Underdispersiveness of EnKF posterior spread when $p \gg K$ as revealed by diagnostics of information content measured by Degrees of Freedom for Signal (DFS)

Daisuke Hotta* and Yoichiro Ota
Japan Meteorological Agency

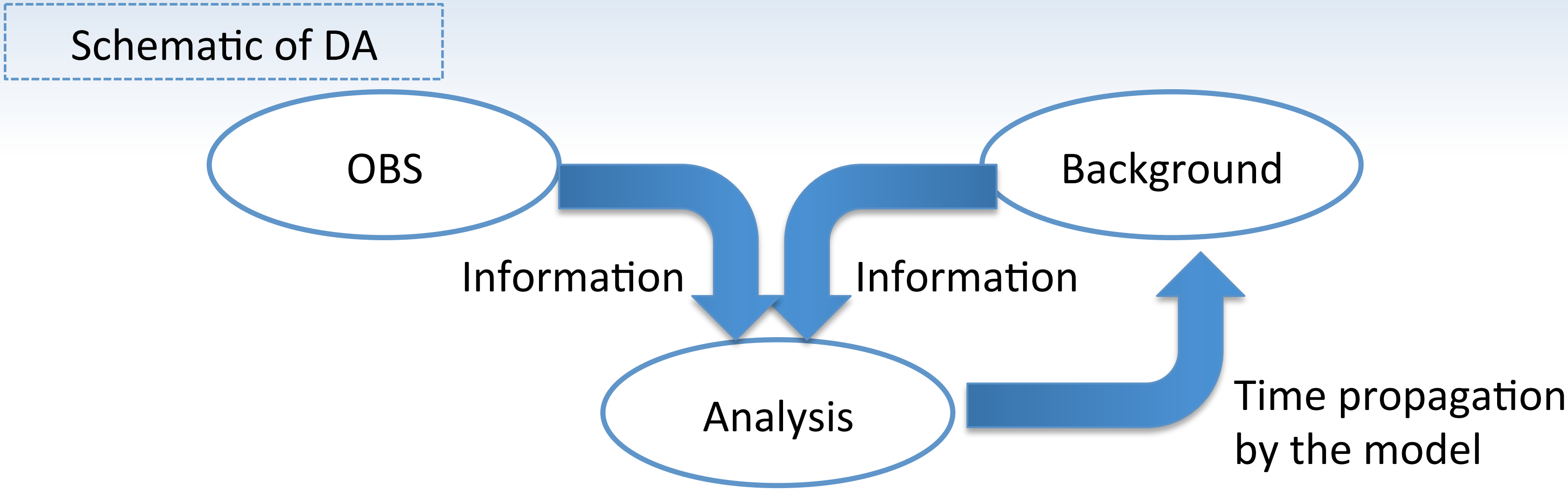
7th EnKF Workshop
May 25, 2016, State College, PA, USA

Use of information theory to understand why posterior spread is almost always underestimated

Daisuke Hotta* and Yoichiro Ota
Japan Meteorological Agency

7th EnKF Workshop
May 25, 2016, State College, PA, USA

1. Motivation



- Information carried to analysis: How much from obs ? How much from background?

→One way to quantify this: **Degrees of Freedom for Signal (DFS, or *information content*).**

2. What is DFS?

- Defined as the trace $\mathrm{tr}(\mathbf{S})$ of the "influence matrix" $\mathbf{S} = (\mathbf{HK})^{\mathbf{T}} = \frac{\partial \mathbf{y^a}}{\partial \mathbf{y^o}}$

- Shown to behave similarly to Shannon entropy reduction under some conditions (Fisher 2003, ECMWF tech memo #397):

$$\mathrm{tr}(\mathbf{S}) \approx \left[\, H\left(\mathbf{x}|\mathbf{x^b}\right) - H\left(\mathbf{x}|\mathbf{x^b}, \mathbf{y^o}\right) \,\right] \times \text{const.}$$

- Two ways to interpret:
  1. Analysis sensitivity to observations measured in obs space.
  2. The amount of information that the analysis extracted from observations.

Simple illustrative examples:

- **Forecast-Forecast cycle**: analysis is always the same as the background.
  - $\mathbf{y^a} \equiv \mathbf{y^b} \rightarrow \mathbf{S}$ is null, DFS=tr(S) = 0 (**0% information from obs.**)
- **Direct Insertion:** background is completely replaced by the obs.
  - $\mathbf{y^a} \equiv \mathbf{y^o} \rightarrow \mathbf{S}$ is identity, DFS = tr(S) = #obs
  - DFS per obs = 1 (**100% information comes from obs.**)

2. What is DFS?

- First introduced to NWP by Fisher (2003) and Cardinali et al. (2004)
- Popular diagnostics for variational DA systems.
  - Routinely monitored by several NWP centers (e.g. ECMWF, Météo-France)
- Liu et al. (2009) derived a simple method to compute DFS for EnKF:

$$\mathbf{S} = \frac{\partial \mathbf{y}^{\mathbf{a}}}{\partial \mathbf{y}^{\mathbf{o}}} = (\mathbf{HK})^{\mathbf{T}} = \mathbf{R}^{-\mathbf{1}}\mathbf{HAH}^{\mathbf{T}} \approx \frac{\mathbf{1}}{\mathbf{K}-\mathbf{1}}\mathbf{R}^{-\mathbf{1}}(\mathbf{Y}^{\mathbf{a}})(\mathbf{Y}^{\mathbf{a}})^{\mathbf{T}}$$

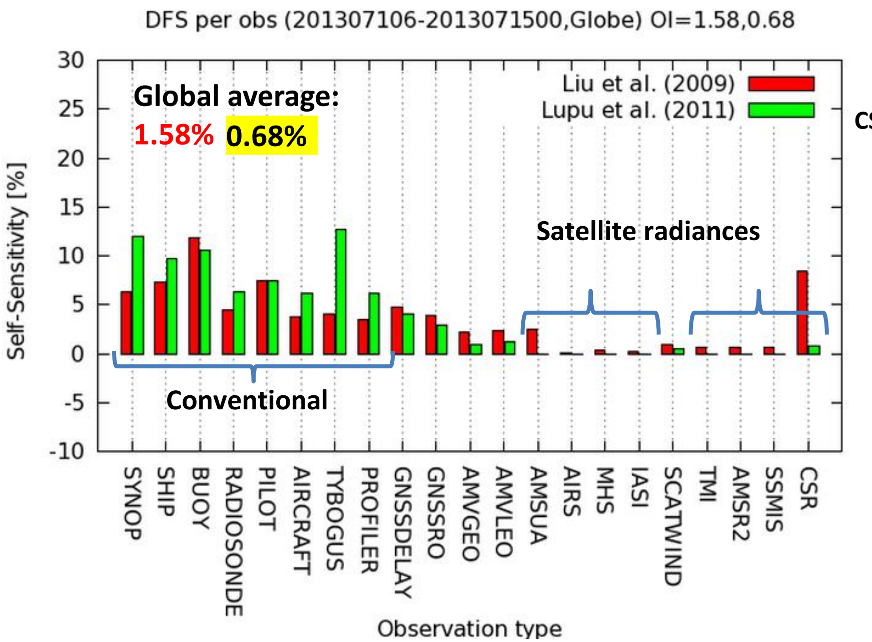
- Verified in Liu et al. (2009) with a simple AGCM (SPEEDY) in an idealized “identical-twin” scenario, but
- Up to present, not yet applied to operational Ensemble DA with real observations.

3. Ensemble-based DFS diagnostics at JMA
3-1. Experimental set-up

- DA system: hybrid LETKF/4D-Var coupled with JMA GSM
  - Resolution: (outer) TL959L100 ; (inner and ensemble) T319L100
  - Window: 6 hours (analysis time +/- 3 hours)
  - B weights: 77% from static, 23% from ensemble
  - Member size: 50
  - Localization scales (e-folding):
    - LETKF: Horizontal: 400km, Vertical: 0.4 scale heights
    - 4D-Var: Horizontal: 800km, Vertical: 0.8 scale heights
  - Covariance Inflation: Adaptive inflation of Miyoshi (2011)
- DFS estimation Algorithms:
  - Liu et al. (2009) $\frac{\mathbf{1}}{\mathbf{K-1}}\mathrm{tr}\left(\mathbf{R}^{-\mathbf{1}}(\mathbf{Y}^{\mathbf{a}})(\mathbf{Y}^{\mathbf{a}})^{\mathbf{T}}\right)$
  - also tried the residual-based method of Lupu et al. (2011) as a double check:
    - $\mathrm{tr}(\mathbf{HK}) = \mathrm{tr}\left(\widetilde{\mathbf{R}}^{-1}\mathbb{E}\left(\mathbf{d}_{\mathrm{b}}^{\mathrm{a}}(\mathbf{d}_{\mathrm{a}}^{\mathrm{o}})^{\mathrm{T}}\right)\right), \widetilde{\mathbf{R}} = \mathbb{E}\left(\mathbf{d}_{\mathrm{a}}^{\mathrm{o}}(\mathbf{d}_{\mathrm{b}}^{\mathrm{o}})^{\mathrm{T}}\right)$ with the expectation evaluated as the average over a period and samples, assuming ergodicity and homogeneity

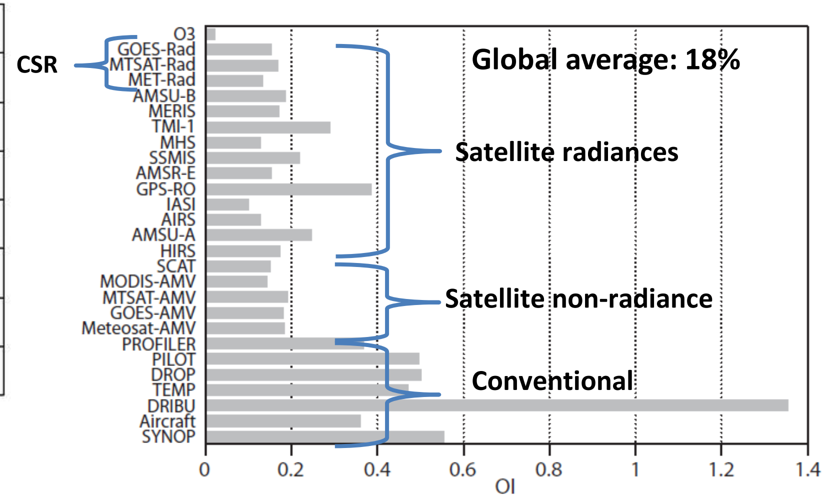
3. Ensemble-based DFS diagnostics at JMA
3-2. Results: **DFS per obs**

LETKF within JMA hybrid DA



c.f. ECMWF 4D-Var (as of 2011)

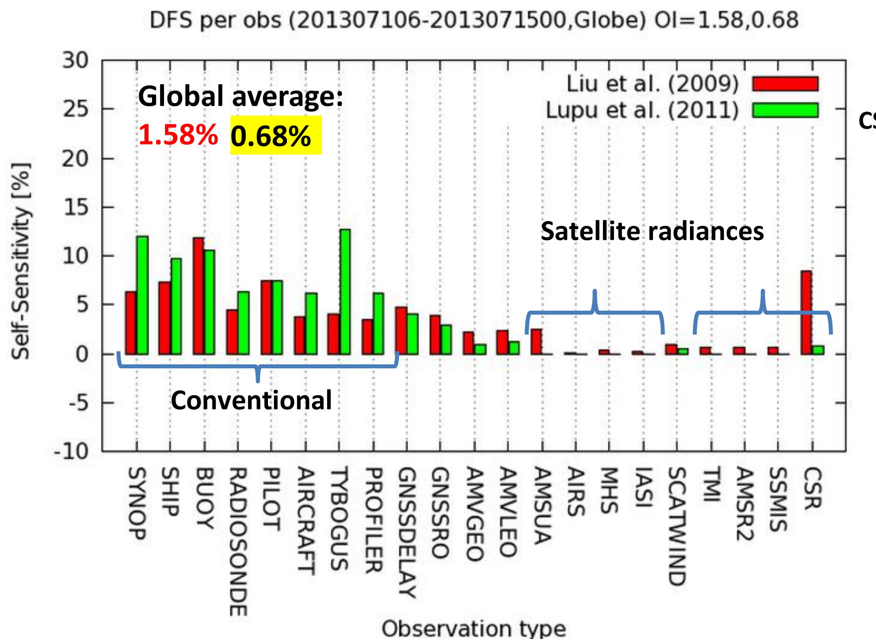
from Cardinali (2013; ECMWF lecture notes)



- Reasonable agreement between the two methods (at least for conventional obs).
- **Shockingly small observational impact**:
  - for JMA only about **1%** of information comes from observations,
  - whereas it is about 20% in ECMWF 4D-Var

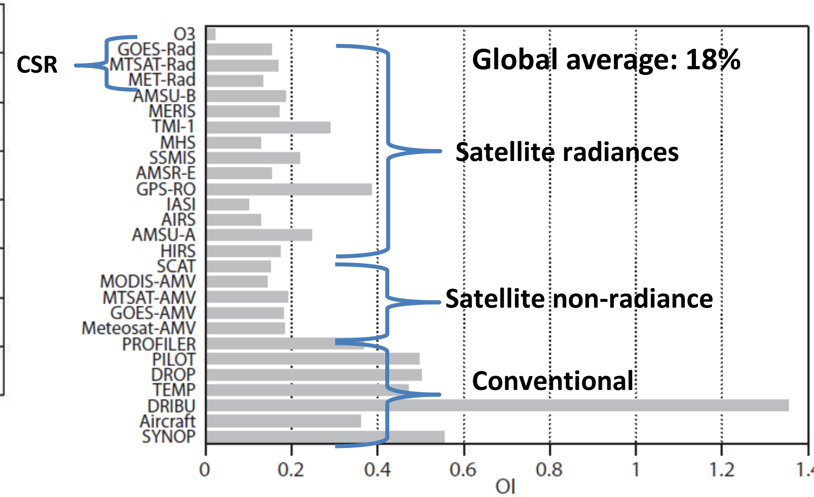
3. Ensemble-based DFS diagnostics at JMA
3-2. Results: **DFS per obs**

LETKF within JMA hybrid DA



c.f. ECMWF 4D-Var (as of 2011)

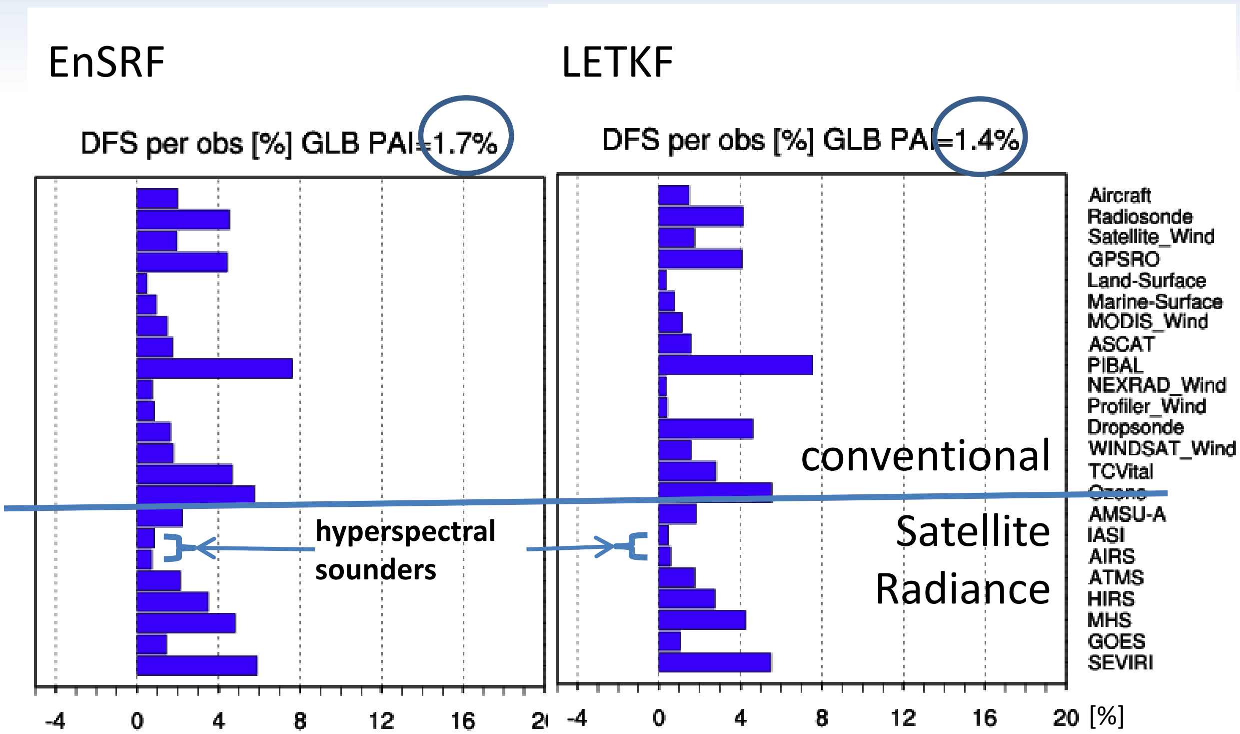
from Cardinali (2013; ECMWF lecture notes)



- DFS particularly small for dense observations, satellite radiances in particular (except AMSU-A and CSR*).

  * CSR: Clear Sky Radiances measured by infrared imagers on geostationary satellites (MTSAT, GOES and Meteosat)

4. Ensemble-based DFS for NCEP GFS hybrid GSI



- To discern if the "very small DFS problem" is merely an idiosyncrasy of JMA, we computed DFS for NCEP's lower-resolution version of GFS/GSI hybrid DA as well.
- Results: DFS is very small for NCEP's system as well.

→ **"Small DFS problem" possibly universal to all EnKF systems.**

5. Why DFS so small for EnKF?

- Our Answer: not enough ensemble size.
- We can show, for a local analysis in LETKF, that:

$$\mathrm{tr}(\mathbf{S}_{\mathrm{loc}}) = \mathrm{tr}\left(\mathbf{K}_{\mathrm{loc}}^{\mathrm{T}}\mathbf{H}_{\mathrm{loc}}^{\mathrm{T}}\right) = \mathrm{tr}(\mathbf{H}_{\mathrm{loc}}\mathbf{K}_{\mathrm{loc}}) \leq K-1$$

- i.e., ***<u>DFS is bounded from above by the degrees of freedom of the background ensemble</u>***. See the next slide for proof.
- The number of observations locally assimilated, p_{loc}, is $\sim O(10^3)$, much larger than the member size $K = 50$.
- Suppose, for convenience, that each observation locally assimilated has comparable DFS, and that the observation density can be assumed homogeneous.
- Then, we can assume that, locally, (DFS per obs) $\sim \frac{K-1}{p_{\mathrm{loc}}}$, which gives:

$$\mathrm{DFS}_{\mathrm{global}} = \sum_{\mathrm{all\ obs}}(\mathrm{DFS\ per\ obs\ })_{\mathrm{local}} \sim p_{\mathrm{global}} \times \frac{K-1}{p_{\mathrm{loc}}}$$

$\rightarrow$ $(\mathrm{DFS\ per\ obs\ })_{\mathrm{global}} = \frac{\mathrm{DFS}_{\mathrm{global}}}{p_{\mathrm{global}}} \sim \frac{K-1}{p_{\mathrm{loc}}}$, which, for our system, is $\frac{49}{4{,}000} \sim O(1\%)$

6. Implications

- We have seen that, for an EnKF with ensemble size K much smaller than the number of the locally assimilated observations p_{loc} ($p_{\mathrm{loc}} \gg K$), DFS is inevitably bounded by the member size K and hence automatically underestimated. **(= smaller than what it should be in true KF)**
- This means that such a system cannot fully extract information from observations.
- We believe this fact has a lot of important implications, e.g., on:
  1. why drastic ***observation thinning does not harm performance***,
  2. why ***covariance inflation*** is necessary,
  3. ***what the localization scale should be***, given the ensemble size and observation density,
  4. how, ***in serial assimilation,*** the ***order*** of assimilating observations affects the accuracy of the analysis ...etc.

6-1. Implication for observation thinning *(highly speculative)*

- Hamrud et al. (2015 Part I; MWR) reports that, **in ECMWF's LETKF local analysis, limiting the number of assimilated observations to only 30 per report type and element *improves* forecast performance** while achieving dramatic computational saving at the same time.
  - Similar result was also obtained with JMA's LETKF (Ota 2015, "adjoint Workshop").
  - Talk by Guo-Yuan Lien on Thursday on Radar assimilation
- This fact "**using less obs is better**" seems **counterintuitive** and difficult to interpret (at least to me).
- DFS discussion could provide a plausible interpretation (justification):
  - **In LETKF local analysis, the amount of information extractable from observations (=DFS) is limited by the ensemble member size.**
  - **Thus, assimilating too many observations beyond this limit only adds noises rather than signal.**
  - **→ Assimilating observations within the limit of DFS imposed by the member size reduces noises and improves analysis.**
- Related to the argument above, in a situation where thinning of observations is necessary (dense obs, e.g. radiance, radar, aircraft etc.) DFS could be used to guide in choosing which obs to assimilate.

6-2. Implication on covariance inflation *(highly speculative)*

- If the ensemble size is insufficient, DFS=$\mathrm{tr}(\mathbf{R}^{-1}\mathbf{HAH}^{\mathrm{T}})$ is underestimated.
- → The analysis error covariance $\mathbf{A}$ is also underestimated.
- → Need to inflate $\mathbf{A}$.

smaller than what we expect from true KF

- Traditionally, nonlinearity and model errors are considered to be the source of necessity for covariance inflation

$B=MBM^{T} + Q$

  - It is **B** rather than **A** that need inflation.
  - This is true for Extended Kalman Filter.
- The inherent underestimation of DFS could be another mechanism behind the need for covariance inflation.
- This argument gives intuitive explanation as to why Relaxation-to-prior methods of Zhang et al. (2004) and Whitaker and Hamil (2012) are so successful:
  - underestimation of DFS (= ***posterior spread in obs space***) is severer when/where observations are denser
  - Relaxation-to-prior methods act to inflate **A** more exactly in such a situation.

6-3. Implication on covariance localization *(highly speculative)*

- Traditionally, it is believed that localization is necessary to filter out spurious correlations in **B** due to sampling errors.
- From this perspective, observation density/distribution does not come into play.

- The fact that DFS is bounded by the member size provides another criterion for optimality of localization:
  - Let $\{\sigma_i\}$ be the singular values of $\mathbf{R}^{-\frac{1}{2}}\mathbf{H}\mathbf{B}^{\frac{1}{2}}\left(=\frac{\mathbf{1}}{\sqrt{\mathrm{K}-1}}\mathbf{R}^{-\frac{1}{2}}\mathbf{Y}^{\mathbf{b}}\right)$. Then, DFS = $\sum_i \frac{\sigma_i^2}{1+\sigma_i^2}$
  - → DFS will not be underestimated if K-th largest singular value σ_K is negligibly small.
- This gives a criterion for the optimal member size K given the observation network ($\mathbf{H}$, $\mathbf{R}$) and background error covariance ($\mathbf{B}$).
- Inversely, given the member size K, we can choose localization scale so that DFS is not artificially bounded. For this, we can require that the observations within the localized area are few enough such that $\sigma_i \ll 1$ for some $i < K$.

6-4. Implication on order of obs. assimilation in serial EnKF *(highly speculative)*

- Given that the total DFS is bounded by the ensemble size, it would make sense to assimilate the observation with the largest DFS first.
- In fact, Dr. Jeff Whitaker showed at ISDA 2015 that, in serial EnSRF, the following procedure improves the analysis:
  - assimilating observations from those with the smallest $\rho \coloneqq \frac{\mathrm{HAH}^T}{\mathrm{HBH}^T}$ to those with the largest,
  - assigning large localization scale to observations whos ρ is small.
- It is easy to see $\rho = \frac{\mathrm{HAH}^T}{\mathrm{HBH}^T} = \frac{\mathrm{H(I-KH)BH}^T}{\mathrm{HBH}^T} = 1 - \mathrm{HK}$ =1-DFS, i.e., Dr. Whitaker's successful method is equivalent to:
  - assimilating observations from largest DFS to those with smallest,
  - assigning a larger localization scale to observations with larger DFS
- → DFS argument could provide a theoretical justification to his somewhat empirical but successful method.

7. Take home messages

- EnKF algorithms inherently underestimate posterior spread if you (locally) assimilate much more obs than the ensemble size.
- This limitation is particularly relevant to dense observations (≈ **Big Data Assimilation issue**)
  - Theory well developed for cases #state >> #obs ~ #ens
  - Not much so for cases #obs >> #ens
- DFS argument provides insight into many important aspects of Ensemble DA (inflation, localization, thinning, superobing …)
- ***So please start thinking about your problem in terms of DFS!***

Thank you!

Questions?

A question from me in case nobody has any questions to me.

- The "Posterior spread underestimation" issue occurs whenever rank($\mathbf{HBH}^T$) < #obs.
- so Reduced Rank Kalman Filter (RRKF) must have suffered from the same issue.

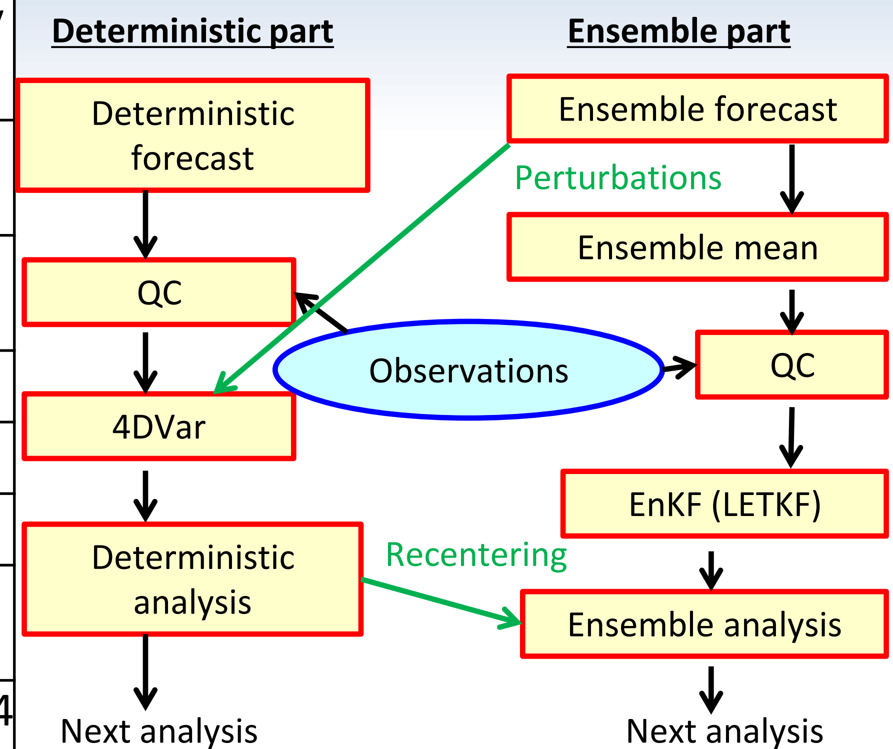
Q: Does anybody know whether this issue had ever been addressed in RRKF context?

Backup slide

from Yoichiro Ota (2015, Adjoint Workshop)

Hybrid 4DVar-LETKF DA developed in JMA

Analysis resolution (outer / inner)	T_L959L100 (~20km, top:0.01hPa) / T_L319L100 (~55km, top:0.01hPa)
Assimilation window	6 hours (analysis time +/- 3 hours)
Hybrid method	Extended control variable method (Lorenc 2003)
Weights on $\mathbf{B}$	$\beta_{stat}{}^2 = 0.77$, $\beta_{ens}{}^2 = 0.23$
LETKF resolution	T_L319L100
Ensemble size	50
Localization scale (4DVar)	Horizontal: 800km Vertical: 0.8 scale heights
Localization scale (LETKF)	Horizontal: 400km, Vertical: 0.4 (0.8 for Ps) scale heights
Covariance inflation	Adaptive inflation (Miyoshi 2011)



$$\mathbf{B} = \beta_{stat}^2 \mathbf{B}_{stat} + \beta_{ens}^2 \mathbf{B}_{ens}$$

Static (Climatological) background error covariance

Ensemble-based background error covariance

Operational global DA at JMA is 4DVar (not hybrid)

Proof of $\mathrm{tr}(\mathbf{S}_{\mathrm{loc}}) \equiv \mathrm{tr}(\mathbf{H}_{\mathrm{loc}}\mathbf{K}_{\mathrm{loc}}) \le K-1$ for LETKF local analysis

- In each local analysis of LETKF, DFS can be expressed as
- $\mathrm{tr}(\mathbf{S}) \equiv \mathrm{tr}(\mathbf{HK}) = \mathrm{tr}(\mathbf{HAH}^T\mathbf{R}^{-1})$ $(\because \mathbf{K} = \mathbf{AH}^T\mathbf{R}^{-1})$
- LETKF estimates the analysis error covariance by :

$$\mathbf{A} = \mathbf{X^b}\widetilde{\mathbf{A}}\mathbf{X^b}^T, \quad \widetilde{\mathbf{A}} = \left[(K-1)\mathbf{I} + \mathbf{Y^b}^T\mathbf{R}^{-1}\mathbf{Y^b}\right]^{-1} = \frac{\mathbf{1}}{K-1}(\mathbf{I} + \mathbf{Z}^T\mathbf{Z})^{-1}, \text{ with } \mathbf{Z} \equiv \frac{\mathbf{1}}{\sqrt{K-1}}\mathbf{R}^{-\frac{1}{2}}\mathbf{Y^b}$$

- $\mathbf{Z}^T\mathbf{Z}$ is a $K \times K$ positive semi-definite symmetric matrix. Its eigenvalue decomposition becomes :

$$\mathbf{Z}^T\mathbf{Z} = \mathbf{U\Lambda U}^{-1}, \mathbf{UU}^{-1} = \mathbf{I}, \qquad \mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_K)$$

- Since $\mathrm{rank}(\mathbf{Z}) = K-1$, $\lambda_K = 0$. From positive semi-definiteness of $\mathbf{Z}^T\mathbf{Z}$, $\lambda_i > 0$ $(1 \le i \le K-1)$.
- Thus:

$$\mathbf{HK} = \mathbf{HAH}^T\mathbf{R}^{-1} = \mathbf{HX^b}\widetilde{\mathbf{A}}\mathbf{X^b}^T\mathbf{H}^T\mathbf{R}^{-1} = \mathbf{Y^b}\widetilde{\mathbf{A}}\mathbf{Y^b}^T\mathbf{R}^{-1}$$

$$= \left(\sqrt{K-1}\mathbf{R}^{\frac{1}{2}}\mathbf{Z}\right)\widetilde{\mathbf{A}}\left(\sqrt{K-1}\mathbf{R}^{\frac{1}{2}}\mathbf{Z}\right)^T\mathbf{R}^{-1} = \mathbf{R}^{\frac{1}{2}}\mathbf{Z}(\mathbf{I} + \mathbf{Z}^T\mathbf{Z})^{-1}\left(\mathbf{R}^{\frac{1}{2}}\mathbf{Z}\right)^T\mathbf{R}^{-1}$$

- Because trace is invariant under cyclic reordering,

$\mathrm{tr}(\mathbf{S}) \equiv \mathrm{tr}(\mathbf{HK}) = \mathrm{tr}\left(\mathbf{R}^{\frac{1}{2}}\mathbf{Z}(\mathbf{I} + \mathbf{Z}^T\mathbf{Z})^{-\mathbf{1}}\left(\mathbf{R}^{\frac{1}{2}}\mathbf{Z}\right)^T\mathbf{R}^{-1}\right) = \mathrm{tr}\left(\mathbf{Z}(\mathbf{I} + \mathbf{Z}^T\mathbf{Z})^{-\mathbf{1}}\mathbf{Z}^T\mathbf{R}^{\frac{1}{2}}\mathbf{R}^{-1}\mathbf{R}^{\frac{1}{2}}\right) =$
$\mathrm{tr}(\mathbf{Z}(\mathbf{I} + \mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T) = \mathrm{tr}\left((\mathbf{I} + \mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^{\mathbf{T}}\mathbf{Z}\right) = \mathrm{tr}\left(\left(\mathbf{I} + \mathbf{U\Lambda U}^{-\mathbf{1}}\right)^{-1}\mathbf{U\Lambda U}^{-1}\right) =$
$\sum_{i=1}^{K}\frac{\lambda_i}{1+\lambda_i} = \frac{\lambda_1}{1+\lambda_1} + \cdots + \frac{\lambda_{K-1}}{1+\lambda_{K-1}} + \frac{0}{1+0} \le K-1$