Data compression in the presence of observation error correlations

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

RIKEN Seminar, Monday 28th January 2019

Introduction

- Numerical weather prediction (NWP) models are moving towards higher resolutions.
- Need high-resolution observations to constrain these models (e.g. PAWR).
- The presence of non-negligible spatial observation error correlations (OECs) has typically meant the observations need to be thinned.
- However, progress is being made, with centres around the world now explicitly accounting for OECs in a variety of observation types.
- This work explores how to make efficient use of this potentially dramatic increase in the amount of data available for assimilation.

Data compression

- An alternative to thinning is to compress the observations such that the maximum amount of information is retained.
- Can define information content of the observations in terms of the sensitivity of the analysis to the observations

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

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

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

$$MI = 0.5 \ln det\left(\mathbf{B}(\mathbf{P}^{\mathrm{a}})^{-1}\right) = -0.5 \ln det(\mathbf{I} - \mathbf{S})$$

Data compression

- Let $\mathbf{M} = \mathbf{R}^{-1/2}\mathbf{H}\mathbf{B}^{1/2} = \mathbf{U}\Lambda^{\mathrm{M}}\mathbf{V}^{\mathrm{T}}$
- Then $DFS = trace(\mathbf{M}\mathbf{M}^{\mathrm{T}}(\mathbf{I}+\mathbf{M}\mathbf{M}^{\mathrm{T}})^{-1})$

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

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

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

Data compression

- Ordering the 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

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

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

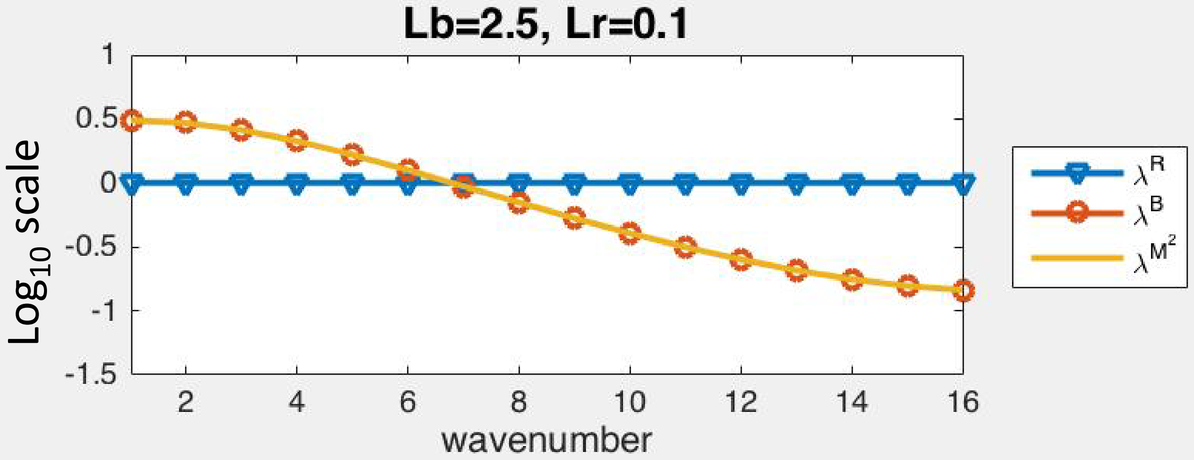
Isotropic, homogenous example

- 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

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

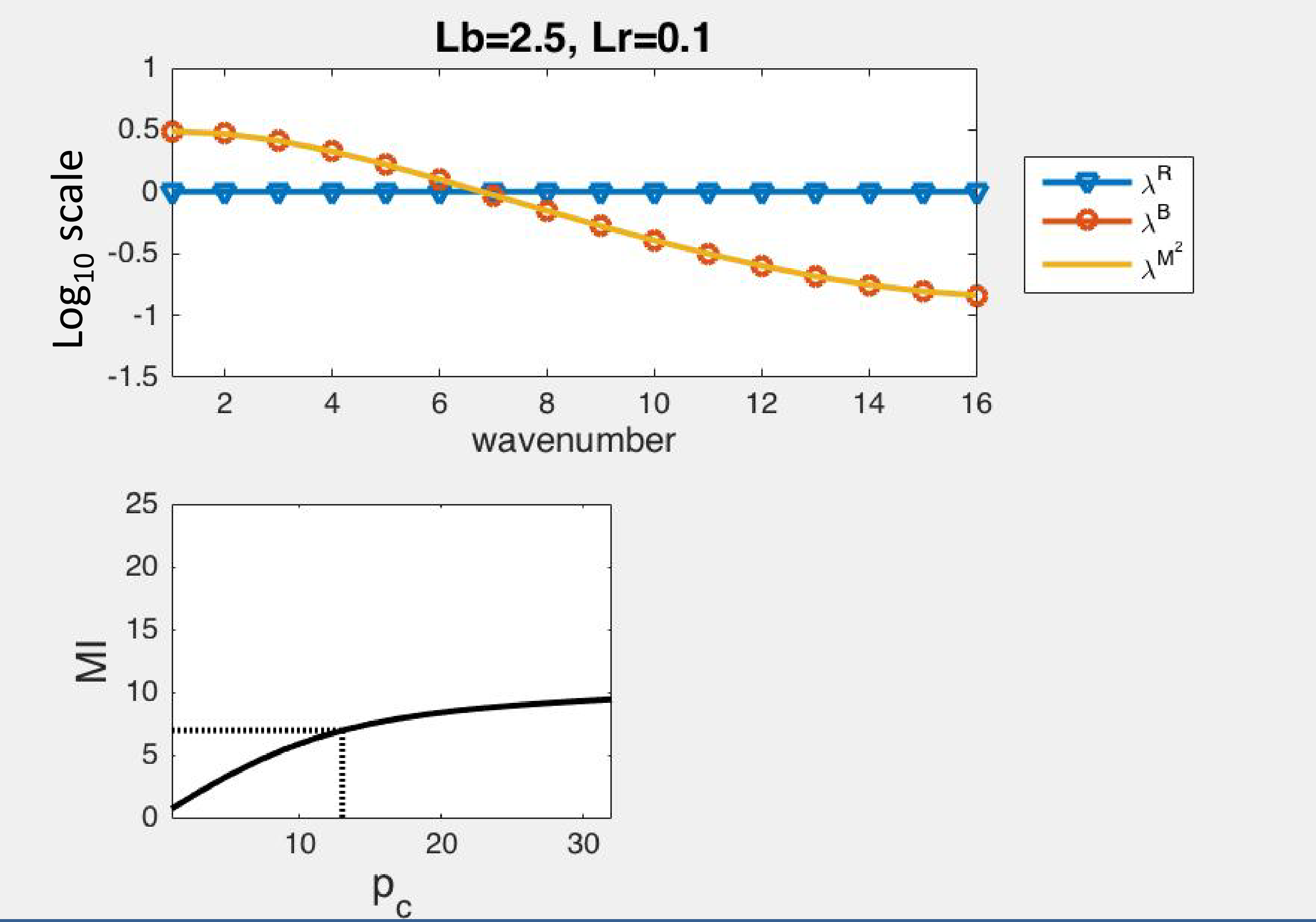
Isotropic, homogenous example…

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



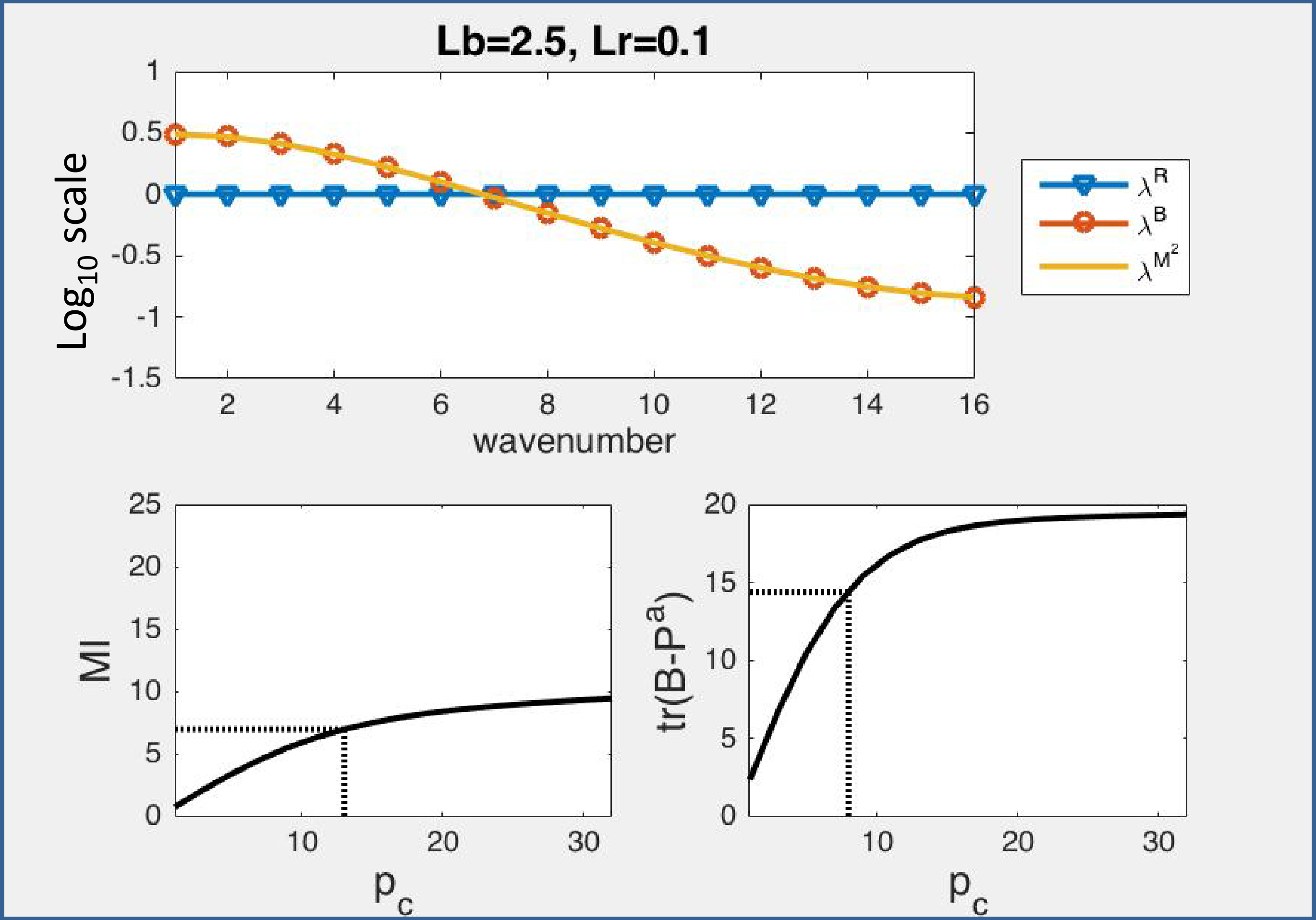
Isotropic, homogenous example…

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, Lr, increase the observations become more informative about the small scales.
- When Lr > Lb, 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

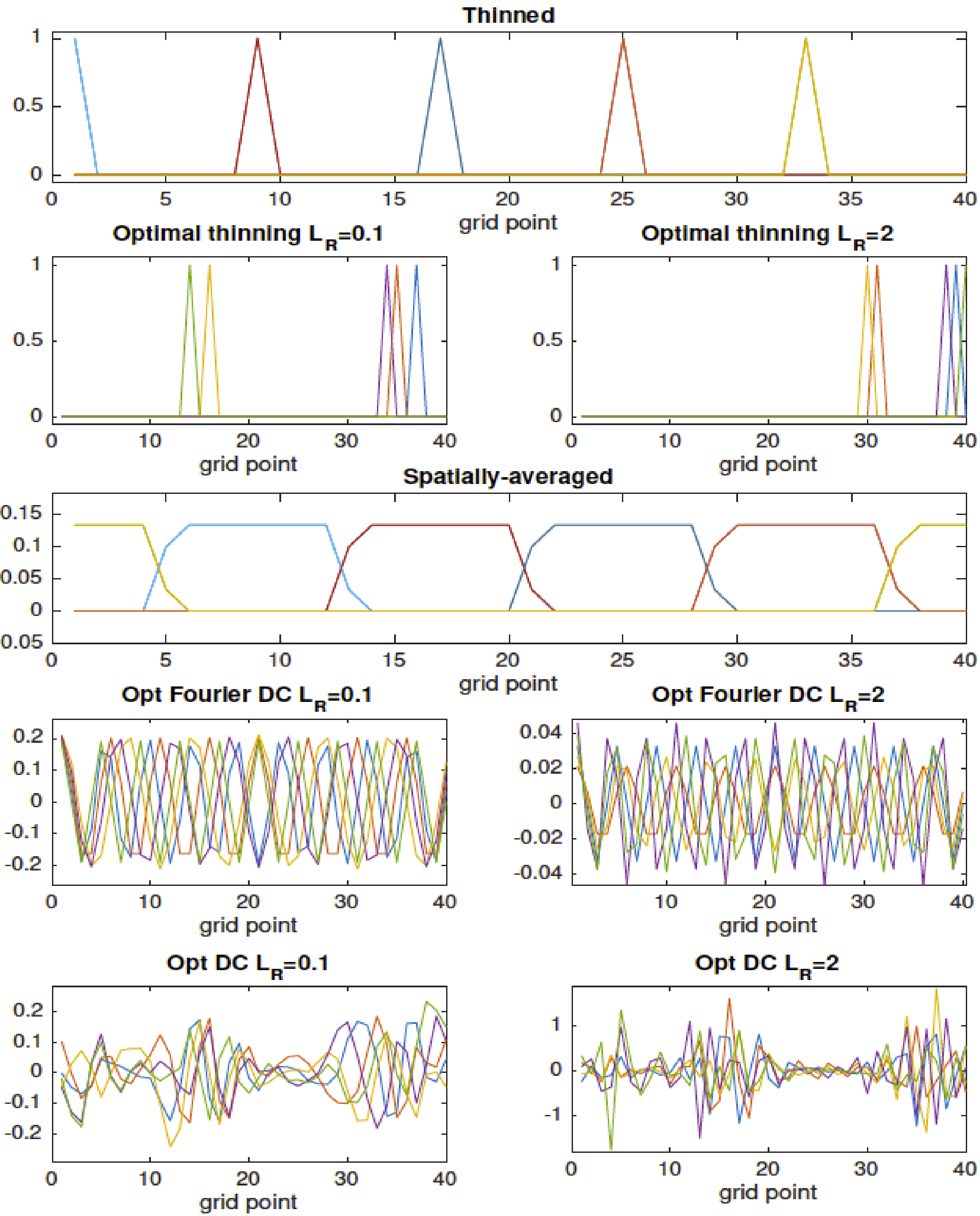
1. **Regular thinning:** to every 16th observation (giving 5 in total at each assimilation time).

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

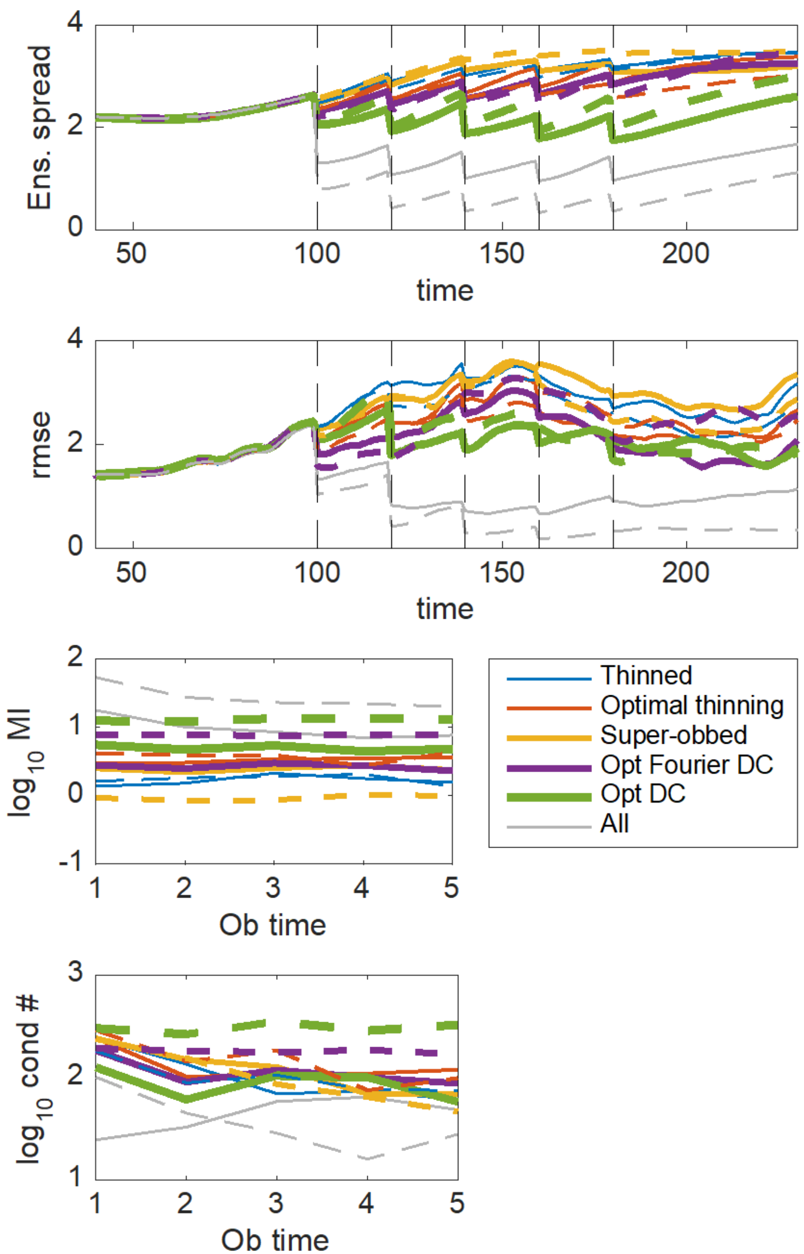


Fig: Results are averaged over 200 experiments with different realisations of the observation and model error.

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.

Open questions

- Reducing the cost of on-line data compression
  - Approximations to the optimal data compression.
  - Need for adaptive compression?
- Sensitivity to accuracy of R and P^f.
- Sensitivity to ensemble size.
  - Small ensemble size restricts info that can be provided by the obs- understand effect of e.g. localisation on data compression.

Thank you for listening