Using reconditioning methods to reduce the cost of using correlated observation error information: theory and practice

Jemima M. Tabeart

Supervised by

Amos Lawless, Sarah Dance, Joanne Waller, Nancy Nichols (University of Reading)
David Simonin (MetOffice@Reading)

Additional collaboration

Stefano Migliorini and Fiona Smith (Met Office, Exeter)

RMetSoc Assimilation of Big Data, September 2019

Overview

1. Including correlated observation errors in NWP
2. Using NLA to develop theory for the linear case
3. Reconditioning methods
4. Case study using the Met Office 1D-Var system
   - Impact of reconditioning on convergence

Variational data assimilation for numerical weather prediction

In variational DA, the most likely state of the atmosphere, $\mathbf{x}$ minimises the objective function:

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

This measures the goodness of fit to the background/prior and observations, weighted by their respective error covariances $\mathbf{B} \in \mathbb{R}^{N \times N}$ and $\mathbf{R} \in \mathbb{R}^{p \times p}$.

Huge problem: $O(10^9)$ state variables, $O(10^7)$ observations multiple times per day

Including correlated observation error information has multiple benefits for NWP

Benefits

- Using uncorrelated OEC matrices means we have to thin observations - this can result in **up to 80%** of obs being discarded!
- Including correlated OEC information allows us to take advantage of dense observation networks to get **high-resolution forecasts**.
- Neglecting correlations where they are present also **limits our skill**.

Including correlated observation error information has multiple benefits for NWP

Benefits

- Using uncorrelated OEC matrices means we have to thin observations - this can result in **up to 80%** of obs being discarded!
- Including correlated OEC information allows us to take advantage of dense observation networks to get **high-resolution forecasts**.
- Neglecting correlations where they are present also **limits our skill**.

Issues

Correlations typically estimated using sampling methods:

- Rank deficient matrices
- Extremely small eigenvalues

How does the variational data assimilation problem change with introduction of **correlated** OEC matrices?

- The Hessian of the objective function is given by

$$\mathbf{S} = \mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}. \quad (2)$$

  $\mathbf{H}$ *is the linearised observation operator, linearised about the optimal solution.*

- Can use **condition number of the Hessian** to investigate speed of convergence and sensitivity of analysis to perturbations.

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

Bounds on $\kappa(\mathbf{S})$ in the case of correlated observation error

Theorem [T. et al, 2018a]

Let $\mathbf{B} \in \mathbb{R}^{N \times N}$ *and* $\mathbf{R} \in \mathbb{R}^{p \times p}$, *with* $p < N$, *be the background and observation error covariance matrices respectively. Additionally, let* $\mathbf{H} \in \mathbb{R}^{p \times N}$ *be the observation operator. Then the following bounds are satisfied by the condition number of the Hessian,* $\mathbf{S} = \mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$,

$$\max\left(\frac{1 + \frac{\lambda_{max}(\mathbf{B})}{\lambda_{min}(\mathbf{R})}\lambda_{min}(\mathbf{H}\mathbf{H}^T)}{\kappa(\mathbf{B})}, \frac{1 + \frac{\lambda_{max}(\mathbf{B})}{\lambda_{max}(\mathbf{R})}\lambda_{max}(\mathbf{H}\mathbf{H}^T)}{\kappa(\mathbf{B})}, \frac{\kappa(\mathbf{B})}{(1 + \frac{\lambda_{max}(\mathbf{B})}{\lambda_{min}(\mathbf{R})}\lambda_{max}(\mathbf{H}\mathbf{H}^T))}\right)$$

$$\leq \kappa(\mathbf{S}) \leq (1 + \frac{\lambda_{min}(\mathbf{B})}{\lambda_{min}(\mathbf{R})}\lambda_{max}(\mathbf{H}\mathbf{H}^T))\kappa(\mathbf{B}).$$

Minimum eigenvalue of **R** appears in upper and lower bound

What is reconditioning?

Methods which can be applied to matrices to reduce their condition number, while retaining underlying matrix structure.

What is reconditioning?

Methods which can be applied to matrices to reduce their condition number, while retaining underlying matrix structure.

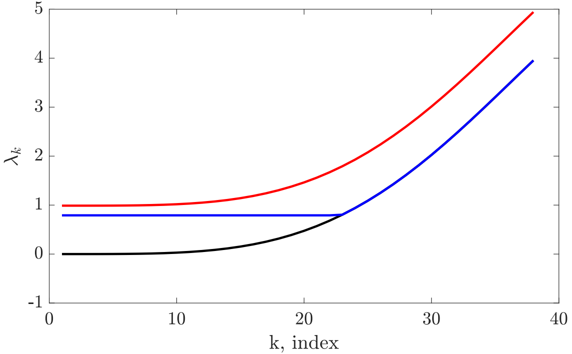

Figure: Illustration of recond methods: original spectrum (black), and spectrum reconditioned via two different methods. We are interested in ridge regression - red line - in this talk.

Recap on covariance vs. correlation, standard deviations

Want to consider how variances and correlations are changed by reconditioning methods. Let

$$\mathbf{R} = \mathbf{\Sigma} \mathbf{C} \mathbf{\Sigma}, \tag{3}$$

where $\mathbf{C}$ is the correlation matrix, and $\mathbf{\Sigma}$ is the diagonal matrix of standard deviations. We calculate $\mathbf{C}$ and $\mathbf{\Sigma}$ via:

$$\mathbf{\Sigma}(i, i) = \sqrt{\mathbf{R}(i, i)} \tag{4}$$

and

$$\mathbf{C}(i, j) = \frac{\mathbf{R}(i, j)}{\sqrt{\mathbf{R}(i, i)}\sqrt{\mathbf{R}(j, j)}}. \tag{5}$$

Ridge regression method

Idea: Add a scalar multiple of identity to $\mathbf{R}$ *to obtain reconditioned* $\mathbf{R}_{RR}$ *with* $\kappa(\mathbf{R}_{RR}) = \kappa_{max}$.

Setting δ

- *Define* $\delta = \frac{\lambda_1(\mathbf{R}) - \lambda_p(\mathbf{R})\kappa_{max}}{\kappa_{max} - 1}$.
- *Set* $\mathbf{R}_{RR} = \mathbf{R} + \delta \mathbf{I}$

We can prove [T. et al, 2019b]:

- Standard deviations are increased by using this method.
- Absolute value of off-diagonal correlations decreased by this method.

Case study: Met Office 1D-Var system [T. et al, 2019c]

We study the impact of introducing correlated observation errors/reconditioning for IASI instrument in the **1D-Var** procedure

- Run prior to every 4D-Var/forecast cycle.
- Assimilates 97330 'observations' individually
- Used as quality control (reject ob if it doesn't converge in 10 iterations)
- Also used to fix values for variables that aren't assimilated in 4D-Var procedure.

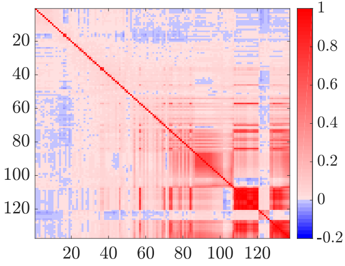

Figure: Estimated IASI OEC matrix

Experimental choices of $\mathbf{R}_{RR}$ - standard deviations

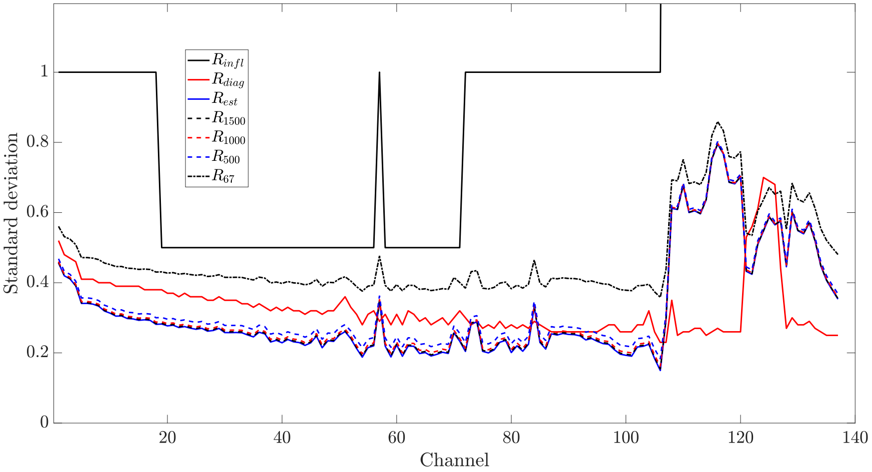

Figure: Standard deviation for each of the experiment choices

Experimental choices of $\mathbf{R}_{RR}$ - correlations

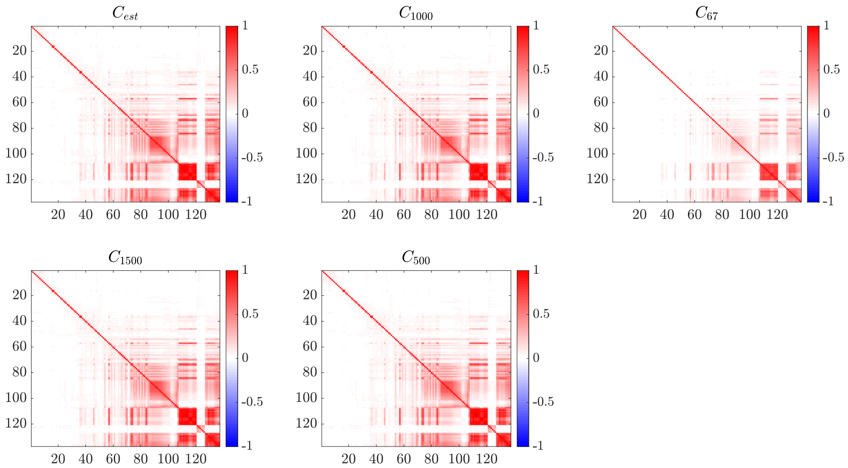

Figure: Changes to correlation with reconditioning for the correlated experiments

Reconditioning improves convergence of 1D-Var

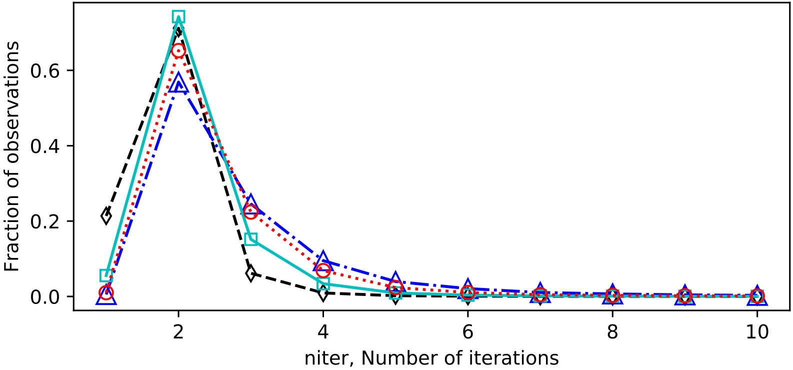

Figure: Number of iterations required to reach convergence of the 1D-Var minimization as a fraction of the total number of observations common to all choices of $\mathbf{R}$. Symbols correspond to: $\triangle = \mathbf{R}_{diag}$, $\circ = \mathbf{R}_{est}$, $\square = \mathbf{R}_{67}$, $\lozenge = \mathbf{R}_{infl}$.

Changes to variables not in 4D-Var state vector are mostly small for correlated choices of $\mathbf{R}_{RR}$

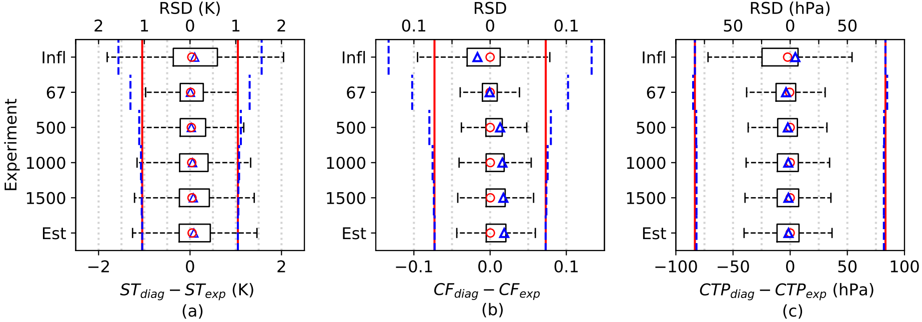

Figure: Change to estimates for skin temperature (left), cloud fraction (centre) and cloud top pressure (right) with reconditioning.

Altering **R** also changes the quality control procedure

- More observations pass QC when **R** is correlated. This number increases with reconditioning. Largest number pass for $\mathbf{R}_{infl}$.
- For ST, 10 observations with differences $>$ 20K (clearly unrealistic!)
  - Our QC should be catching these observations.
  - Although small, can have large effect on 4D-Var.

Key results from case study

- Convergence improves with reconditioning - obtain better performance than current choice of diagonal matrix
- Reducing κ_{max} increases the number of observations that converge in less than 10 iterations.
- The impact to variables that aren't assimilated in 4D-Var is small for most observations.
- However, some very large (unphysical) changes occur - these are possibly due to changes in cloud but further work is required.

Conclusions

- We want to include correlated OEC matrices - but there are problems associated with sampling.
- Bounds on $\kappa(\mathbf{S})$ show that minimum eigenvalue of observation error covariance is important term.
- We study reconditioning methods as way to improve convergence
- Tests in the Met Office 1D-Var system show that:
  - the ridge regression method improves convergence.
  - the quality control process is altered.

References I

J. M. Tabeart, S. L. Dance, S. A. Haben, A. S. Lawless, N. K. Nichols, and J. A. Waller (2018)

The conditioning of least squares problems in variational data assimilation.

Numerical Linear Algebra with Applications http://dx.doi.org/10.1002/nla.2165

J. M. Tabeart, S. L. Dance, A. S. Lawless, N. K. Nichols, and J. A. Waller (2019a)

Improving the condition number of estimated covariance matrices

Submitted, Tellus A https://arxiv.org/abs/1810.10984

J. M. Tabeart, S. L. Dance, F. Hilton, A. S. Lawless, S. Migliorini, N. K. Nichols, and J. A. Waller (2019b)

The impact of reconditioning of the correlated observation error covariance matrix on the Met Office system

Submitted, Quarterly Journal Royal Meteorological Society

Stephen Haben (2011)

Conditioning and Preconditioning of the Minimisation Problem in Variational Data Assimilation

PhD thesis University of Reading

References II

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

Niels Bormann, Massimo Bonavita, Rossana Dragani, Reima Eresmaa, Marco Matricardi, and Anthony McNally (2016)
Enhancing the impact of IASI observations through an updated observation error covariance matrix
doi: 10.1002/qj.2774

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

O. Ledoit and M. Wolf (2004)
A well-conditioned estimator for large-dimensional covariance matrix.
J. Multivariate Anal. 88:365-411

References III

J. H. Wilkinson (1965)
The Algebraic Eigenvalue Problem
Clarendon Press

V. B. Lidskii, C. D. Benster, and G. E. Forsythe. (1953)
The proper values of the sum and product of symmetric matrices
Office of Naval Research

B. Wang and F. Zhang (1992)
Some Inequalities for the Eigenvalues of the Product of Positive Semidefinite Hermitian Matrices
Linear Algebra and Its Applications 160:113–118

M. Tanaka and K. Nakata (2014)
Positive definite matrix approximation with condition number constraint.
Optim. Lett. 8:939947.

Laura Stewart (2010)
Correlated observation errors in data assimilation
PhD thesis University of Reading

Theory of the ridge regression method

Effect of RR on standard deviations:

$$\Sigma_{RR} = (\mathbf{\Sigma}^2 + \delta \mathbf{I}_p)^{1/2}. \tag{6}$$

i.e. variances are increased by the reconditioning constant, δ.

Theory of the ridge regression method

Effect of RR on standard deviations:

$$\Sigma_{RR} = (\mathbf{\Sigma}^2 + \delta \mathbf{I}_p)^{1/2}. \quad (6)$$

i.e. variances are increased by the reconditioning constant, δ.

Effect of RR on correlations:

$$\text{For } i \neq j, |\mathbf{C}_{RR}(i,j)| < |\mathbf{C}(i,j)| \quad (7)$$

i.e. the magnitude of all off-diagonal correlations is strictly decreased.

Minimum eigenvalue method

Idea: Fix a threshold, T, below which all eigenvalues of the reconditioned matrix, $\mathbf{R}_{ME}$, are set equal to T to yield $\kappa(\mathbf{R}_{ME}) = \kappa_{max}$.

Setting T:

- *Set $\lambda_1(\mathbf{R}_{ME}) = \lambda_1(\mathbf{R})$*
- *Define $T = \lambda_1(\mathbf{R})/\kappa_{max}$.*
- *Set the remaining eigenvalues of $\mathbf{R}_{ME}$ via*

$$\lambda_k(\mathbf{R}_{ME}) = \begin{cases} \lambda_k(\mathbf{R}) & \text{if } \lambda_k(\mathbf{R}) > T \\ T & \text{if } \lambda_k(\mathbf{R}) \leq T \end{cases}. \tag{8}$$

- *We define $\mathbf{\Gamma}(k,k) = max\{T - \lambda_i, 0\}$.*

Minimum eigenvalue method

Idea: Fix a threshold, T, below which all eigenvalues of the reconditioned matrix, $\mathbf{R}_{ME}$, are set equal to T to yield $\kappa(\mathbf{R}_{ME}) = \kappa_{max}$.

Setting T:

- *Set $\lambda_1(\mathbf{R}_{ME}) = \lambda_1(\mathbf{R})$*
- *Define $T = \lambda_1(\mathbf{R})/\kappa_{max}$.*
- *Set the remaining eigenvalues of $\mathbf{R}_{ME}$ via*

$$\lambda_k(\mathbf{R}_{ME}) = \begin{cases} \lambda_k(\mathbf{R}) & \text{if } \lambda_k(\mathbf{R}) > T \\ T & \text{if } \lambda_k(\mathbf{R}) \leq T \end{cases}. \tag{8}$$

- *We define $\mathbf{\Gamma}(k,k) = max\{T - \lambda_i, 0\}$.*

A variant of this method is used at the European Centre for Medium-Range Weather Forecasts (ECMWF) [Bormann et al, 2016].

Theory of the minimum eigenvalue method

Effect of ME on standard deviations:

$$\mathbf{\Sigma}_{ME}(i,i) = \left(\mathbf{\Sigma}(i,i)^2 + \sum_{k=1}^{p} \mathbf{V}_R(i,k)^2 \mathbf{\Gamma}(k,k) \right)^{1/2} \tag{9}$$

Theory of the minimum eigenvalue method

Effect of ME on standard deviations:

$$\boldsymbol{\Sigma}_{ME}(i,i) = \left(\boldsymbol{\Sigma}(i,i)^2 + \sum_{k=1}^{p} \mathbf{V}_R(i,k)^2 \boldsymbol{\Gamma}(k,k)\right)^{1/2} \tag{9}$$

This can be bounded by

$$\boldsymbol{\Sigma}(i,i) \leq \boldsymbol{\Sigma}_{ME}(i,i) \leq \left(\boldsymbol{\Sigma}(i,i)^2 + T - \lambda_p(\mathbf{R})\right)^{1/2}. \tag{10}$$

Theory of the minimum eigenvalue method

Effect of ME on standard deviations:

$$\mathbf{\Sigma}_{ME}(i,i) = \left(\mathbf{\Sigma}(i,i)^2 + \sum_{k=1}^{p} \mathbf{V}_R(i,k)^2 \mathbf{\Gamma}(k,k)\right)^{1/2} \tag{9}$$

This can be bounded by

$$\mathbf{\Sigma}(i,i) \leq \mathbf{\Sigma}_{ME}(i,i) \leq \left(\mathbf{\Sigma}(i,i)^2 + T - \lambda_p(\mathbf{R})\right)^{1/2}. \tag{10}$$

Effect of ME on correlations:

It is not evident how correlation entries are altered in general.
This is due to the fact that the spectrum of **R** is not altered uniformly by this method.

Comparison of both methods

- Both methods increase (or maintain) standard deviations

Comparison of both methods

- Both methods increase (or maintain) standard deviations
- We can show that $T - \lambda_p(\mathbf{R}) < \delta$ which yields:

$$\boldsymbol{\Sigma}_{ME}(i,i) \leq \left(\boldsymbol{\Sigma}(i,i)^2 + T - \lambda_p(\mathbf{R})\right)^{1/2} < (\boldsymbol{\Sigma}(i,i)^2 + \delta)^{1/2} = \Sigma_{RR}(i,i)$$

- Therefore RR increases standard deviations more than ME

Impact of reconditioning on $\kappa(\mathbf{S})$

	$\mathbf{R}_{diag}$	$\mathbf{R}_{raw}$	$\mathbf{R}_{1500}$	$\mathbf{R}_{1000}$	$\mathbf{R}_{500}$	$\mathbf{R}_{67}$	$\mathbf{R}_{old}$
max	3.01×10^{12}	7.546×10^{11}	7.469×10^{11}	7.30×10^{11}	7.02×10^{11}	3.71×10^{11}	1.74×10^{11}
mean	2.78×10^{10}	6.71×10^{9}	6.62×10^{9}	6.43×10^{9}	6.00×10^{9}	4.01×10^{9}	2.83×10^{9}
median	2.09×10^{8}	1.31×10^{8}	1.32×10^{8}	1.33×10^{8}	1.37×10^{8}	1.78×10^{8}	2.89×10^{8}

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

Impact on temperature and humidity

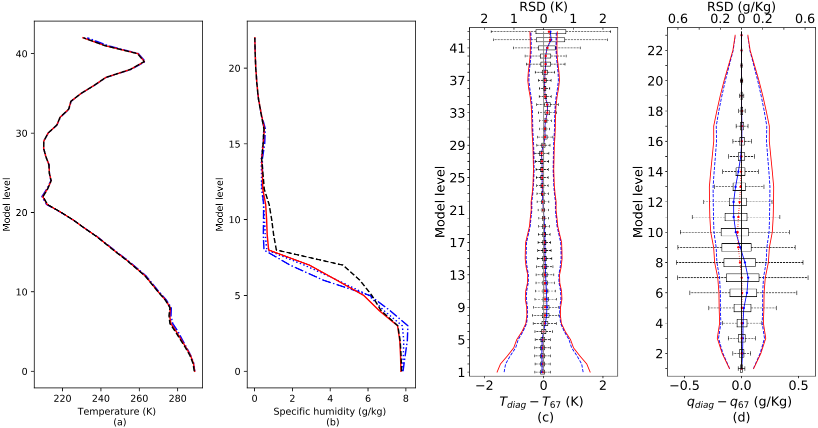

Figure: Example retrieved profiles of temperature (a) and specific humidity (b), and differences in retrievals between E_{diag} and E_{67} for temperature (c) and specific humidity (d) for 97330 observations.

Impact of reconditioning on quality control procedure

Set	No. of accepted obs	No of obs accepted by both E_{diag} and E_{exp}
$\mathbf{R}_{diag}$	100686	99039
$\mathbf{R}_{est}$	100655	99175
$\mathbf{R}_{1000}$	101002	99352
$\mathbf{R}_{500}$	101341	99656
$\mathbf{R}_{67}$	102333	100382
$\mathbf{R}_{infl}$	102859	100679

Table: Change to number of accepted observations with reconditioning