Data Assimilation Research Centre

NERC SCIENCE OF THE ENVIRONMENT

University of Reading

National Centre for Earth Observation NATURAL ENVIRONMENT RESEARCH COUNCIL

Data compression in the presence of observation error correlations

Alison Fowler (University of Reading and National Centre for Earth Observation)

ModSW model workshop, Leeds May 2019

Introduction

- Numerical weather prediction (NWP) models are moving towards higher resolutions to capture the rapid development of convective-scale systems.
- Need high-resolution observations to allow for frequent update of these models (via data assimilation)
  – E.g. the next generation of hyper-spectral and geostationary satellites, developments in ground-based remote sensing and the exploitation of existing sources of information such as mobile phone data
- The unprecedented volume of data provided by these new observing systems will bring huge opportunities for convective-scale forecasting but also many challenges.
  – high volume of data makes it difficult to transmit, store and assimilate the data in a timely manner.
  – data may have complicated error characteristics, such as non-negligible error correlations, that need to be represented.

Observation Error Correlations (OECs)

- Many ob errors have significant spatial correlations, e.g. ~200km for AMV, ~20km DRWs
- OECs are difficult to estimate
  - often attributed to the uncertainty in the comparison of the observations to the model variables, known as representation error, rather than instrument noise (Janjic et al., 2017).
  - can be state and model dependent (Waller et al., 2014).
- Neglecting OECs means the observations cannot be assimilated optimally at the resolutions needed, this can have a particularly detrimental impact on the analysis of small-scale structures (Fowler et al. 2018).
- However, progress is being made, with centres around the world now explicitly accounting for OECs in a variety of observation types allowing for the optimal use of denser observations (e.g. Simonin et al. 2019.
- This work explores how to make efficient use of this potentially dramatic increase in the amount of data available for assimilation.

Data compression (DC)

- Instead of regular thinning, the compression of the data provides an objective way for identifying data redundancy within the observations. DC, therefore, allows for a reduction in the volume of the data while retaining much of the information within the observations.
- Can define information content of the observations, $\mathbf{y}$, in terms of the sensitivity of the analysis, $\mathbf{x}^{\mathrm{a}}$, to the observations

$$\mathbf{S} = \frac{\partial h(\mathbf{x}^{\mathrm{a}})}{\partial \mathbf{y}} = \mathbf{K}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}$$

$\mathbf{K}$ = Kalman gain matrix (weighting given to the observations)
h $(\mathbf{H})$ = (linearised) observation operator (mapping from state to observed variables)

- This can be summarised in terms of the degrees of freedom for signal, or mutual information.

$$DFS = trace(\mathbf{S})$$

$$MI = -0.5 \ln det(\mathbf{I} - \mathbf{S})$$

Data compression

- Could compress the observations in terms of the eigenvectors of $\mathbf{S}$, but a neater way is the following:
- Let $\mathbf{M} = \mathbf{R}^{-1/2}\mathbf{H}\mathbf{B}^{1/2} = \mathbf{U}\boldsymbol{\Lambda}^{\mathrm{M}}\mathbf{V}^{\mathrm{T}}$
- Then $DFS = trace(\mathbf{M}\mathbf{M}^{\mathrm{T}}(\mathbf{I}+\mathbf{M}\mathbf{M}^{\mathrm{T}})^{-1})$

$$MI = 0.5\ln det(\mathbf{I}+\mathbf{M}\mathbf{M}^{\mathrm{T}})$$

$\mathbf{R}$ and $\mathbf{B}$ are the observation and prior error covariances respectively

- Can compress the observations using $\mathbf{C} = \mathbf{I}^{\mathrm{c}}\mathbf{U}^{\mathrm{T}}\mathbf{R}^{-1/2}$, where $\mathbf{I}^{\mathrm{c}} \in \mathbb{R}^{p_c \times p}$ and p_c is the number of compressed observations retained for assimilation.

- The compressed observations are given by $\mathbf{y}^{\mathrm{c}} = \mathbf{C}\mathbf{y}$
- The error covariance matrix is given by $\mathbf{R}^{\mathrm{c}} = \mathbf{C}\mathbf{R}\mathbf{C}^{\mathrm{T}}$. Can see that $\mathbf{R}^{\mathrm{c}}$ reduces to $\mathbf{I}^{\mathrm{c}}(\mathbf{I}^{\mathrm{c}})^{\mathrm{T}} = \mathbf{I}_{p_c}$

Data compression

- Ordering the transformed observations w.r.t the singular values of $\mathbf{M}$ allows for the first p_c observations with the maximum information to be selected for assimilation.
- The information content of the remaining compressed observations becomes:

$$DFS^{c} = \sum_{k=1}^{p_c} \lambda_k^{M^2} / (1 + \lambda_k^{M^2})$$

$$MI^{c} = \sum_{k=1}^{p_c} \ln(1 + \lambda_k^{M^2})^{1/2}$$

Isotropic, homogenous example

- Circulant matrices have the property that eigenvectors are given by the Fourier basis, $\mathbf{F}$.
- Let $\mathbf{B} = \mathbf{F\Gamma F^T}$, $\mathbf{R} = \mathbf{F\Psi F^T}$ and $\mathbf{H} = \mathbf{I}$ (direct observations of the state)
- Then $\mathbf{M} = \mathbf{F\Psi^{-1/2}\Gamma^{1/2}F^T}$, $\mathbf{C} = \mathbf{I^c\Psi^{-1/2}F^T}$

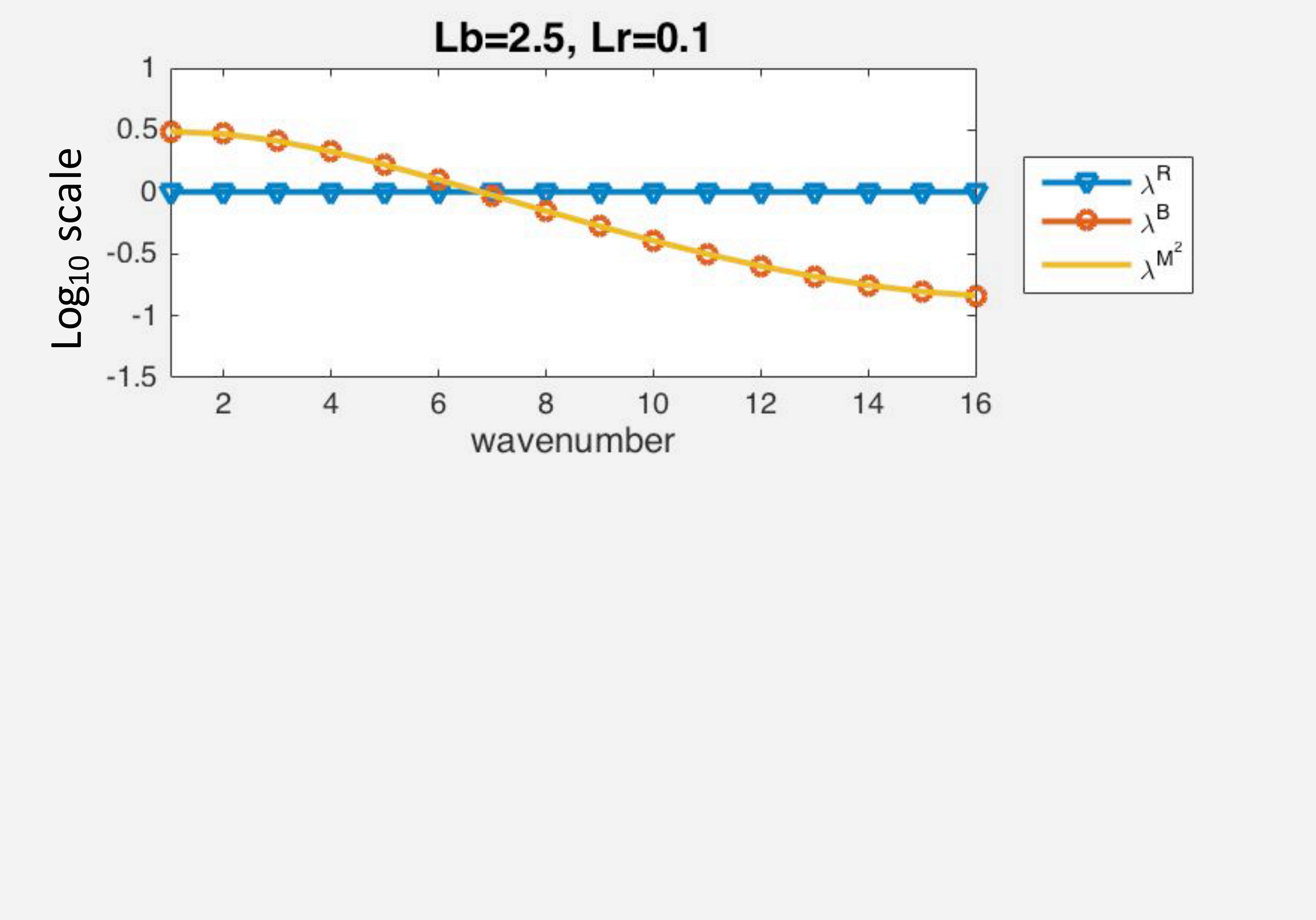
  and $\lambda_i^{M^2} = \gamma_i/\psi_i,$

  where γ_i and ψ_i are the ith eigenvalue of $\mathbf{B}$ and $\mathbf{R}$ respectively.
- The most informative compressed observations are those associated with the scales at which the prior uncertainty is relatively large compared to the observation uncertainty.
- The reduction in the analysis error variance compared to the prior is given by

$$trace(\mathbf{B} - \mathbf{P}^{\mathrm{a}})^{\mathrm{c}} = \sum_{k=1}^{p_c} \frac{\gamma_k \lambda_k^{M^2}}{1 + \lambda_k^{M^2}}$$

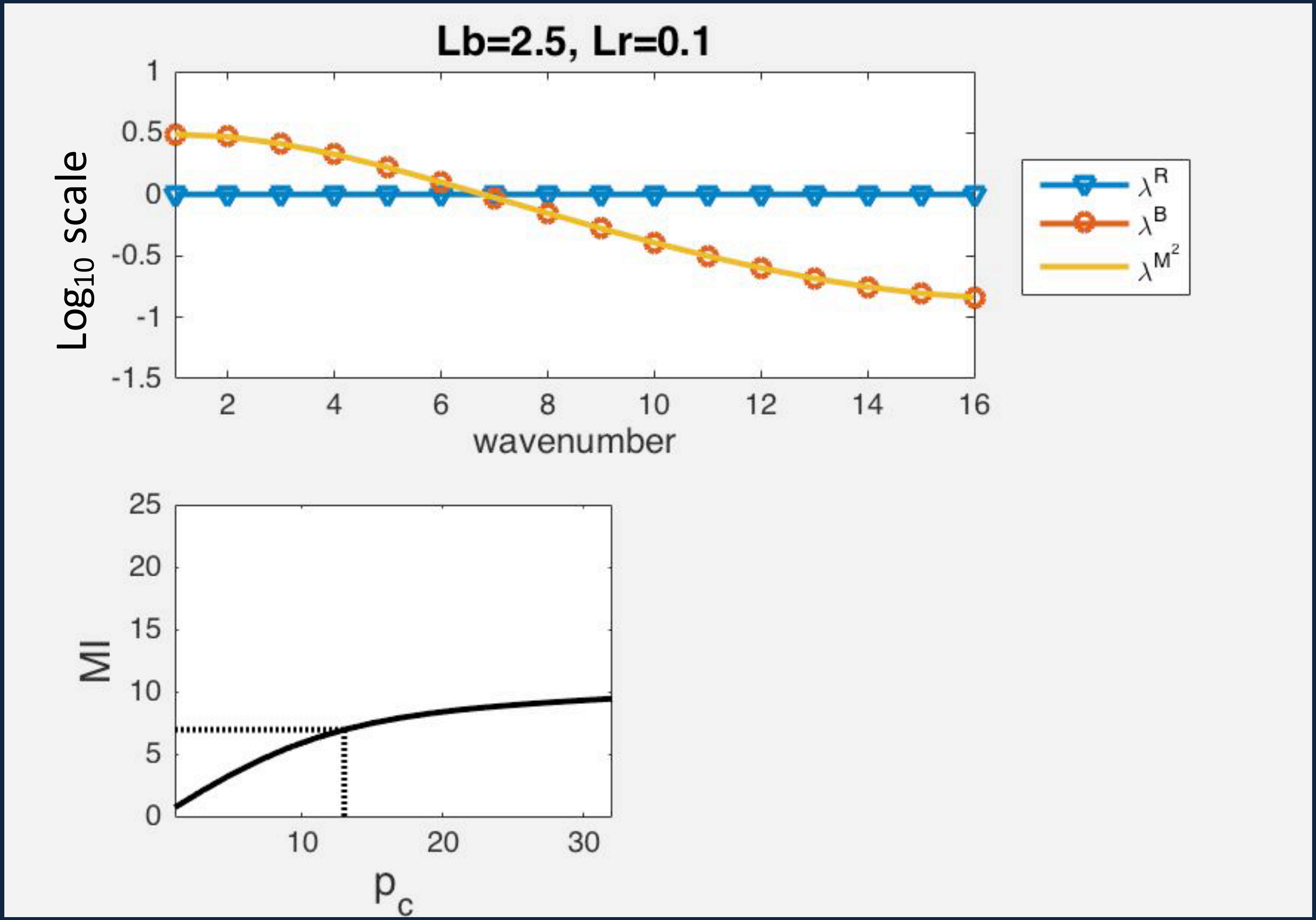
Isotropic, homogenous example…

circular grid discretised into 32 grid points. SOAR correlation structure.



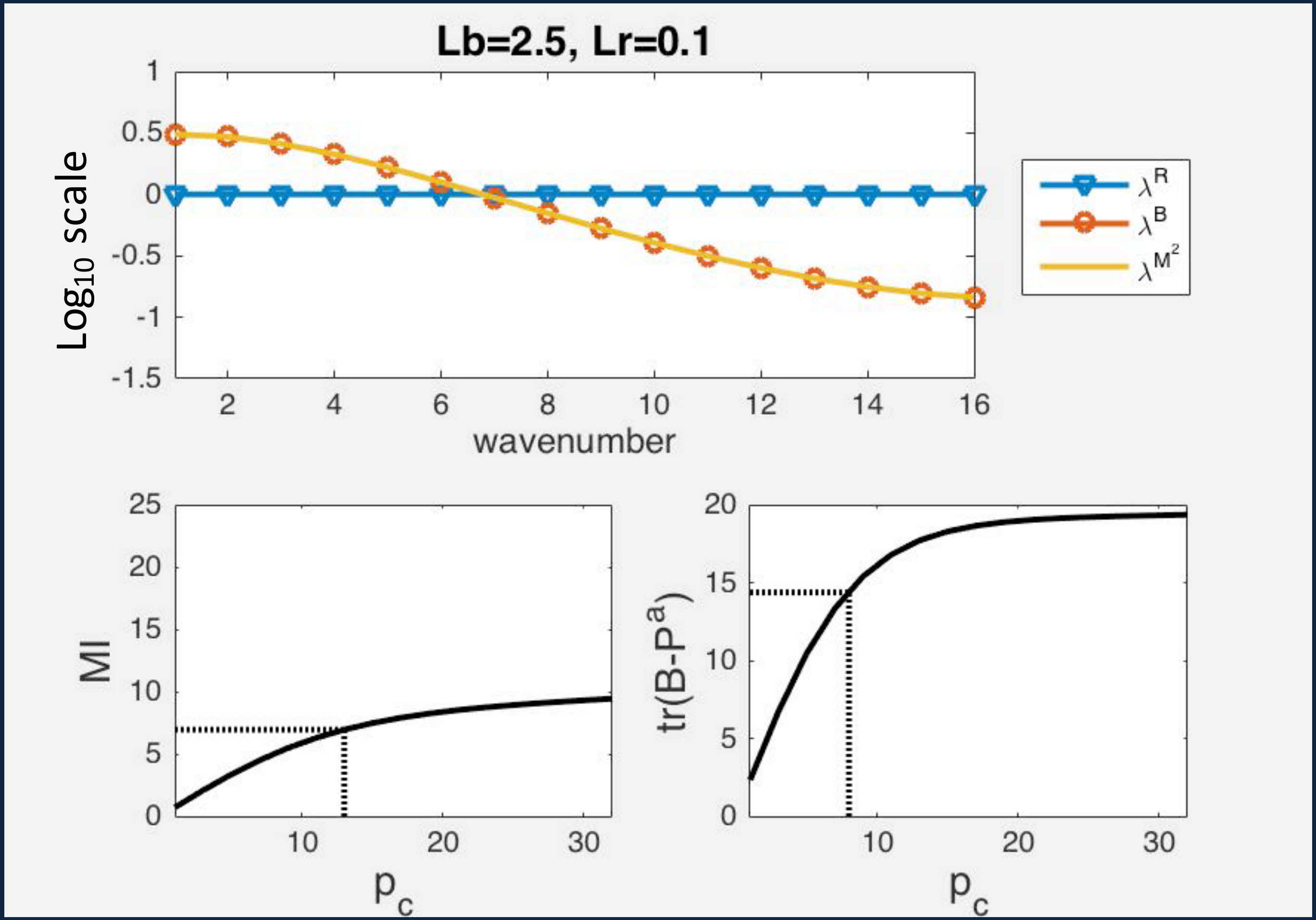
Isotropic, homogenous example…

circular grid discretised into 32 grid points. SOAR correlation structure.



Isotropic, homogenous example…

circular grid discretised into 32 grid points. SOAR correlation structure.



Observation network design Conclusions

- As the length-scales in the observation errors, L_r, increase the observations become more informative about the small scales.
- When $L_r > L_b$, the observations are more certain at small scale than the prior and so the benefit of denser observations increases.
  - Data compression can be used to help reduce the amount of data while retaining the small scale information
  - Assimilating just the small-scale information may not result in the greatest reduction in analysis error variance
    - is this an issue for nested models?
    - use a metric which focuses on accuracy of small scales?

Lorenz 96 example: Comparison of data thinning strategies

- Circular domain with 40 grid points

$$\frac{dx_j}{dt} = (x_{j+1} - x_{j-2})x_{j-1} - x_j + F,$$

- F=8.
- 80 direct, regularly distributed observations of the state are simulated
- Assimilation using EnSRF (Hunt et al. 2007).
- 100 ensemble members
- $\mathbf{P}^{\mathrm{f}}$ is no longer circulant

Data Reduction methods

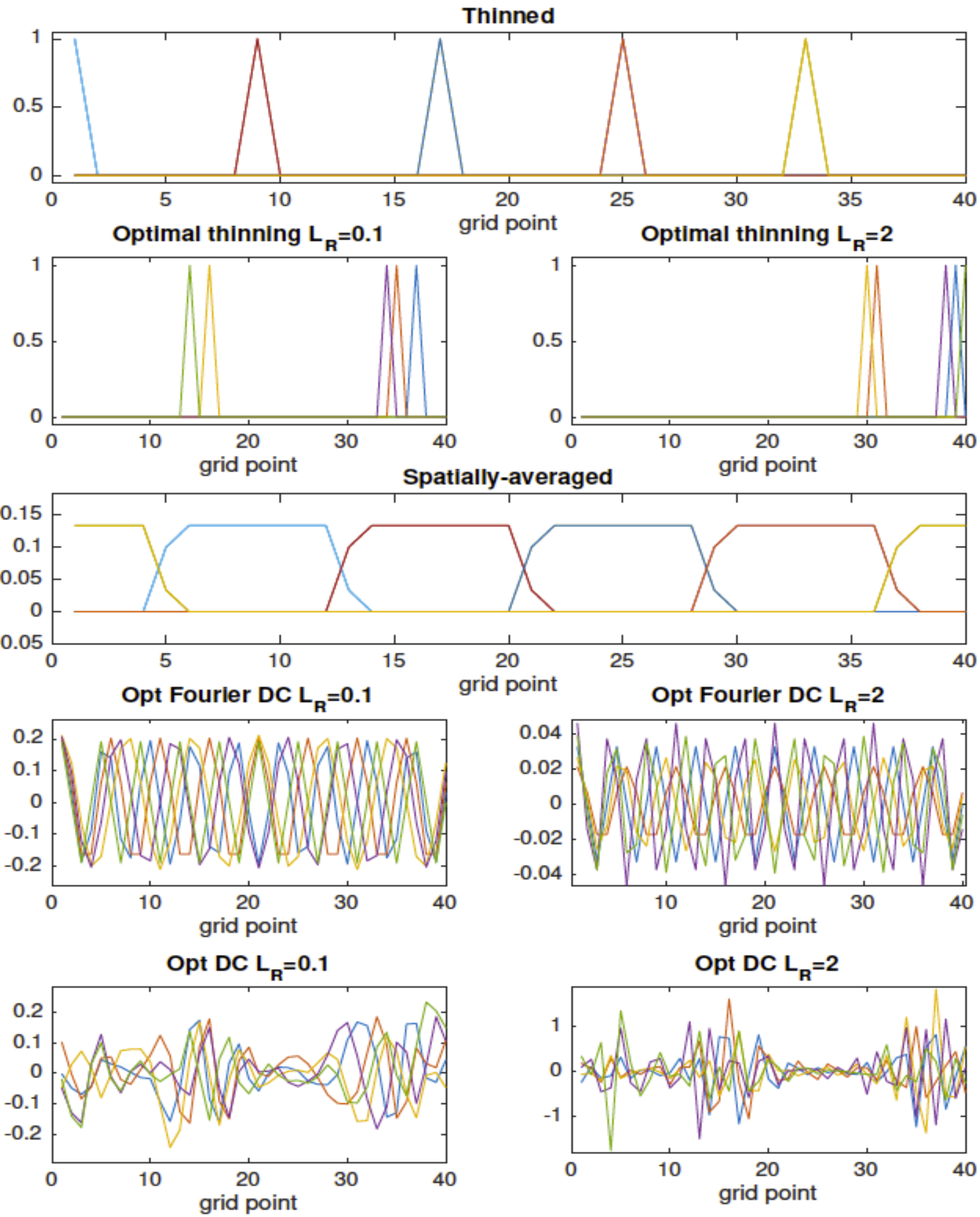
1. **Regular thinning:** to every 16th observation (giving 5 in total at each assimilation time).

2. **Optimal thinning**: obs corresponding to the 5 largest diagonal values of $S = dx^a/dy$

3. **Spatial averaging**: Observations are averaged over 8 grid-points

4. **Optimal Fourier Data Compression** (DC): Observations are compressed using a Fourier transform with wavelengths chosen corresponding to the 5 largest diagonal values of $\mathbf{FSF}^T$.

5. **Optimal DC**: Observations are compressed using the method described earlier, again assimilating just the 5 most informative observations.



Observations are simulated from a truth run:

1: R is diag (left)

2: R is spatially correlated (SOAR function), significant to 7 grid points (right).

Fig: **Rows of the observation operator matrix** for the five strategies for reducing the observation data. The optimal strategies are **illustrated for the first observation time**.

ob time	$L_R = 0.1$	$L_R = 2$
1	10,9,7,8,11	31,33,32,29,30
2	12,11,10,9,7	29,31,30,28,32
3	11,9,10,12,6	31,29,33,30,32
4	10,9,11,8,5	31,29,30,32,33
5	9,10,13,7,8	31,29,30,32,28

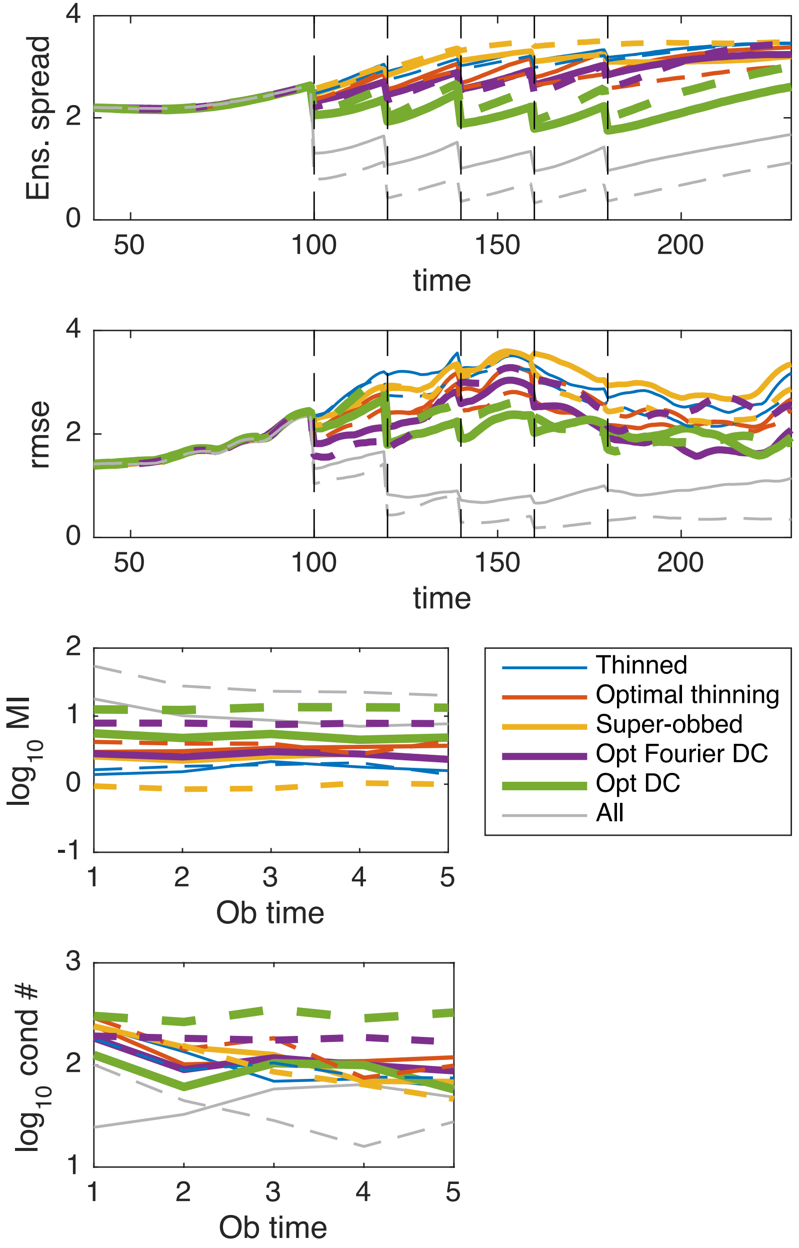


Fig: Results are averaged over 200 experiments with different realisations of the observation and model error.

- - - - correlated ob error

——— uncorrelated ob error

When observations have correlated error they are more sensitive to the form of data reduction.

Can see that largest MI does not necessarily correspond to smallest ens spread

Selecting the observations with the greatest information increases the condition number of the Hessian.

Conclusions

- Recent advances in the estimation and inclusion of OECs in data assimilation means that we are getting closer to assimilating observations optimally at their full resolution.
- The potential large increase in the number of observations available for assimilation carries a large computational and storage burden with it.
- Important to justify any increase in the amount of data assimilated and give careful thought to data reduction strategies.

- Submitted to Tellus A.

Future Work

Outstanding questions:

- How does the importance of the observations depend on the initialisation and evolution of convection? Is this correctly quantified using traditional metrics of information content such as DFS and MI?
- How can the optimum DC strategy be simplified so that it is practically feasible for convective-scale rapid update forecasting?
- How can we modify the DC strategy to take into account that the error covariances are only known approximately?

- Plan is to develop technique using idealised convective scale models like the modSW model.
- Use PAWR data assimilated using the Rapid-update-forecasting (RUF) system at RIKEN, Japan as a proof of concept. The RUF system exploits the massive power of the K computer to assimilate data every 30 seconds into a 100-m mesh limited area model (Miyoshi et al. 2016).

Thank you for listening