreover, we find that after the localization length and the number of singular vectors reach certain values, further increasing them brings little benefit. The optimal value for them varies from application to application and depends on, for example, the configuration of boxes and the observation error correlation lengthscale. In a simple variational data assimilation experiment, the local SVD-FMM is found to dramatically reduce the computational cost of the variational minimization problem while maintaining the accuracy of the analysis. In particular, the use of the local SVD-FMM can preserve the improvement in the analysis accuracy at all scales.

The local SVD-FMM has a wide range of applicability. It is applicable to both regularly and irregularly distributed observations and the observation domain can be partitioned into a regular or irregular mesh of boxes. We may only need to have each box contain a similar number of observations in order to achieve good efficiency and load balancing. Given the efficiency, accuracy and broad applicability, the local SVD-FMM has the potential to be used in practical data assimilation applications, such as convection-permitting NWP (Dance et al., 2019; Hu et al., 2023), where it may be important to assimilate a large volume of dense observations with strong mutual spatial error correlations.

Data Availability Statement

The numerical experiments in the study were carried out using MATLAB. The MATLAB code written by authors was published in Hu and Dance (2021b).

Acknowledgments
We thank David Simonin and Joanne A. Waller for their useful discussions. We also appreciate the valuable suggestions and constructive comments the reviewers and the associate editor provided, which have helped improve this work. G. Hu was funded in part by the UK Engineering and Physical Sciences Research Council (EPSRC) DARE project (EP/P002331/1), Met Office and the National Centre for Earth Observation (NCEO). S. L. Dance was funded in part by the UK EPSRC DARE project (EP/P002331/1) and the NCEO.

References

Anderson, J. L. (2003). A local least squares framework for ensemble filtering. *Monthly Weather Review*, *131*(4), 634–642. https://doi.org/10.1175/1520-0493(2003)131⟨0634:ALLSFF⟩2.0.CO;2

Ballard, S. P., Li, Z., Simonin, D., & Caron, J.-F. (2016). Performance of 4D-Var NWP-based nowcasting of precipitation at the Met Office for summer 2012. *Quarterly Journal of the Royal Meteorological Society*, *142*(694), 472–487. https://doi.org/10.1002/qj.2665

Barrett, R., Berry, M., Chan, T. F., Demmel, J., Donato, J., Dongarra, J., et al. (1994). Templates for the solution of linear systems: Building blocks for iterative methods. *Society for Industrial and Applied Mathematics*. https://doi.org/10.1137/1.9781611971538

Bédard, J., & Buehner, M. (2020). A practical assimilation approach to extract smaller-scale information from observations with spatially correlated errors: An idealized study. *Quarterly Journal of the Royal Meteorological Society*, *146*(726), 468–482. https://doi.org/10.1002/qj.3687

Bell, Z., Dance, S. L., & Waller, J. A. (2020). Accounting for observation uncertainty and bias due to unresolved scales with the Schmidt-Kalman filter. *Tellus A: Dynamic Meteorology and Oceanography*, *72*(1), 1–21. https://doi.org/10.1080/16000870.2020.1831830

Bernstein, D. S. (2009). *Matrix mathematics: Theory, facts, and formulas* (2nd ed.). Princeton University Press. https://doi.org/10.1515/9781400833344

Bollmeyer, C., Keller, J. D., Ohlwein, C., Wahl, S., Crewell, S., Friederichs, P., et al. (2015). Towards a high-resolution regional reanalysis for the European CORDEX domain. *Quarterly Journal of the Royal Meteorological Society*, *141*(686), 1–15. https://doi.org/10.1002/qj.2486

Bormann, N., Saarinen, S., Kelly, G., & Thépaut, J.-N. (2003). The spatial structure of observation errors in atmospheric motion vectors from geostationary satellite data. *Monthly Weather Review*, *131*(4), 706–718. https://doi.org/10.1175/1520-0493(2003)131⟨0706:TSSOOE⟩2.0.CO;2

Brand, M. (2006). Fast low-rank modifications of the thin singular value decomposition. *Linear Algebra and its Applications*, *415*(1), 20–30. https://doi.org/10.1016/j.laa.2005.07.021

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

Chabot, V., Nodet, M., Papadakis, N., & Vidard, A. (2015). Accounting for observation errors in image data assimilation. *Tellus A: Dynamic Meteorology and Oceanography*, *67*(1), 23629. https://doi.org/10.3402/tellusa.v67.23629

Chabot, V., Nodet, M., & Vidard, A. (2020). Multiscale representation of observation error statistics in data assimilation. *Sensors*, *20*(5), 1–20. https://doi.org/10.3390/s20051460

Clark, P., Roberts, N., Lean, H., Ballard, S. P., & Charlton-Perez, C. (2016). Convection-permitting models: A step-change in rainfall forecasting. *Meteorological Applications*, *23*(2), 165–181. https://doi.org/10.1002/met.1538

Cordoba, M., Dance, S. L., Kelly, G. A., Nichols, N. K., & Waller, J. A. (2017). Diagnosing atmospheric motion vector observation errors for an operational high-resolution data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *143*(702), 333–341. https://doi.org/10.1002/qj.2925

Courtier, P., Andersson, E., Heckley, W., Vasiljevic, D., Hamrud, M., Hollingsworth, A., et al. (1998). The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation. *Quarterly Journal of the Royal Meteorological Society*, *124*(550), 1783–1807. https://doi.org/10.1002/qj.49712455002

Courtier, P., Thépaut, J.-N., & Hollingsworth, A. (1994). A strategy for operational implementation of 4D-Var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society*, *120*(519), 1367–1387. https://doi.org/10.1002/qj.49712051912

Daley, R. (1994). *Atmospheric data analysis*. Cambridge University Press.

Dance, S. L., Ballard, S. P., Bannister, R. N., Clark, P., Cloke, H. L., Darlington, T., et al. (2019). Improvements in forecasting intense rainfall: Results from the FRANC (forecasting rainfall exploiting new data assimilation techniques and novel observations of convection) project. *Atmosphere*, *10*(3), 125. https://doi.org/10.3390/atmos10030125

Deng, Y. (2012). *Applied parallel computing*. World Scientific Publishing Co., Inc.

Desroziers, G., Berre, L., Chapnik, B., & Poli, P. (2005). Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society*, *131*(613), 3385–3396. https://doi.org/10.1256/qj.05.108

Eijkhout, V., & Polman, B. (1988). Decay rates of inverses of banded M-matrices that are near to Toeplitz matrices. *Linear Algebra and its Applications*, *109*, 247–277. https://doi.org/10.1016/0024-3795(88)90211-X

Fisher, M. (2005). Accounting for correlated observation error in the ECMWF analysis. In *ECMWF technical memoranda, ECMWF, reading UK, MF/05106*.

Fowler, A. M. (2019). Data compression in the presence of observational error correlations. *Tellus A: Dynamic Meteorology and Oceanography*, *71*(1), 1634937. https://doi.org/10.1080/16000870.2019.1634937

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

Fujita, T., Seko, H., Kawabata, T., Ikuta, Y., Sawada, K., Hotta, D., & Kunii, M. (2020). Variational data assimilation with spatial and temporal observation error correlations of Doppler radar radial winds. WMO. (Section 1: Assimilation of atmospheric and land observations). Data impact and sensitivity studies. Methodological advances. Research activities in Earth system modelling. In *Working group on numerical experimentation. Report No. 50. WCRP report no. 12/2020*. WMO.

Gimbutas, Z., & Rokhlin, V. (2003). A generalized fast multipole method for nonoscillatory Kernels. *SIAM Journal on Scientific Computing*, *24*(3), 796–817. https://doi.org/10.1137/S1064827500381148

Golub, G. H., & Van Loan, C. F. (1996). *Matrix computations*. Johns Hopkins University Press.

Grama, A., Karypis, G., Kumar, V., & Gupta, A. (2003). *Introduction to parallel computing*. Addison-Wesley.

Greybush, S. J., Kalnay, E., Miyoshi, T., Ide, K., & Hunt, B. R. (2011). Balance and ensemble Kalman filter localization techniques. *Monthly Weather Review*, *139*(2), 511–522. https://doi.org/10.1175/2010MWR3328.1

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

19422466, 2024, 6, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023MS003871 by Test, Wiley Online Library on [13/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Haben, S. A. (2011). *Conditioning and preconditioning of the minimisation problem in variational data assimilation (Unpublished doctoral dissertation)*. University of Reading.
Haben, S. A., Lawless, A., & Nichols, N. (2011). Conditioning of incremental variational data assimilation, with application to the Met Office system. *Tellus A: Dynamic Meteorology and Oceanography*, *63*(4), 782–792. https://doi.org/10.1111/j.1600-0870.2011.00527.x
Hackbusch, W. (2016). Survey on the technique of hierarchical matrices. *Vietnam Journal of Mathematics*, *44*(1), 71–101. https://doi.org/10.1007/s10013-015-0168-5
Hamill, T. M., Whitaker, J. S., & Snyder, C. (2001). Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Monthly Weather Review*, *129*(11), 2776–2790. https://doi.org/10.1175/1520-0493(2001)129⟨2776:DDFOBE⟩2.0.CO;2
Healy, S. B., & White, A. A. (2005). Use of discrete Fourier transforms in the 1D-Var retrieval problem. *Quarterly Journal of the Royal Meteorological Society*, *131*(605), 63–72. https://doi.org/10.1256/qj.03.193
Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803
Hodyss, D., & Satterfield, E. (2017). The treatment, estimation, and issues with representation error modelling. In S. K. Park & L. Xu (Eds.), *Data assimilation for atmospheric, oceanic and hydrologic applications* (Vol. iii, pp. 177–194). Springer International Publishing. https://doi.org/10.1007/978-3-319-43415-5_8
Houtekamer, P. L., & Mitchell, H. L. (1998). Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review*, *126*(3), 796–811. https://doi.org/10.1175/1520-0493(1998)126⟨0796:DAUAEK⟩2.0.CO;2
Hu, G., Bódai, T., & Lucarini, V. (2019). Effects of stochastic parametrization on extreme value statistics. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *29*(8), 083102. https://doi.org/10.1063/1.5095756
Hu, G., & Dance, S. L. (2021a). Efficient computation of matrix-vector products with full observation weighting matrices in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *147*(741), 4101–4121. https://doi.org/10.1002/qj.4170
Hu, G., & Dance, S. L. (2021b). MATLAB code for the localized singular value decomposition approach of the fast multipole method (the local SVD-FMM) [Software]. *University of Reading*. https://doi.org/10.17864/1947.329
Hu, G., & Dance, S. L. (2024). Sampling and misspecification errors in the estimation of observation-error covariance matrices using observation-minus-background and observation-minus-analysis statistics. *Quarterly Journal of the Royal Meteorological Society*. https://doi.org/10.1002/qj.4750
Hu, G., Dance, S. L., Bannister, R. N., Chipilski, H. G., Guillet, O., Macpherson, B., et al. (2023). Progress, challenges, and future steps in data assimilation for convection-permitting numerical weather prediction: Report on the virtual meeting held on 10 and 12 November 2021. *Atmospheric Science Letters*, *24*(1), e1130. https://doi.org/10.1002/asl.1130
Hu, G., & Franzke, C. L. E. (2020). Evaluation of daily precipitation extremes in reanalysis and gridded observation-based data sets over Germany. *Geophysical Research Letters*, *47*(18), e2020GL089624. https://doi.org/10.1029/2020GL089624
Janjić, T., Bormann, N., Bocquet, M., Carton, J. A., Cohn, S. E., Dance, S. L., et al. (2018). On the representation error in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *144*(713), 1257–1278. https://doi.org/10.1002/qj.3130
Janjić, T., Nerger, L., Albertella, A., Schröter, J., & Skachko, S. (2011). On domain localization in ensemble-based kalman filter algorithms. *Monthly Weather Review*, *139*(7), 2046–2060. https://doi.org/10.1175/2011MWR3552.1
Le, T.-M., & Zhong, P.-S. (2022). High-dimensional precision matrix estimation with a known graphical structure. *Stat*, *11*(1), e424. https://doi.org/10.1002/sta4.424
Lorenc, A. C. (2003). The potential of the ensemble kalman filter for NWP—A comparison with 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, *129*(595), 3183–3203. https://doi.org/10.1256/qj.02.132
Lorenc, A. C., Ballard, S. P., Bell, R. S., Ingleby, N. B., Andrews, P. L. F., Barker, D. M., et al. (2000). The Met Office global three-dimensional variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, *126*(570), 2991–3012. https://doi.org/10.1002/qj.49712657002
MATLAB routine fft2.m. (2022). The MathWorks Inc., Natick, MA, USA [Software]. Retrieved from https://www.mathworks.com
MATLAB routine inv.m. (2022). The MathWorks Inc., Natick, MA, USA [Software]. Retrieved from https://www.mathworks.com
MATLAB routine mvnrnd.m. (2022). The MathWorks Inc., Natick, MA, USA [Software]. Retrieved from https://www.mathworks.com
MATLAB routine pcg.m. (2022). The MathWorks Inc., Natick, MA, USA [Software]. Retrieved from https://www.mathworks.com
MATLAB routine sqrtm.m. (2022). The MathWorks Inc., Natick, MA, USA [Software]. Retrieved from https://www.mathworks.com
Michel, Y. (2018). Revisiting Fisher's approach to the handling of horizontal spatial correlations of observation errors in a variational framework. *Quarterly Journal of the Royal Meteorological Society*, *144*(716), 2011–2025. https://doi.org/10.1002/qj.3249
Morrison, R., Baptista, R., & Basor, E. (2022). Diagonal nonlinear transformations preserve structure in covariance and precision matrices. *Journal of Multivariate Analysis*, *190*, 104983. https://doi.org/10.1016/j.jmva.2022.104983
Nichols, N. K. (2010). Mathematical concepts of data assimilation. In W. Lahoz, B. Khattatov, & R. Menard (Eds.), *Data assimilation: Making sense of observations* (pp. 13–39). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-74703-1_2
Nino-Ruiz, E. D., Sandu, A., & Deng, X. (2019). A parallel implementation of the ensemble Kalman filter based on modified Cholesky decomposition. *Journal of Computational Science*, *36*, 100654. https://doi.org/10.1016/j.jocs.2017.04.005
Oke, P. R., Sakov, P., & Corney, S. P. (2007). Impacts of localisation in the EnKF and EnOI: Experiments with a small model. *Ocean Dynamics*, *57*(1), 32–45. https://doi.org/10.1007/s10236-006-0088-8
Periáñez, Á., Reich, H., & Potthast, R. (2014). Optimal localization for ensemble Kalman filter systems. *Journal of the Meteorological Society of Japanese Series II*, *92*(6), 585–597. https://doi.org/10.2151/jmsj.2014-605
Petrie, R. E., & Dance, S. L. (2010). Ensemble-based data assimilation and the localisation problem. *Weather*, *65*(3), 65–69. https://doi.org/10.1002/wea.505
Rainwater, S., Bishop, C. H., & Campbell, W. F. (2015). The benefits of correlated observation errors for small scales. *Quarterly Journal of the Royal Meteorological Society*, *141*(693), 3439–3445. https://doi.org/10.1002/qj.2582
Simonin, D., Waller, J. A., Ballard, S. P., Dance, S. L., & Nichols, N. K. (2019). A pragmatic strategy for implementing spatially correlated observation errors in an operational system: An application to Doppler radial winds. *Quarterly Journal of the Royal Meteorological Society*, *145*(723), 2772–2790. https://doi.org/10.1002/qj.3592
Stewart, L. M. (2010). *Correlated observation errors in data assimilation (Unpublished doctoral dissertation)*. University of Reading.
Stewart, L. M., Dance, S. L., & Nichols, N. K. (2008). Correlated observation errors in data assimilation. *International Journal for Numerical Methods in Fluids*, *56*(8), 1521–1527. https://doi.org/10.1002/fld.1636
Stewart, L. M., Dance, S. L., & Nichols, N. K. (2013). Data assimilation with correlated observation errors: Experiments with a 1-D shallow water model. *Tellus A: Dynamic Meteorology and Oceanography*, *65*(1), 19546. https://doi.org/10.3402/tellusa.v65i0.19546

19422466, 2024, 6, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023MS003871 by Test, Wiley Online Library on [13/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Tabeart, J. M. (2019). *On the treatment of correlated observation errors in data assimilation (Doctoral dissertation, School of Mathematical, Physical and Computational Sciences)*. University of Reading. https://doi.org/10.48683/1926.00088830

Tabeart, J. M., Dance, S. L., Haben, S. A., Lawless, A. S., Nichols, N. K., & Waller, J. A. (2018). The conditioning of least-squares problems in variational data assimilation. *Numerical Linear Algebra with Applications*, *25*(5), e2165. https://doi.org/10.1002/nla.2165

Tabeart, J. M., Dance, S. L., Lawless, A. S., Migliorini, S., Nichols, N. K., Smith, F., & Waller, J. A. (2020). The impact of using reconditioned correlated observation-error covariance matrices in the Met Office 1D-Var system. *Quarterly Journal of the Royal Meteorological Society*, *146*(728), 1372–1390. https://doi.org/10.1002/qj.3741

Tabeart, J. M., Dance, S. L., Lawless, A. S., Nichols, N. K., & Waller, J. A. (2022). New bounds on the condition number of the Hessian of the preconditioned variational data assimilation problem. *Numerical Linear Algebra with Applications*, *29*(1), e2405. https://doi.org/10.1002/nla.2405

Trefethen, L. N., & Bau, D. (1997). *Numerical linear algebra*. Society for Industrial and Applied Mathematics.

Waller, J. A., Ballard, S. P., Dance, S. L., Kelly, G., Nichols, N. K., & Simonin, D. (2016). Diagnosing horizontal and inter-channel observation error correlations for SEVIRI observations using observation-minus-background and observation-minus-analysis statistics. *Remote Sensing*, *8*(7), 581. https://doi.org/10.3390/rs8070581

Waller, J. A., Bauernschubert, E., Dance, S. L., Nichols, N. K., Potthast, R., & Simonin, D. (2019). Observation error statistics for Doppler radar radial wind superobservations assimilated into the DWD COSMO-KENDA system. *Monthly Weather Review*, *147*(9), 3351–3364. https://doi.org/10.1175/MWR-D-19-0104.1

Waller, J. A., Dance, S. L., & Lean, H. W. (2021). Evaluating errors due to unresolved scales in convection-permitting numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, *147*(738), 2657–2669. https://doi.org/10.1002/qj.4043

Waller, J. A., Simonin, D., Dance, S. L., Nichols, N. K., & Ballard, S. P. (2016). Diagnosing observation error correlations for Doppler radar radial winds in the Met Office UKV model using observation-minus-background and observation-minus-analysis statistics. *Monthly Weather Review*, *144*(10), 3533–3551. https://doi.org/10.1175/MWR-D-15-0340.1

Weston, P. P., Bell, W., & Eyre, J. R. (2014). Accounting for correlated error in the assimilation of high-resolution sounder data. *Quarterly Journal of the Royal Meteorological Society*, *140*(685), 2420–2429. https://doi.org/10.1002/qj.2306

Yeh, H.-L., Yang, S.-C., Terasaki, K., Miyoshi, T., & Liou, Y.-C. (2022). Including observation error correlation for ensemble radar radial wind assimilation and its impact on heavy rainfall prediction. *Quarterly Journal of the Royal Meteorological Society*, *148*(746), 2254–2281. https://doi.org/10.1002/qj.4302

Ying, Y. (2020). Assimilating observations with spatially correlated errors using a serial ensemble filter with a multiscale approach. *Monthly Weather Review*, *148*(8), 3397–3412. https://doi.org/10.1175/MWR-D-19-0387.1

Yuan, M. (2010). High dimensional inverse covariance matrix estimation via linear programming. *Journal of Machine Learning Research*, *11*, 2261–2286.

Zeng, Y., Janjic, T., Feng, Y., Blahak, U., de Lozar, A., Bauernschubert, E., et al. (2021). Interpreting estimated observation error statistics of weather radar measurements using the ICON-LAM-KENDA system. In *Atmospheric measurement techniques discussions, 2021* (pp. 1–28). https://doi.org/10.5194/amt-2021-95

19422466, 2024, 6, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023MS003871 by Test, Wiley Online Library on [13/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License