Department of Meteorology

University of Reading

Data Assimilation Research Centre

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

Information Aware Data Compression of High-Resolution Observations

Alison Fowler | a.m..fowler@reading.ac.uk | www.met.reading.ac.uk/~gj901587

Introduction

Numerical weather prediction models are moving towards higher resolutions to capture the rapid development of convective-scale systems. To ensure the realism of high-resolution simulations there is a need to constrain the models with frequent high-resolution observations. The over-whelming volume of data provided by these new observing systems will bring many challenges:

- **high volume of data will make it difficult to transmit, store and assimilate the data in a timely manner.**
- **the data may have complicated error characteristics, such as non-negligible error correlations, that need to be represented.**

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.

Information Aware Data Compression (IADC)

The information content of the observations, $\mathbf{y}$, can be defined in terms of the sensitivity of the analysis, $\mathbf{x}^a$, to the observations

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

where $\mathbf{K}$ is a function the error covariances of the observations and prior, $\mathbf{R}$ and $\mathbf{B}$ respectively: $\mathbf{K}=\mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T+\mathbf{R})^{-1}$ and governs the weighting given to the observations versus the prior. h, $\mathbf{H}$ are the non-linear and linearised mapping from model to observation space.

$\mathbf{S}$ can be summarised in terms of mutual information (or equivalently the reduction in the posterior entropy compared to the prior entropy):

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

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 $\mathbf{M} = \mathbf{R}^{-1/2}\mathbf{H}\mathbf{B}^{1/2} = \mathbf{U}\Lambda^M\mathbf{V}^T$. Then $MI = 0.5\ln det(\mathbf{I} + \mathbf{M}\mathbf{M}^T)$.

The observations can be compressed using $\mathbf{C} = \mathbf{I}^c\mathbf{U}^T\mathbf{R}^{-1/2}$. $\mathbf{I}^c \in \mathbb{R}^{p_c \times p}$ is a matrix that picks out the p_c compressed observations to be retained for assimilation.

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

Implications of correlated observation errors

The information in the observations depends on how the uncertainty in the observations (represented by $\mathbf{R}$) relates to the uncertainty in the prior (represented by $\mathbf{B}$). In a nutshell observations that are accurate in regions of high prior uncertainty are the most informative.

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 $\mathbf{B}$, the observations provide relatively more (less) information at the large scales compared to the small scales.

Therefore**, the scales represented by the observations that are retained for assimilation using IADC will depend upon how the length scales in B and R relate**.

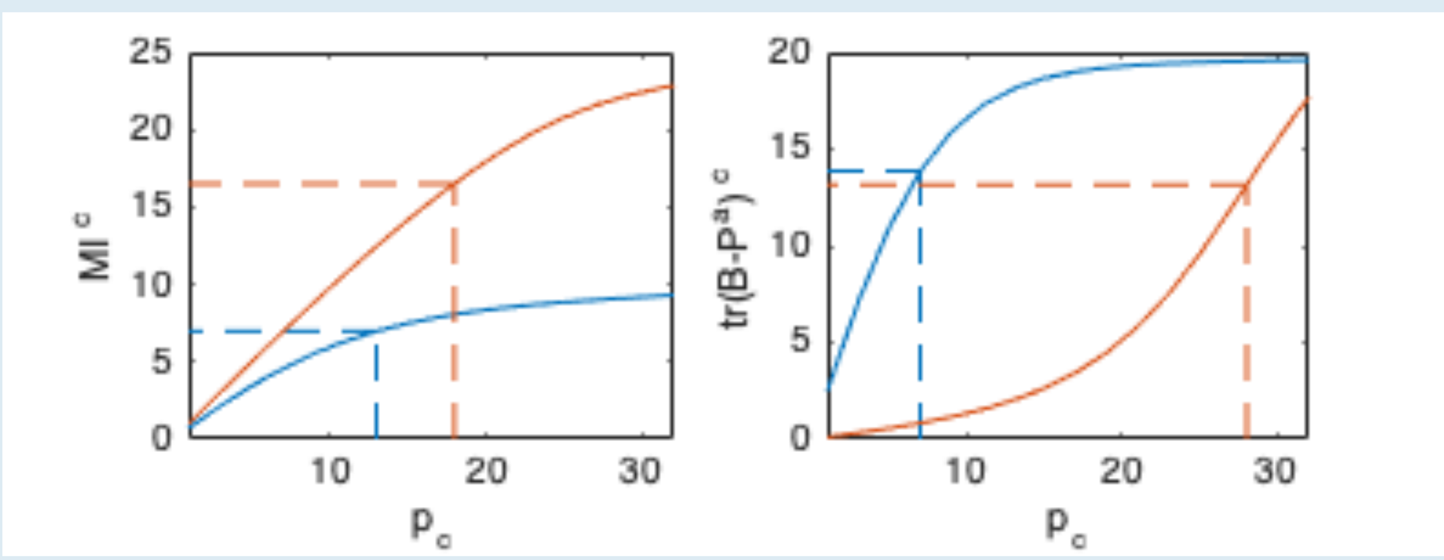

Fig 1. MI^c (left) and $tr(\mathbf{B}\text{-}\mathbf{P}^a)^c$ (right) vs number of compressed observations assimilated (p_c). $\mathbf{B}$ and $\mathbf{R}$ are given by circulant matrices with SOAR correlation functions. The length scales in $\mathbf{B}$ are 5 and the length scales in $\mathbf{R}$ are 0.1 (blue) or 10 (red). *Adapted from Fowler 2019.*

References * Fowler at al. 2018. On the interaction of observation and prior error correlations. QJRMS.
* Fowler 2019. Data compression in the presence of observation error correlations. Tellus.

Discrepancy between information content and error variance reduction

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 retained 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 due to the assimilation of the retained compressed observations is given by:

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

We see that the $trace(\mathbf{B}\text{-}\mathbf{P}^a)^c$ is not only a function of the singular values of $\mathbf{M}$ but also depends on the eigenvalues of $\mathbf{B}$, γ. Therefore, **the observations with the maximum information content will not necessarily provide the greatest reduction in analysis error variance**. This is illustrated in fig. 1: when the length scales in $\mathbf{R}$ are shorter than in $\mathbf{B}$ (blue line) large scale observations are more informative and also result in the greatest reduction in error variance. However, when the length scales in $\mathbf{R}$ are longer than in $\mathbf{B}$ (red line) small scale observations are more informative but these do not result in the greatest reduction in error variance.

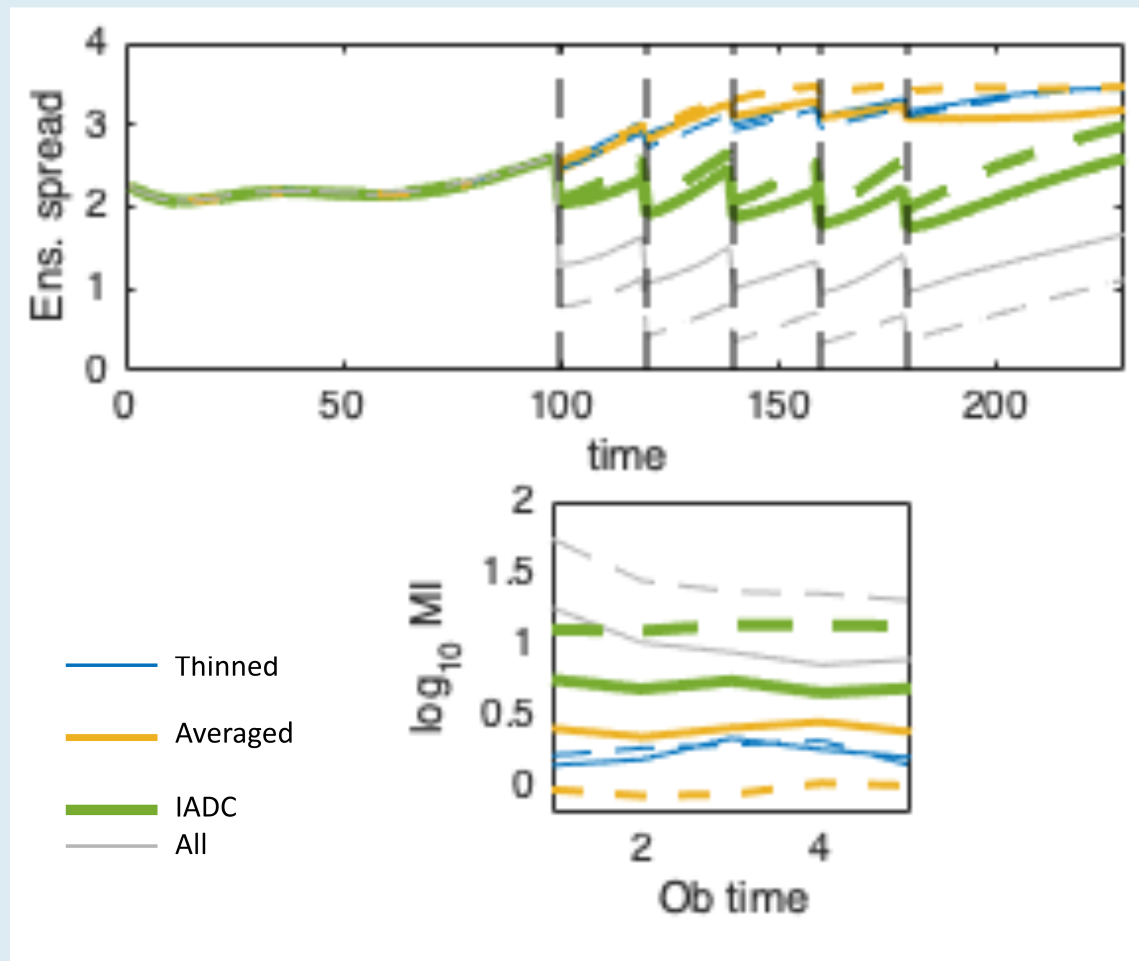

Fig 2. Information aware data compression compared to thinning and averaging the data when the observations have uncorrelated (sold line) and correlated (dashed line) errors. In each case the data is reduced from 80 to 5. *Adapted from Fowler 2019.*

Comparison of data reduction techniques

In fig. 2 IADC is applied to observations of Lorenz 96 model solved on the circular domain with 40 grid points, F=8. Two different sets of 80 direct, regularly distributed observations of the state are simulated at 5 time steps intervals.

In the top panel the ensemble spread resulting from the EnSRF (Hunt et al. 2007) with 100 ensemble members averaged over the 40 variables is plotted. In the bottom the *MI* is shown. The results are averaged over 200 experiments.

Can see that **the information content of the reduced data set is more sensitive to the reduction technique when the observations have correlated errors (dashed lines) compared to uncorrelated errors (solid lines).**

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, justification of which data to retain is necessary. 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?