Received: 24 January 2018 | Revised: 30 April 2018 | Accepted: 22 May 2018

DOI: 10.1002/qj.3336

NOTES AND CORRESPONDENCE

Quarterly Journal of the Royal Meteorological Society RMetS

On the approximation of the inverse error covariances of high-resolution satellite altimetry data

Max Yaremchuk[1] | Joseph M. D'Addezio[2] | Gleb Panteleev[1] | Gregg Jacobs[1]

[1]Naval Research Laboratory, Stennis Space Center, Mississippi, USA
[2]Division of Marine Science, University of Southern Mississippi, Stennis Space Center, Mississippi, USA

Correspondence
Max Yaremchuk, Naval Research Laboratory, 1009 Balch Blvd., Stennis Space Center, 39522 MS, USA
Email: max.yaremchuk@nrlssc.navy.mil

Funding information
Office of Naval Reseach program elements 0646352N, 0602435N, and Naval Research Laboratory cooperative agreement BAA-N00173-03-13-01

High-resolution (swath) altimeter missions scheduled to monitor the ocean surface in the near future have observation-error covariances (OECs) with slowly decaying off-diagonal elements. This property presents a challenge for the majority of the data assimilation algorithms which were designed under the assumption of the diagonal OECs being easily inverted. In this note, we present a method of approximating the inverse of a dense OEC by a sparse matrix represented by the polynomial of spatially inhomogeneous differential operators, whose coefficients are optimized to fit the target OEC by minimizing a quadratic cost function. Explicit expressions for the cost function gradient and the Hessian are derived. The method is tested with an OEC model generated by the SWOT simulator.

KEYWORDS
covariance modelling, data assimilation, observational data analysis, wide-swath altimetry

1 | INTRODUCTION

Over the last several decades, representation of the background-error covariances (BECs) by the polynomials of the diffusion operator has been extensively studied in both meteorological and oceanographic data assimi9lation (DA) applications (e.g. Derber and Rosati, 1989; Weaver *et al.*, 2003; Xu, 2005; Yaremchuk and Smith, 2011; Yaremchuk *et al.*, 2013). Among the advantages are the computational efficiency of the approach and its ability to preserve the positive semi-definite (psd) property of the resulting BEC estimates. The method proves to be especially useful in heuristic modelling of the climatological (static) BECs which supplement the ensemble-based estimates of the background errors.

In contrast to the BECs, observational-error covariances (OECs) are conventionally represented by diagonal matrices under the implicit assumption that observation errors are weakly correlated at spatial scales exceeding the grid step of the numerical models. This assumption, being realistic for most current observational platforms, provides an additional convenience of inexpensive computation of the inverse OECs and of their square roots currently employed by the majority of DA algorithms (e.g. Cummings, 2005; Hunt *et al.*, 2007; Fairbairn *et al.*, 2014).

Recent developments in high-resolution ("swath") altimetry (Durand *et al.*, 2010; Ito *et al.*, 2014; Ubelmann *et al.*, 2015; Gaultier *et al.*, 2016) introduce challenges for DA. Novel issues emerge due to both the data density that may approach model grid scales in the horizontal, and because observational errors at such high resolutions appear to be highly correlated in space Ruggiero *et al.* (2016) (hereinafter R16) due to the design of the satellite and sensor. At the same time, recent studies (e.g. Stewart *et al.*, 2013; Miyoshi *et al.*, 2013; Waller *et al.*, 2014) demonstrate substantial benefits of accounting for spatial correlations of the observation errors even in low-dimensional DA systems. As a consequence, these newly arriving data require special treatment in order to maintain skill and retain the computational efficiency of the DA schemes.

In most of the DA algorithms, the difference between the model sea surface height (SSH) and the data has to be multiplied by either the inverse of the observation-error covariance $\mathbf{R}$ or by its inverse square root $\mathbf{R}^{-1/2}$, so a computationally efficient representation $\mathbf{R}_i$ of these matrices (e.g. by a

sparse matrix) is highly desirable. This note contributes to the methodology of approximating $\mathbf{R}^{-1}$ (or its square root) using differential operators. Specifically, with the forthcoming Surface Water and Ocean Topography (SWOT) altimeter mission in mind, we postulate that an estimate of $\mathbf{R}$ is available on a regular two-dimensional set of $N = n_x \times n_y$ observation grid points from the simulating software of Gaultier *et al.* (2017).

2 | APPROXIMATING THE INVERSE COVARIANCE

2.1 | The matrix parametrization model

In what follows, we present a methodology of parametrizing $\mathbf{R}_i$ by a linear combination of sparse matrices with matrix-valued coefficients represented by discretized differential operators. To simplify the notation, the method is illustrated by a particular example relevant to the SWOT application.

Specifically, we consider $\mathbf{R}_i$ of the form

$$\mathbf{R}_i = \mathbf{A} + \nabla^{\mathrm{T}}\mathbf{B}\nabla + \Delta\mathbf{C}\Delta, \tag{1}$$

where ∇ is the $2N - n_x - n_y \times N$ matrix, representing the first-order approximation of the gradient operator on a 2D grid, $\Delta = \nabla^{\mathrm{T}}\nabla$ is the Laplacian, and $\mathbf{A}, \mathbf{B}, \mathbf{C}$ are sparse control matrices. Their M non-zero elements populating the vector $\mathbf{x} \in \mathbb{R}^M$ have to be optimized by minimization of the quadratic cost function, measuring the Frobenius norm $||\cdot||_{\mathrm{F}}$ of the respective residual:

$$J = ||\mathbf{R}_i\mathbf{R} - \mathbf{I}||_{\mathrm{F}}^2 = \mathrm{tr}\left[\mathbf{P}\mathbf{P}^{\mathrm{T}}\right] \longrightarrow \min_{\mathbf{x}}. \tag{2}$$

Here $\mathbf{I}$ is the identity matrix and notation $\mathbf{P} = \mathbf{R}_i\mathbf{R} - \mathbf{I}$ is introduced. The first variation of Equation 2 has the form

$$\delta J = 2\,\mathrm{tr}\left[\delta\mathbf{P}\mathbf{P}^{\mathrm{T}}\right] = 2\,\mathrm{tr}\left[\delta\mathbf{R}_i\,\mathbf{R}\mathbf{P}^{\mathrm{T}}\right]. \tag{3}$$

Taking into account that $\delta\mathbf{R}_i = \delta\mathbf{A} + \nabla^{\mathrm{T}}\delta\mathbf{B}\nabla + \Delta\delta\mathbf{C}\Delta$, and introducing the notation $\mathbf{Q} = 2\mathbf{P}\mathbf{R}$, transforms δJ to

$$\delta J = \mathrm{tr}\left[\delta\mathbf{A}\mathbf{Q}^{\mathrm{T}} + \delta\mathbf{C}\Delta\mathbf{Q}^{\mathrm{T}}\Delta\right] + \mathrm{tr}\left[\delta\mathbf{B}\nabla\mathbf{Q}^{\mathrm{T}}\nabla^{\mathrm{T}}\right], \tag{4}$$

so that the respective expressions for the gradient are:

$$\frac{\delta J}{\delta\mathbf{A}} = \mathbf{Q}, \qquad \frac{\delta J}{\delta\mathbf{B}} = \nabla\mathbf{Q}\nabla^{\mathrm{T}}, \qquad \frac{\delta J}{\delta\mathbf{C}} = \Delta\mathbf{Q}\Delta, \tag{5}$$

and the system of equations $\delta J/\delta\mathbf{x} = 0$ defining the minimum of J takes the form

$$\mathbf{R}_i\mathbf{R}^2 = \mathbf{R}, \tag{6}$$

$$\nabla\mathbf{R}_i\mathbf{R}^2\nabla^{\mathrm{T}} = \nabla\mathbf{R}\nabla^{\mathrm{T}}, \tag{7}$$

$$\Delta\mathbf{R}_i\mathbf{R}^2\Delta = \Delta\mathbf{R}\Delta. \tag{8}$$

Equations 6–8 can be rewritten explicitly in terms of the Hessian matrix $\mathbf{H} \in \mathbb{R}^{M\times M}$ and the r.h.s. vector $\mathbf{r} \in \mathbb{R}^M$

$$\mathbf{H}\mathbf{x} = \mathbf{r}, \tag{9}$$

where $\mathbf{r}$ contains non-zero elements in the r.h.s.s of Equations 6–8 listed columnwise as in Equation A7 of the Appendix.

In this note we consider the simplest sparsity pattern for $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$, assuming that the matrices are diagonal so that their sparsity patterns are the identity matrices of the respective size. In this case, the general expression for the Hessian (Equation A8 in the Appendix) simplifies to

$$\mathbf{H} = \begin{bmatrix} \mathbf{R}^2\circ\mathbf{I} & \mathbf{R}^2\nabla^{\mathrm{T}}\circ\nabla^{\mathrm{T}} & \mathbf{R}^2\Delta\circ\Delta \\ \nabla\mathbf{R}^2\circ\nabla & \nabla\mathbf{R}^2\nabla^{\mathrm{T}}\circ\nabla\nabla^{\mathrm{T}} & \nabla\mathbf{R}^2\Delta\circ\nabla\Delta \\ \Delta\mathbf{R}^2\circ\Delta & \Delta\mathbf{R}^2\nabla^{\mathrm{T}}\circ\Delta\nabla^{\mathrm{T}} & \Delta\mathbf{R}^2\Delta\circ\Delta^2 \end{bmatrix}, \tag{10}$$

where $\circ$ denotes Hadamard (element-wise) matrix product. The latter relationship could be useful for constructing block-diagonal preconditioners for the iterative solvers of Equation 9, or for direct solution of Equation 9 on the moderate-size ($N < 10^4$) grids.

2.2 | Model reduction

The system of Equations 9 may not be well conditioned, so its solution should be sought using a certain parametrizaion of the original control variables $\mathbf{x}$. In the linear case, such parametrization can be expressed in terms of a projection operator $\mathbf{\Pi}$

$$\mathbf{x} = \mathbf{\Pi}\widetilde{\mathbf{x}}, \tag{11}$$

where m columns of $\mathbf{\Pi}$ contain the "structure functions", parametrizing spatial variability of $\mathbf{x}$, and $\widetilde{\mathbf{x}}$ stands for the vector of the reduced control variables. For instance, to enforce smoothness of the diagonal elements of the control matrices, these functions can be represented by m smoothest harmonics, the first one being independent on horizontal coordinates. Although in the present study we did not employ any model reduction ($\mathbf{\Pi} = \mathbf{I}$), more sophisticated projectons (e.g. Brankart *et al.*, 2009, R16) can be employed (Appendix). However, using non-trivial structure of $\mathbf{\Pi}$ requires prior information on the spatial variability of the control fields in order to maintain a reasonable balance bteween the accuracy of the approximation of the target matrix and numerical efficiency.

The reduced normal system $\widetilde{\mathbf{H}}\widetilde{\mathbf{x}} = \widetilde{\mathbf{r}}$ is characterized by the reduced Hessian $\widetilde{\mathbf{H}} \in \mathbb{R}^{m\times m}$ and the reduced rhs $\widetilde{\mathbf{r}} \in \mathbb{R}^m$:

$$\widetilde{\mathbf{H}} = \mathbf{\Pi}^{\mathrm{T}}\mathbf{H}\mathbf{\Pi}, \qquad \widetilde{\mathbf{r}} = \mathbf{\Pi}^{\mathrm{T}}\mathbf{r}. \tag{12}$$

Since $\delta J/\delta\mathbf{x} = \mathbf{H}\mathbf{x} - \mathbf{r}$, the cost function gradient (4) can be conveniently estimated by applying the projection operator to Equation 5:

$$\frac{\delta J}{\delta\widetilde{\mathbf{x}}} = \mathbf{\Pi}^{\mathrm{T}}\frac{\delta J}{\delta\mathbf{x}}. \tag{13}$$

Note that, since the relationships (5) are valid for arbitrary control matrices, Equations 11–13 could be used in optimization algorithms employing gradient information under arbitrary linear constraints on spatial variability of the control matrix elements. In particular, the approach could be useful for maintaining the psd property of $\mathbf{R}_i$ in the process of minimization.

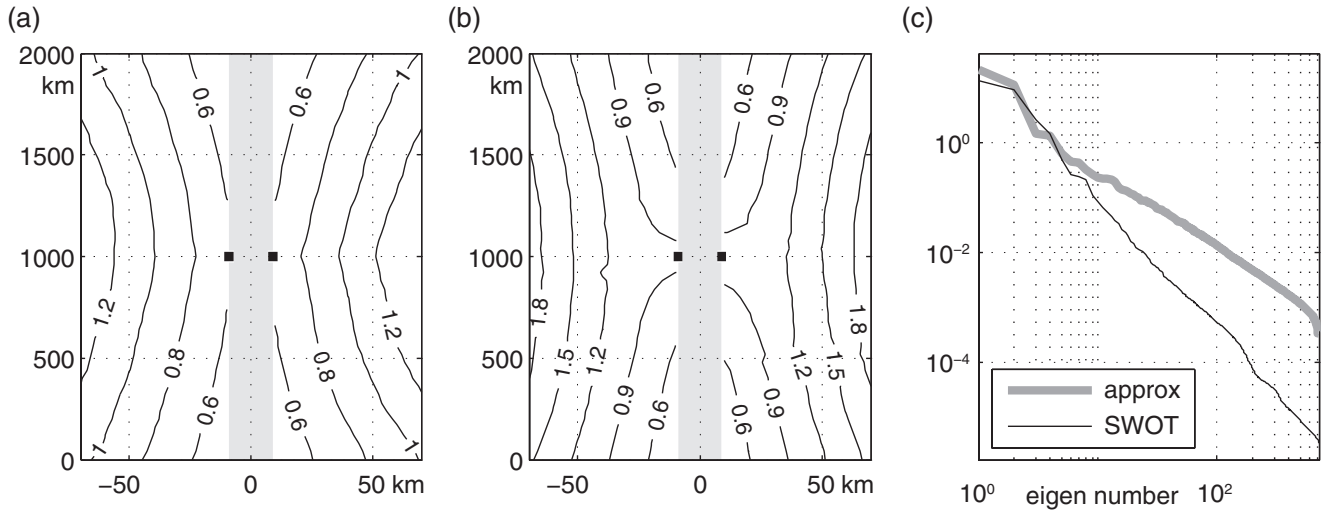

FIGURE 1 (a) Map of the sum of two columns of the SWOT covariance matrix $\mathbf{R}$ (column positions shown by squares) and its approximation (b) by $\mathbf{R}_i^{-1}$. (c) shows the spectra of the SWOT covariance (thin black line) and its approximation. Covariance values are divided by 100 cm^2

3 | NUMERICAL TESTING

3.1 | SWOT covariance model

The ansatz (1) for the approximation of the inverse observation error covariance was tested with the target OEC generated by the Jet Propulsion Laboratory's (JPL) SWOT simulator of Gaultier *et al.* (2017). The simulator generates realizations of SSH observation-error fields based on the latest estimate of the SWOT error budget of Esteban-Fernandez (2013). The error field contains six constituents: Ka-band radar interferometer noise, wet tropospheric error, and errors associated with uncertainties in the estimation of roll, phase, baseline, and timing of the SWOT observational platform. With a reasonable degree of accuracy, the first two error fields can be considered to be uncorrelated. The remaining four error sources are of particular interest because they are highly correlated over large spatial scales: As shown by Ruggiero *et al.* (2016), these errors are characterized by typical decorrelation scales of several hundred kilometres along the swath and approximately a hundred across, with the marginal pointwise p.d.f.s being very close to Gaussian.

3.2 | Results

In generating the target OEC matrix, we used the SWOT simulator version 2.0.0 with the following parameters: the default cut-off wavelength of 40,000 km and a two-beam wet tropospheric error correction. Additionally, anticipating large decorrelation scales (compared to the projected SWOT resolution of 1–2 km) and the absence of smaller-scale spatial variability in the matrix columns, we elected 40 and 10 km sampling in the along- and across-track directions respectively. This selection also decreased the influence of uncorrelated Ka-band noise on the OEC structure. 5,000 random realizations of all error sources summed together were generated by having the simulator recursively sample the same 21 day repeat orbit over a subdomain of the Western Pacific (116–133°E, 18–34°N) with a total sampled track length of 2,000 km and width of 140 km. Excluding the grid points in the 20 km wide nadir gap, the sampled OEC field dimensions were $n_x = 14$, $n_y = 51$, ($N = n_x n_y = 714$) with the total number of adjusted degrees of freedom $4N - n_x - n_y = 2791$ and the number of the independent elements in the target covariance matrix $N(N+1)/2 = 255,255$. The resulting error fields were characterized by approximately Gaussian pointwise p.d.f.s with the average magnitude of the means $\sim 10^{-4}$ m and the standard deviations varying between of 0.03 m near the nadir gap and 0.31 m at the swath periphery.

Figure 1a shows half the sum of the OEC fields corresponding to the pair of SSH observations located on both edges of the nadir gap in the centre of the sampled track. Similar to the results of R16, covariance structures at intermediate scales are barely visible. However, there is a strong anisotropy of the covariance with the typical OEC spatial scales in the along- and across-track directions differing by an order in magnitude (600 and 60 km respectively).

Figures 1a,b demonstrate the result of approximating $\mathbf{R}$ by the inverse covariance model (1). Due to the limited number of ensemble members, a slight asymetry (of the order of 1%) has been observed in the structure of the mirror rows of $\mathbf{R}$. Figure 1b shows that this asymetry is considerably enhanced in the approximating matrix $\mathbf{R}_i^{-1}$ (cf. Figure 1a,b). The effect is caused by the coarse resolution of the nadir gap which is only two grid steps wide, and associated errors in the finite-difference approximation of $\mathbf{R}^{-1}$ by the ansatz (1).

Due to the modest dimension of the control space ($N = 2791$) and low condition number (cond($\mathbf{H}$)=2×10^4) of the Hessian matrix (10), the optimization took a few seconds on a single CPU of a PC using the MatLab sparse system solver. As it is seen, the algorithm provides a reasonably accurate fit to the leading eigenmodes of $\mathbf{R}$ (Figure 1c) with the relative error $\mathrm{tr}(\mathbf{R}_i^{-1} - \mathbf{R})/\mathrm{tr}(\mathbf{R})$ of 22%.

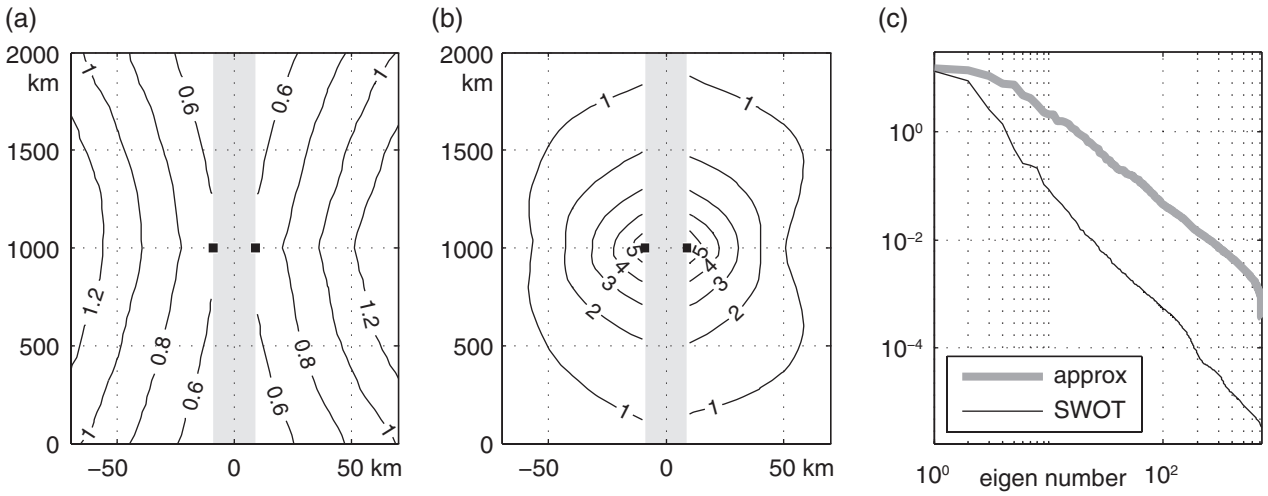

FIGURE 2 As Figure 1, but the inverse covariance model is described by Equations 12 and A1–A7

As a matter of comparison, we performed approximation of $\mathbf{R}^{-1}$ in the reduced five-dimensional space proposed by Ruggiero *et al.* (2016), who assigned a fixed spatial variability to the diagonals of the control matrices and minimized Equation 2 by varying five diagonal scaling factors $\widetilde{\mathbf{x}}$ (Appendix). In this procedure, we employed the technique of sections 2.1 and 2.2, which can be viewed as a generalization of the computational approach of Ruggiero *et al.* (2016) who used five-fold expansion of the data space by computing the derivatives of the error fields in SWOT simulator output instead of explicit computation of the Hessian (Equation A6 in Appendix) and its projection on the reduced control space (Equation 12).

Figure 2 shows the results of the reduced space optimization. As it is seen, the reduced method provides a poorer fit to the SWOT spectrum being tested (cf. Figures 1c and 2c) and a larger error in approximating the columns of the SWOT covariance matrix (cf. Figures 1b, 2b and 1a). This should be attributed to lesser flexibility of the reduced procedure, as the number of adjusted parameters is approximately 2791/5≈560 times smaller than the case of full optimization involving solution of Equations 6–8.

4 | SUMMARY AND DISCUSSION

In recent years, there has been an increased interest in inverse OEC modelling due to the high-resolution swath altimetry missions planned in the near future Durand *et al.* (2010), Ito *et al.* (2014), Ichikawa (2014). Although this new type of observational platform is characterized by improved accuracy (1–2 cm) and higher spatial resolution (1–2 km), the respective OECs are expected to be highly correlated in space. This property presents a computational challenge for many operational DA algorithms that are based on the diagonal OECs.

This note proposes a methodology of approximating the inverse OECs by a polynomial in differential operators acting on sparse control matrices whose non-zero elements are adjusted to minimize the Frobenius norm of the approximation error. Explicit relationships for the cost function gradient and the Hessian matrix of the optimization problem have been obtained for control matrices with fixed sparsity patterns. A method of reduction of the optimization problem has been demonstrated for the case of the degenerate Hessian. The proposed approach could be used in realistic DA systems by replacing the code normalizing model–data misfits by observation-error variances with the code multiplying the misfits by a sparse matrix retrieved from an estimate of the respective error covariance.

Further developments of the approach can be foreseen in several directions. First, the method does not maintain the positive semi-definite (psd) property of the approximation matrix in the process of optimization. The psd constraint can be imposed in many ways if the method is restricted to the diagonal control matrices. A straightforward way is to constrain all the components of the control vector to be positive in the process of optimization. This approach (combined with the projection technique) was used by Ruggiero *et al.* (2016) to ensure the psd property. A somewhat more sophisticated methodology is based on factorizing $\mathbf{R}_i = \mathbf{L}\mathbf{L}^{\mathrm{T}}$ and representing $\mathbf{L}$ as a composite of (sparse) control matrices: $\mathbf{L} = [\mathbf{L}_0 \quad \nabla^{\mathrm{T}}\mathbf{L}_1 \quad ...]^{\mathrm{T}}$ with diagonal controls $\mathbf{L}_i, i = 0, 1...,$. However, this option destroys the attractive quadratic property of the optimization problem. More general approaches going beyond the *fixed* sparsity patterns of the controls can also be explored (e.g. Hsieh *et al.*, 2014). However, our experience with the presented version of the method have shown that optimal $\mathbf{R}_i$ was very close to psd with only a few negative eigenvalues that contributed less than 0.1% to the trace of $\mathbf{R}_i$.

The cost function could also be defined by $J = ||\mathbf{I} - \mathbf{R}_i\mathbf{R}\mathbf{R}_i||_{\mathrm{F}}^2$ to directly retrieve a sparse approximation to $\mathbf{R}^{-1/2}$ which may be more useful in the DA applications. Our numerical experiments with this formulation have shown that one has to pay more attention to initialization of the control variables, because starting the quasi-Newtonian descent from $\mathbf{x} = 0$ proved inefficient for several simulated classes

of OECs. In contrast, the considered quadratic/diagonal formulation (1) performed well and never encountered convergence/conditioning problems for the same classes of OECs.

Elaboration of an efficient reduction scheme also remains an important issue. Ruggiero *et al.* (2016) have shown that certain OECs can be efficiently approximated with just a few parameters, if an appropriate projection method is elected. In particular, useful information on the structure of $\mathbf{\Pi}$ could be retrived from the structure of the diagonal cells of the Hessian matrices. An alternative way of regularizing the problem is to augment (2) with the terms which penalize high-frequency variations of the control variables. However, the respective low-pass filter should be designed with caution, as the high-frequency variations of the inverse matrix elements (partly simulated by the differential operators) are a key component of the optimized matrix.

We believe that further studies of the matrix approximation methodologies in application to the class of psd matrices with slowly varying spatial structure has good prospects in the future development of DA techniques in geophysical applications and may benefit more general areas such as the search for efficient preconditioners.

ACKNOWLEDGEMENTS

This study was supported by Office of Naval Reseach projects (program elements 0646352N, 0602435N). J. D'Addezio was supported by the Naval Research Laboratory cooperative agreement BAA-N00173-03-13-01 awarded to the University of Southern Mississippi. Helpful discussions with Prof. C. Beattie are acknowledged.

REFERENCES

Brankart, J., Ubelmann, C., Testut, C., Cosme, E., Brasseur, P. and Verron, J. (2009) Efficient parameterization of the observation error covariance matrix for square root or ensemble Kalman filters: application to ocean altimetry. *Monthly Weather Review*, 137, 1908–1927. https://doi.org/10.1175/2008MWR2693.1.

Cummings, J. (2005) Operational multivariate data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 131, 3583–3604. https://doi.org/10.1256/qj.05.105.

Derber, J. and Rosati, A. (1989) A global oceanic data assimilation system. *Journal of Physical Oceanography*, 19, 1333–1347.

Durand, M., Fu, L.L., Lettenemaier, D., Alsdorf, D., Rodriguers, E. and Esteban-Fernandez, D. (2010) The surface water and ocean topography mission: observing terrestrial surface water and oceanic submesoscale eddies. *Proceedings of the IEEE*, 98, 766–779. https://doi.org/10.1109/JPROC.2010.2043031.

Esteban-Fernandez, D. (2013) *SWOT project: mission performance and error budget. Revision A*. NASA/JPL Technical Report JPL D-79084. Available at: https://swot.jpl.nasa.gov/docs/SWOT_D-79084_v10Y_FINAL_REVA_06082017.pdf [Accessed 3rd September 2018].

Fairbairn, D., Pring, S.R., Lorenc, A.C. and Roulstone I. (2014) A comparison of 4D-Var with ensemble data assimilation methods. *Quarterly Journal of the Royal Meteorological Society*, 140, 281–294.

Gaultier, L., Ubelmann, C. and Fu, L.-L. (2016) The challenge of using future SWOT data for oceanic field reconstruction. *Journal of Atmospheric and Oceanic Technology*, 33, 119–126.

Gaultier, L., Ubelmann, C. and Fu, L.-L. (2017) *SWOT simulator for ocean science*. Documentation, release 2.3.0. Available at: https://swot.oceansciences.org/docs/documentation_swotsimulator.pdf [Accessed 3rd September 2018].

Hsieh, C.-J., Sustik, M.A., Dhillon, I.S. and Ravikumar, P. (2014) Quadratic approximation for sparse inverse covariance estimation. *Journal of Machine Learning Research*, 15, 2911–2947.

Hunt, B.R., Kostelich, E.J. and Szunyogh, E.K. (2007) Assimilation for spatio-temporal chaos: a local ensemble transform filter. *Physica D*, 230, 112–126.

Ichikawa, K. (2014) Satellite altimeters in the early 21st Century. *Oceanography in Japan*, 23, 13–27. (in Japanese, English abstract/figures).

Ito, N., Uematsu, A., Yajima, Y. and Isoguchi, O. (2014) A Japanese new altimetry mission COMPIRA – towards high temporal and spatial sampling of sea surface height. In: *AGU Fall Meeting*. 15–19 December 2014, San Francisco, CA, Abstract OS34B-05.

Miyoshi, T., Kalnay, E. and Li, H. (2013) Estimating and including observation-error correlations in data assimilation. *Inverse Problems in Science and Engineering*, 21, 387–398. https://doi.org/10.1080/17415977.2012.712527.

Ruggiero, G.A., Cosme, E., Brankart, J.-M., Le Sommer, J. and Ubelmann, C. (2016) An efficient way to account for observation error correlations in the assimilation of data from the future SWOT high-resolution altimeter mission. *Journal of Atmospheric and Oceanic Technology*, 33, 2755–2768.

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

Ubelmann, C., Klein, P. and Fu, L.-L. (2015) Dynamic interpolation of sea surface height and potential applications for future high-resolution altimetry mapping. *Journal of Atmospheric and Oceanic Technology*, 32, 177–184.

Waller, J.A., Dance, S.L., Lawless, A.S. and Nichols, N.K. (2014) Estimating correlated observation error statistics using an ensemble transform Kalman filter. *Tellus*, 66A, 23294. https://doi.org/10.3402/tellusa.v66.23294.

Weaver, A.T., Vialard, J. and Anderson, D.L.T. (2003) Three- and four-dimensional variational assimilation with a general circulation model of the Tropical Pacific Ocean. Part I: formulation, internal diagnostics and consistency checks. *Monthly Weather Review*, 131, 1360–1378.

Xu, Q. (2005) Representations of inverse covariances by differential operators. *Advances in Atmospheric Sciences*, 22, 181–198.

Yaremchuk, M. and Smith, S. (2011) On the correlation functions associated with polynomials of the diffusion operator. *Quarterly Journal of the Royal Meteorological Society*, 137, 1927–1032.

Yaremchuk, M., Carrier, M., Smith, S. and Jacobs, G. (2013) Background error correlation modeling with diffusion operators. In: Park, S.K. and Xu, L. (Eds.) *Data assimilation for Atmospheric, Oceanic and Hydrologic Applications* Vol. 3. Berlin: Springer, pp. 83–114.

How to cite this article: Yaremchuk M, D'Addezio JM, Panteleev G, Jacobs G. On the approximation of the inverse error covariances of high-resolution satellite altimetry data. *Q J R Meteorol Soc*. 2018;1–6. https://doi.org/10.1002/qj.3336

APPENDIX

Model reduction and the Hessian matrix

Ruggiero *et al.* (2016) proposed a simplified method of estimating the inverse of $\mathbf{R}$ through the adjustment of only five free parameters. In the notation of section 2, their scheme involves a combination of the inverse covariance model containing four diagonal matrices

$$\mathbf{R}_i = \mathbf{A} + \nabla^{\mathrm{T}}\mathbf{B}\nabla + \Delta_x\mathbf{C}\Delta_x + \Delta_y\mathbf{D}\Delta_y, \qquad \text{(A1)}$$

with a projection scheme which assigns a certain spatial structure to the matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$ and $\mathbf{D}$. In Equation A1 the operators Δ_x and Δ_y stand for the cross- and along-track constituents of the Laplacian: $\Delta = \Delta_x + \Delta_y$.

Specifically,the projection adopted in R16 is defined by

$$\mathbf{A} = \alpha_0(\mathbf{R}\circ\mathbf{I})^{-1}, \tag{A2}$$

$$\mathbf{B} = \begin{matrix}\alpha_{1c}(\partial_x\mathbf{R}\,\partial_x^{\mathrm{T}}\circ\mathbf{I})^{-1} \\ \alpha_{1a}(\partial_y\mathbf{R}\,\partial_y^{\mathrm{T}}\circ\mathbf{I})^{-1},\end{matrix} \tag{A3}$$

$$\mathbf{C} = \alpha_{2c}(\Delta_x\mathbf{R}\Delta_x\circ\mathbf{I})^{-1}, \tag{A4}$$

$$\mathbf{D} = \alpha_{2a}(\Delta_y\mathbf{R}\Delta_y\circ\mathbf{I})^{-1}, \tag{A5}$$

with the control vector $\widetilde{\mathbf{x}} = [\alpha_0\ \ \alpha_{1c}\ \ \alpha_{1a}\ \ \alpha_{2c}\ \ \alpha_{2a}]^{\mathrm{T}}$ and the projection operator $\mathbf{\Pi}$ represented by a $(5N - n_x - n_y) \times 5$ block-diagonal matrix containing the inverse matrices in the r.h.s.s of Equations A2–A5 on the diagonal.

The system of equations $\mathbf{H}\boldsymbol{x} = \mathbf{r}$ is

$$\mathbf{H} = \begin{bmatrix} \mathbf{R}^2\circ\mathbf{I} & \mathbf{R}^2\nabla^{\mathrm{T}}\circ\nabla^{\mathrm{T}} & \mathbf{R}^2\Delta_x\circ\Delta_x & \mathbf{R}^2\Delta_y\circ\Delta_y \\ \nabla\mathbf{R}^2\circ\nabla & \nabla\mathbf{R}^2\nabla^{\mathrm{T}}\circ\nabla\nabla^{\mathrm{T}} & \nabla\mathbf{R}^2\Delta_x\circ\nabla\Delta_x & \nabla\mathbf{R}^2\Delta_y\circ\nabla\Delta_y \\ \Delta_x\mathbf{R}^2\circ\Delta_x & \Delta_x\mathbf{R}^2\nabla^{\mathrm{T}}\circ\Delta_x\nabla^{\mathrm{T}} & \Delta_x\mathbf{R}^2\Delta_x\circ\Delta_x^2 & \Delta_x\mathbf{R}^2\Delta_y\circ\Delta_{xy} \\ \Delta_y\mathbf{R}^2\circ\Delta_y & \Delta_y\mathbf{R}^2\nabla^{\mathrm{T}}\circ\Delta_y\nabla^{\mathrm{T}} & \Delta_y\mathbf{R}^2\Delta_x\circ\Delta_{xy} & \Delta_y\mathbf{R}^2\Delta_y\circ\Delta_y^2 \end{bmatrix}, \tag{A6}$$

$$\mathbf{r} = \begin{bmatrix} \mathbf{R}\circ\mathbf{I} \\ \nabla\mathbf{R}\nabla^{\mathrm{T}}\circ\mathbf{I} \\ \Delta_x\mathbf{R}\Delta_x\circ\mathbf{I} \\ \Delta_y\mathbf{R}\Delta_y\circ\mathbf{I} \end{bmatrix}. \tag{A7}$$

It it noteworthy, that the ansatz (A1) produced a degenerate Hessian (A6) which did not allow us to compare the results of full optimizations with the inverse OEC models (1) and (A1). Reduction of the control space regularized the problem, but resulted in a relatively poor fit to the spectrum of the tested covariance (cf. Figures 1c and 2c).

As a final note, we present the general expression for the Hessian matrix associated with the column-vectorized form of Equations 6–8. Defining the sparsity pattern $\mathbf{S_A}$ of a matrix $\mathbf{A}$ by replacing non-zero elements of $\mathbf{A}$ with ones, and adoptng the notation $\mathbf{S_{AB}} = \mathrm{vec}(\mathbf{S_A}) \otimes \mathrm{vec}(\mathbf{S_B^{\mathrm{T}}})$ for mutual Kronecker products of the sparsity patterns, the Hessian is given by

$$\mathbf{H} = \begin{bmatrix} (\mathbf{R}^2\otimes\mathbf{I})\circ\mathbf{S_{AA}} & (\mathbf{R}^2\nabla^{\mathrm{T}}\otimes\nabla^{\mathrm{T}})\circ\mathbf{S_{AB}} & (\mathbf{R}^2\Delta\otimes\Delta)\circ\mathbf{S_{AC}} \\ (\nabla\mathbf{R}^2\otimes\nabla)\circ\mathbf{S_{BA}} & (\nabla\mathbf{R}^2\nabla^{\mathrm{T}}\otimes\nabla\nabla^{\mathrm{T}})\circ\mathbf{S_{BB}} & (\nabla\mathbf{R}^2\Delta\otimes\nabla\Delta)\circ\mathbf{S_{BC}} \\ (\Delta\mathbf{R}^2\otimes\Delta)\circ\mathbf{S_{CA}} & (\Delta\mathbf{R}^2\nabla^{\mathrm{T}}\otimes\Delta\nabla^{\mathrm{T}})\circ\mathbf{S_{CB}} & (\Delta\mathbf{R}^2\Delta\otimes\Delta^2)\circ\mathbf{S_{CC}} \end{bmatrix}. \tag{A8}$$