Using reconditioning to study the impact of correlated observation errors in the Met Office 1D-Var system

Jemima M. Tabeart
jemima.tabeart@pgr.reading.ac.uk
@jemimatabeart

Supervised by
Sarah L. Dance, Nancy K. Nichols, Amos S. Lawless, Joanne A. Waller (University of Reading) David Simonin (MetOffice@Reading)

Additional collaboration
Stefano Migliorini and Fiona Smith (Met Office, Exeter)

Overview

1. Motivation
   - Why do we want to include correlated observation error information?
2. Reconditioning is one solution!
   - What is reconditioning?
   - Theory of reconditioning
3. Implementation in the Met Office system
   - IASI - operational interchannel correlations
   - Impact of reconditioning on convergence
   - Impact of reconditioning on quality control procedure

Cost function for 3DVar

We want to minimise

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

where,

- $\mathbf{B} \in \mathbb{R}^{N \times N}$ background error covariance
- $\mathbf{R} \in \mathbb{R}^{p \times p}$ observation error covariance
- $h : \mathbb{R}^N \to \mathbb{R}^p$ observation operator
- $\mathbf{y} \in \mathbb{R}^p$ vector of observations
- $\mathbf{x}_b \in \mathbb{R}^N$ vector representing the background

What happens to the convergence of (1) if we introduce correlated **R**?.

Why include correlated observation error information?

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

Why is it hard to include correlation information?

- For high-dimensional problems (e.g. big data) we can only estimate correlation information by sampling.

Why is it hard to include correlation information?

- For high-dimensional problems (e.g. big data) we can only estimate correlation information by sampling.

→ Sample covariance matrices that aren't full rank.

→ Sample matrices that are **extremely** ill-conditioned.

- We use $\mathbf{R}^{-1}$ in (1) - calculating this inverse is expensive/impossible

→ May have to do this online

Why is it hard to include correlation information?

- For high-dimensional problems (e.g. big data) we can only estimate correlation information by sampling.

$\rightarrow$ Sample covariance matrices that aren't full rank.

$\rightarrow$ Sample matrices that are **extremely** ill-conditioned.

- We use $\mathbf{R}^{-1}$ in (1) - calculating this inverse is expensive/impossible

$\rightarrow$ May have to do this online

Reminder: If $\mathbf{S} \in \mathbb{R}^{p \times p}$ *is a symmetric and positive definite matrix with eigenvalues* $\lambda_1(\mathbf{S}) \geq \ldots \geq \lambda_p(\mathbf{S}) > 0$ *then we write the condition number*

$$\kappa(\mathbf{S}) = \frac{\lambda_1(\mathbf{S})}{\lambda_p(\mathbf{S})}.$$

If $\mathbf{S}$ *is singular, we take* $\kappa(\mathbf{S}) = \infty$.

Motivation II - Satellite observations

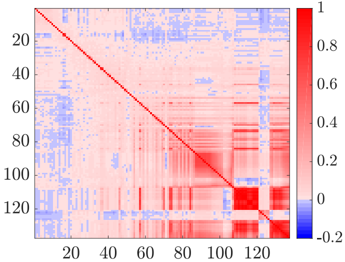


Figure: Diagnosed correlation matrix for IASI

- Satellite observation errors are known to have correlated observation errors [Stewart, 2010].

Motivation II - Satellite observations

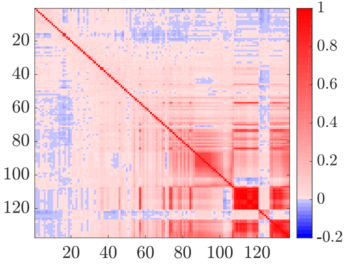


Figure: Diagnosed correlation matrix for IASI

- Satellite observation errors are known to have correlated observation errors [Stewart, 2010].
- Methods to diagnose covariance matrices exist but yield matrices that are highly ill-conditioned and caused problems with convergence.

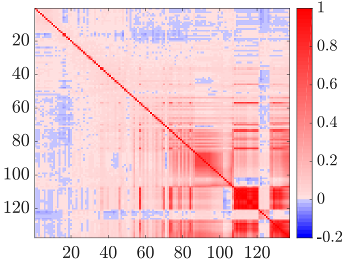


Figure: Diagnosed correlation matrix for IASI

- Satellite observation errors are known to have correlated observation errors [Stewart, 2010].
- Methods to diagnose covariance matrices exist but yield matrices that are highly ill-conditioned and caused problems with convergence.
- [Weston et al, 2014] suggested these problems were due to very small eigenvalues and tested two methods of ‘reconditioning’.

Motivation II - Satellite observations

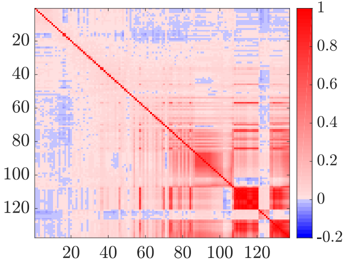


Figure: Diagnosed correlation matrix for IASI

- Satellite observation errors are known to have correlated observation errors [Stewart, 2010].
- Methods to diagnose covariance matrices exist but yield matrices that are highly ill-conditioned and caused problems with convergence.
- [Weston et al, 2014] suggested these problems were due to very small eigenvalues and tested two methods of ‘reconditioning’.
- [Tabeart et al, 2018] proved that the minimum eigenvalue of the observation error covariance matrix is important for the conditioning of the general data assimilation problem.

Reconditioning is one solution!

We want to reduce the impact of small eigenvalues of $\mathbf{R}$ (experimental and theoretical evidence) - one way is by using reconditioning methods.

What is reconditioning?

Reconditioning is one solution!

We want to reduce the impact of small eigenvalues of $\mathbf{R}$ (experimental and theoretical evidence) - one way is by using reconditioning methods.

What is reconditioning?

- *Methods which can be applied to matrices to reduce their condition number, while retaining underlying matrix structure.*
- Examples of methods:
  - Thresholding
  - Tapering
  - General regularisation methods.

Reconditioning is one solution!

We want to reduce the impact of small eigenvalues of $\mathbf{R}$ (experimental and theoretical evidence) - one way is by using reconditioning methods.

What is reconditioning?

- *Methods which can be applied to matrices to reduce their condition number, while retaining underlying matrix structure.*
- Examples of methods:
  - Thresholding
  - Tapering
  - General regularisation methods.
- We will focus on a method that is used at the Met Office for numerical weather prediction. This method works by altering the eigenvalues of the original covariance matrix $\mathbf{R}$.

Reminder about variances and correlations

Want to also consider how variances and correlations are changed by the two methods. Let

$$\mathbf{R} = \mathbf{\Sigma} \mathbf{C} \mathbf{\Sigma}, \tag{2}$$

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:

$$\mathbf{\Sigma}(i,i) = \sqrt{\mathbf{R}(i,i)} \tag{3}$$

and

$$\mathbf{C}(i,j) = \frac{\mathbf{R}(i,j)}{\sqrt{\mathbf{R}(i,i)}\sqrt{\mathbf{R}(j,j)}}. \tag{4}$$

The ridge regression (RR) and minimum eigenvalue (ME) methods

Both methods improve the condition number of a covariance matrix by altering their eigenvalues to yield a reconditioned matrix with a user-defined condition number κ_{max}.

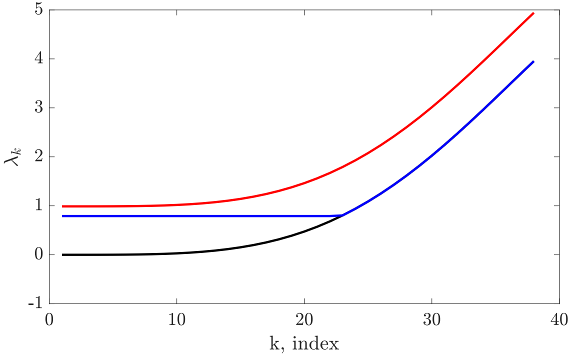


Figure: Illustration of recond methods: original spectrum (black), and spectrum reconditioned via ME and RR

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* $\mathbf{R}_{RR} = \mathbf{R} + \delta \mathbf{I}$

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 theoretically:

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

IASI matrix - framework

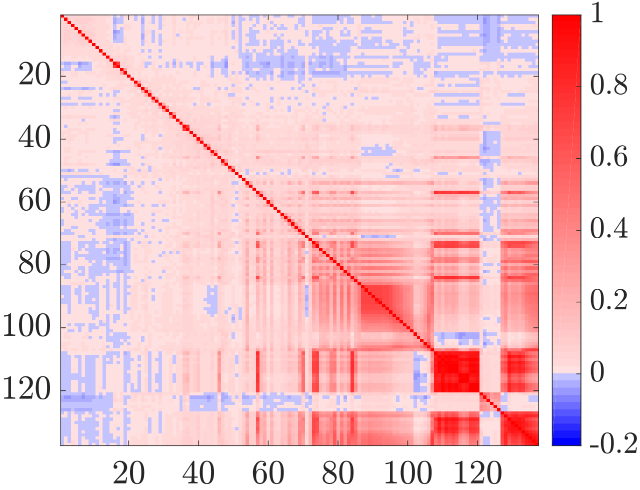
Interchannel correlations for a covariance matrix of satellite observation errors

- The UK Met Office diagnosed a correlated observation error covariance matrix in 2010.
- This was extremely ill-conditioned and crashed the system when used directly.
- Use 137 channels - **Original condition number:** 27703**.**

We study the impact of reconditioning in the **1D-Var** procedure.

- Run prior to every 4D-Var/forecast cycle.
- Assimilates each observation 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.

Diagnosed IASI correlation matrix



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

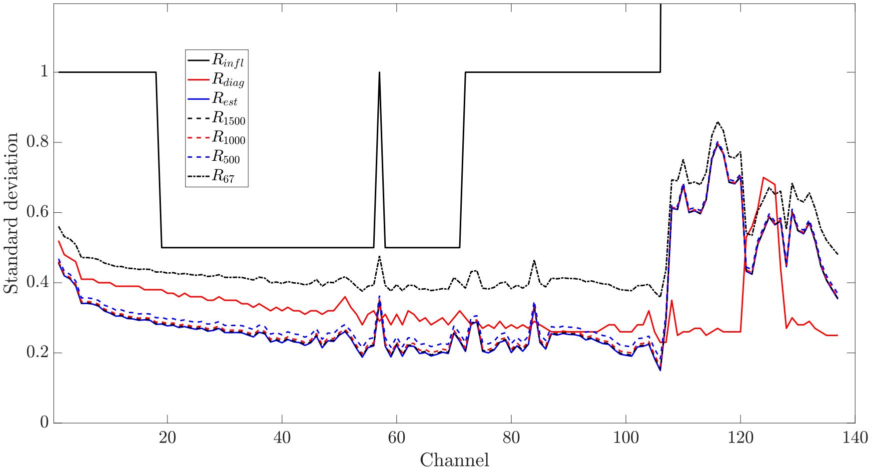


Figure: Standard deviation for each of the experiment choices

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

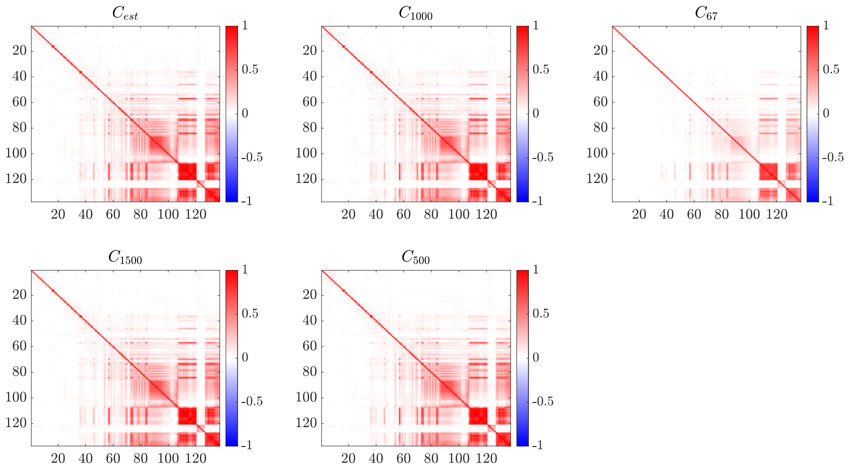


Figure: Changes to correlation with reconditioning for the correlated experiments

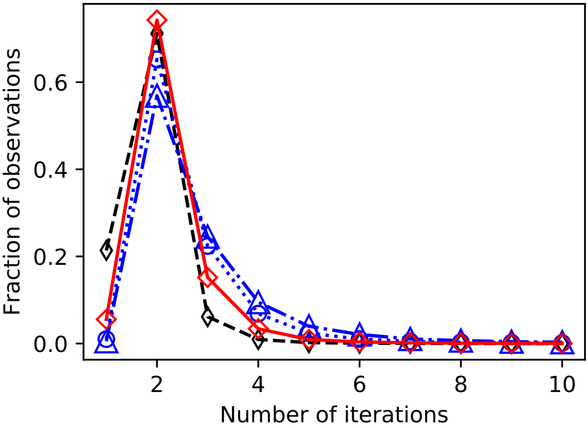


Figure: Number of iterations required to reach convergence of the 1D-Var minimization as a fraction of the total number of observations common to all choices of **R**. Symbols correspond to: $\triangle = \mathbf{R}_{diag}$, $\circ = \mathbf{R}_{est}$, $\Diamond = \mathbf{R}_{67}$, $\Diamond = \mathbf{R}_{infl}$.

Impact on temperature and humidity

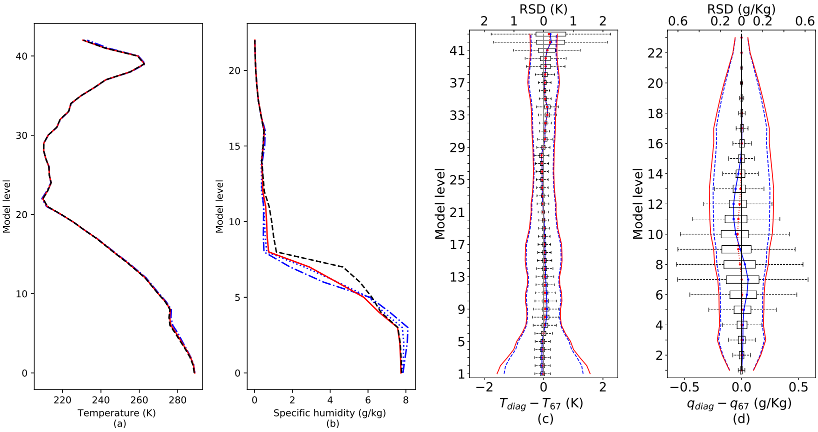


Figure: Example retrieved profiles of temperature (a) and specific humidity (b), and differences in retrievals between E_{diag} and E_{67} for temperature (c) and specific humidity (d) for 97330 observations.

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

Impact of reconditioning on variables not in 4D-Var state vector

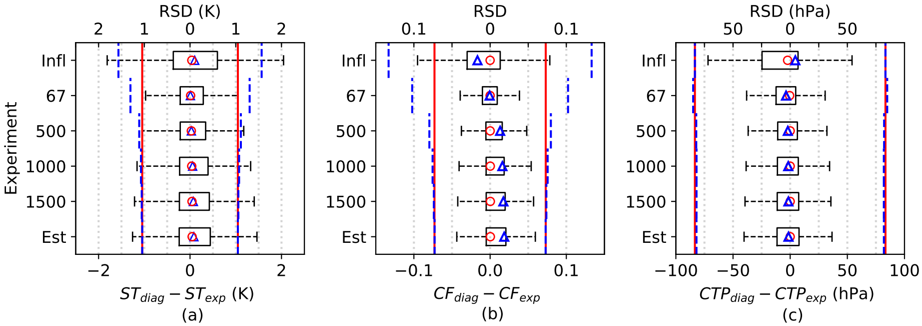


Figure: Change to estimates for skin temperature (left), cloud fraction (centre) and cloud top pressure (right) with reconditioning.

What about the outliers?

	E_{est}	E_{1500}	E_{1000}	E_{500}	E_{67}	E_{infl}
% outliers (CF)	23.9	24.0	24.2	24.6	25.3	21.4
% outliers (CTP)	22.8	22.8	23.0	22.9	21.4	18.8
% outliers (ST)	15.1	15.3	15.6	16.3	17.6	15.9
Max diff (ST (K))	21.67	21.12	21.14	22.38	21.03	26.83
Min diff (ST (K))	-33.52	-33.01	-32.14	-29.76	-23.82	-20.88

What about the outliers?

	E_{est}	E_{1500}	E_{1000}	E_{500}	E_{67}	E_{infl}
% outliers (CF)	23.9	24.0	24.2	24.6	25.3	21.4
% outliers (CTP)	22.8	22.8	23.0	22.9	21.4	18.8
% outliers (ST)	15.1	15.3	15.6	16.3	17.6	15.9
Max diff (ST (K))	21.67	21.12	21.14	22.38	21.03	26.83
Min diff (ST (K))	-33.52	-33.01	-32.14	-29.76	-23.82	-20.88

Think about **extreme outliers** - defined here as mean $\pm$ 0.25 $\times$ max difference

	E_{est}	E_{1500}	E_{1000}	E_{500}	E_{67}	E_{infl}
% $\|CF\| > 0.25$	4.9	4.7	4.4	3.9	3.2	7.5
% $\|CTP\| > 225hPa$	3.3	3.3	3.3	3.3	2.7	4.4
% $\|ST\| > 5K$	1.6	1.5	1.5	1.4	1.4	3.6

Conclusions

- Including correlated observation error in data assimilation methods is important for high-resolution forecasts and to make the best use of observation information.
- Convergence problems can be mitigated by using reconditioning methods.
- Tests in the Met Office 1D-Var system show that:
  - the ridge regression method improves convergence.
  - the quality control process is altered.
- Future work is needed to understand why some retrieved values change by a large amount.

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

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

References II

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

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