Information-based data selection for ensemble data assimilation

Stefano Migliorini

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

Outline

University of Reading

Introduction (1/2)

- A necessary condition for an ideal assimilation is that the true state should be statistically indistinguishable from any analysis ensemble members that are randomly sampled from the posterior (or analysis) *pdf* resulting from data assimilation.
- A key shortcoming of ensemble filtering, due to its representation of the posterior *pdf* with a limited number of analysis ensemble members, is that 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($\tilde{\mathbf{P}}^f$).

University of Reading

Introduction (2/2)

- 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.
- Sampling error may also lead to a misrepresentation of forecast error covariance values between two different locations, and this can be particularly detrimental when long-range spatial correlations are overestimated, leading to spurious analysis increments.
- To minimize these shortcomings, ensemble filtering usually make use of procedures such as covariance inflation and covariance localization.

Localisation and Inflation

- Inflation methods aim to enlarge the spread of the forecast ensemble either by multiplying the ensemble member perturbations from the mean by $\alpha > 1$ (Anderson and Anderson 1999) or by adding random perturbations to the ensemble members (e.g., Whitaker et al. 2008; Houtekamer et al. 2009). Adaptive multiplicative inflation schemes have also been introduced (Anderson, 2007b; Anderson, 2009 and Li et al., 2009).
- Localization techniques aim to eliminate long-range correlations either via an element-wise multiplication (or Schur product) of $\tilde{\mathbf{P}}^f$ with compactly supported correlation matrix (see Houtekamer and Mitchell 1998 and 2001; Janijc et al., 2011 for square-root filters) or by estimating the analysis on a local domain (e.g., Ott et al. 2004, Hunt et al., 2007) using only observations within a given radius of influence from each grid point.
- Adaptive localization techniques have also been proposed (Anderson, 2007a and Bishop and Hodyss (2009a,b))

University of Reading

Observations and ensemble size (1/2)

- 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.
- Both distance-dependent or Schur-product 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.

Observations and ensemble size (2/2)

- However, the 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 proposed, which ensures that only the observational components that are able to constrain the analysis are assimilated using ensemble filtering techniques.

University of Reading

Characterisation of the measurements (1/2)

- The relationship between a measurement vector $\mathbf{y}^o \in \mathbb{R}^q$ and the true state $\mathbf{x}^t \in \mathbb{R}^n$ of a system (e.g., the atmosphere) can be expressed as
$$\mathbf{y}^o = H(\mathbf{x}^t) + \epsilon^o \tag{1}$$
where $H(\mathbf{x}^t)$ is the observation operator calculated in $\mathbf{x}^t$ and where $\epsilon^o \in \mathbb{R}^q$ is the measurement error, assumed Gaussian, unbiased and with covariance $\mathbf{R} \in \mathbb{R}^{q \times q}$.
- The observation operator can be linearised about a given $\mathbf{x}_i$ and write
$$\mathbf{y}^o \simeq H(\mathbf{x}_i) + \mathbf{H}^{(i)}(\mathbf{x}^t - \mathbf{x}_i) + \epsilon^o \tag{2}$$
where $\mathbf{H}^{(i)} \equiv (\partial H / \partial \mathbf{x})_{\mathbf{x}=\mathbf{x}_i} \in \mathbb{R}^{q \times n}$ is the Jacobian matrix of $H(\mathbf{x})$ calculated in $\mathbf{x} = \mathbf{x}_i$. We can also define $\mathbf{y}^{(i)}$ as (e.g., Migliorini, 2011)
$$\mathbf{y}^{o(i)} \equiv \mathbf{y}^o - H(\mathbf{x}_i) + \mathbf{H}^{(i)}\mathbf{x}_i \simeq \mathbf{H}^{(i)}\mathbf{x}^t + \epsilon^o. \tag{3}$$

University of Reading

Characterisation of the measurements (1/2)

- $\mathbf{R}$ can be expressed in terms of its eigenvector decomposition as $\mathbf{R} = \mathbf{L}\boldsymbol{\Lambda}\mathbf{L}^T$. When the number m of non-zero (or not too small) eigenvalues is less than q, we can write $\mathbf{E} \simeq \mathbf{L}_m\boldsymbol{\Lambda}_m\mathbf{L}_m^T$. For $m \leq q$ we define $\mathbf{y}^{o\prime} = \boldsymbol{\Lambda}_m^{-1/2}\mathbf{L}_m^T\mathbf{y}^o$ so that from (3) we can write

$$\mathbf{y}^{o\prime} = \boldsymbol{\Lambda}_m^{-1/2}\mathbf{L}_m^T\mathbf{H}\mathbf{x}^t + \boldsymbol{\Lambda}_m^{-1/2}\mathbf{L}_m^T\epsilon^o = \mathbf{H}'\mathbf{x}^t + \epsilon^{o\prime} \tag{4}$$

  where $\mathbf{H}' \in \mathbb{R}^{m\times n}$ is defined as $\mathbf{H}' \equiv \boldsymbol{\Lambda}_m^{-1/2}\mathbf{L}_m^T\mathbf{H}'$ and where the covariance $\epsilon^{o\prime}$ is the unit matrix of rank m.

- Finally, an alternative definition of $\mathbf{y}^{o\prime}$ that preserves the nonlinear relationship with $\mathbf{x}^t$ (when applicable) is given by

$$\mathbf{y}^{o\prime} = \boldsymbol{\Lambda}_m^{-1/2}\mathbf{L}_m^T H(\mathbf{x}^t) + \boldsymbol{\Lambda}_m^{-1/2}\mathbf{L}_m^T\epsilon^o = H'(\mathbf{x}^t) + \epsilon^{o\prime}. \tag{5}$$

University of Reading

Ensemble square-root filtering (1/3)

- The ensemble transform Kalman filter (ETKF, Bishop et al., 2001), which is the ensemble square root filter we will concentrate on, provides an approximation of $\mathbf{X}^a$ by means of the analysis perturbations matrix $\mathbf{X}'^a$, calculated as

$$\mathbf{X}'^a = \mathbf{X}'^f \tilde{\mathbf{T}} \in \mathbb{R}^{n \times K} \tag{6}$$

where

$$\mathbf{X}'^f = \frac{1}{\sqrt{K-1}}(\mathbf{x}_1^f - \overline{\mathbf{x}^f}, \mathbf{x}_2^f - \overline{\mathbf{x}^f}, \cdots, \mathbf{x}_i^f - \overline{\mathbf{x}^f}, \cdots, \mathbf{x}_K^f - \overline{\mathbf{x}^f}) \in \mathbb{R}^{n \times K} \tag{7}$$

with K being the number of ensemble forecast members $\mathbf{x}_i^f$, and where $\tilde{\mathbf{T}}$ is a suitable approximation of $\mathbf{T}$.

- In this way, approximations of $\mathbf{P}^f$ and $\mathbf{P}^a$ are given by $\tilde{\mathbf{P}}^f = \mathbf{X}'^f \mathbf{X}'^{fT}$ and $\tilde{\mathbf{P}}^a = \mathbf{X}'^f \tilde{\mathbf{T}} \tilde{\mathbf{T}}^T \mathbf{X}'^{fT}$, respectively.

University of Reading

Ensemble square-root filtering (2/3)

- 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 stochastic-dynamic system and given by

$$\mathbf{P}^a = \mathbf{P}^f - \mathbf{P}^f \mathbf{H}'^T (\mathbf{H}'\mathbf{P}^f \mathbf{H}'^T + \mathbf{I}_m)^{-1} \mathbf{H}'\mathbf{P}^f \tag{8}$$

- To determine an expression for $\tilde{\mathbf{T}}$, we define $\tilde{\mathbf{S}} \equiv \mathbf{H}'\mathbf{X}'^f \in \mathbb{R}^{m \times K}$ so that $\tilde{\mathbf{P}}^a \simeq \mathbf{P}^a$ can be written as

$$\tilde{\mathbf{P}}^a = \mathbf{X}'^f (\mathbf{I}_K - \tilde{\mathbf{S}}^T (\tilde{\mathbf{S}}\tilde{\mathbf{S}}^T + \mathbf{I}_m)^{-1} \tilde{\mathbf{S}}) \mathbf{X}'^{fT}. \tag{9}$$

- Note that it is possible to avoid linearising the observation operator as in (3) if we define $\mathbf{y}^{o\prime}$ and $H'(\mathbf{x}^t)$ as in (5) and

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

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

Ensemble square-root filtering (3/3)

It is possible to express $\tilde{\mathbf{S}}$ as $\tilde{\mathbf{S}} = \tilde{\mathbf{E}}\tilde{\boldsymbol{\Gamma}}\tilde{\mathbf{V}}^T$, where $\tilde{\mathbf{E}} \in \mathbb{R}^{m\times m}$, $\tilde{\boldsymbol{\Gamma}} \in \mathbb{R}^{m\times K}$ and $\tilde{\mathbf{V}} \in \mathbb{R}^{K\times K}$. In this way, (9) can be expressed as

$$\tilde{\mathbf{P}}^a = \mathbf{X}'^f\tilde{\mathbf{V}}(\tilde{\mathbf{Y}}_K + \mathbf{I}_K)^{-1}\tilde{\mathbf{V}}^T\mathbf{X}'^{fT} \tag{10}$$

where

$$\tilde{\boldsymbol{\Gamma}} = \begin{pmatrix} \tilde{\boldsymbol{\Gamma}}_{\tilde{p}} & \mathbf{0}_{\tilde{p}\times(K-\tilde{p})} \\ \mathbf{0}_{(m-\tilde{p})\times\tilde{p}} & \mathbf{0}_{(m-\tilde{p})\times(K-\tilde{p})} \end{pmatrix} \tag{11}$$

and

$$\tilde{\mathbf{Y}}_K \equiv \begin{pmatrix} \tilde{\boldsymbol{\Gamma}}^2_{\tilde{p}} & \mathbf{0}_{\tilde{p}\times(K-\tilde{p})} \\ \mathbf{0}_{(K-\tilde{p})\times\tilde{p}} & \mathbf{0}_{K-\tilde{p}} \end{pmatrix} \tag{12}$$

with $\tilde{p} = \mathrm{rank}(\tilde{\mathbf{S}}) \leq \min(K-1, m)$. From (6) it follows that $\tilde{\mathbf{T}}$ can be written as

$$\tilde{\mathbf{T}} = \tilde{\mathbf{V}}(\tilde{\mathbf{Y}}_K + \mathbf{I}_K)^{-1/2}\tilde{\mathbf{V}}^T \in \mathbb{R}^{K\times K} \tag{13}$$

where we have chosen a symmetric form of the ensemble transform matrix $\tilde{\mathbf{T}}$ so as to ensure that $\mathbf{X}'^a$ is unbiased (e.g., Wang et al., 2004; Sakov and Oke, 2008; Livings et al., 2008).

University of Reading

Information considerations (1/2)

- 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)

$$\tilde{d}_s = \mathrm{tr}(\tilde{\mathbf{S}}^T(\tilde{\mathbf{S}}\tilde{\mathbf{S}}^T + \mathbf{I}_m)^{-1}\tilde{\mathbf{S}}) = \sum_{i=1}^{\tilde{p}} \frac{\tilde{\gamma}_i^2}{1+\tilde{\gamma}_i^2} \tag{14}$$

- It follows that for a given number of ensemble members K, there are at most $K-1$ components of the measurement vector $\mathbf{y}^{o\prime}$ that can provide information. Note that the above result is consistent with the discussion provided in Lorenc (2003), where the special case of a perfect observation is considered.

Information considerations (2/2)

- 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.
- 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 (1/2)

- Let us define $\mathbf{y}^{o''} \in \mathbb{R}^r$ as $\mathbf{y}^{o''} \equiv \tilde{\mathbf{E}}_r^T \mathbf{y}^{o'}$, 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 $r \leq \tilde{p}$. From (4) we can write

$$\mathbf{y}^{o''} = \tilde{\mathbf{E}}_r^T \mathbf{H}' \mathbf{x}^t + \tilde{\mathbf{E}}_r^T \epsilon^{o'} = \mathbf{H}'' \mathbf{x}^t + \epsilon^{o''} \tag{15}$$

  where $\mathbf{H}'' \in \mathbb{R}^{r \times n}$ is defined as $\mathbf{H}'' \equiv \tilde{\mathbf{E}}_r^T \mathbf{H}'$. Note that the covariance of $\epsilon^{o''}$ is $\mathbf{I}_r$, the unit matrix of rank r.
- From (9), the analysis error covariance can now be written as

$$\tilde{\mathbf{P}}^a = \mathbf{X}'^f (\mathbf{I}_K - \tilde{\mathbf{S}}'^T (\tilde{\mathbf{S}}' \tilde{\mathbf{S}}'^T + \mathbf{I}_r)^{-1} \tilde{\mathbf{S}}') \mathbf{X}'^{fT} \tag{16}$$

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

University of Reading

Data selection strategy (2/2)

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

$$\mathbf{X}'^a = \mathbf{X}'^f \tilde{\mathbf{V}} (\tilde{\mathbf{Y}}'_K + \mathbf{I}_K)^{-1/2} \tilde{\mathbf{V}}^T \tag{17}$$

where, in analogy with (12), $\tilde{\mathbf{Y}}_K$ is defined as

$$\tilde{\mathbf{Y}}'_K \equiv \begin{pmatrix} \tilde{\mathbf{\Gamma}}_r^2 & \mathbf{0}_{r\times(K-r)} \\ \mathbf{0}_{(K-r)\times r} & \mathbf{0}_{K-r} \end{pmatrix}. \tag{18}$$

- The analysis ensemble mean can be calculated as

$$\begin{aligned} \overline{\mathbf{x}}^a &= \overline{\mathbf{x}}^f + \mathbf{X}'^f \tilde{\mathbf{S}}'^T (\tilde{\mathbf{S}}' \tilde{\mathbf{S}}'^T + \mathbf{I}_r)^{-1} (\mathbf{y}^{o\prime\prime} - \mathbf{H}'' \overline{\mathbf{x}}^f) && (19) \\ &= \overline{\mathbf{x}}^f + \mathbf{X}'^f \tilde{\mathbf{V}}_r \tilde{\mathbf{\Gamma}}_r (\tilde{\mathbf{\Gamma}}_r^2 + \mathbf{I}_r)^{-1} (\mathbf{y}^{o\prime\prime} - \mathbf{H}'' \overline{\mathbf{x}}^f). && (20) \end{aligned}$$

University of Reading

Discussion

- The data selection strategy presented here is compatible with any existing localisation procedure that may be used for ensemble data assimilation. When localisation 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 can then have a larger ROI or correlation functions whose support spans a larger part of their domain, while always keeping the number of measurements to be considered for assimilation below K.
- 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

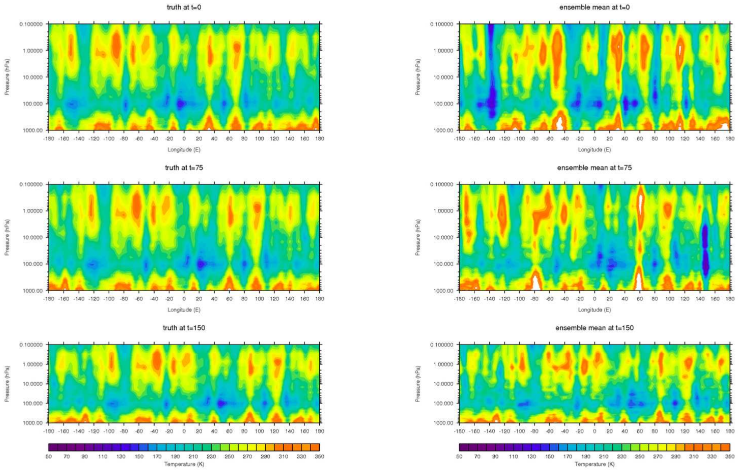
- Two-dimensional linear temperature advection model on a circle of latitude. Constant zonal-only advection speed $u = 1$, $\Delta x = 1$ and $\Delta t = 1$ ($C = 1$). Forward-upstream finite difference scheme (no model error). The zonal length of the domain is 1000, with 43 vertical levels (0.1 – 1013.25 hPa), 150 time steps.
- Initial condition for the truth from random field with Gaussian horizontal correlation function ($\sigma = 10\sqrt{2}$) 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.

Assimilation strategy

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

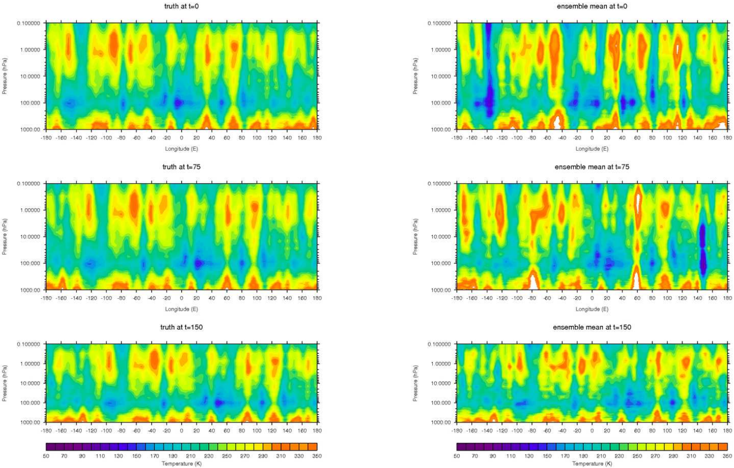
Results with data selection

- $n = 1001 \times 43 = 43043$, $K = 300$, no localisation, $SNR > 1$, $43 \times 8 = 344$ obs every $5\ \Delta t$, $T = 150\Delta t$



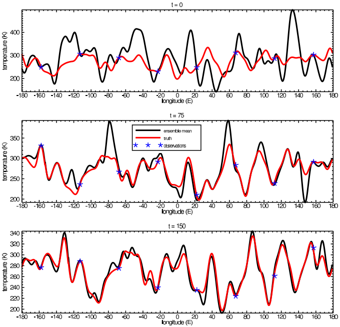
Results without data selection

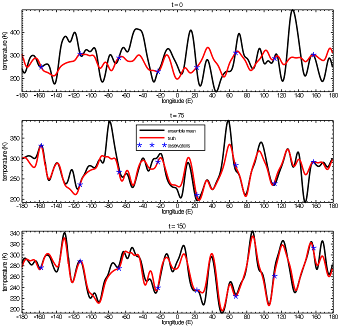
- $n = 1001 \times 43 = 43043$, $K = 300$, no localisation, $43 \times 8 = 344$ obs every $5\,\Delta t$, $T = 150\Delta t$



Results at a model level

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

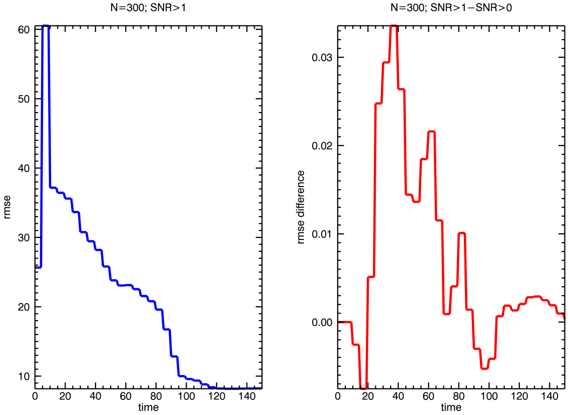




University of Reading

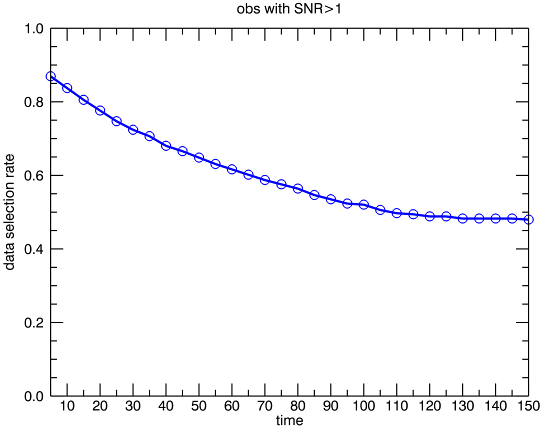
RMSE

- $SNR > 1$ (left), $SNR > 1 - SNR > 0$ (right), $K = 300$, no localisation



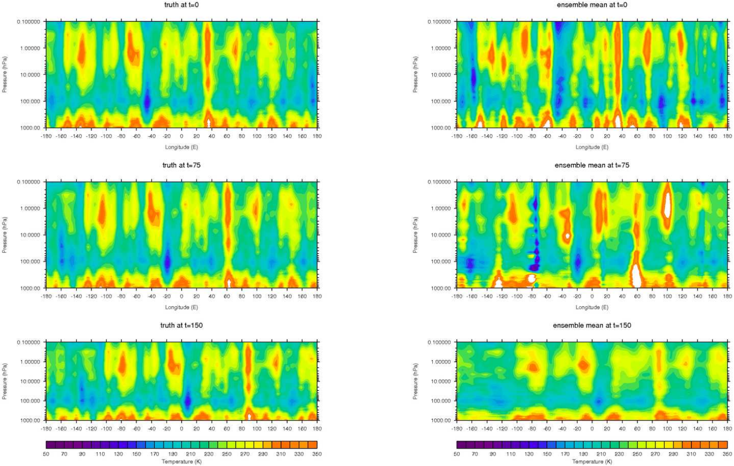
Data selection rate

- $SNR > 1$, $K = 300$, no localisation, 344 obs



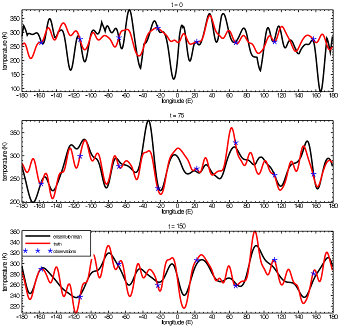
Results with data selection

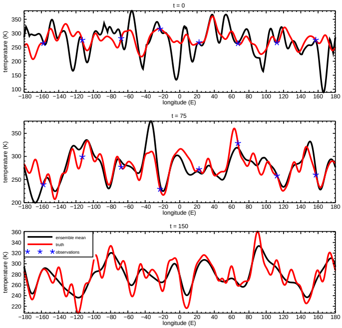
- $n = 1001 \times 43 = 43043$, $K = 100$, $SNR > 1$, localisation ROI=200



Results at a model level

- $SNR > 1$ (left), all data (right), $K = 100$, $\sim$ 500 hPa, localisation ROI=200

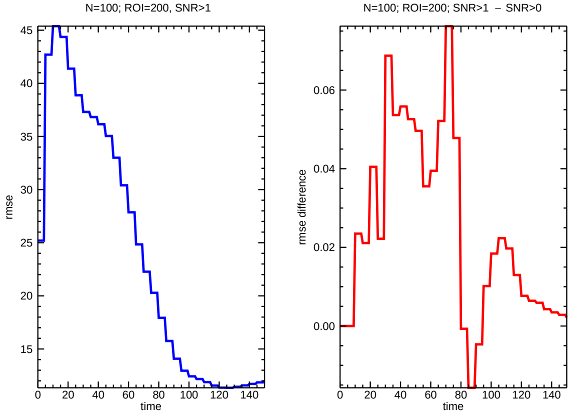




University of Reading

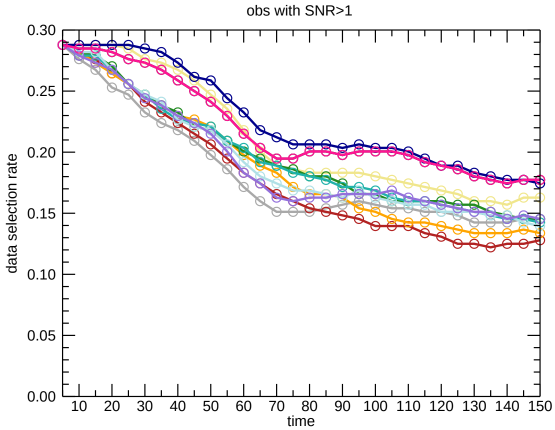
RMSE

- $SNR > 1$ (left), $SNR > 1 - SNR > 0$ (right), $K = 100$, localisation ROI=200



Data selection rate

- $SNR > 1$, $K = 100$, localisation ROI=200
- number of local obs: (172, 129, 172, 129, 129, 129, 129, 129, 172, 129, 172)



Conclusions

- A effective and physically-based method to address the ensemble filtering shortcomings in the case when $m \gg K$ is described.
- Results with an idealized model show that it is possible to use only about 30% of the components of the observation vector from the first assimilation cycle without significantly affecting the results.
- Can be used with both in situ and remote sounding data, and is particularly suited for operational NWP applications.
- QJ paper about to be submitted.

University of Reading