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

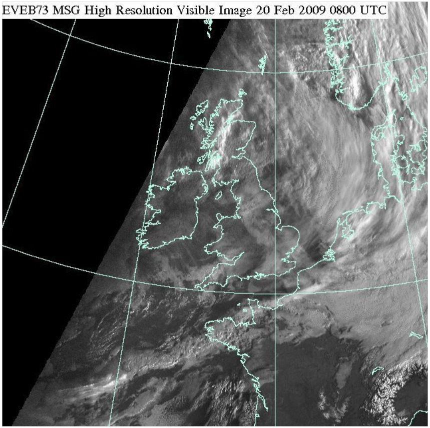
What are observation errors?

In data assimilation, we consider the observation equation

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

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

$$\mathbf{R}_{ij} = \mathbb{E}(\varepsilon_i \varepsilon_j).$$

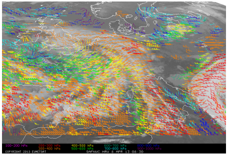


Where do observation errors come from?

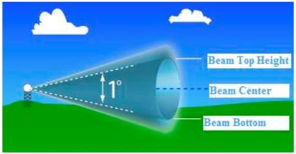
The error vector, ε, contains errors from four main sources: Janjić et al (2017)

Instrument noise

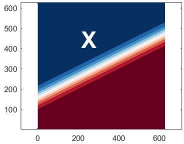
Observation pre-processing

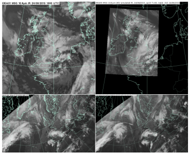


Observation operator error



Scale mis-match



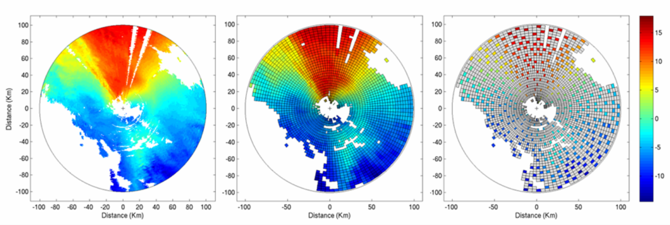


Problems dealing with observation error correlations

- 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.



Outline

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:

$$\mathbf{d}_b^o = \mathbf{y} - \mathcal{H}(\mathbf{x}^b),$$

DBCP diagnostic, Desroziers et al., (2005)

Use the background innovations and analysis residuals:

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

DBCP diagnostic, Desroziers et al., (2005)

Use the background innovations and analysis residuals:

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

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

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

where

- $\mathbf{R}^e$ is the estimated observation error covariance matrix

DBCP diagnostic, Desroziers et al., (2005)

Use the background innovations and analysis residuals:

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

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

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

where

- $\mathbf{R}^e$ is the estimated observation error covariance matrix
- $\mathbf{B}$ and $\mathbf{R}$ are the exact background and observation covariance matrices.

DBCP diagnostic, Desroziers et al., (2005)

Use the background innovations and analysis residuals:

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

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

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

where

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

DBCP diagnostic, Desroziers et al., (2005)

Use the background innovations and analysis residuals:

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

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

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

where

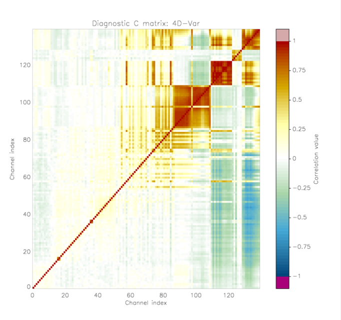
- $\mathbf{R}^e$ is the estimated observation error covariance matrix
- $\mathbf{B}$ and $\mathbf{R}$ are the exact background and observation covariance matrices.
- $\widetilde{\mathbf{R}}$ and $\widetilde{\mathbf{B}}$ are the assumed statistics used in the assimilation.

If $\widetilde{\mathbf{R}} = \mathbf{R}$ and $\widetilde{\mathbf{B}} = \mathbf{B}$, 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.



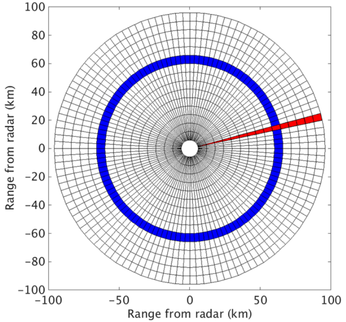
  - Non-symmetric structure
  - Index 86-108 are surface channels; 109-121, 122-127, 128-138 are sensitive to water vapour

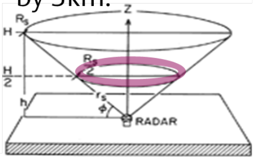
- Including interchannel correlations has improved the NWP skill score (e.g., Weston et al 2014).
- BUT results are sensitive to the stats used in the assimilation (Waller et al, 2016a).

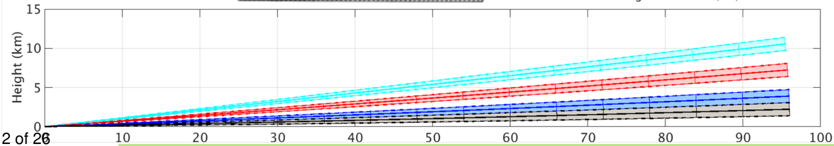
Doppler radar winds and Met Office UKV

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 3º by 3km.
- Thinned to 6km.



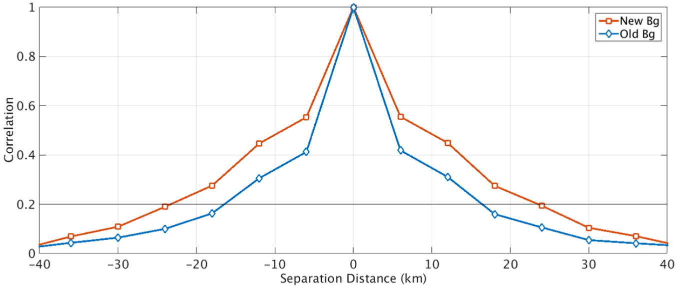




Horizontal Correlations, sensitivity to $\widetilde{\mathbf{B}}$

Waller et al. (2016b)

Case	B statistics	Superobs	Observation operator	Standard deviation (m/s)
New Bg	New	Yes	Old	1.97
Old Bg	Old	Yes	Old	1.57



- Increasing variance and lengthscale in $\widetilde{\mathbf{B}}$ reduces variance and lengthscale in diagnosed $\mathbf{R}^e$.
- 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

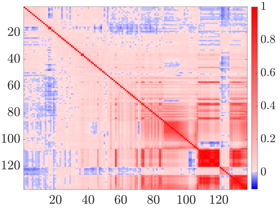


Figure 0.1. *Diagnosed correlation matrix for IASI for the subset of 137 channels with non-zero off-diagonal entries.*

Diagnosed interchannel observation error correlations for IASI (for the Met Office global model)

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

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

where λ denotes the eigenvalue and the Hessian is

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

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

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 e.g.,

- 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.

Reconditioning results with IASI inter-channel error matrix (Tabeart et al, submitted)

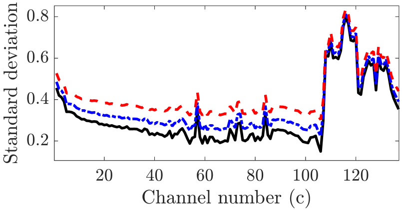


Figure 4.2. *Standard deviations* $\mathbf{\Sigma}$ *(black solid),* $\mathbf{\Sigma}_{RR}$ *(red dashed) and* $\mathbf{\Sigma}_{ME}$ *(blue dot-dashed) for* $\kappa_{max} = 100$.

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

Experiments with Ridge Regression in 1D-Var - Tabeart et al.

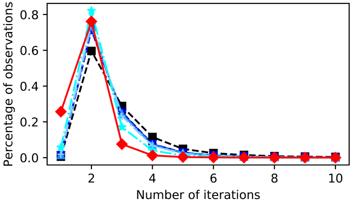


Figure 1. Number of iterations required for convergence of the minimization of the 1D-Var cost function as a fraction of the total number of observations common to all choices of **R**. Symbols correspond to: $\square = \mathbf{R}_{ctrl}$, $\circ = \mathbf{R}_{raw}$, $\triangledown = \mathbf{R}_{1500}$, $\triangle = \mathbf{R}_{1000}$, $+ = \mathbf{R}_{500}$, $\star = \mathbf{R}_{67}$, $\diamond = \mathbf{R}_{old}$.

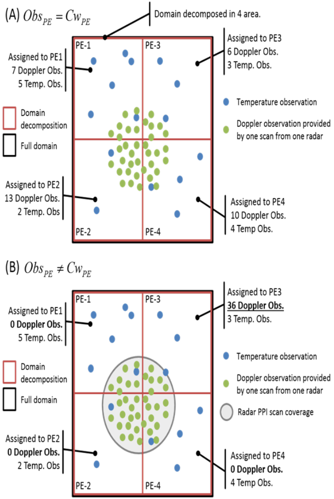
- Reconditioning increases convergence speed

Spatial correlations

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

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

This might require expensive communication between processors.

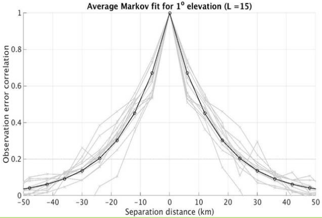


Expt: Doppler radar wind assimilation (Simonin et al, submitted)

- 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 **R** ($\sim$ 2000 radar obs per cycle)

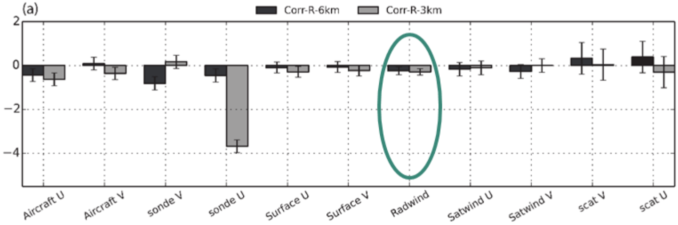
Corr-R-6km: 6km thinning with correlated **R** ($\sim$ 2000 radar obs per cycle)

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

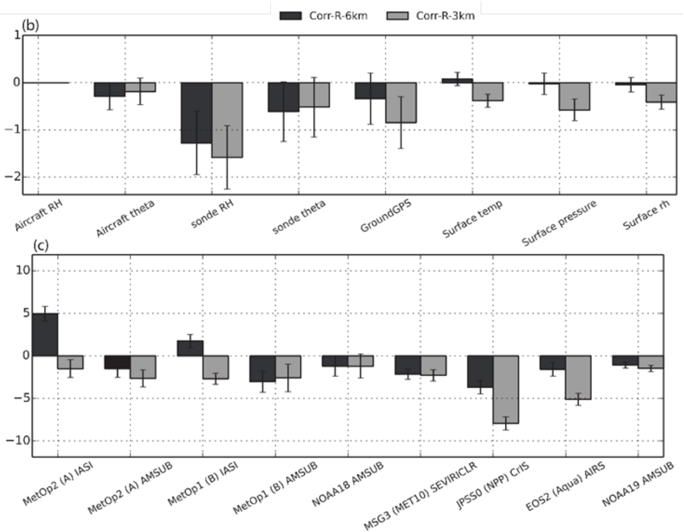
Results

- 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.

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



O-B Forecast skill cont



Conclusions

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

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

G. Desroziers, L. Berre, B. Chapnik, and P. Poli. Diagnosis of observation, background and analysis-error statistics in observation space. *Q.J.R. Meteorol. Soc.*, 131:3385-3396, 2005.

Fowler, A. M., Dance, S. L. and Waller, J. A. (2018), On the interaction of observation and prior error correlations in data assimilation. Q.J.R. Meteorol. Soc., 144: 48-62. doi:10.1002/qj.3183

Janjić, T., Bormann, N., Bocquet, M., Carton, J. A., Cohn, S. E., Dance, S. L., Losa, S. N., Nichols, N. K., Potthast, R., Waller, J. A. and Weston, P. (2017), On the representation error in data assimilation. *Q.J.R. Meteorol. Soc.*. doi:10.1002/qj.3130

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

L. M. Stewart, J. Cameron, S. L. Dance, S. English, J. R. Eyre, and N. K. Nichols. Observation error correlations in IASI radiance data. Technical report, University of Reading, 2009. Mathematics reports series, www.reading.ac.uk/web/FILES/maths/obs_error_IASI_radiance.pdf.

L. M. Stewart, S. L. Dance, N. K. Nichols, J. R. Eyre, and J. Cameron. Estimating interchannel observation-error correlations for IASI radiance data in the Met Office system. *Quarterly Journal of the Royal Meteorological Society*, 140:1236-1244, 2014. doi: 10.1002/qj.2211.

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* doi:10.1002/nla.2165

Waller, J. A., Dance, S. L., Lawless, A. S., Nichols, N. K. and Eyre, J. R. (2014), Representativity error for temperature and humidity using the Met Office high-resolution model. *Q.J.R. Meteorol. Soc.*, doi: 10.1002/qj.2207

Waller, J. A., Dance, S. L. and Nichols, N. K. (2016a) Theoretical insight into diagnosing observation error correlations using observation-minus-background and observation-minus-analysis statistics. *Q.J.R. Meteorol. Soc.* doi: 10.1002/qj.2661

Waller, J. A., Simonin, D., Dance, S. L., Nichols, N. K. and Ballard, S. P. (2016b) Diagnosing observation error correlations for Doppler radar radial winds in the Met Office UKV model using observation-minus-background and observation-minus-analysis statistics. *Monthly Weather Review*. doi: 10.1175/MWR-D-15-0340.1

P. P. Weston, W. Bell, and J. R. Eyre. Accounting for correlated error in the assimilation of high-resolution sounder data. *Quarterly Journal of the Royal Meteorological Society*, 2014. doi: 10.1002/qj.2306.