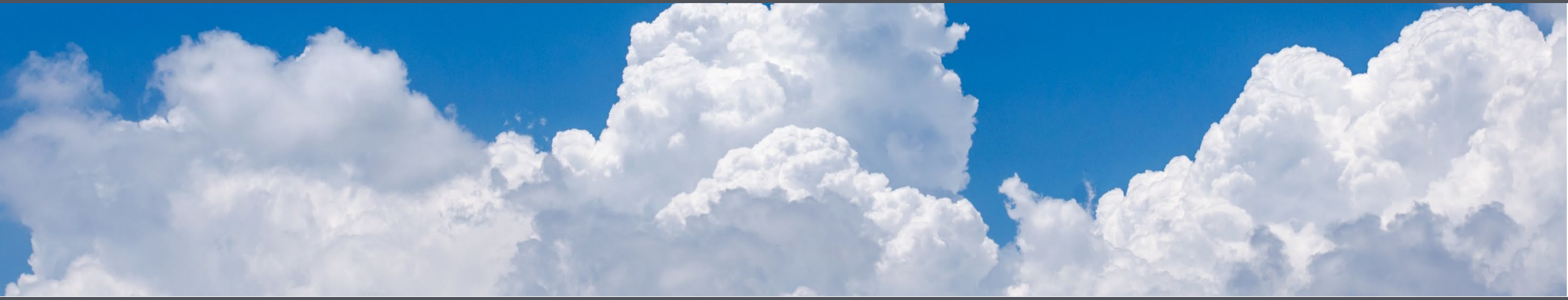
Department of Meteorology
Department of Mathematics & Statistics

Advancing the Frontiers of Earth System Prediction

University of Reading

Approaches for diagnosing random and systematic errors in observations



Sarah L Dance

Thanks to: Rishabh Bhatt, Massimo Bonavita, Niels Bormann, Alison Fowler, Elias Holm, Guannan Hu, Chris Merchant, Jonathan Mittaz, Stuart Newman, Nancy K. Nichols, Jennifer Scott, David Simonin, Joanne Waller

University of Reading

ECMWF
EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS

Met Office

NPL
National Physical Laboratory

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

Outline

University of Reading

- What is observation uncertainty?
- Why should we care about observation uncertainty?
- How can we estimate observation uncertainty?
- What now? conclusion

What are observation errors?

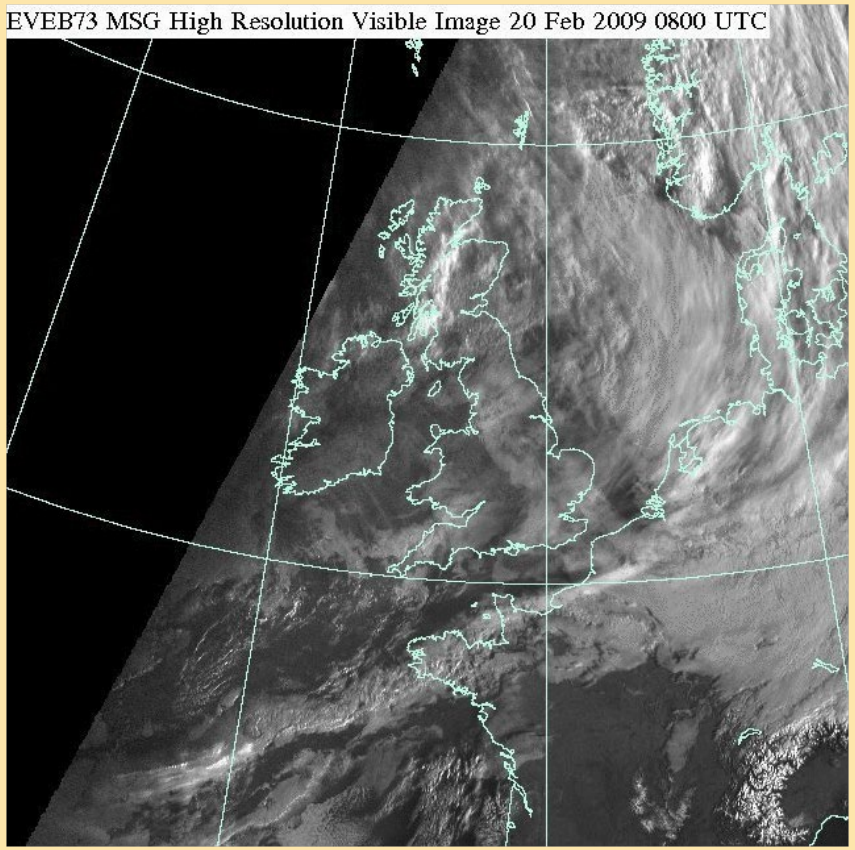
University of Reading

In data assimilation, we consider the observation equation

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

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

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



Where do observation errors come from?

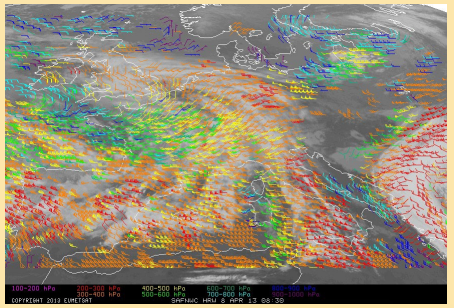
University of Reading

The error vector, ε, contains errors from four main sources: Janjic et al (2018)

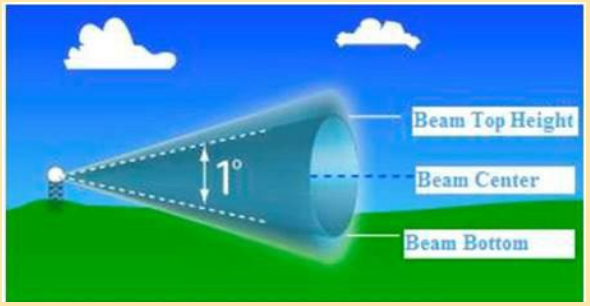
Instrument noise



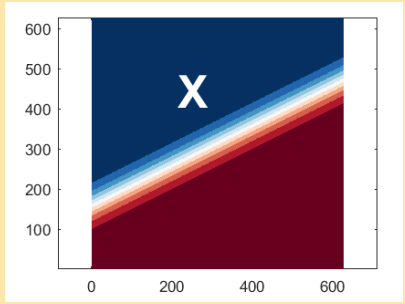
Observation pre-processing

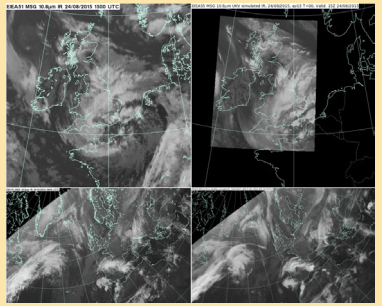


Observation operator error



Scale mis-match





The role of R in the assimilation
(Fowler et al 2018 10.1002/qj.3183)

University of Reading

Two variable problem:

$$\mathbf{B} = \beta \begin{bmatrix} 1 & \chi_b \\ \chi_b & 1 \end{bmatrix}$$

where $-1 < \chi_b < 1$,

$$\mathbf{R} = \rho \begin{bmatrix} 1 & \chi_r \\ \chi_r & 1 \end{bmatrix}$$

where $-1 < \chi_r < 1$ and

$$\mathbf{H} = \begin{bmatrix} 1 & a \\ a & 1 \end{bmatrix}.$$

- This system has two scales (large and small) corresponding to two eigenvectors

Large scale error variance | Small scale error variance

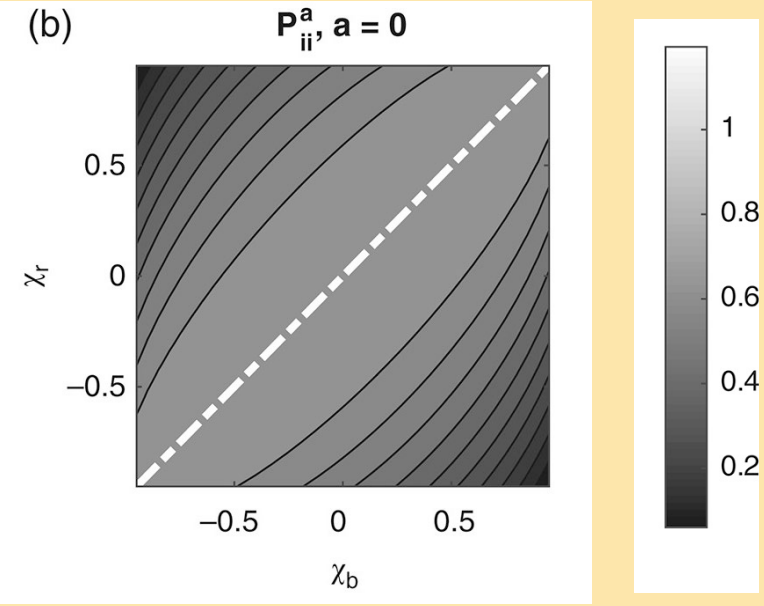
B $\gamma_0 = 1 + \chi_b, \quad \gamma_1 = 1 - \chi_b,$

R $\psi_0 = 1 + \chi_r, \quad \psi_1 = 1 - \chi_r,$

- Large χ - larger error variance at large scales (more accurate at small scales)
- Small χ – larger error variance at small scales (more accurate at large scales)

Analysis error variance

Analysis error variance as a function of correlations with β=2, ρ=1 and **H** = **I**



- Maximum analysis error variance when

$$\chi_b = \chi_r$$

- Background and observations both most accurate at the same scale - accurate analysis only at that scale
- When

$$\chi_b \neq \chi_r$$

- We can weight the data appropriately to provide an accurate analysis at both scales

But….

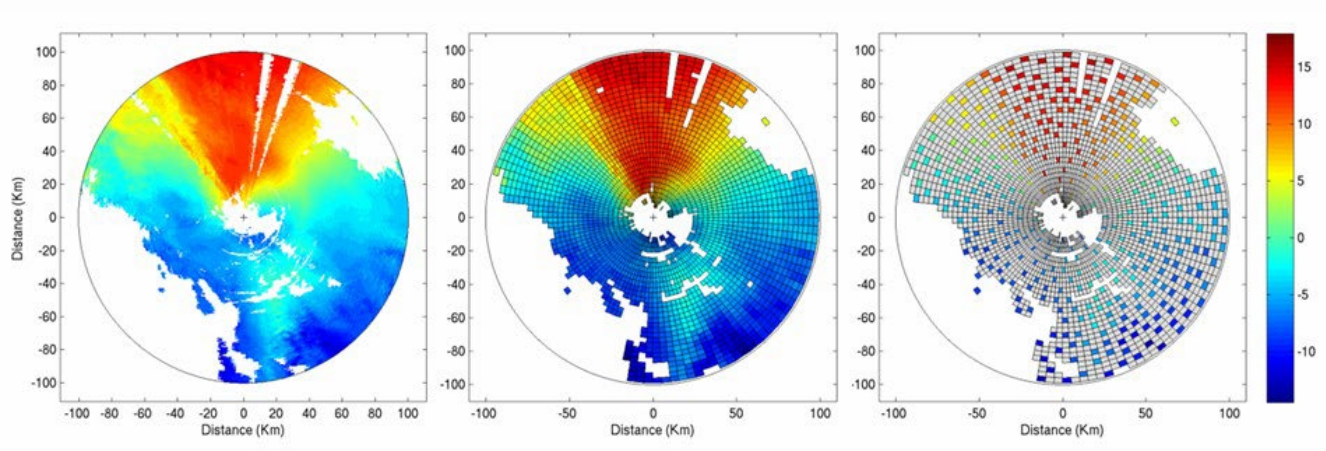
- Usually we treat observation errors as uncorrelated
- Only use **B** to filter and spread observation information
- We avoid the computationally expensive treatment of observation error correlations in the assimilation
- The analysis is suboptimal

Taking proper account of observation error correlations

University of Reading

- Use higher proportion of observations

Raw (~600m) Superobbed (~3km) Thinned (~6km)



Example: Doppler radar winds (*old* Met Office system from Jo Waller/David Simonin)

- Improve analysis accuracy and forecast skill (e.g., Stewart et al. 2013; Weston et al., 2014, Simonin et al 2019)
- Changes to scales of observation information content in analysis depending on both the prior and observation error correlations (Fowler et al, 2018)

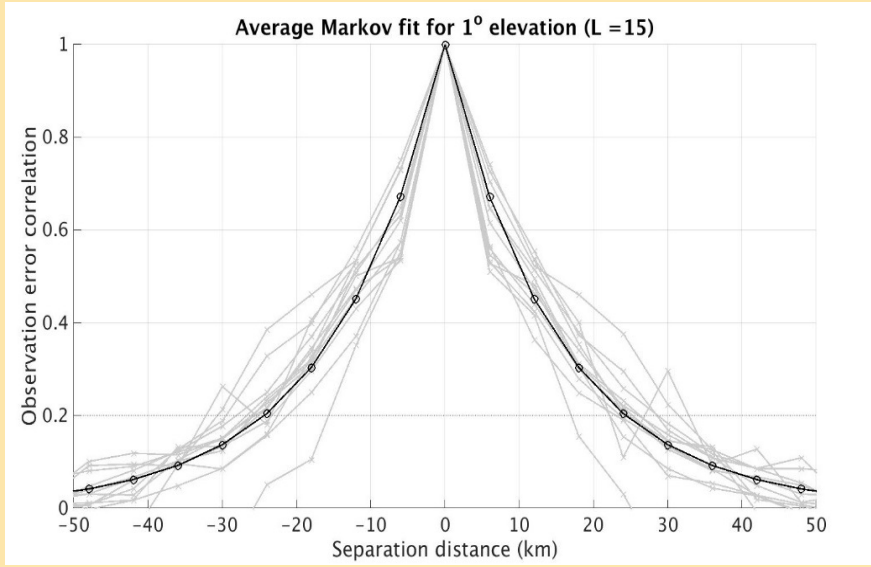
Treating correlated observation error for Doppler radar winds (Simonin et al, 2019)

University of Reading

Met Office

- Current 6km thinning reduced to 3km by treating spatial observation error correlations
- Spatial correlations estimated offline using Desroziers et al (2005) method
- New parallelization and smart load balancing

⇒ can assimilate 4x as many observations without an increase in wall-clock time



Joint work using Met Office UKV

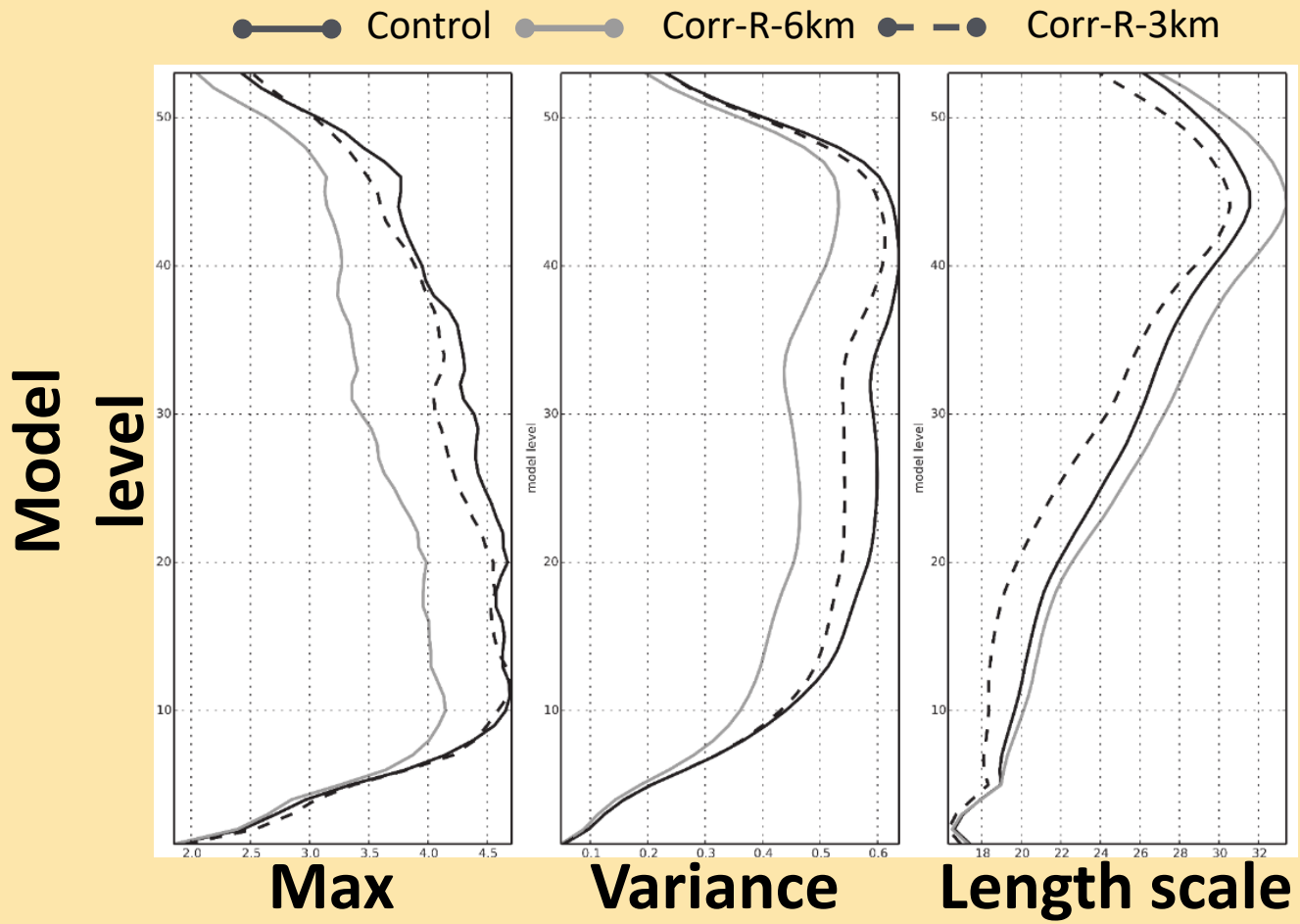
Effect on wind increments (Simonin et al 2019)

Compared to the Control with diagonal **R** and 6km thinning

The **Corr-R-3km** wind's increments show more small scale features with smaller range

Compares well with theory (Fowler et al 2018)

We also found improvements in forecast skill – particularly for convective rainfall.



Estimating observation uncertainty

Three main approaches

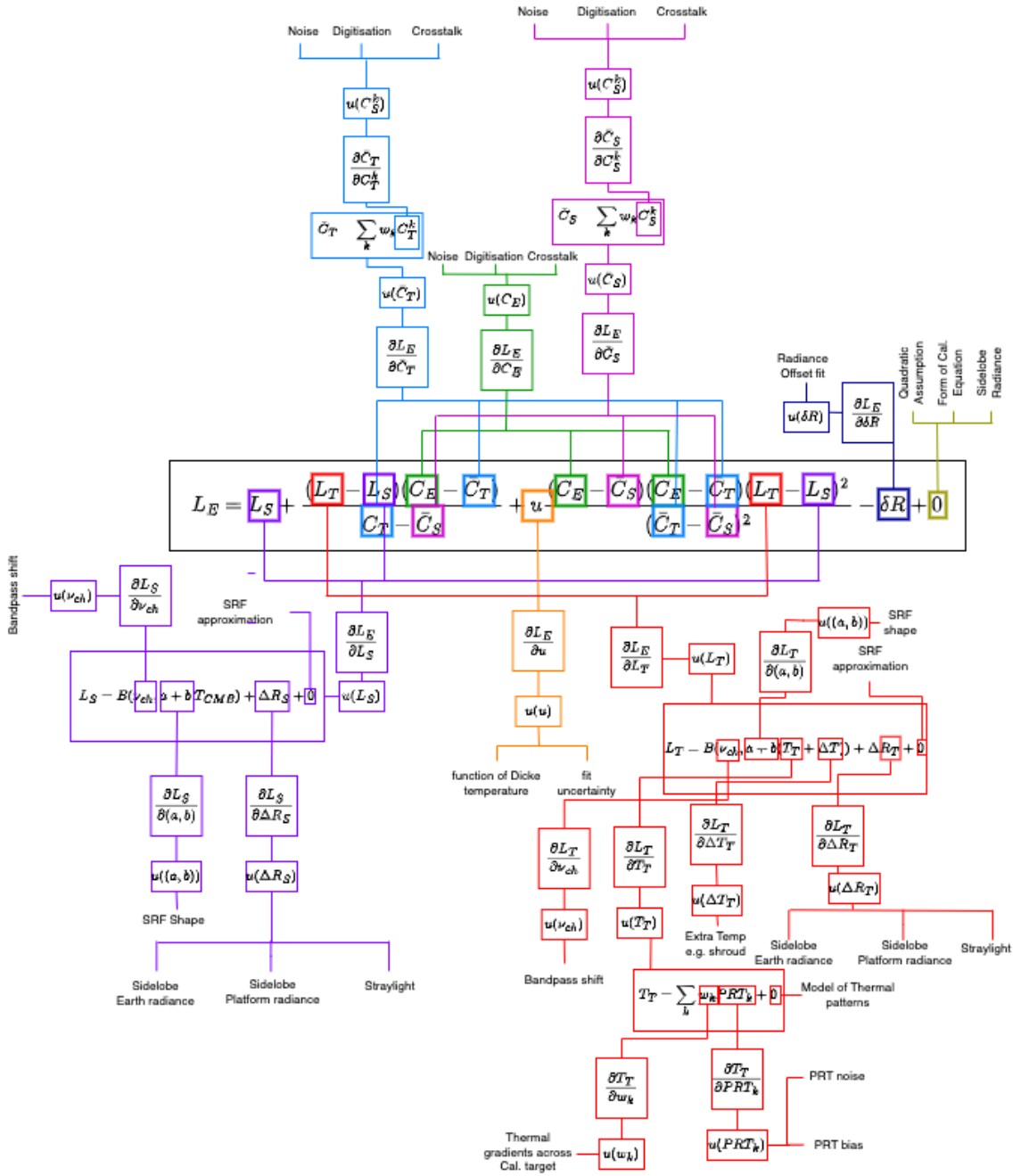
- Metrological approach – estimation from first principles

- 3-Cornered Hat – using other collocated data-types

- Diagnosis from assimilation – using innovation and residual statistics (O-A, O-B)

Metrological approach MSU Uncertainty Tree

- Example of measurement equation for MSU (AMSU pre-cursor)
  - Metrological approach to uncertainty e.g. FIDUCEO or QA4EO (www.qa4eo.org) or Mittaz, Merchant & Woolliams (2019) doi: 10.1088/1681-7575/ab1705
- For each term in the measurement equation traces back to sources of error
  - Including +0 term designed to highlight assumptions
- Each 'twig' comes with an effects table which also includes error correlation scales



Slide from Jonathan Mittaz

Metrological approach

University of Reading

- Estimates instrument and processing errors

- Takes no account of
  - Quality control errors
  - Observation operator errors
  - Model-observation scale mismatch errors

- Can only provide a lower-bound on the error estimate that we need

3-Cornered Hat *(similar to Triple Collocation)*

University of Reading

- 3 datasets, collocated in space and time

[x, y, z]

- Error variance of x is

$$\hat{\mathbf{X}} = \frac{1}{2}\{\mathrm{cov}(\mathbf{x}-\mathbf{y}) + \mathrm{cov}(\mathbf{x}-\mathbf{z}) - \mathrm{cov}(\mathbf{y}-\mathbf{z})\} + \left[E(\epsilon^{\mathbf{x}} \odot \epsilon^{\mathbf{y}}) + E(\epsilon^{\mathbf{x}} \odot \epsilon^{\mathbf{z}}) - E(\epsilon^{\mathbf{y}} \odot \epsilon^{\mathbf{z}})\right],$$

$$\epsilon^{\mathbf{x}} = \mathbf{x} - E[\mathbf{x}] - \mathbf{t},$$

- If datasets are independent, cross terms disappear and you have an eq. calculable from sample data



Image generated using MS Copilot

3CH

- Extremely difficult to find 3 types of independent observations, measuring the same thing, co-located in space and time

- Typically observations will have different “footprints” and this is ignored

- Even harder to find collocated networks (to enable calculation of error correlations) as well as variances

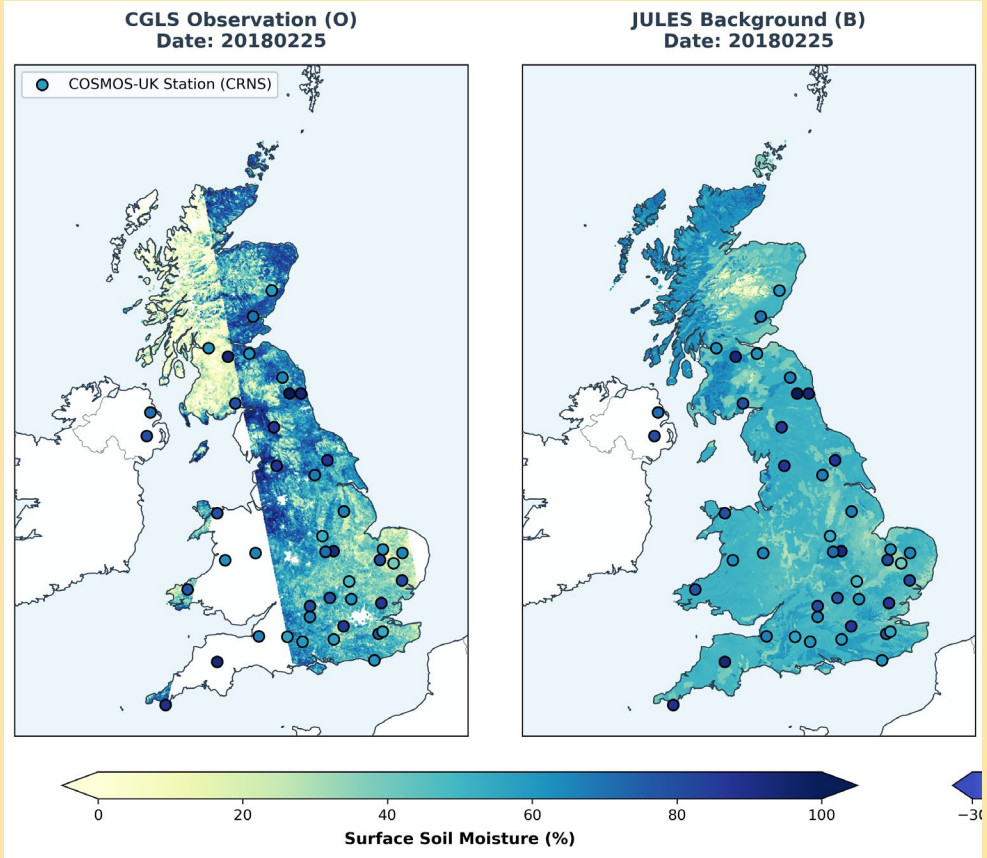


Figure from Suping Ni

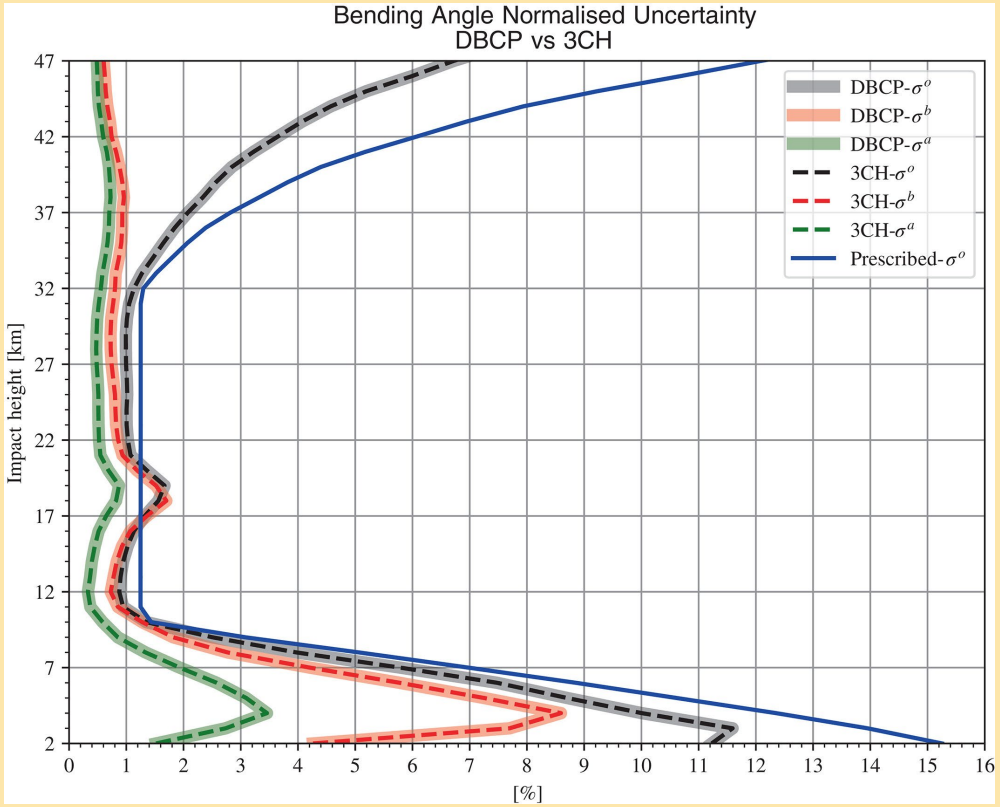
CGLS soil moisture, COSMOS soil moisture stations, JULES model

3CH equivalent to Desroziers et al (2005)

University of Reading

- If you use observations, analysis and background as the three sources of information, there is an equivalence with Desroziers et al (2005) method for diagnosing **R** (Todling et al 2022).

- Example for COSMIC-2 bending angles with ERA-5 background and analysis



Todling et al (2022) 10.1002/qj.4343

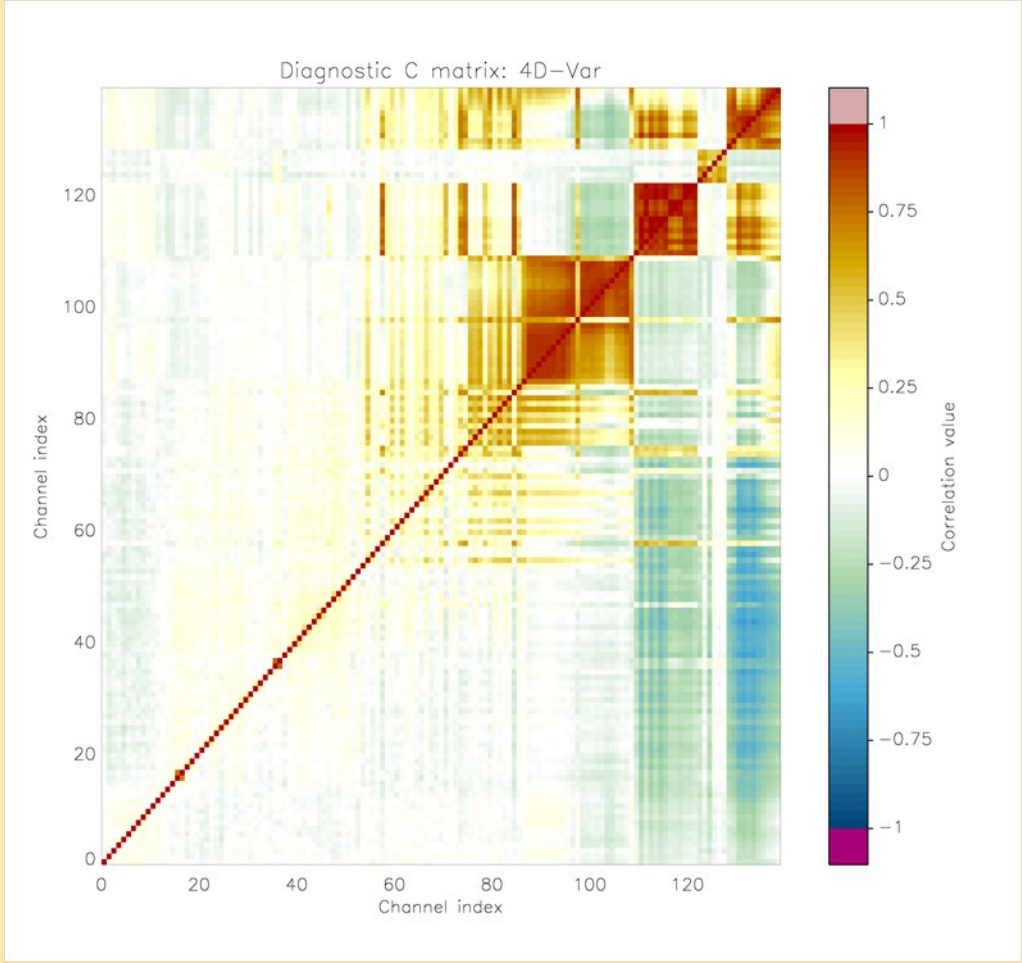
Desroziers et al (2005) diagnostic (DBCP)

University of Reading

- Become popular because it is very easy to compute from readily available quantities
- It does NOT give a PERFECT estimate but a USEFUL one

- Early estimate of IASI interchannel correlations (Stewart et al 2009, 2014)

- Note it is not even symmetric



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...

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

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...

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

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...

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

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:

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

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

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

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

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

Diagnostic Pitfalls and Possibilities



This Photo by Unknown Author is licensed under CC BY



Errors expected with the Desroziers et al (2005) diagnostic (Hu and Dance 2024, 10.1002/qj.4750)

$$\frac{1}{m}\mathbb{E}\left[\|\hat{\mathbf{R}}^{\mathrm{e}}-\mathbf{R}\|_{\mathrm{F}}^{2}\right]=\frac{1}{m}\mathbb{E}\left[\|\hat{\mathbf{R}}^{\mathrm{e}}-\mathbf{R}^{\mathrm{e}}\|_{\mathrm{F}}^{2}\right]+\frac{1}{m}\|\mathbf{R}^{\mathrm{e}}-\mathbf{R}\|_{\mathrm{F}}^{2}+\frac{2}{m}\left\langle\mathbb{E}\left[\hat{\mathbf{R}}^{\mathrm{e}}-\mathbf{R}^{\mathrm{e}}\right],\mathbf{R}^{\mathrm{e}}-\mathbf{R}\right\rangle_{\mathrm{F}}. \quad (30)$$

$\hat{\mathbf{R}}^{\mathrm{e}}$ is the sample estimate using the diagnostic

$\mathbf{R}^{\mathrm{e}}$ is the theoretical diagnosis estimate in exact arithmetic

Errors expected with the Desroziers et al (2005) diagnostic (Hu and Dance 2024, 10.1002/qj.4750)

$$\frac{1}{m}\mathbb{E}\left[\|\hat{\mathbf{R}}^{\mathrm{e}}-\mathbf{R}\|_{\mathrm{F}}^{2}\right]=\frac{1}{m}\mathbb{E}\left[\|\hat{\mathbf{R}}^{\mathrm{e}}-\mathbf{R}^{\mathrm{e}}\|_{\mathrm{F}}^{2}\right]+\frac{1}{m}\|\mathbf{R}^{\mathrm{e}}-\mathbf{R}\|_{\mathrm{F}}^{2}+\frac{2}{m}\left\langle\mathbb{E}\left[\hat{\mathbf{R}}^{\mathrm{e}}-\mathbf{R}^{\mathrm{e}}\right],\mathbf{R}^{\mathrm{e}}-\mathbf{R}\right\rangle_{\mathrm{F}}. \quad (30)$$

Small

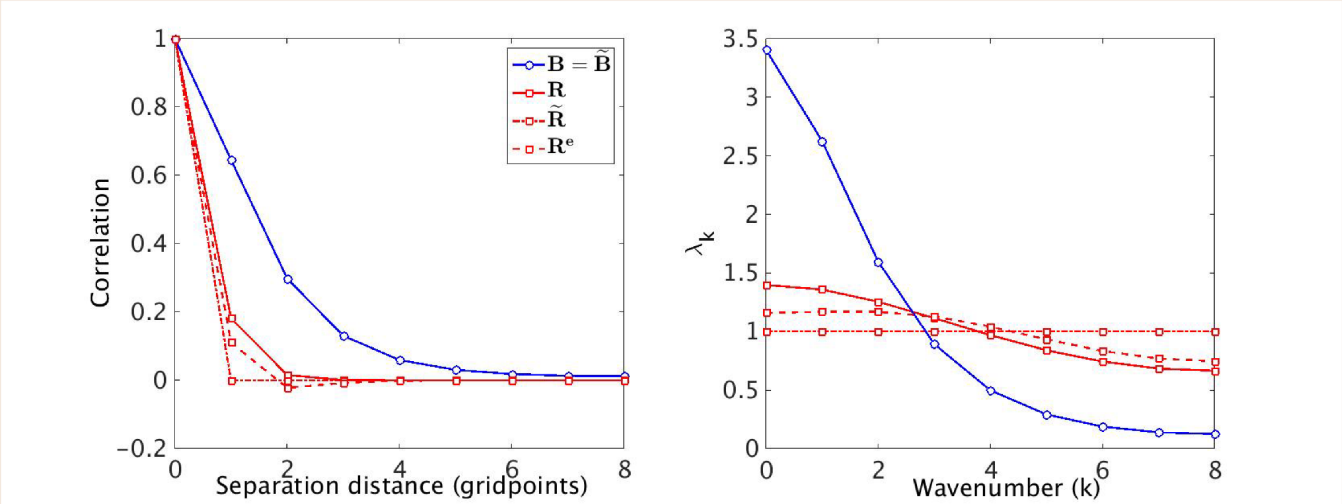
Error = indirect sampling error + error due to misspecification

Misspecification error

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

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

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

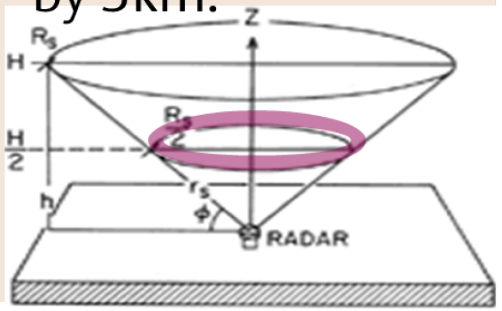


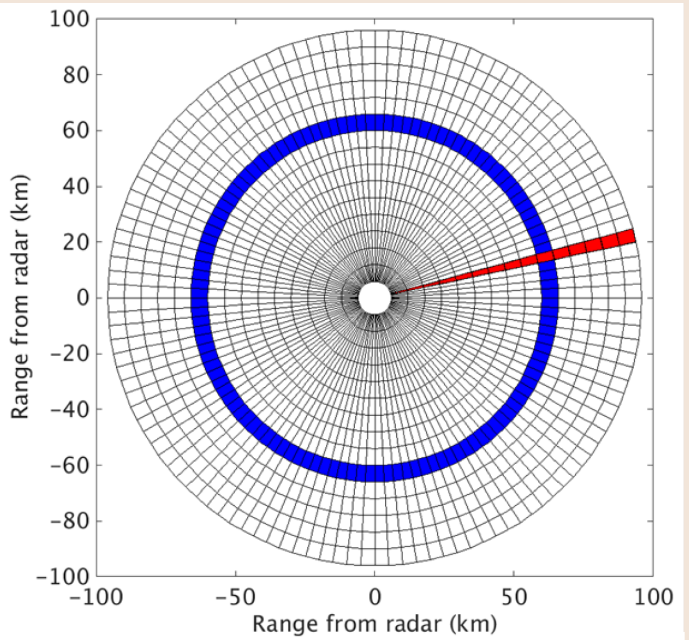
$\mathbf{R}^e$ has an underestimated variance and correlation lengthscale, but is a better approximation than $\widetilde{\mathbf{R}}$.

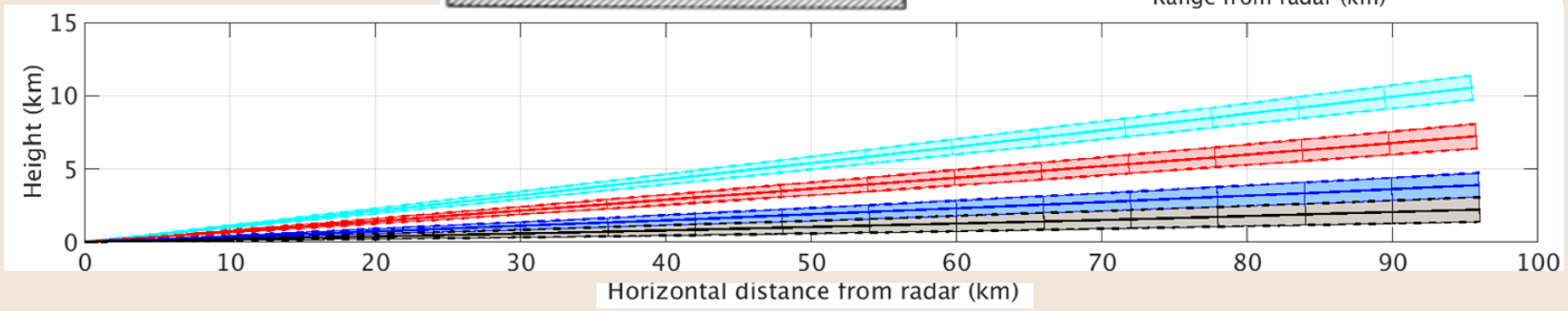
Doppler radar winds and Met Office UKV

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.



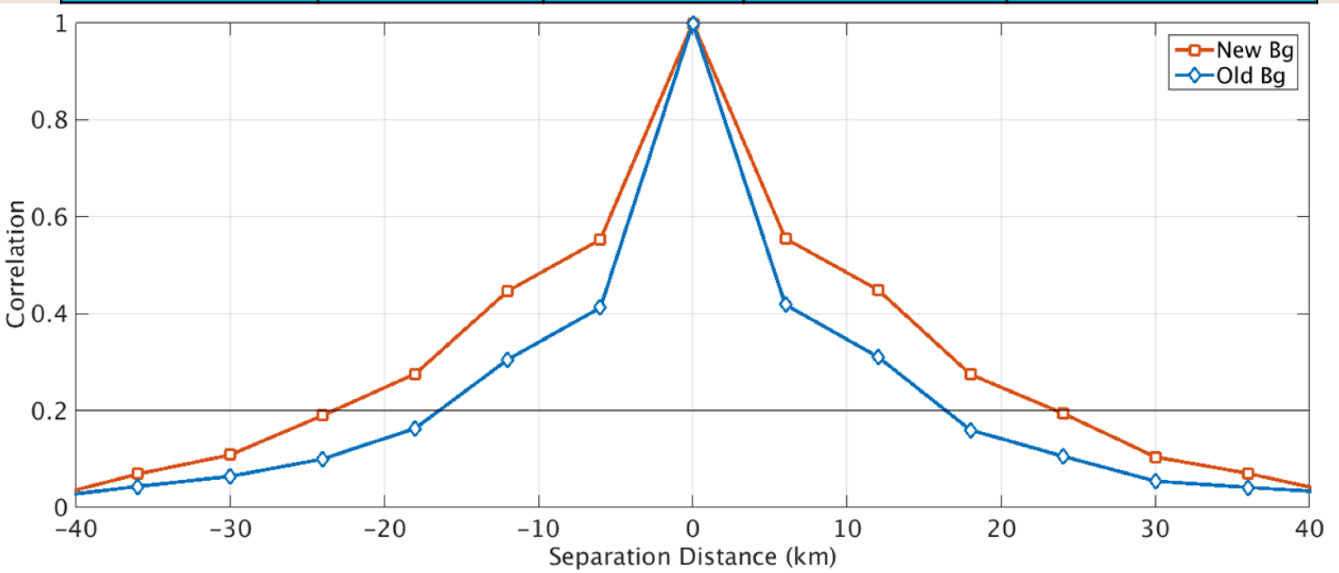




Horizontal Correlations, sensitivity to $\widetilde{\mathbf{B}}$

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

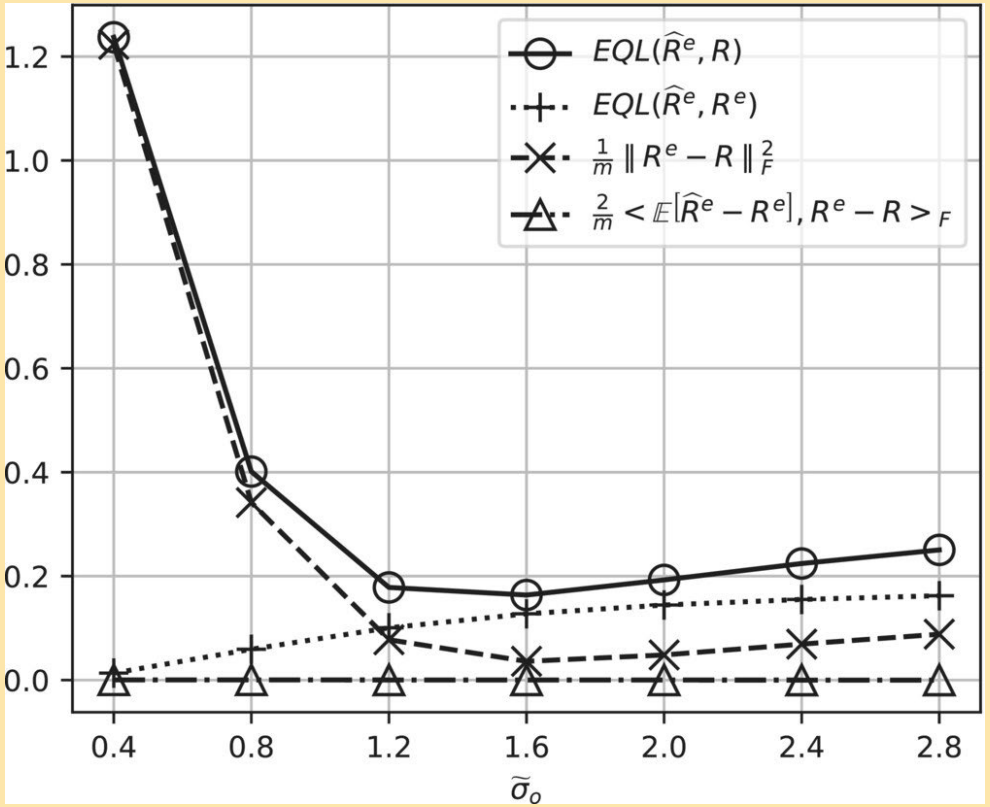


- Increasing variance and lengthscale in $\widetilde{\mathbf{B}}$ reduces variance and lengthscale in diagnosed $\mathbf{R}^e$.
- Consistent with Waller et al (2016) QJ theory.

Sampling error (Hu and Dance 2024)

Sampling error depends on

- Number of samples
- Assumed error statistics – variances and lengthscales



Example – assumed **R** is diagonal but should be correlated, **B** is correct.

\+ Indirect sampling errors increase with assumed observation error variance

x misspecification error decreases and then increases with assumed observation error variance

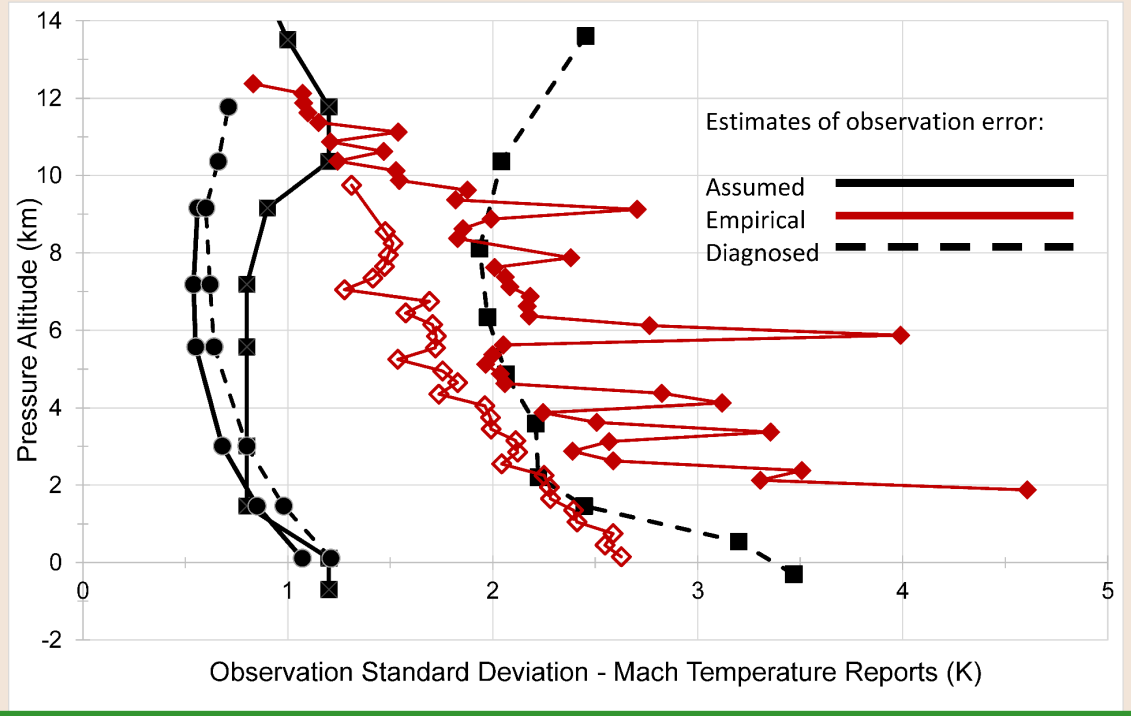
O overall errors decrease and then increase with assumed observation error variance

University of Reading

Examples of practical use

Comparison of approaches (Mirza et al, 2020)

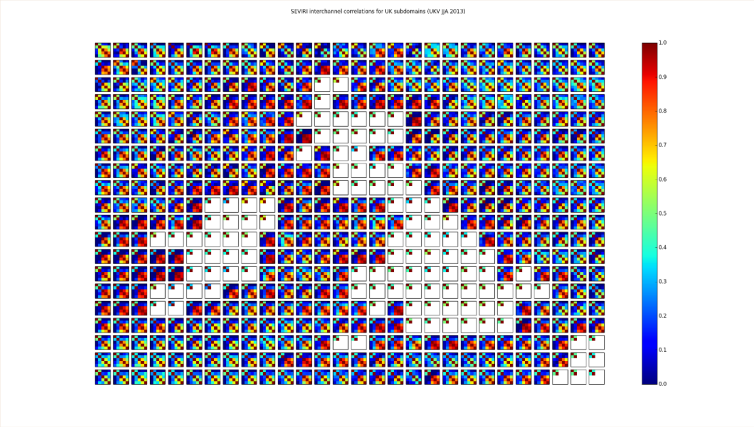
- Mode-S EHS temperatures - errors from lack of precision in Mach number
- Diagnosed std (black-dashed-squares) compare well with metrological estimates (red diamonds)



Identifying sources of error - Examples

Waller et al (2016) Rem. sens.

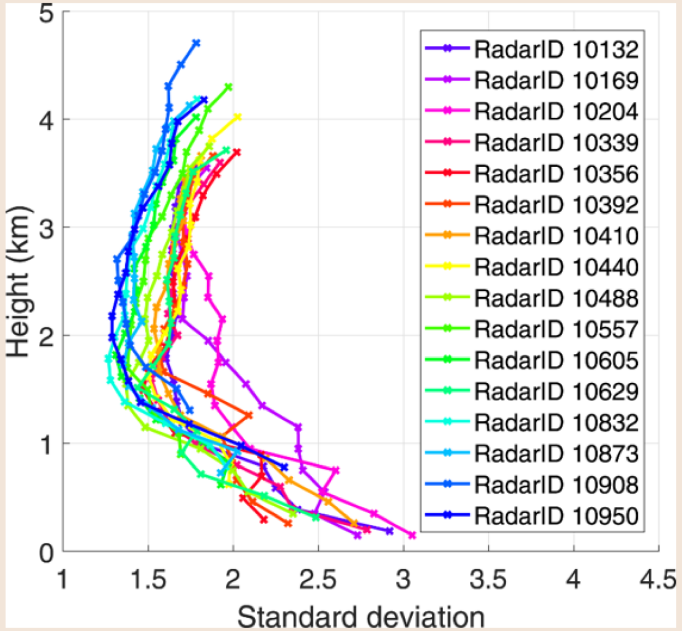
SEVIRI interchannel error covariances over different subdomains



Land-sea QC issue

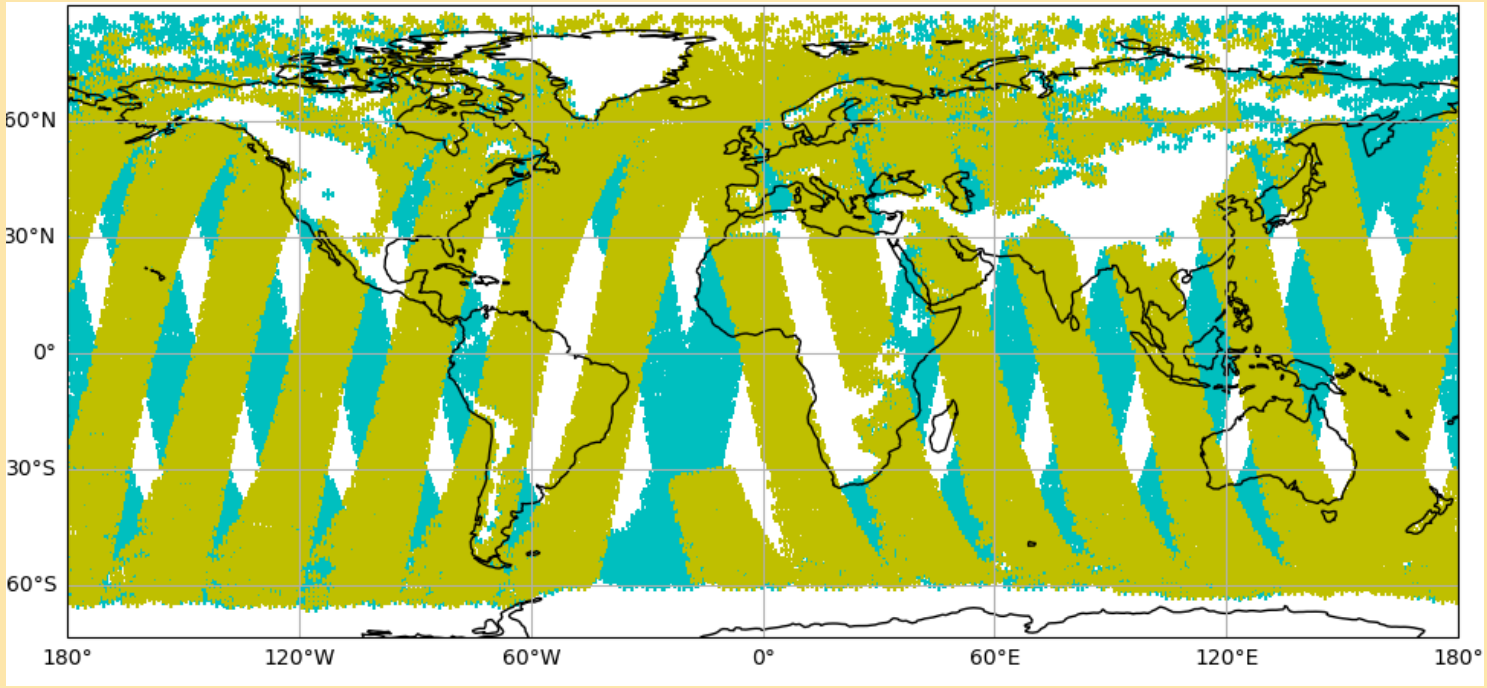
Bauernschubert et al (2019)

Doppler radar wind error std



Radars 10169 and 10204 contaminated by wind turbines and ship tracks

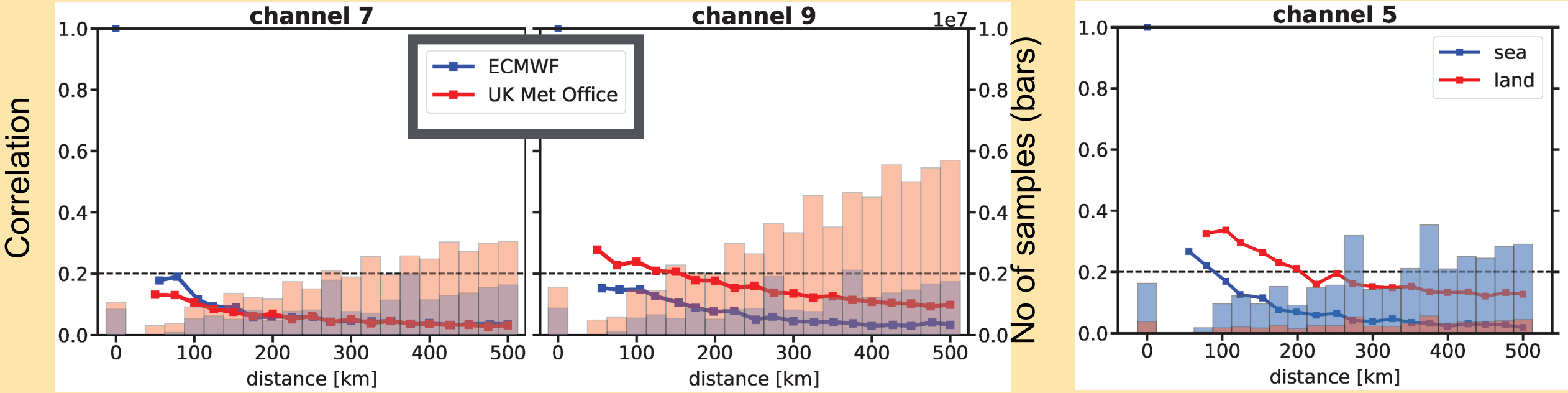
Polar orbiter – AMSU-A



AMSU-A **(Advanced Microwave Sounding Unit)**

Chosen as high impact, all-sky data, but correlations expected to be simple(ish)

AMSU-A on MetOp-C (Bhatt et al, 2026)



- Some channels could be assimilated with diagonal R at higher density as correlations are small (e.g., channel 7LEFT)
- Some evidence that some correlations are due to imperfect bias correction (e.g. MO channel 9MIDDLE), issues with surface emissivity over land (channel 5 RIGHT) Can we fix some of these problems? Can we account for correlations in DA?

Conclusions and questions

University of Reading

- Properly accounting for observation error correlations in data assimilation can allow more observations to be used, and result in more accurate analyses on appropriate scales – better forecasts of high impact weather

- Methods for estimating uncertainty have varying levels of rigour, difficulty and practical usefulness

- The Desroziers et al (2005) method is easy to calculate and useful, BUT it is sensitive to the assumed error statistics (used in the assimilation) and the results must be analysed carefully

- Q: can ML algorithms solve this problem in a different way and allow us to use more data more efficiently?