Information Aware Data Compression of High-Resolution Observations

Alison Fowler (Data Assimilation Research Centre, University of Reading)

NCEO & CEOI conference, Nottingham, Sept 2019.

Motivation

- Numerical weather prediction 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 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.
- We already rely on thinning observation data available for these reasons!

Questions to address

This work explores how to reduce the volume of high-resolution observations so that they can feasibly be assimilated frequently, but without throwing out the useful information they provide.

- How does the information content of the observations depend on the uncertainty in the observations?
- How can this be used to identify redundancy in the observation to reduce the volume of data with minimum information loss?
- What are the benefits of information-aware data compression versus simply thinning or averaging the data?

What are the implications of correlated observation errors?

- Correlated observation errors are only just starting to be accounted for in DA.
- They result from the way the observations are used in DA rather than uncertainty in the instrument, i.e. the uncertainty in relating what is observed to what is modeled (see Janjic et al. 2018).
- Many observations have significant spatially correlated errors, e.g. 200km for AMV (Corboda 2016), 20km for DRWs (Waller 2016) estimated in the Met Office's UKV 1.5km system.
- The longer the correlation length scales the more certain the observations are about the small scales and the less about the large-scales.

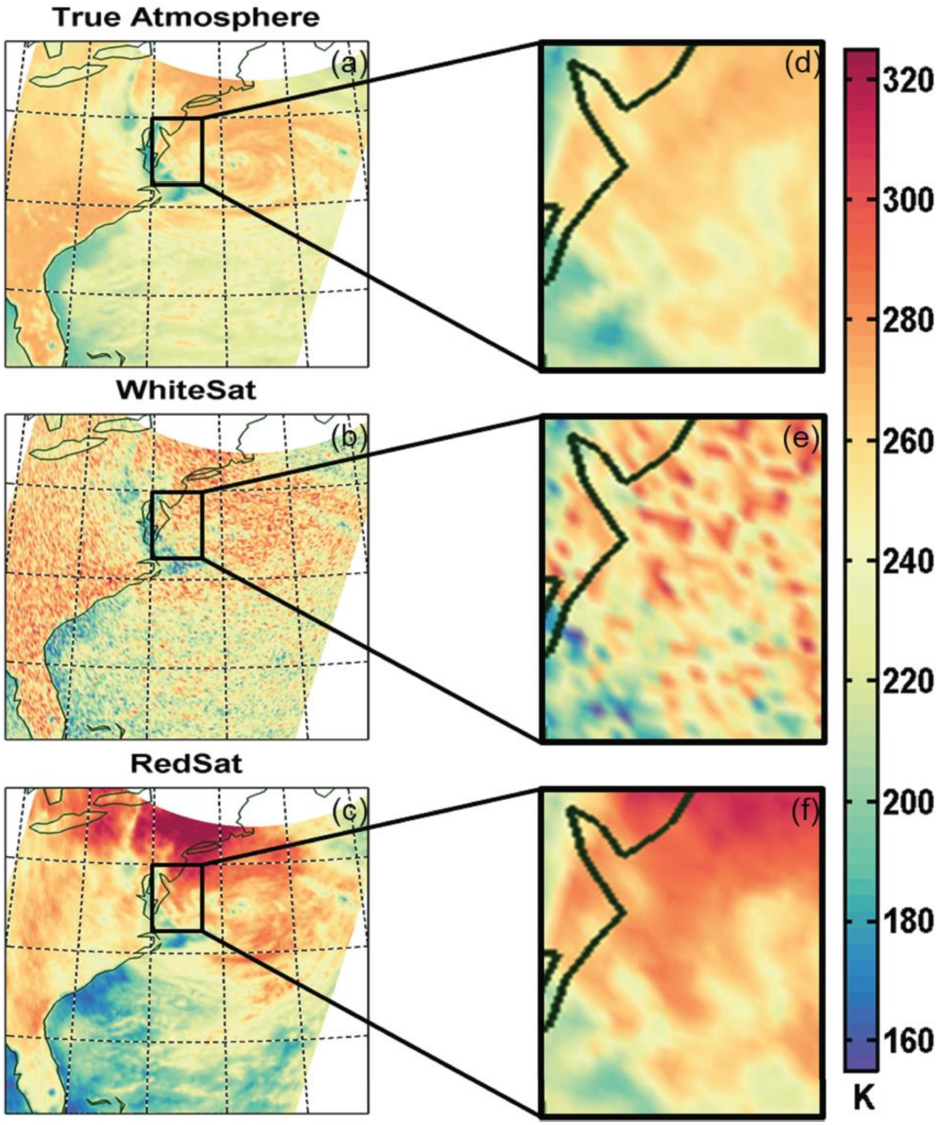


Fig: Microwave image of a hurricane.

Rainwater et al. 2015 QJRMS

Measuring Information Content

- Can define the information content of the observations, $\mathbf{y}$, in terms of the sensitivity of the analysis, $\mathbf{x}^{\mathrm{a}}$, to the observations

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

  - $\mathbf{K}$ is a function the error covariances of the observations and prior ($\mathbf{K}=\mathbf{B}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}+\mathbf{R})^{-1}$) and governs the weighting given to the observations versus the prior.
  - h, $\mathbf{H}$ are the non-linear, linearised mapping from model to observation space.
- **S** can be summarised in terms of mutual information (entropy reduction):

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

- The information in the observation depends on how the uncertainty in the observations (represented by $\mathbf{R}$) relates to the uncertainty in the prior (represented by $\mathbf{B}$).
- As well as the variances, Fowler et al. 2018 showed how the length scales in $\mathbf{B}$ and $\mathbf{R}$ interact to govern the information content of the observations. When the length scales in $\mathbf{R}$ are smaller (greater) than in **B**, the observations provide relatively more (less) information at the large scales compared to the small scales.

Information Aware Data Compression

- Instead of regular thinning, metrics of information content can be used to systematically identify redundancy in the data and provide an `information-aware' approach to reducing the volume of the data.
- This gives us the following approach to data compression:
  - Let $\mathbf{M} = \mathbf{R}^{-1/2}\mathbf{H}\mathbf{B}^{1/2} = \mathbf{U}\mathbf{\Lambda}^{\mathrm{M}}\mathbf{V}^{\mathrm{T}}$
  - Then $MI = 0.5 \ln det(\mathbf{I} + \mathbf{M}\mathbf{M}^{\mathrm{T}})$
  - Can compress the observations using $\mathbf{C} = \mathbf{I}^{\mathrm{c}}\mathbf{U}^{\mathrm{T}}\mathbf{R}^{-1/2}$. where $\mathbf{I}^{\mathrm{c}} \in \mathbb{R}^{p_{\mathrm{c}} \times p}$ 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 $\mathbf{I}^{\mathrm{c}}(\mathbf{I}^{\mathrm{c}})^{\mathrm{T}} = \mathbf{I}_{p_{\mathrm{c}}}$

Alison Fowler| a.m.fowler@reading.ac.uk | NCEO (University of Reading, Dept. Meteorology)

Discrepancy between information content and reduction in analysis error variance

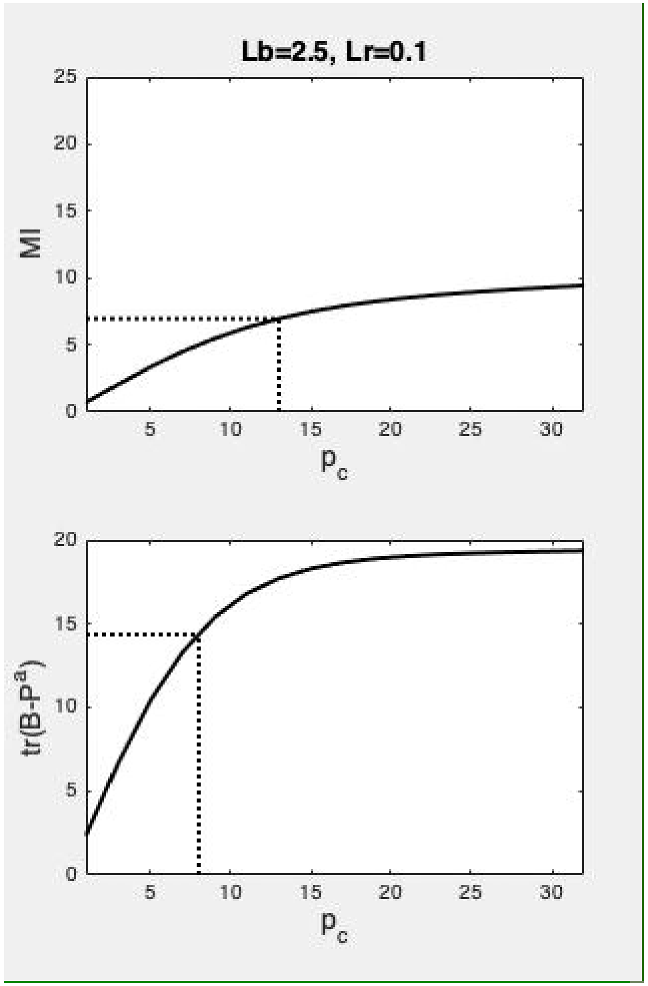
- Ordering the transformed observations w.r.t the singular values of $\mathbf{M}$ allows for the first p_c observations with the maximum information to be selected for assimilation.
- The information content of the remaining compressed observations becomes:

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

- The reduction in the analysis error variance compared to the prior is given by

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

  - This is not only a function of the singular values of **M**. Therefore the observation with the maximum information content will not necessarily provide the greatest reduction in analysis error variance.
- Is trace($\mathbf{B}$-$\mathbf{P}^a$) a good measure of the performance of the assimilation? Small scales may be important when the model is nested and the BC are not updated.



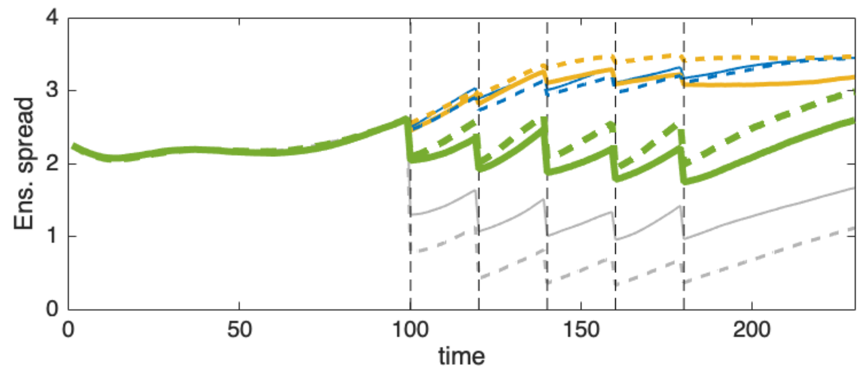
Comparison of data reduction techniques: Illustration with the EnKF and the Lorenz 1996 models

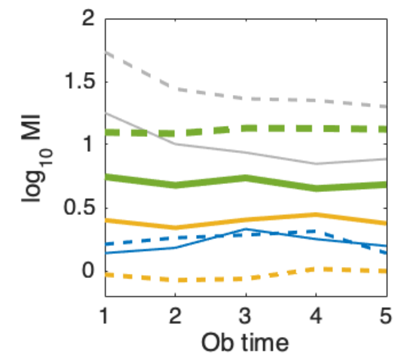
- Circular domain with 40 grid points

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

- F=8.
- 80 direct, regularly distributed observations of t state are simulated at 5 timesteps.
- Two different sets of observations with differen error characteristics:
  - Uncorrelated
  - Correlated
- Assimilation using EnSRF (Hunt et al. 2007).
- 100 ensemble members
- Averaged of 200 experiments

- Comparison of different methods to reduce the observations to 5 pieces of data.





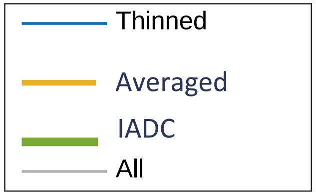


Figure: Adapted from Fowler, 2019. Tellus A.

Summary and Conclusions

- The assimilation of high-resolution observations is essential for successful high-resolution forecasting.
- When the volume of observational data is too high for assimilation, need to justify which data is retained.
- In experiments found:
    - The benefits of information aware data compression over thinning or averaging is greater when observations have significant error correlations compared to those in the model.
- When the small-scales are favoured by the data compression, there is a discrepancy between the observations with the highest information and those that provide the greatest reduction in the analysis error variances (equivalently analysis RMSE/ensemble spread). Are these useful metrics for high-resolution applications?

Alison Fowler| a.m.fowler@reading.ac.uk | NCEO (University of Reading, Dept. Meteorology)