Incorporating Correlated Observation Errors in Variational Data Assimilation

Nancy Nichols*
Jemima Tabeart*, Sarah Dance*, Amos Lawless*, Joanne Waller*
Stefano Migliorini**, Fiona Smith**, Sue Ballard**
Observation Errors

• Observation errors assumed uncorrelated in data assimilation

• Observation errors in real data are found to be correlated (Stewart et al, 2009, 2013; Bormann et al, 2010; Waller et al, 2013, 2014a.)

• Using observation error correlations in data assimilation is shown to improve the analysis (Stewart et al, 2008, 2010, 2014; Weston, 2014.)
 Observation Errors

It is important to be able to account for observation error correlations:

- More of available data used (avoids thinning)
- More information content
- Better analysis accuracy
- Improved NWP forecast skill scores
Observation Errors

It is important to be able to account for observation error correlations:

• More of available data used (avoids thinning)
• More information content
• Better analysis accuracy
• Improved NWP forecast skill scores

Error correlations can be diagnosed by techniques such as Deroziers et al (DBCP) or Hollingsworth-Lönnberg
Problems for DA:

Diagnosed correlation matrices:
- Non-symmetric
- Variances too small
- Not positive-definite
- Very ill-conditioned
Problems for DA:

Diagnosed correlation matrices:
- Non-symmetric
- Variances too small
- Not positive-definite
- Very ill-conditioned

Result: including observation error correlations in 4DVar slows convergence catastrophically
Problems for DA:

Diagnosed correlation matrices:
- Non-symmetric
- Variances too small
- Not positive-definite
- Very ill-conditioned

Aim: to understand / enable the use of (diagnosed) correlated observation errors in DA
Minimize with respect to initial state $x_0$:

$$J(x_0) = \frac{1}{2} (x_0 - x_0^b)^T B^{-1} (x_0 - x_0^b) + \frac{1}{2} (\mathcal{H}(x_0) - y)^T R^{-1} (\mathcal{H}(x_0) - y)$$

- Background state $x_0^b$
- Observations $y$
- Observation operator $\mathcal{H}$
- Error covariance matrices $B$, $R$

The solution at the minimum, $x^a$, is the analysis.
Rate of convergence and accuracy of the solution are bounded in terms of the condition number of the Hessian of the variational cost function:

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

where $\lambda$ denotes an eigenvalue and the Hessian is:

$$S = B^{-1} + (H)^T R^{-1} H$$
Conditioning of Hessian

We can establish the following theorem:

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

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

Haben et al, 2011; Haben 2011, Tabeart, 2016, Tabeart et al, 2018
Conditioning of Hessian

We can establish the following theorem:

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, $\mathbf{S} = \mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}$,

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

Now the upper bound grows as $\frac{1}{\lambda_{\min}(\mathbf{R})}$ grows and depends also on the observation operator.

Haben et al, 2011; Haben 2011, Tabeart, 2016, Tabeart et al, 2018
Summary: Conditioning of the Problem

We find that the condition number of $S$ increases as:

- the observations become more accurate
- the observation spacing decreases
- the prior (background) becomes less accurate
- the prior error correlation length scales increase
- the observation error covariance becomes ill-conditioned

Haben et al, 2011; Haben 2011, Tabeart, 2016, Tabeart et al, 2018
Reconditioning R

To improve the conditioning of $R$ (and $S$) we alter the eigenstructure of $R$ so as to obtain a specified condition number for the modified covariance matrix by:

- **Ridge regression** - add constant to all diagonal elements.

- **Eigenvalue modification**: increase the smallest eigenvalues of $R$ to a threshold value that ensures the desired condition number, keeping the rest unchanged.

*Details given in talk by Jemima Tabeart.*
Operational Tests - Met Office

Experiments using the Met Office 1-DVar Observation Pre-processing System (OPS) for retrievals:

• Aim to test qualitative conclusions in an operational system.

• Focus on observations from IASI (Infrared Atmospheric Sounding Interferometer) instrument (on MetOp-A satellite). Note the observation operator is non-linear in this case.

• Investigate how changing the minimum eigenvalue of R affects the convergence of the iterations – we only show results using the ridge regression method.
Results - 1

The method of reconditioning is that described in Algorithm 1. Figure (a) depicts how the ridge regression method changes the eigenvalues of the covariance matrix $R_{137}$. Eigenvalues of the raw Desroziers $R_{unpre}^{137}$ are shown in blue, and those of $R_{RC}^{137}$ reconditioned so that $\kappa(R_{RC}) = 67$ are shown in green. Figure (b) compares the standard deviations of $R_{old}^{137}$ (red), $R_{unpre}^{137}$ (blue), and $R_{67}^{137}$ (green).
Results - 2

\( R_{raw} \) – Raw (symmetrised) matrix;
\( R_{ctrl} \) – Current MO diagonal;
\( R_{1500} \) – Reconditioned with \( \kappa = 1500 \);
\( R_{1000} \) – Reconditioned with \( \kappa = 1000 \);
\( R_{500} \) – Reconditioned with \( \kappa = 500 \);
\( R_{67} \) – Reconditioned with \( \kappa = 67 \);
\( R_{old} \) – Old MO diagonal matrix;
Shown are the retrieved temperature and humidity profiles for 4 different choices of $R$: $R_{\text{oper}}$, $R_{\text{unpre}}$, R500 and R67.
Summary: Operational Experiments

- Investigated the effects including observation error correlations in Met Office 1-D Var system.
- Impact on temperature retrievals was minimal, the impact on humidity retrievals much larger.
- Reducing the condition number of R reduces the number of iterations required for convergence.
- Decreasing the observation error variance increases the required number of iterations.

_Tabeart, 2016; Tabeart et al, in prep_
Future

Many more challenges left!
References

• Stewart LM, Dance SL, Nichols NK. 2013. Data assimilation with correlated observation errors: experiments with a 1-D shallow water model *Tellus A*, 65;19546 (14 pp). doi:10.3402/tellusa.v65i0.19546
• Waller JA. 2013, Using observations at different spatial scales in data assimilation for environmental prediction, PhD thesis, Dept of Mathematics & Statistics, University of Reading.