Information-based data selection for ensemble data assimilation

Stefano Migliorini

National Centre for Earth Observation
University of Reading
s.migliorini@reading.ac.uk

International Conference on Ensemble Methods in Geophysical Sciences, Toulouse, 12-16 November 2012

Outline

Introduction

- It is well known that ensemble-based data assimilation can provide a spatially-inhomogeneous, time-dependent (and cyclical) estimate of the forecast error covariance, which can be used to achieve a more accurate assimilation of available observations.
- For increasing ensemble size, estimate converges to the BLUE (linear unbiased estimate with minimum variance) regardless whether the dynamical model or the observation operator are linear and whether the forecast or observation errors are Gaussian (Snyder, 2011). Note that conditional expectation of the posterior pdf may have smaller error variance.

Introduction (cont.)

- In practice, however, the analysis error variance underestimates the optimal analysis error variance estimated using an infinite number of ensemble members (Sacher and Bartello, 2007).
- Also, the sample covariance of forecast error $\tilde{\mathbf{P}}^f$ is rank deficient when $K < n + 1$, where K is the number of ensemble members and n is the dimension of the state space. This implies that the analysis increments can only belong to ran($\tilde{\mathbf{P}}^f$).
- It follows that ensemble filtering can lead to filter divergence , where the magnitude of the true analysis error becomes much larger than its estimate, as a result of the fact that observations are progressively ignored by the filter.

Observations and ensemble size

- Another consequence of using a rank-deficient forecast error covariance matrix is that at most $K-1$ degrees of freedom are available to ensemble-based data assimilation schemes in order to fit the observations (Lorenc, 2003).
- Observations that are sensitive to components of the state vector that do not belong to the range of $\tilde{\mathbf{P}}^f$ do not improve the analysis estimate.
- Localization procedures ease the rank-deficiency problem as the localized $\tilde{\mathbf{P}}^f$ is only supposed to represent the covariance of the local forecast error.

University of Reading

Observations and ensemble size (cont.)

- The localization radius of influence should be large enough not to disturb the balances that act at given spatial scales and that are well represented by the ensemble error covariance (e.g., Lorenc, 2003).
- The radius of influence should also be large enough to include enough observations to constrain the analysis effectively. At the same time, a radius of influence that is too large may not substantially reduce the number of assimilated observations, particularly over data-dense areas.
- A data selection strategy based on the information content of the measurements is here illustrated, which ensures that only the observational components that are able to constrain the analysis are assimilated using ensemble filtering techniques.

University of Reading

Ensemble square-root filtering

- Consider observation $\mathbf{y}^o = H(\mathbf{x}^t) + \epsilon^o$ with unit error covariance, to determine estimate $\mathbf{x}^a$ (analysis) of $\mathbf{x}$. The analysis error covariance $\mathbf{P}^a$ is related to the forecast error covariance $\mathbf{P}^f$ according to the Kalman filter solution of the cycling problem for a linear(ized) stochastic-dynamic system and given by

$$\mathbf{P}^a = \mathbf{P}^f - \mathbf{P}^f\mathbf{H}^T(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{I}_m)^{-1}\mathbf{H}\mathbf{P}^f \tag{1}$$

- We now approximate $\mathbf{P}^{f,a}$ with $\tilde{\mathbf{P}}^{f,a} \equiv \mathbf{X}'^{f,a}\mathbf{X}'^{f,a^T}$, where

$$\mathbf{X}'^{f,a} = \frac{1}{\sqrt{K-1}}(\mathbf{x}_1^{f,a} - \overline{\mathbf{x}^{f,a}}, \mathbf{x}_2^{f,a} - \overline{\mathbf{x}^{f,a}}, \cdots, \mathbf{x}_i^{f,a} - \overline{\mathbf{x}^{f,a}}, \cdots, \mathbf{x}_K^{f,a} - \overline{\mathbf{x}^{f,a}}) \tag{2}$$

- The ensemble transform Kalman filter (ETKF, Bishop et al., 2001) expresses $\mathbf{X}'^a$ as

$$\mathbf{X}'^a = \mathbf{X}'^f\mathbf{T} \in \mathbb{R}^{n\times K} \tag{3}$$

where $\mathbf{T} \in \mathbb{R}^{K\times K}$ is the ensemble transform matrix, to be determined.

Ensemble square-root filtering (cont.)

- To determine an expression for $\mathbf{T}$, we define $\tilde{\mathbf{S}} \in \mathbb{R}^{m\times K}$ as

$$\tilde{\mathbf{S}} = \frac{1}{\sqrt{K-1}}\left(H(\mathbf{x}_1^f)-\overline{H(\mathbf{x}^f)}, \cdots, H(\mathbf{x}_i^f)-\overline{H(\mathbf{x}^f)}, \cdots, H(\mathbf{x}_K^f)-\overline{H(\mathbf{x}^f)}\right)$$

$$\overline{H(\mathbf{x}^f)} \equiv \frac{1}{K}\sum_{i=1}^{K} H(\mathbf{x}_i^f).$$

so that $\tilde{\mathbf{S}}\tilde{\mathbf{S}}^T \simeq \mathbf{H}\mathbf{P}^f\mathbf{H}^T \equiv \mathbf{S}\mathbf{S}^T$.

- we can write

$$\tilde{\mathbf{P}}^a = \mathbf{X}'^f(\mathbf{I}_K - \tilde{\mathbf{S}}^T(\tilde{\mathbf{S}}\tilde{\mathbf{S}}^T + \mathbf{I}_m)^{-1}\tilde{\mathbf{S}})\mathbf{X}'^{fT}. \tag{4}$$

Ensemble square-root filtering (cont.)

It is possible to express $\tilde{\mathbf{S}}$ as $\tilde{\mathbf{S}} = \tilde{\mathbf{E}}\tilde{\mathbf{\Gamma}}\tilde{\mathbf{V}}^T$, where $\tilde{\mathbf{E}} \in \mathbb{R}^{m\times m}$, $\tilde{\mathbf{\Gamma}} \in \mathbb{R}^{m\times K}$ and $\tilde{\mathbf{V}} \in \mathbb{R}^{K\times K}$. In this way, we can write

$$\tilde{\mathbf{P}}^a = \mathbf{X}'^f\tilde{\mathbf{V}}(\tilde{\mathbf{Y}}_{K,\tilde{p}} + \mathbf{I}_K)^{-1}\tilde{\mathbf{V}}^T\mathbf{X}'^{fT} \tag{5}$$

where

$$\tilde{\mathbf{\Gamma}} = \begin{pmatrix} \tilde{\mathbf{\Gamma}}_{\tilde{p}} & \mathbf{0}_{\tilde{p}\times(K-\tilde{p})} \\ \mathbf{0}_{(m-\tilde{p})\times\tilde{p}} & \mathbf{0}_{(m-\tilde{p})\times(K-\tilde{p})} \end{pmatrix} \tag{6}$$

and

$$\tilde{\mathbf{Y}}_{K,\tilde{p}} \equiv \begin{pmatrix} \tilde{\mathbf{\Gamma}}^2_{\tilde{p}} & \mathbf{0}_{\tilde{p}\times(K-\tilde{p})} \\ \mathbf{0}_{(K-\tilde{p})\times\tilde{p}} & \mathbf{0}_{K-\tilde{p}} \end{pmatrix} \in \mathbb{R}^{K\times K} \tag{7}$$

with $\tilde{p} = \mathrm{rank}(\tilde{\mathbf{S}}) \leq \min(K-1, m)$. It follows that $\mathbf{T}$ can be written as

$$\mathbf{T} = \tilde{\mathbf{V}}(\tilde{\mathbf{Y}}_{K,\tilde{p}} + \mathbf{I}_K)^{-1/2}\tilde{\mathbf{V}}^T \in \mathbb{R}^{K\times K} \tag{8}$$

Information considerations

- From 8 follows that error reduction due to informative obs is only along the directions of the state space spanned by the $\tilde{p}$ right singular vectors of $\tilde{\mathbf{S}}$ with positive singular values. When $\mathbf{S}$ is approximated by $\tilde{\mathbf{S}}$ there are only $\tilde{p} \leq \min(m, K-1)$ measurements that provide information, i.e., with $\tilde{\gamma}_i > 0$, so that the effective number of degrees of freedom for signal $\tilde{d}_s$ resulting from the use of a reduced-rank forecast error covariance can be written as (Rodgers, 2000; D. Zupanski et al., 2007)

$$\tilde{d}_s = \mathrm{tr}(\tilde{\mathbf{S}}^T(\tilde{\mathbf{S}}\tilde{\mathbf{S}}^T + \mathbf{I}_m)^{-1}\tilde{\mathbf{S}}) = \sum_{i=1}^{\tilde{p}} \frac{\tilde{\gamma}_i^2}{1+\tilde{\gamma}_i^2} \tag{9}$$

- It follows that for a given number of ensemble members K, there are at most $K-1$ components of the measurement vector $\mathbf{y}^o$ that can provide information.

Information considerations (cont.)

- The importance of this consideration is that it is now possible to decide whether a given observational component is worth assimilating, according to whether one of these equivalent conditions are met:
  - its signal-to-noise ratio $\tilde{\gamma}_i$ is greater than about 1,
  - its information content $H_i = \frac{1}{2}\log_2(1 + \tilde{\gamma}_i^2)$ is greater than about 0.5 or
  - it provides more than about half a degree of freedom for signal.

  The effects of choosing different threshold values should be tested (see later).
- It follows that when $m \gg K$, only the $r < K$ leading singular values and vectors of $\tilde{\mathbf{S}}$ need to be determined for assimilation

Data selection strategy

- Let us define $\mathbf{y}^{o\prime} \in \mathbb{R}^r$ as $\mathbf{y}^{o\prime} \equiv \tilde{\mathbf{E}}_r^T \mathbf{y}^o$, where $\tilde{\mathbf{E}}_r \in \mathbb{R}^{m \times r}$ is the matrix whose columns are the r left singular vectors corresponding to the r positive singular values of $\tilde{\mathbf{S}}$ that are greater than about unity, with $r \leq \tilde{p}$. We can write

$$\mathbf{y}^{o\prime} = \tilde{\mathbf{E}}_r^T H(\mathbf{x}^t) + \tilde{\mathbf{E}}_r^T \epsilon^o = H'(\mathbf{x}^t) + \epsilon^{o\prime} \tag{10}$$

  where $H'(\mathbf{x}^t) \in \mathbb{R}^{r \times n}$ is defined as $H'(\mathbf{x}^t) \equiv \tilde{\mathbf{E}}_r^T H(\mathbf{x}^t)$. Note that the covariance of $\epsilon^{o\prime}$ is $\mathbf{I}_r$, the unit matrix of rank r.

- The analysis error covariance can now be written as

$$\tilde{\mathbf{P}}^a = \mathbf{X}'^f (\mathbf{I}_K - \tilde{\mathbf{S}}'^T (\tilde{\mathbf{S}}' \tilde{\mathbf{S}}'^T + \mathbf{I}_r)^{-1} \tilde{\mathbf{S}}') \mathbf{X}'^{fT} \tag{11}$$

  where $\tilde{\mathbf{S}}' \in \mathbb{R}^{r \times K}$ is defined as $\tilde{\mathbf{S}}' \equiv \tilde{\mathbf{E}}_r^T \tilde{\mathbf{S}}$.

University of Reading

Data selection strategy (cont.)

- It follows that the analysis perturbation matrix can be written as

$$\mathbf{X}'^{a} = \mathbf{X}'^{f}\tilde{\mathbf{V}}(\tilde{\mathbf{Y}}_{K,r} + \mathbf{I}_K)^{-1/2}\tilde{\mathbf{V}}^T \tag{12}$$

where $\tilde{\mathbf{Y}}_{K,r} \in \mathbb{R}^{K\times K}$ is defined as

$$\tilde{\mathbf{Y}}_{K,r} \equiv \begin{pmatrix} \tilde{\mathbf{\Gamma}}_r^2 & \mathbf{0}_{r\times(K-r)} \\ \mathbf{0}_{(K-r)\times r} & \mathbf{0}_{K-r} \end{pmatrix}. \tag{13}$$

- The analysis ensemble mean can be calculated as

$$\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^f + \mathbf{X}'^{f}\tilde{\mathbf{S}}'^T(\tilde{\mathbf{S}}'\tilde{\mathbf{S}}'^T + \mathbf{I}_r)^{-1}(\mathbf{y}^{o\prime\prime} - \mathbf{H}''\overline{\mathbf{x}}^f) \tag{14}$$

$$= \overline{\mathbf{x}}^f + \mathbf{X}'^{f}\tilde{\mathbf{V}}_r\tilde{\mathbf{\Gamma}}_r(\tilde{\mathbf{\Gamma}}_r^2 + \mathbf{I}_r)^{-1}(\mathbf{y}^{o\prime\prime} - \mathbf{H}''\overline{\mathbf{x}}^f). \tag{15}$$

An analogous expression can be found for the ensemble Kalman filter algorithm.

Localization considerations

- The data selection strategy presented here is compatible (and should be used) with localization procedures for ensemble data assimilation. When localization is used, the data selection procedure will result in a further data reduction over the local domain or over the domain where the compactly-supported correlation function is different from zero.
- Localization procedures over data-dense areas do not need to restrict the magnitude of their ROI as a way to reduce the amount of observations to be assimilated.
- Appropriate dimension of the local domain from trade off between the need of reducing the rank deficiency of the forecast error covariance matrix for a given K and of avoiding shortening the natural correlation length scales that may lead to unbalanced initial conditions.

Numerical experiments

- Two-dimensional temperature advection model on a circle of latitude. Zonal-only advection speed $u(\mathbf{x}, t)$ can be constant or dependent on $T(\mathbf{x}, t)$ via the thermal wind equation. Forward-upstream finite difference scheme. The zonal length of the domain is 1000, with 43 vertical levels (0.1 – 1013.25 hPa), 150+ time steps (longer exp for nonlinear case).
- Initial condition for the truth from random field with Gaussian horizontal correlation function ($\sigma^2 = 20$) and an exponential vertical correlation with 50 km de-correlation length.
- Initial conditions for the "background" trajectory are defined from the same random field, but with expectation given by the true state at initial time. The K members of the initial ensemble are then created in a similar manner, with expectation given by the background state.

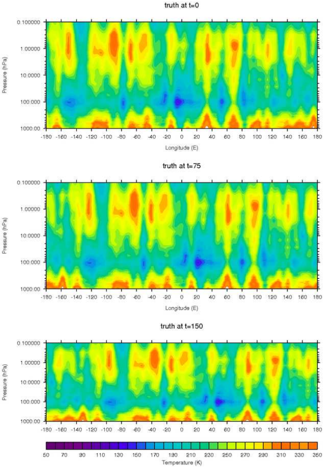
University of Reading

Assimilation strategy

- Each initial condition propagated forward in time until observation time, when an analysis scheme based either on a standard (e.g., Evensen, 2004, 2009) or on the data-selective ensemble square-root method generates a new set of initial conditions.
- Two sets of observations: a) 8 regularly-spaced vertical temperature profiles with 43 elements; b) satellite radiances (66 channels) over 8 regularly-spaced locations; all at $5\Delta t$ observation frequency.
- All observations are simulated from the truth and zero-mean random noise with given standard deviation. For temperature profiles: $\sigma^o_{T_j} = 0.1\% T^f_j$ at initial time.

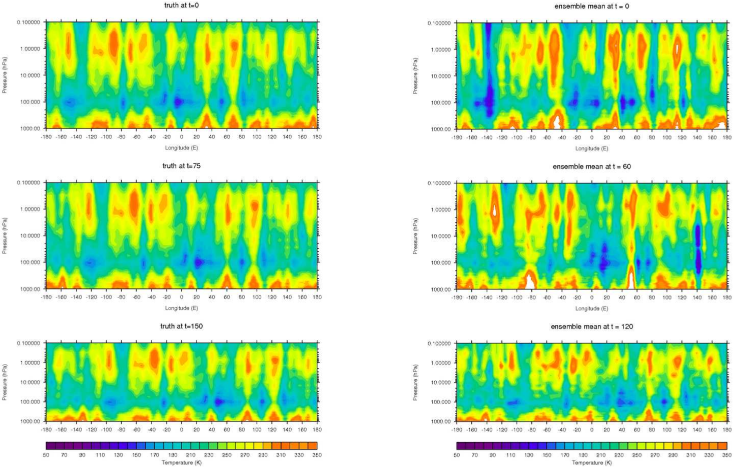
Results without data selection

- $n = 1001 \times 43 = 43043$, $K = 300$, no localization, $43 \times 8 = 344$ obs every 5 Δt, $T = 120\Delta t$



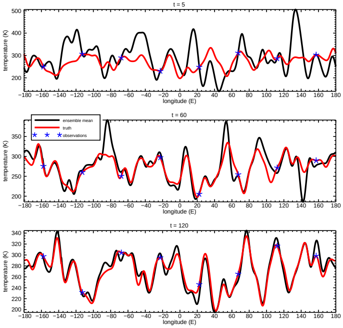
Results with data selection

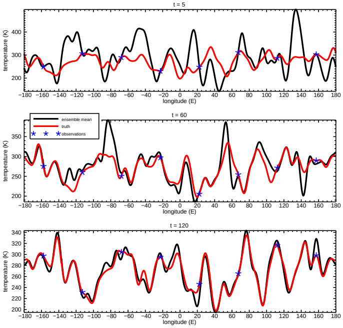
- $n = 1001 \times 43 = 43043$, $K = 300$, no localization, $SNR > 1$, $43 \times 8 = 344$ obs every $5\ \Delta t$, $T = 120\Delta t$



Results at a model level

- $SNR > 1$ (left), all data (right), $K = 300$, no localization, ~ 500 hPa

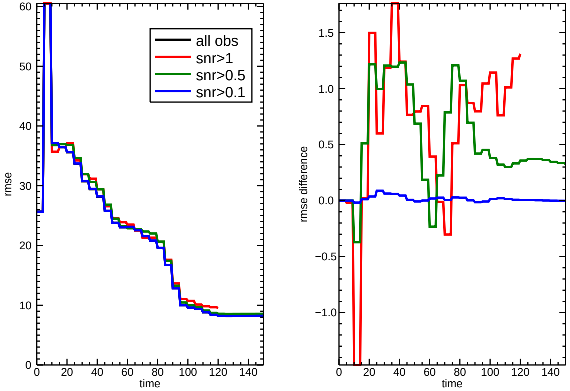




University of Reading

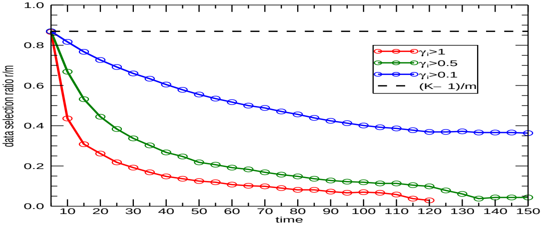
RMSE

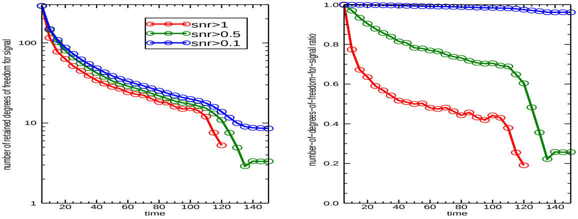
- $SNR > \alpha$ (left), $SNR > \alpha - SNR > 0$ (right), $\alpha = 0.1, 0.5, 1$, $K = 300$, no localization



Data selection and d_s ratio

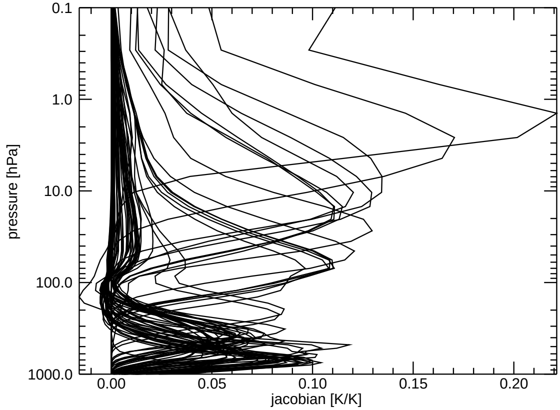
- $SNR > \alpha$, $\alpha = 0.1, 0.5, 1$, $K = 300$, no localization, 344 obs





Remote sounding data

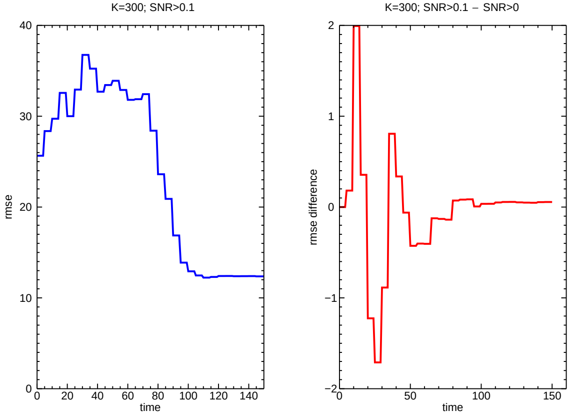
- IASI temperature jacobians (66 channels), channel selection from Collard (2007); noise stddev 0.22-0.37 K.



University of Reading

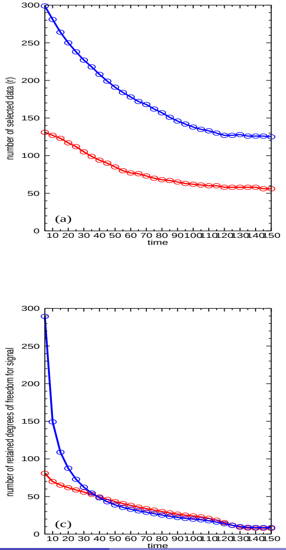
Remote sounding data RMSE

- $SNR > 0.1$ (left), $SNR > 0.1 - SNR > 0$ (right), $K = 300$, no localization



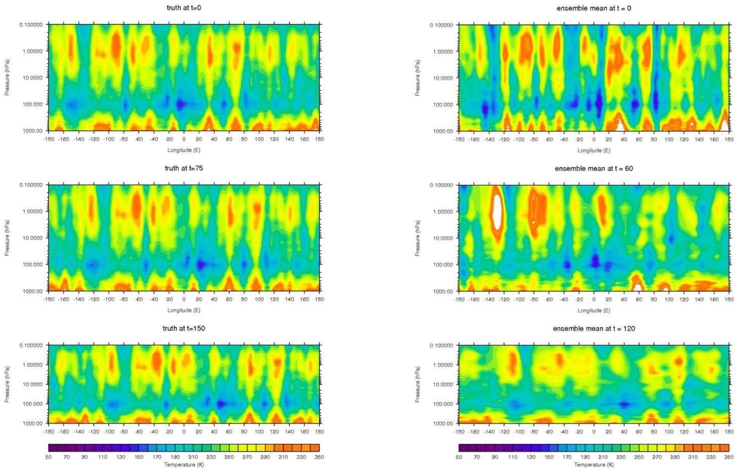
IASI data selection and d_s ratio

- $SNR > 0.1$, $K = 300$, no localization, 528 obs (66×8)



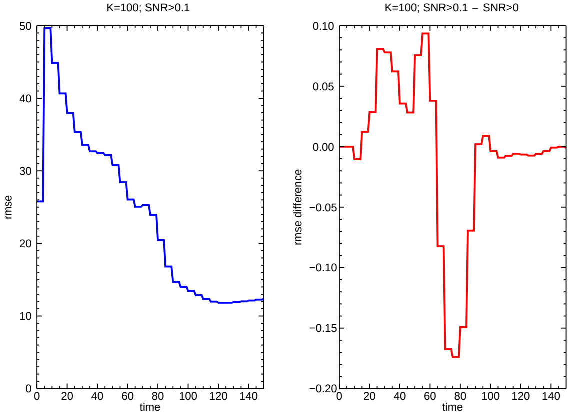
Localization results with data selection

- Temperature profiles, $n = 1001 \times 43 = 43043$, $K = 100$, $SNR > 0.1$, ROI=200



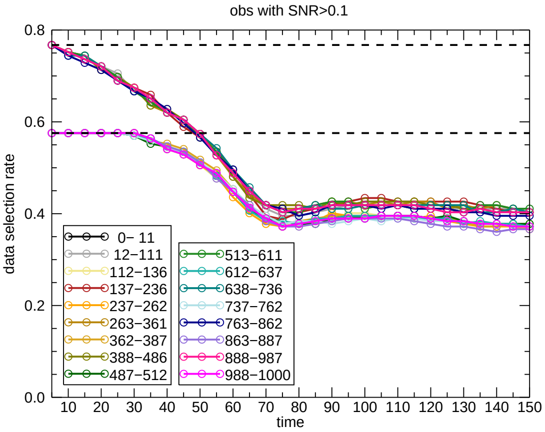
Localization results: RMSE

- Temperature profiles, $SNR > 0.1$ (left), $SNR > 0.1 - SNR > 0$ (right), $K = 100$, ROI=200



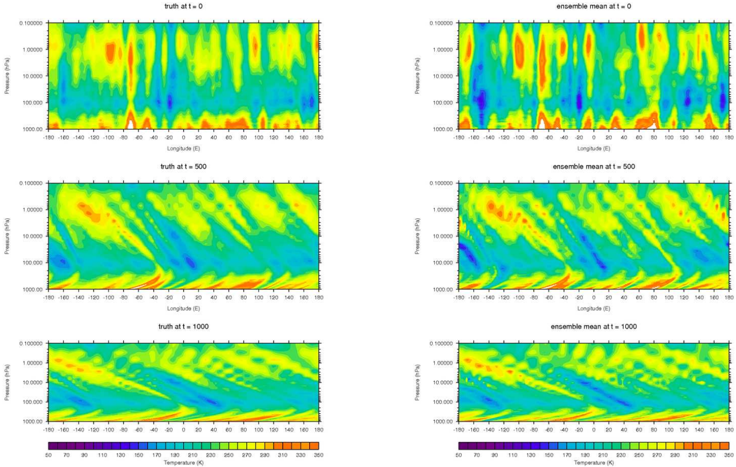
Data selection rate

- $SNR > 0.1$, $K = 100$, localization ROI=200
- For ROI=200 and 125 obs separation, domain subdivided in 17 regions including either 129 (3×43) or 172 (4×43) obs



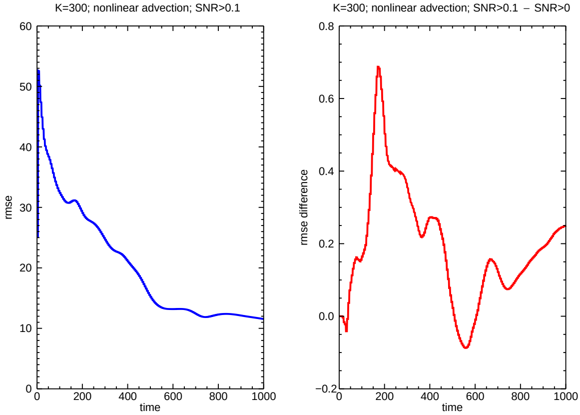
Results with nonlinear advection

- $n = 1001 \times 43 = 43043$, $K = 300$, no localization, $43 \times 8 = 344$ obs every $5\,\Delta t$, $T = 1000\Delta t$



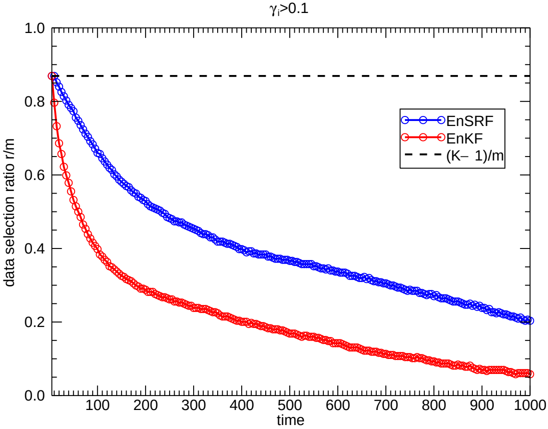
Nonlinear advection results: RMSE

- Temperature profiles, $SNR > 0.1$ (left), $SNR > 0.1 - SNR > 0$ (right), $K = 300$, no localization



Data selection ratio: EnSRF v. EnKF

- Nonlinear advection, $SNR > 0.1$, $K = 300$, no localization



Conclusions

- An effective and physically-based method to address the ensemble filtering shortcomings in the case when $m \gg K$ is described.
- Results with a temperature advection model when assimilating remote sounding data using the data-selection procedure show that it is possible to discard more than 75% of the components of the observation vector without significantly affecting the accuracy of the results. Only about 40% of the in situ data components are retained for assimilation towards the end of the run, even at low SNR thresholds.
- Can be used with both in situ and remote sounding data, and it is attractive for operational NWP applications with ensemble-based or hybrid DA schemes as it may lead to more balanced initial conditions at larger scales, over data-dense areas.
- QJ paper under revision.

University of Reading