Characterising and modelling observation errors

Sarah L. Dance

With thanks to collaborators at UoR, DWD and the Met Office including the late Sue Ballard, Elisabeth Bauernschubert, John Eyre, Alison Fowler, Graeme Kelly, Amos Lawless, Stefano Migliorini, Nancy Nichols, Roland Potthast, David Simonin, Fiona Smith, Laura Stewart, Jemima Tabeart, Jo Waller....
Outline

What are observation errors?

Why estimate observation uncertainty?

How can we estimate observation uncertainty?

How can we implement correlated observation statistics?

Conclusions

References
What are observation errors?

In data assimilation, we consider the observation equation

\[ \mathbf{y} = H(\mathbf{x}) + \mathbf{\varepsilon}. \]

We assume \( \mathbf{\varepsilon} \) is unbiased, \( \mathbb{E}(\varepsilon) = 0 \), and has covariance \( \mathbf{R} \) such that

\[ R_{ij} = \mathbb{E}(\varepsilon_i \varepsilon_j). \]
Where do observation errors come from?

The error vector, $\varepsilon$, contains errors from four main sources: Janjić et al (2017)

- Instrument noise
- Observation pre-processing
- Observation operator error
- Scale mis-match
Problems dealing with observation error correlations

- Magnitude and character of observation error correlations largely unknown - can only be estimated in a statistical sense, not observed directly
Problems dealing with observation error correlations

- Magnitude and character of observation error correlations largely unknown - can only be estimated in a statistical sense, not observed directly
- Observations thinned spatially to reduce the correlations, and the $R$ matrix treated as diagonal.
What are observation errors?

Why estimate observation uncertainty?

How can we estimate observation uncertainty?

How can we implement correlated observation statistics?

Conclusions

References
Why do we want to estimate observation uncertainty?

- Currently only use 5% of some obs types due to thinning
- High resolution forecasting
- Improve analysis accuracy and forecast skill (e.g., Stewart et al. 2013; Weston et al., 2014)
- Changes to scales of observation information content in analysis depending on both the prior and observation error correlations (Fowler et al, 2018)
Estimating observation uncertainty

- Observation uncertainty depends on YOUR observation operator, model resolution etc and is state dependent (Waller et al., 2014)
- Approximations are still useful and can give improved forecast skill (Stewart et al, 2013)
- Approaches:
  - Error inventory/Metrological approach
  - Diagnosis from assimilation statistics (here)
Outline

What are observation errors?

Why estimate observation uncertainty?

How can we estimate observation uncertainty?

How can we implement correlated observation statistics?

Conclusions

References
DBCP diagnostic, Desroziers et al., (2005)

Use the background innovations and analysis residuals:

\[
d^o_b = y - \mathcal{H}(x^b),
\]

Taking the statistical expectation, and after some calculations...

\[
E[d^o_ad^o_b]^T = \tilde{R}(\tilde{H}B\tilde{B}^T + \tilde{R})^{-1}(HBH^T + R) = R_e,
\]
DBCP diagnostic, Desroziers et al., (2005)

Use the background innovations and analysis residuals:

\[ d_b^o = y - \mathcal{H}(x^b), \]
\[ d_a^o = y - \mathcal{H}(x^a). \]
DBCP diagnostic, Desroziers et al., (2005)

Use the background innovations and analysis residuals:

\[ \begin{align*}
    d_b^o &= y - \mathcal{H}(x^b), \\
    d_a^o &= y - \mathcal{H}(x^a).
\end{align*} \]

Taking the statistical expectation, and after some calculations...

\[
E[\mathbf{d}_a^o \mathbf{d}_b^o^T] = \mathbf{\tilde{R}} (\mathbf{HBH}^T + \mathbf{\tilde{R}})^{-1} (\mathbf{HBH}^T + \mathbf{R}) = \mathbf{R}^e,
\]

where

- \( \mathbf{R}^e \) is the estimated observation error covariance matrix.
DBCP diagnostic, Desroziers et al., (2005)

Use the background innovations and analysis residuals:

\[ \mathbf{d}_b^o = \mathbf{y} - \mathcal{H}(\mathbf{x}_b^b), \]
\[ \mathbf{d}_a^o = \mathbf{y} - \mathcal{H}(\mathbf{x}_a^a). \]

Taking the statistical expectation, and after some calculations...

\[ E[\mathbf{d}_a^o \mathbf{d}_b^o^T] = \tilde{\mathbf{R}}(\mathbf{HBH}^T + \tilde{\mathbf{R}})^{-1}(\mathbf{HBH}^T + \mathbf{R}) = \mathbf{R}^e, \]

where

- \( \mathbf{R}^e \) is the estimated observation error covariance matrix
- \( \mathbf{B} \) and \( \mathbf{R} \) are the exact background and observation covariance matrices.
DBCP diagnostic, Desrozières et al., (2005)

Use the background innovations and analysis residuals:

\[
\begin{align*}
    d^o_b & = y - \mathcal{H}(x^b), \\
    d^o_a & = y - \mathcal{H}(x^a).
\end{align*}
\]

Taking the statistical expectation, and after some calculations...

\[
E[d^o_a d^o_b^T] = \tilde{R}(HBH^T + \tilde{R})^{-1}(HBH^T + R) = R^e,
\]

where

- \(R^e\) is the estimated observation error covariance matrix
- \(B\) and \(R\) are the exact background and observation covariance matrices.
- \(\tilde{R}\) and \(\tilde{B}\) are the assumed statistics used in the assimilation.
DBCP diagnostic, Desroziers et al., (2005)

Use the background innovations and analysis residuals:

\[
\begin{align*}
\mathbf{d}_b^o &= \mathbf{y} - \mathcal{H}(\mathbf{x}^b), \\
\mathbf{d}_a^o &= \mathbf{y} - \mathcal{H}(\mathbf{x}^a).
\end{align*}
\]

Taking the statistical expectation, and after some calculations...

\[
E[\mathbf{d}_a^o \mathbf{d}_b^o^T] = \tilde{\mathbf{R}}(\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}})^{-1}(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}) = \mathbf{R}^e,
\]

where

- \( \mathbf{R}^e \) is the estimated observation error covariance matrix.
- \( \mathbf{B} \) and \( \mathbf{R} \) are the exact background and observation covariance matrices.
- \( \tilde{\mathbf{R}} \) and \( \tilde{\mathbf{B}} \) are the assumed statistics used in the assimilation.

If \( \tilde{\mathbf{R}} = \mathbf{R} \) and \( \tilde{\mathbf{B}} = \mathbf{B} \), then

\[
E[\mathbf{d}_a^o \mathbf{d}_b^o^T] = \mathbf{R}.
\]
How well does the diagnostic work in practice?

- Gives plausible results e.g. Stewart et al., 2009, 2014. UK Met Office
  global 4D-Var, 139 channels of IASI data (clear sky, sea surface observations only), 2073 observations, 17 July 2008.

- Non-symmetric structure
- Index 86-108 are surface channels; 109-121, 122-127, 128-138 are sensitive to water vapour

- Including interchannel correlations has improved the NWP skill score (e.g., Weston et al 2014).
- **BUT** results are sensitive to the stats used in the assimilation (Waller et al, 2016a).
Doppler radar winds and Met Office UKV

Each radar beam produces observations of radial velocity out to a range of 100km with measurements taken:

- Every 75m along the beam.
- Every degree.
- At five different elevation angles.
- Superobbed to $3^\circ$ by 3km.
- Thinned to 6km.
Horizontal Correlations, sensitivity to $\tilde{B}$
Waller et al. (2016b)

<table>
<thead>
<tr>
<th>Case</th>
<th>B statistics</th>
<th>Superobs</th>
<th>Observation operator</th>
<th>Standard deviation (m/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>New Bg</td>
<td>New</td>
<td>Yes</td>
<td>Old</td>
<td>1.97</td>
</tr>
<tr>
<td>Old Bg</td>
<td>Old</td>
<td>Yes</td>
<td>Old</td>
<td>1.57</td>
</tr>
</tbody>
</table>

- Increasing variance and lengthscale in $\tilde{B}$ reduces variance and lengthscale in diagnosed $R^e$.
- Consistent with Waller et al (2016a) theory.
Outline

What are observation errors?

Why estimate observation uncertainty?

How can we estimate observation uncertainty?

How can we implement correlated observation statistics?

Conclusions

References
Using diagnosed covariances

Problems: Diagnosed covariances typically
- Not symmetric
- Not positive definite
- Variances too small
- Ill-conditioned

Can prevent convergence of variational minimization (Weston et al. 2014)
Convergence of minimization

The sensitivity and accuracy of the solution of the minimization depend on the condition number of the Hessian

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

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

$$S = B^{-1} + H^T R^{-1} H.$$

Tabeart et al. (2018) showed that $\kappa(S)$ increases with $\lambda_{\text{min}}(R)$. 
Reconditioning $\mathbf{R}$

To improve the conditioning of $\mathbf{R}$ (and $\mathbf{S}$) we alter the eigenstructure of $\mathbf{R}$ so as to obtain a specified condition number for the modified covariance matrix by e.g.,

- **Ridge regression** - add constant to all diagonal elements.
- **Eigenvalue modification**: increase the smallest eigenvalues of $\mathbf{R}$ to a threshold value that ensures the desired condition number, keeping the rest unchanged.
Reconditioning results with IASI inter-channel error matrix (Tabeart et al, submitted)

- Ridge regression method increases standard deviations more than minimum eigenvalue method
- Ridge regression method decreases correlations, but minimum eigenvalue method has non-uniform behaviour

![Standard deviation vs. Channel number](image)

**Figure 4.2.** Standard deviations $\Sigma$ (black solid), $\Sigma_{RR}$ (red dashed) and $\Sigma_{ME}$ (blue dot-dashed) for $\kappa_{max} = 100$. 
Experiments with Ridge Regression in 1D-Var - Tabeart et al.

Figure 1. Number of iterations required for convergence of the minimization of the 1D-Var cost function as a fraction of the total number of observations common to all choices of $R$. Symbols correspond to: $\square = R_{ctrl}$, $\circ = R_{raw}$, $\nabla = R_{1500}$, $\triangle = R_{1000}$, $+ = R_{500}$, $\star = R_{07}$, $\diamond = R_{old}$.

- Reconditioning increases convergence speed
Spatial correlations

We need to be able to compute the matrix-vector product

$$R^{-1}v.$$

This might require expensive communication between processors.
Expt: Doppler radar wind assimilation  
(Simonin et al, submitted)

- Assume only horizontal correlations within a family
- $R$ is derived on-the-fly (different observations each assimilation)
- Correlation matrix is determined by calculating the distance between each pair of observations in the family

$$C_{ij} = \exp \left( -\frac{D_{ij}}{L_r} \right)$$

- Lengthscale determined by fitting to diagnosed horizontal correlations
Experiments

Three experiments run for 20 days (3 hourly cycling 3D-Var, UKV 1.5km model)

**Control**: 6km thinning with diagonal $\mathbf{R}$ ($\sim$ 2000 radar obs per cycle)

**Corr-R-6km**: 6km thinning with correlated $\mathbf{R}$ ($\sim$ 2000 radar obs per cycle)

**Corr-R-3km**: 6km thinning with correlated $\mathbf{R}$ ($\sim$ 8000 radar obs per cycle)
Results

- No significant difference in iteration count or wall-clock time
- Corr-R-3km increments are smaller scale and smaller magnitude
- Parameters for experiments have not been tuned, but most O-Bs show a small benefit from the introduction of correlations.

\[
\frac{\sigma_{O-B,exp}}{\sigma_{O-B,ctrl}} - 1 \text{[\%]}
\]

![Graph showing comparison between Corr-R-6km and Corr-R-3km increments]
O-B Forecast skill cont

(b)

Corr-R-6km  Corr-R-3km

Aircraft RH  Aircraft theta  sonde RH  sonde theta  GroundGPS  Surface temp  Surface pressure  Surface rh

(c)

MetOp2 (A) IASI  MetOp2 (B) AMSUB  MetOp1 (B) IASI  MetOp1 (B) AMSUB  NOAA18 AMSUB  MSG3 (MET)  SEVIRI/CL  JPSS0 (NPP) CRIS  EOS2 (Aqua) AMSU  NOAA19 AMSUB
Conclusions

- It is important to be able to account for observation error correlations
  - Avoid thinning (high resolution forecasting)
  - Improved forecast skill score

- First we need to estimate correlations
  - Desroziers et al (2005) diagnostic can be used with caution
  - Can understand sensitivity to the assumed stats in the assimilation (Waller et al. 2016a)
  - Can help us to understand sources of correlations (e.g., Waller et al 2016b)

- Then we need to be able to account for the errors in the assimilation
  - Sample matrices need reconditioning
  - Appropriate software needs to be in place
Conclusions

- It is important to be able to account for observation error correlations
  - Avoid thinning (high resolution forecasting)
  - Improved forecast skill score
- First we need to estimate correlations
  - Desroziers et al (2005) diagnostic can be used with caution
  - Can understand sensitivity to the assumed stats in the assimilation (Waller et al. 2016a)
  - Can help us to understand sources of correlations (e.g., Waller et al 2016b)
Conclusions

• It is important to be able to account for observation error correlations
  ◦ Avoid thinning (high resolution forecasting)
  ◦ Improved forecast skill score

• First we need to estimate correlations
  ◦ Desroziers et al (2005) diagnostic can be used with caution
  ◦ Can understand sensitivity to the assumed stats in the assimilation (Waller et al. 2016a)
  ◦ Can help us to understand sources of correlations (e.g., Waller et al 2016b)

• Then we need to be able to account for the errors in the assimilation
  ◦ Sample matrices need reconditioning
  ◦ Appropriate software needs to be in place
References


doi:10.1002/qj.3130


www.reading.ac.uk/web/FILES/maths/obs_error_IASI_radiance.pdf.


