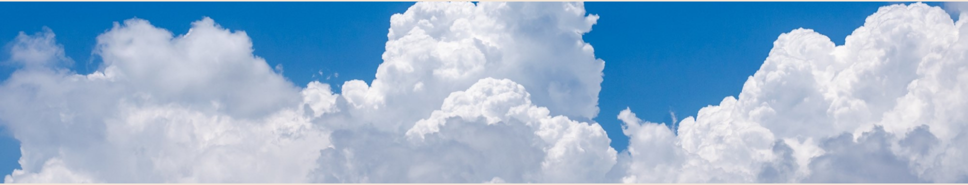
Observation uncertainty in data assimilation

Sarah L. Dance 
email: s.l.dance@reading.ac.uk twitter: @DrSarahDance



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

UKRI Natural Environment Research Council | University of Reading | UKRI Engineering and Physical Sciences Research Council

Outline

What are observation errors?

Why estimate observation uncertainty?

How can we estimate observation uncertainty?

What are the pitfalls?

What are the possibilities?

Conclusions

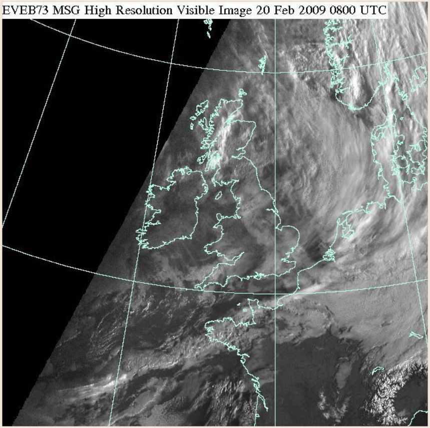
What are observation errors?

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

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

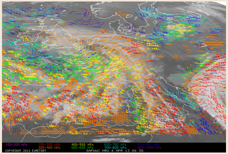


Where do observation errors come from?

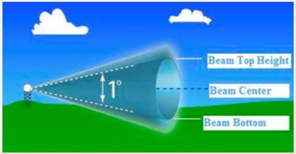
The error vector, ε, contains errors from four main sources: Janjić et al (2017)

Instrument noise

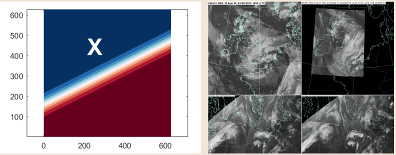
Observation pre-processing



Observation operator error

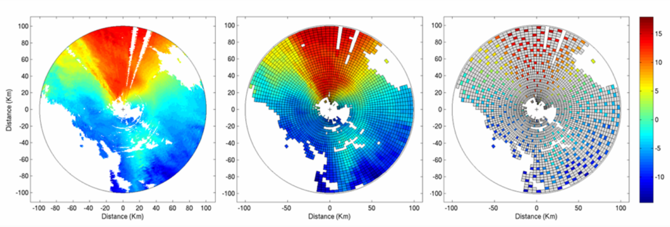


Scale mis-match



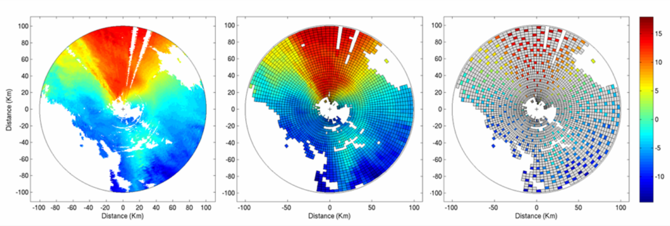
Why do we want to estimate observation uncertainty?

- Only use 5% of some obs types due to thinning



Why do we want to estimate observation uncertainty?

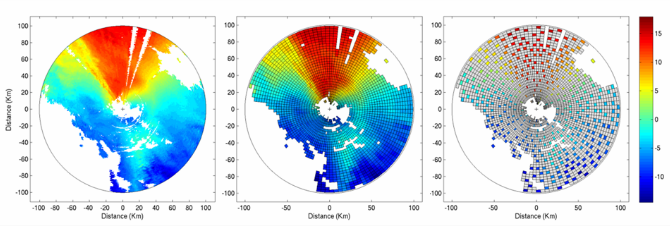
- Only use 5% of some obs types due to thinning



- Improve analysis accuracy and forecast skill (e.g., Stewart et al. 2013; Weston et al., 2014)

Why do we want to estimate observation uncertainty?

- Only use 5% of some obs types due to thinning



- 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

- In DA, observation uncertainty depends on YOUR observation operator, model resolution etc and is state dependent (Waller et al., 2014; Janjić et al, 2018)
- Approximations are still useful and can give improved forecast skill (Healy and White, 2005; Stewart et al, 2013)

How can we estimate observation uncertainty?

- Error inventory/Metrological approach

How can we estimate observation uncertainty?

- Error inventory/Metrological approach
- Collocation with other observations (but rep. error?)

How can we estimate observation uncertainty?

- Error inventory/Metrological approach
- Collocation with other observations (but rep. error?)
- Diagnosis from assimilation (review by Tandeo et al, 2020)
  - Moment based methods (e.g., using innovation and residual statistics, Desroziers et al, 2005)

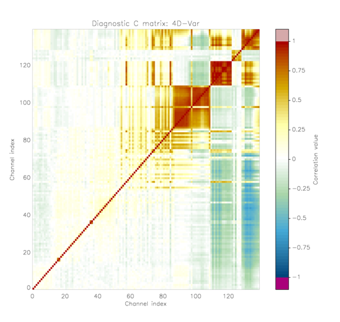
How can we estimate observation uncertainty?

- Error inventory/Metrological approach
- Collocation with other observations (but rep. error?)
- Diagnosis from assimilation (review by Tandeo et al, 2020)
  - Moment based methods (e.g., using innovation and residual statistics, Desroziers et al, 2005)
  - Likelihood based methods (e.g., expectation maximization, Pulido et al, 2018)

DBCP diagnostic (Desroziers et al 2005)

- Easy to compute from standard innovations and analysis residuals
- Proven useful in NWP

Early IASI example (Stewart et al., 2009, 2014.)



- Non-symmetric structure

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:

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

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
- **B** and **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...

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

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.

DBCP diagnostic, Desroziers et al., (2005)

Use the background innovations and analysis residuals:

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

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

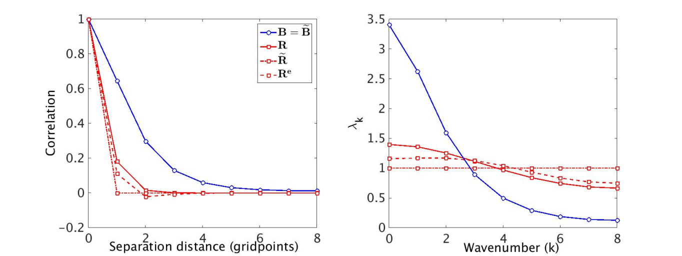
$$E[\mathbf{d}_a^o \mathbf{d}_b^{o\,T}] = \mathbf{R}.$$

What are the pitfalls?

Sensitivity to Assumed Statistics (Waller et al, 2016 QJ)

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

Example: True background error stats, $\widetilde{\mathbf{B}} = \mathbf{B}$; diagonal $\widetilde{\mathbf{R}}$

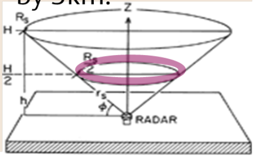


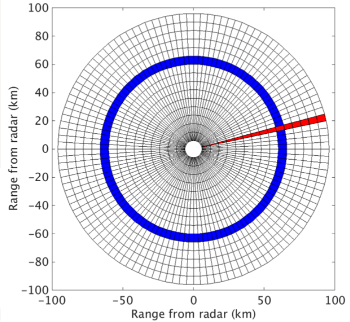
$\mathbf{R}^e$ has an underestimated variance and correlation lengthscale, but is a better approximation than $\widetilde{\mathbf{R}}$.

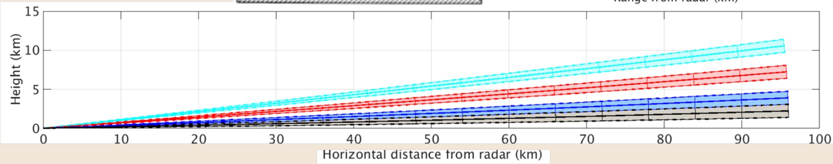
Doppler radar winds and Met Office UKV

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.



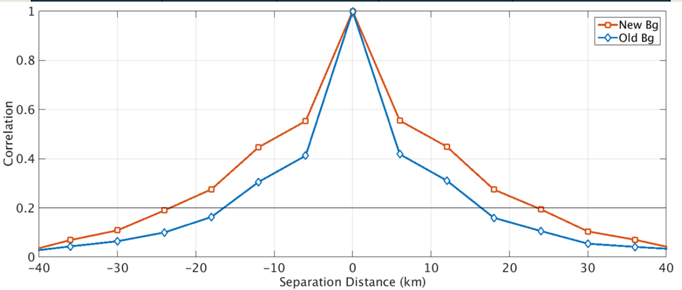




Horizontal Correlations, sensitivity to $\widetilde{\mathbf{B}}$

Waller et al. (2016) MWR

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 (2016) QJ theory.

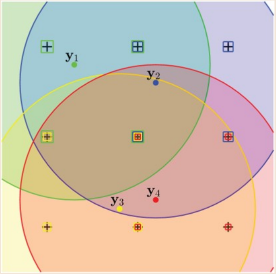
DBCP and Local DA (Waller et al, 2017)

- DBCP does not always give the right answers. Must only calculate with the right set of points.

Regions of observation influence

The region of influence of an observation is the set of analysis states that are updated in the assimilation using the observation.

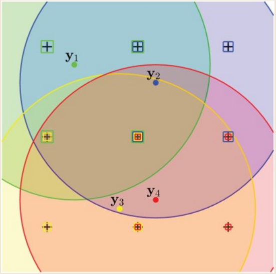
Grid points (pluses) and observations (dots), with observations coloured with corresponding regions of observation influence (shaded coloured circles).



DBCP and Local DA Cont(Waller et al, 2017)

The domain of dependence of an observation y_i is the set of elements of the model state that are used to calculate the model equivalent of y_i

Example: The coloured squares around grid points select the points that would be utilized by the observation operator for the observation of the corresponding colour.

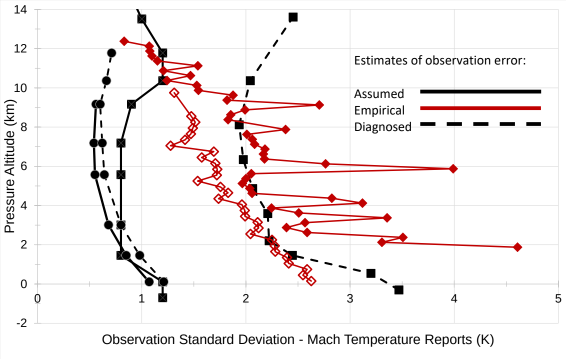


The correlation between the errors of observations y_i and y_j can be estimated using the DBCP diagnostic only if the domain of dependence for observation y_i lies within the region of influence of observation y_j.

What are the possibilities?

Comparison of approaches (Mirza et al, 2021)

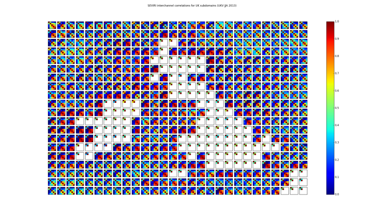
- Mode-S EHS temperatures - errors from lack of precision in Mach number
- Diagnosed std (black-dashed-squares) compare well with metrological estimates (red diamonds)



Identifying sources of error - Examples

Waller et al (2016) Rem. sens.

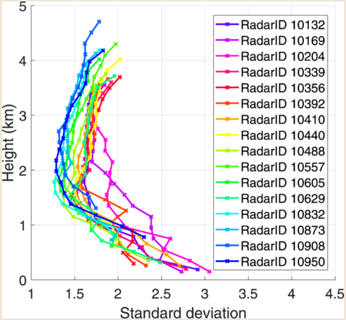
SEVIRI interchannel error covariances over different subdomains



Land-sea QC issue

Bauernschubert et al (2019)

Doppler radar wind error std



Radars 10169 and 10204 contaminated by wind turbines and ship tracks

Using diagnosed covariances in operational systems

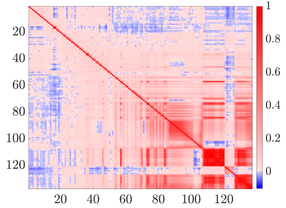


Figure 0.1. *Diagnosed correlation matrix for IASI for the subset of 137 channels with non-zero off-diagonal entries.*

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 $\lambda_{\mathsf{min}}(\mathbf{R})$.

Tabeart et al. (2021) showed similar results with pre-conditioned form (control var transform).

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, 2020)

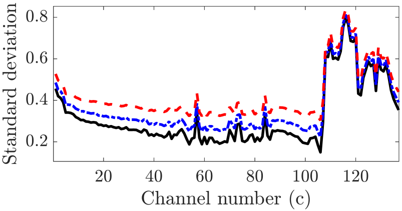


Figure 4.2. *Standard deviations* $\boldsymbol{\Sigma}$ *(black solid),* $\boldsymbol{\Sigma}_{RR}$ *(red dashed) and* $\boldsymbol{\Sigma}_{ME}$ *(blue dot-dashed) for* $\kappa_{max} = 100$.

- 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 Met Office 1D-Var (Tabeart et al. 2020)

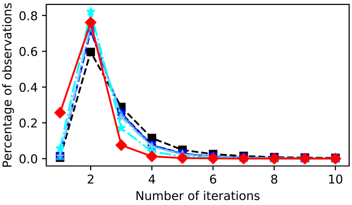


Figure 1. Number of iterations required for convergence of the minimization of the 1D-Var cost function as a fraction of the total number of observations common to all choices of **R**. Symbols correspond to: $\square = \mathbf{R}_{ctrl}$, $\circ = \mathbf{R}_{raw}$, $\triangledown = \mathbf{R}_{1500}$,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

- Reconditioning increases convergence speed

Experiments with Ridge Regression in Met Office 1D-Var (Tabeart et al. 2020)

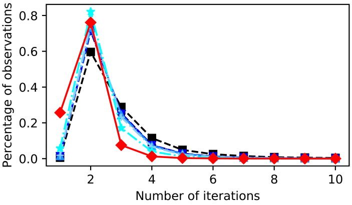


Figure 1. Number of iterations required for convergence of the minimization of the 1D-Var cost function as a fraction of the total number of observations common to all choices of **R**. Symbols correspond to: $\square = \mathbf{R}_{ctrl}$, $\circ = \mathbf{R}_{raw}$, $\nabla = \mathbf{R}_{1500}$, $\triangle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

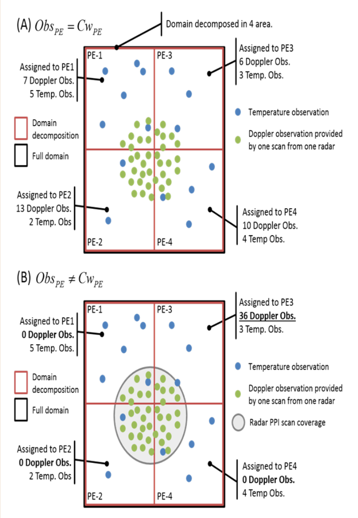
- Reconditioning increases convergence speed
- Lots of examples of improved forecast skill from taking account of interchannel error covariances (Met Office, ECMWF, NRL, ECCC, NASA, NCEP, Meteo France...)

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.

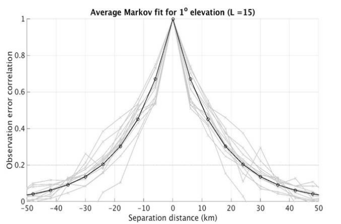


Doppler radar wind assimilation (Simonin et al, 2019)

- 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

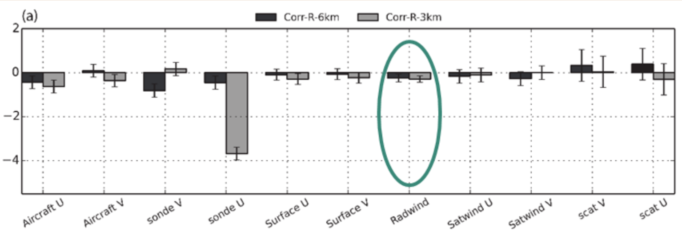
Results

- No significant difference in iteration count or wall-clock time
- Corr-R-3km increments are smaller scale and smaller magnitude

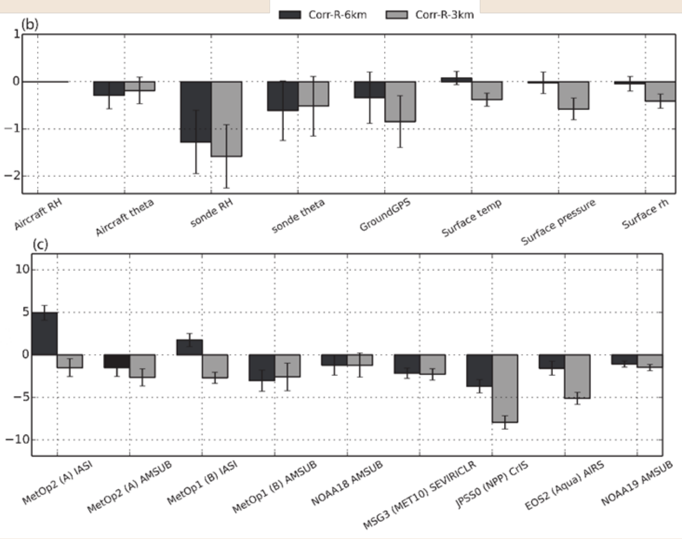
Results

- 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



Work underway to implement in 4D-Var

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 to deal efficiently with full matrices

References Page 1

Bathmann, K., 2018. Justification for estimating observation-error covariances with the Desroziers diagnostic. Quarterly Journal of the Royal Meteorological Society, 144(715), pp.1965-1974.

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

Healy, S.B. and White, A.A., 2005. Use of discrete Fourier transforms in the 1D-Var retrieval problem. Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography, 131(605), pp.63-72.

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. and Weston, P., 2018. On the representation error in data assimilation. Quarterly Journal of the Royal Meteorological Society, 144(713), pp.1257-1278.

Menard R. 2016. Error covariance estimation methods based on analysis residuals: theoretical foundation and convergence properties derived from simplified observation networks. Quarterly Journal of the Royal Meteorological Society 142: 257-273.

Mirza, A.K., S.L. Dance, G. G. Rooney, D. Simonin, E. K. Stone and J. A. Waller Comparing diagnosed observation uncertainties with independent estimates: a case study using aircraft-based observations and a convection-permitting data assimilation system. Submitted

Pulido, M., P. Tandeo, M. Bocquet, A. Carrassi, and M. Lucini, 2018: Stochastic parameterization identification using ensemble Kalman filtering combined with maximum likelihood methods. Tellus A: Dynamic Meteorology and Oceanography, 70 (1), 1442 099.

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.

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

References Page 2

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

Jemima M. Tabeart Sarah L. Dance Amos S. Lawless Stefano Migliorini Nancy K. Nichols Fiona Smith and Joanne A. Waller (2020) The impact of using reconditioned correlated observation-error covariance matrices in the Met Office 1D-Var system. QJR Meteorol Soc.146: 1372-1390. doi:10.1002/qj.3741

Jemima M. Tabeart, Sarah L. Dance, Amos S. Lawless, Nancy K. Nichols & Joanne A. Waller (2020) Improving the condition number of estimated covariance matrices, Tellus A: Dynamic Meteorology and Oceanography, 72:1, 1-19 doi: 10.1080/16000870.2019.1696646

Jemima M. Tabeart, Sarah L. Dance, Amos S. Lawless, Nancy K. Nichols, Joanne A. Waller (2021) The conditioning of least squares problems in preconditioned variational data assimilation, submitted

Tandeo, P., P. Ailliot, M. Bocquet, A. Carrassi, T. Miyoshi, M. Pulido, and Y. Zhen, 2020: A Review of Innovation-Based Methods to Jointly Estimate Model and Observation Error Covariance Matrices in Ensemble Data Assimilation. Mon. Wea. Rev., 148, 3973-3994.

Waller JA, Ballard SP, Dance SL, Kelly G, Nichols NK, Simonin D. 2016. Diagnosing horizontal and inter-channel observation-error correlations for SEVIRI observations using observation-minus-background and observation-minus-analysis statistics. Remote Sens. 8: 581, doi: 10.3390/rs8070581.

Waller, J. A., Dance, S. L. and Nichols, N. K. (2016) 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

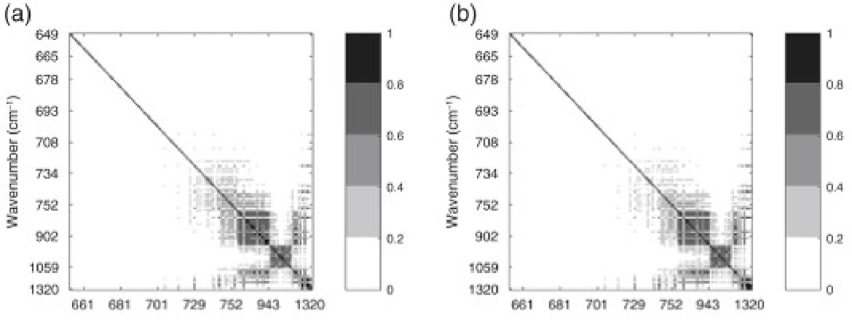
Waller, J. A., Simonin, D., Dance, S. L., Nichols, N. K. and Ballard, S. P. (2016b) Diagnosing observation error correlations for Doppler radar radial winds in the Met Office UKV model using observation-minus-background and observation-minus-analysis statistics. *Monthly Weather Review*. doi: 10.1175/MWR-D-15-0340.1

Waller, J.A., Dance, S.L. and Nichols, N.K. (2017), On diagnosing observation-error statistics with local ensemble data assimilation. Q.J.R. Meteorol. Soc., 143: 2677-2686. doi:10.1002/qj.3117

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.

Iteration

- Iteration converges to the correct estimate only when assumed $\widetilde{\mathbf{B}}$ is correct (Menard, 2016; Bathmann, 2018)
- More often, the first iterate is used. Experience shows little difference between iterates.



Bathmann(2018) a) First iterate b) Sixth iterate correlation matrix for IASI with NCEP global system