Diagnostic methods for ensemble data assimilation

Daisuke Hotta*[1] and **Yoichiro Ota**[1]

[1]Numerical Prediction Division, Japan Meteorological Agency

Advice and advising: Profs. Eugenia Kalnay (UMD), Takemasa Miyoshi (UMD/RIKEN) and Daryl Kleist (UMD)

Collaborators: Guo-Yuan Lien (RIKEN), Tse-Chun Chen (UMD)

Acknowledgements: NCEP/JCSDA (Dr. Sid Boukabara, Dr. Jim Jung, Dr. Krishna Kumar)

RIKEN/AICS Data Assimilation Seminar
Sep 2, 2015

Contents

- Part I:
  1. **Proactive QC** based on EFSO
  2. Ensemble Forecast Sensitivity to **R** (**EFSR**)

Based on my Ph.D. dissertation defended at Univ. of Maryland;
Built upon NCEP's global DA

- Part II: (NEW!)
  1. EFSO at JMA
  2. Degrees of Freedom for Signal (**DFS**)

Built upon JMA's global DA system

Part I-1:
Proactive QC
based on EFSO

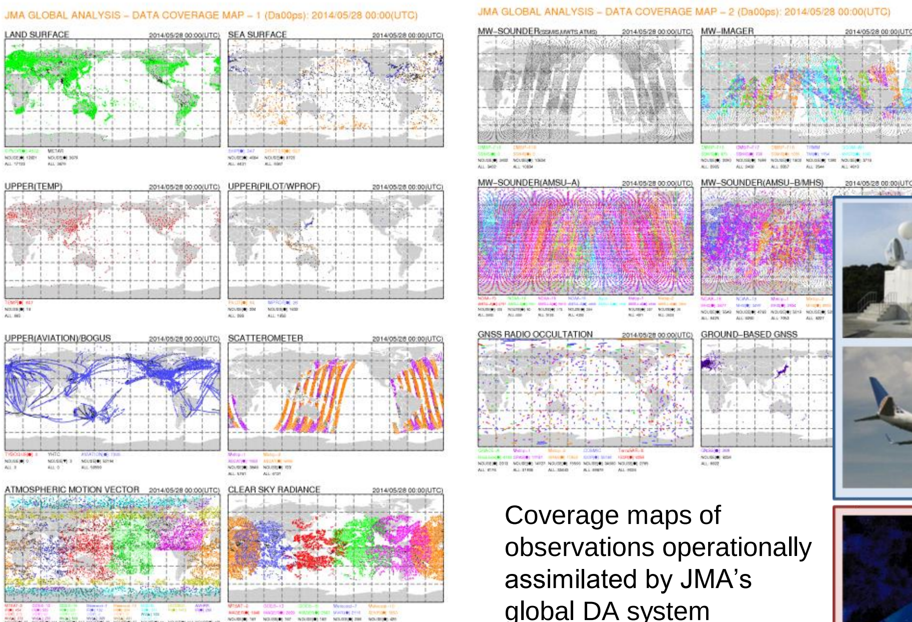
Motivation:
forecast skill dropout problem

- Forecast skills of major NWP centers are very high *on average*.
- However, they occasionally suffer from abrupt "drops" (*forecast skill dropouts, forecast busts*).
- Recent studies (e.g. Kumar et al. 2009) have shown that they are caused by assimilation of ***flawed* observations.**
- → Need to improve Quality Control (QC) so that ***flawed* obs** can be detected and dropped.

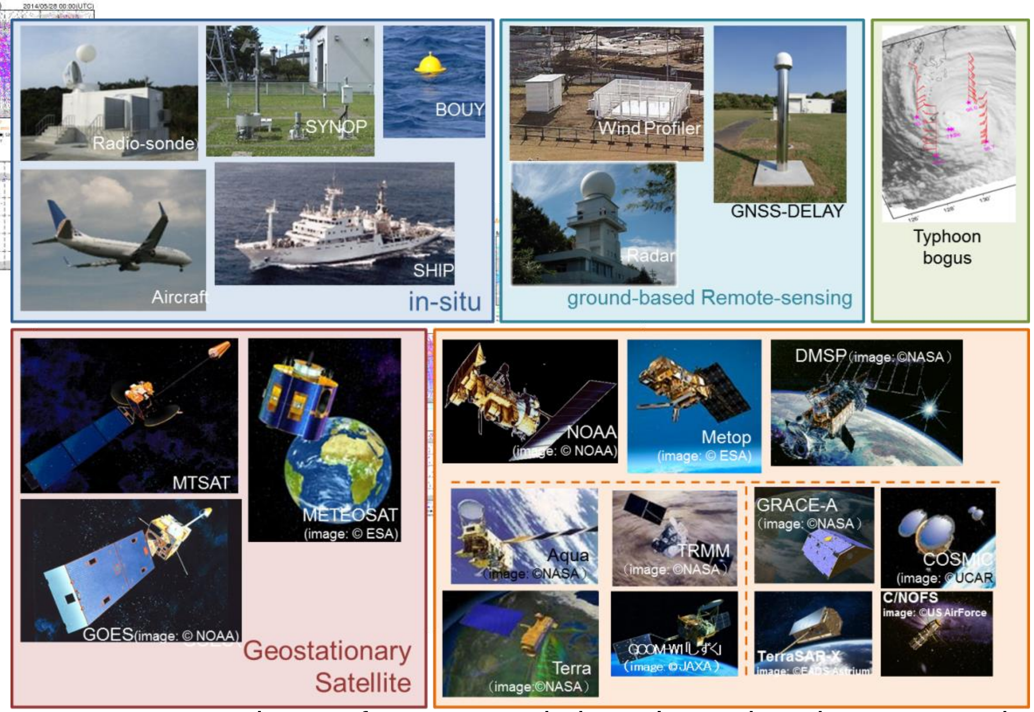
However….

Operational NWP assimilates millions of observations from dozens of different diverse sources.

→ How can we detect the observations that degrade the forecast?

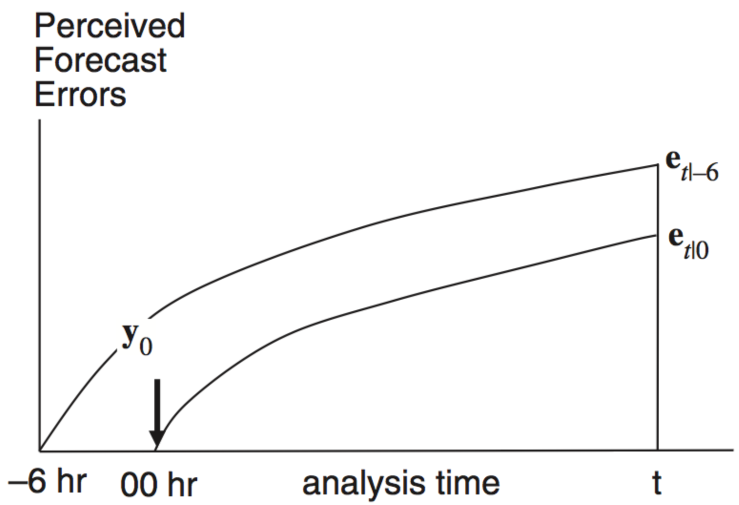


Coverage maps of observations operationally assimilated by JMA's global DA system



Pictures from JMA website unless otherwise annotated

EFSO: Ensemble Forecast Sensitivity to Observations



- Inspired by the FSO ***adjoint method*** of Langland and Baker (2004)
- Kalnay et. al (2012): improved, simpler formulation for ***EnKF***.
- Ota et al. (2013) implemented the new EFSO into the NCEP's operational GFS system

$$\Delta e^2 = \mathbf{e}_{t|0}^T C \mathbf{e}_{t|0} - \mathbf{e}_{t|-6}^T C \mathbf{e}_{t|-6} \approx \frac{1}{K-1} \delta \mathbf{y}_0^T \mathbf{R}^{-1} \mathbf{Y}_0^a \mathbf{X}_{t|0}^{fT} C(\mathbf{e}_{t|0} + \mathbf{e}_{t|-6})$$

Reduction of forecast error by the assimilation of obs.

O-B of ens. mean

forecast ptbs.

analysis spread in obs. space

EFSO enables us to estimate how much *<u>each</u>* observation improved/degraded forecast

Proactive QC: Find the obs. that make the 6hr forecast worse using EFSO, then reject them

Algorithm:

Suppose we wish to identify and delete “flawed” obs. at 00h.

① Run regular DA cycle from -06h to 00h.

② Run regular DA cycle from 00h to 06h.

③ Detect “regional dropouts” using the information available from ① and ②.

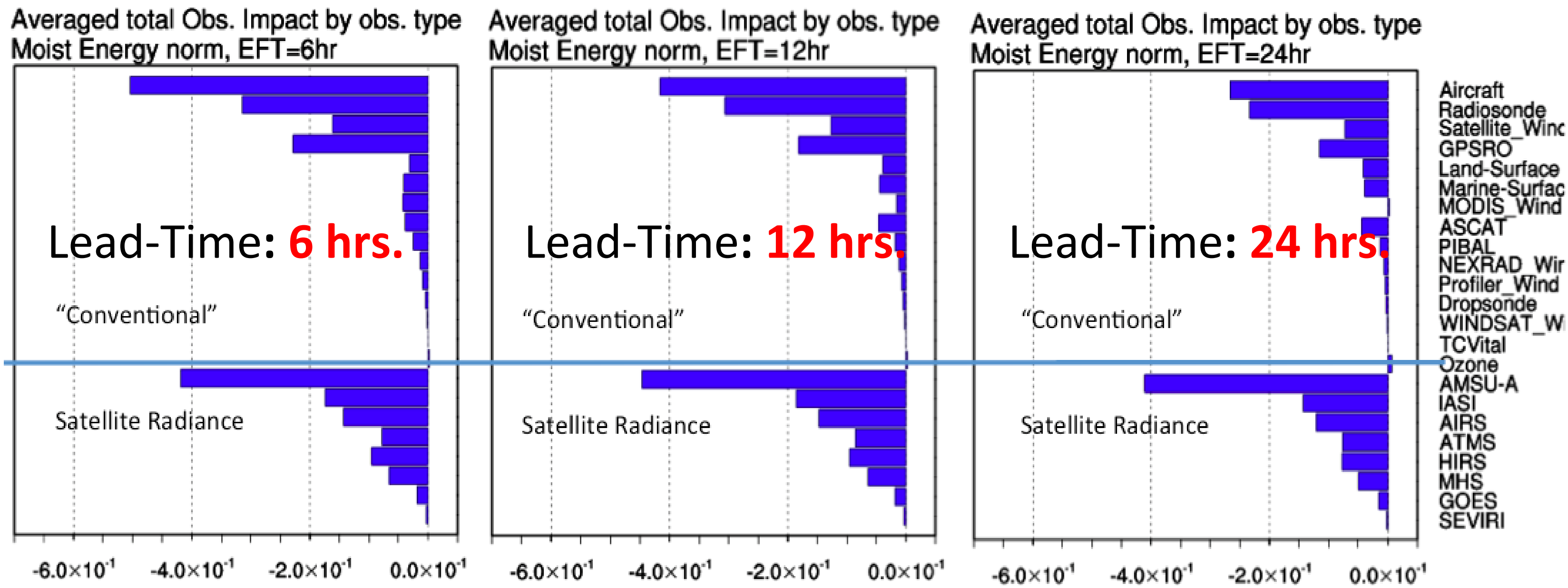
④ Perform **6-hour EFSO** to identify *flawed* obs at 00h.

⑤ If *flawed* obs are identified, repeat 00h analysis without using the detected *flawed* obs.

Are 6 hours long enough

to capture forecast impacts?

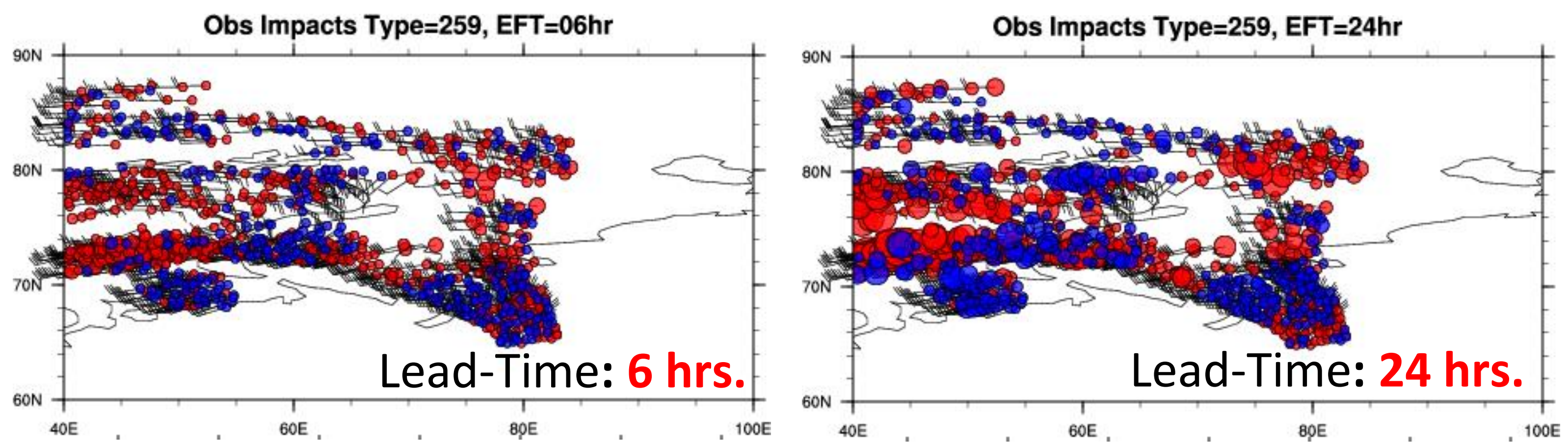
Average net observation impact for each observation type



EFSO **average** results are **not** very sensitive to the choice of evaluation lead time.

Are 6 hours long enough
to capture forecast impacts?

MODIS winds near the North Pole on Feb 06 18UTC, 2012



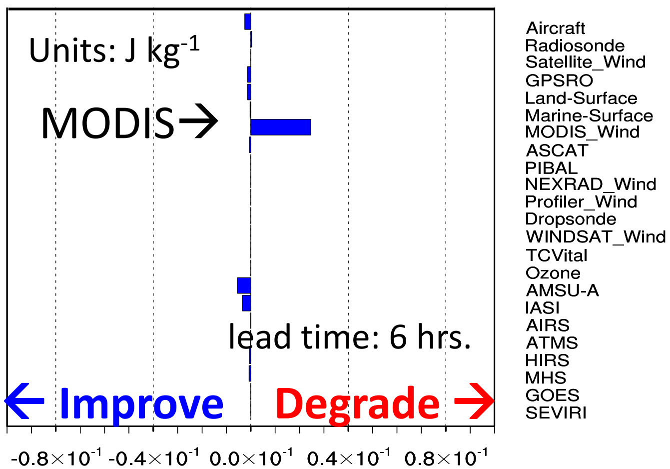
Red: negative impact   Blue: positive impact

EFSO results are **not** very sensitive to the choice of evaluation lead time, **even for individual cases**.

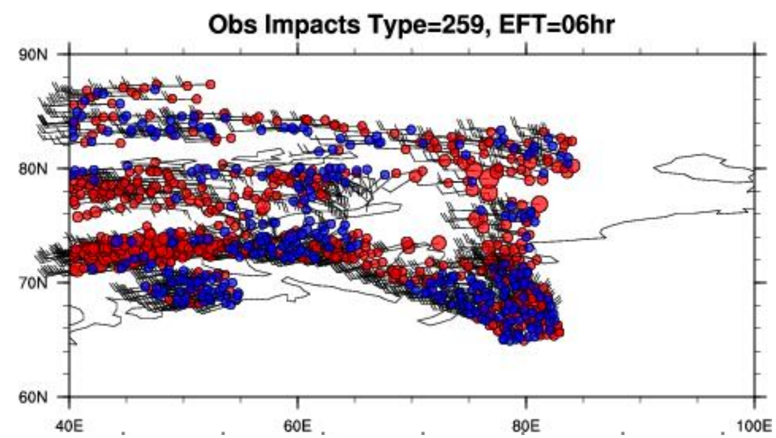
→ **6 hours are enough to detect flawed observations!**

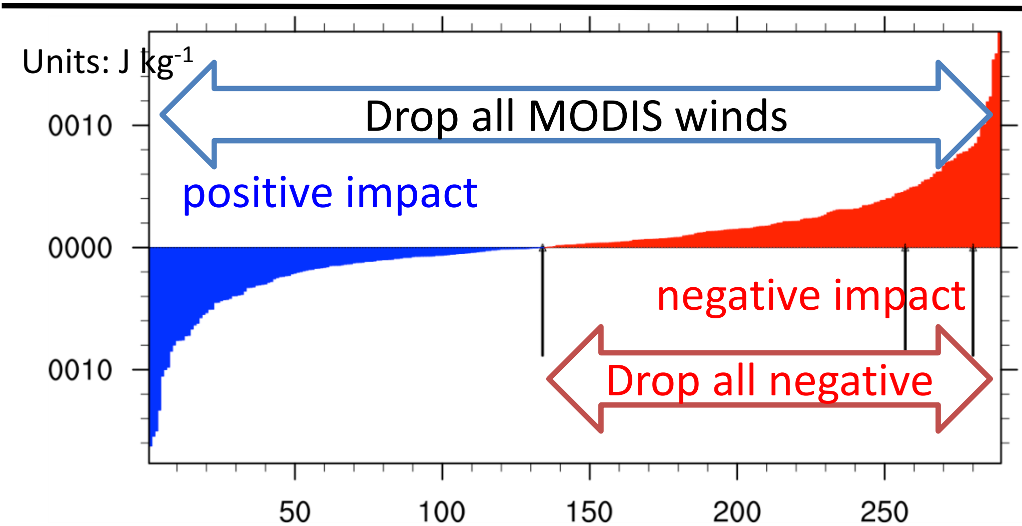
How many of the good/bad obs should we reject? Does the rejection actually improve forecast?

Net EFSO Impact by obs. types
measured with moist total energy norm



EFSO impact from
each MODIS wind observation





Distribution of the impact in one region

How many bad MODIS winds should we drop?

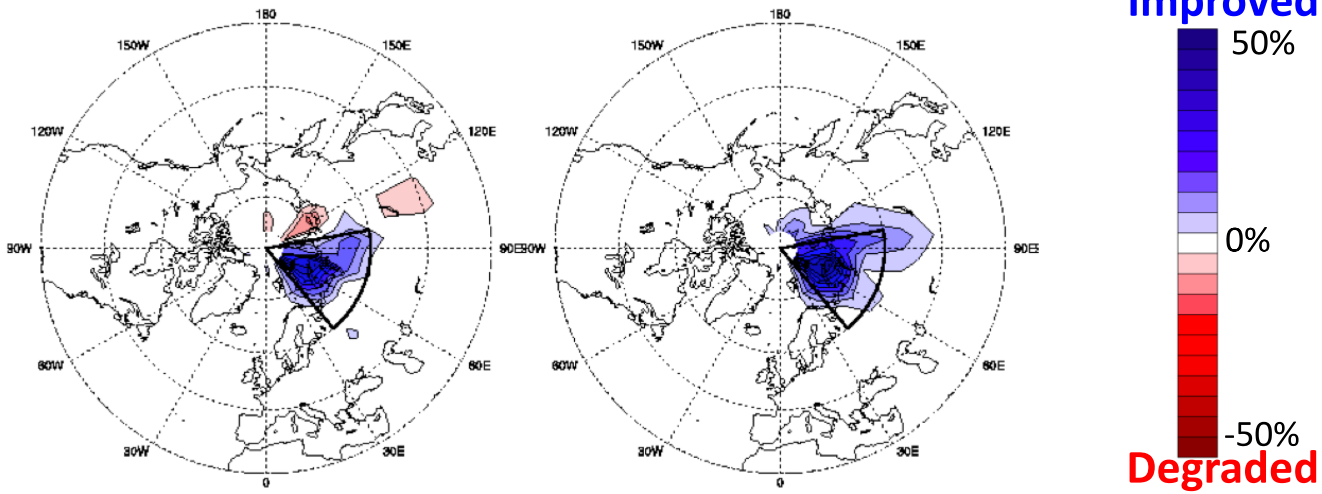
- All of them?
- The ones with negative impact?
- Only the worst ones?

How many of the good/bad obs should we reject? Does the rejection actually improve forecast?

Relative **24-hr** fcst. improvement:= $(e^f_{beforeQC} - e^f_{afterQC}) / e^f_{beforeQC} \times 100$ [%]

Data selection based on **6-hour** EFSO

delete all MODIS winds   delete all negative impact



- Rejection of observations based on **6-hr** impact improves **24-hr** forecast !
- Best results obtained by rejecting **all negative**ly impacting obs.

Data denial experiments: Summary of 20 cases

local maximal improvement/degradation
and improvement averaged over a hemisphere

Case #		6-hour allobs	allneg	one-sigma	netzero
	max.imp.	12%	11%	4%	5%
	max.deg.	-9%	-1%	-1%	-1%
1	avg.imp.	0.0%	0.2%	0.1%	0.1%
	max.imp.	14%	11%		4%
	max.deg.	-5%	-4%		0%
2	avg.imp.	-0.1%	0.3%	N/A	0.2%
	max.imp.	13%	7%	2%	4%
	max.deg.	-15%	-5%	-1%	-2%
3	avg.imp.	0.0%	0.2%	0.0%	0.0%
	max.imp.	25%	27%	15%	13%
	max.deg.	-5%	-5%	-2%	-2%
4	avg.imp.	0.6%	0.7%	0.3%	0.2%
	max.imp.	15%	19%	23%	22%
	max.deg.	-32%	-81%	-30%	-13%
5	avg.imp.	-0.2%	-0.2%	0.2%	0.3%
	max.imp.	9%	15%	12%	3%
	max.deg.	-9%	-6%	-3%	-1%
6	avg.imp.	0.0%	0.4%	0.3%	0.1%
	max.imp.	17%	13%	2%	0%
	max.deg.	-9%	-5%	-3%	0%
7	avg.imp.	-0.0%	0.4%	0.0%	0.0%
	max.imp.	41%	41%	21%	10%
	max.deg.	-18%	-14%	-5%	-2%
8	avg.imp.	0.9%	1.1%	0.8%	0.4%
	max.imp.	7%	8%	8%	8%
	max.deg.	-21%	-16%	-3%	-4%
9	avg.imp.	-0.6%	-0.4%	0.0%	0.1%
	max.imp.	25%	19%		6%
	max.deg.	-6%	-6%		0%
10	avg.imp.	1.1%	0.7%	N/A	0.2%

Case #		6-hour allobs	allneg	one-sigma	netzero
	max.imp.	11%	9%	2%	3%
	max.deg.	-6%	-5%	-2%	0%
11	avg.imp.	0.5%	0.3%	0.1%	0.1%
	max.imp.	37%	39%	19%	19%
	max.deg.	-14%	-12%	-2%	-2%
12	avg.imp.	0.7%	0.7%	0.5%	0.5%
	max.imp.	24%	30%	18%	19%
	max.deg.	-9%	-9%	-10%	-12%
13	avg.imp.	1.4%	0.8%	0.3%	0.4%
	max.imp.	5%	3%	1%	1%
	max.deg.	0%	0%	0%	0%
14	avg.imp.	0.3%	0.1%	0.0%	0.1%
	max.imp.	3%	1%	1%	1%
	max.deg.	-2%	-1%	-1%	-1%
15	avg.imp.	0.1%	0.1%	-0.0%	0.0%
	max.imp.	27%	30%	23%	16%
	max.deg.	-15%	-21%	-4%	-2%
16	avg.imp.	1.9%	1.8%	1.3%	0.7%
	max.imp.	39%	48%	26%	20%
	max.deg.	-15%	-4%	-2%	-2%
17	avg.imp.	0.8%	2.1%	1.2%	0.8%
	max.imp.	46%	46%	25%	21%
	max.deg.	-9%	-8%	-3%	-2%
18	avg.imp.	2.4%	2.2%	1.0%	0.8%
	max.imp.	44%	37%	17%	14%
	max.deg.	-24%	-10%	-1%	-1%
19	avg.imp.	2.2%	2.2%	1.0%	1.0%
	max.imp.	12%	10%	5%	3%
	max.deg.	-3%	-1%	-1%	-1%
20	avg.imp.	0.2%	0.3%	0.2%	0.0%

Denying all neg obs (within the region of 6-hour dropouts):

- **Hemispheric-scale forecast error reduced in 18 out of 20 cases.**
- **Local improvement over 30% in 7 cases.**
- **<u>Improvement continues after 24hrs.</u>**

Key questions answered

- Are 6 hours long enough for detecting *flawed* observations?
  - **→ Yes. 6-hr EFSO is equally capable of detecting *flawed* obs. as 24-hr EFSO is.**
- What is the best criterion for rejection of observations?
  - **→ Best results by rejecting all negatives (allneg) of the identified *flawed*-type inside the identified dropout region.**
- Does rejection of those observations really improve analysis and forecast?
  - **→ Yes, with >30% local improvement in 7 out of 20 cases. <u>Improvement continues after 24hrs</u>.**

Operational feasibility: *Can we wait 6 hours?* → *We don't have to wait!*

- We can exploit time lag between early analysis (GFS) and final analysis (GDAS) (suggested by Dr. John Derber)
  - c.f. backup slide for detail

- The change in analysis by the rejection of a subset of observations can be cheaply approximated by

$$\bar{\mathbf{x}}_0^{a,\mathrm{deny}} - \bar{\mathbf{x}}_0^{a} \approx -\mathbf{K}\delta\bar{\mathbf{y}}_0^{ob,\mathrm{deny}} \approx -\frac{1}{K-1}\mathbf{X}_0^a\mathbf{Y}_0^{aT}\mathbf{R}^{-1}\delta\bar{\mathbf{y}}_0^{ob,\mathrm{deny}}$$

- No need to repeat costly analysis!

Part I-2:
Ensemble Forecast Sensitivity to R matrix (EFSR)

Motivation

- Data Assimilation combines information from background and observations with an "*optimal weight*" which depends on the background- and observation- error covariances **B** and **R.**
- In EnKF, **B** (=$\mathbf{P}^b$) is dynamically estimated, but **R** is still an external parameter.
  – Truth is unknown. → True **R** is also unknown.
  – **R is specified empirically and subjectively.**

- → We need a systematic method for tuning **R**.

EFSR Formulation

- Daescu and Langland (2013) proposed an ***adjoint-based*** formulation of forecast sensitivity to **R** matrix.
- We can formulate an ***ensemble*** version based on **EFSO** by Kalnay et al. (2012) :

$$\left[\frac{\partial e}{\partial \mathbf{R}}\right]_{ij} \approx -\frac{2}{K-1}\left[\mathbf{R^{-1}Y_0^aX_{t|0}^{fT}Ce_{t|0}}\right]_i\left[\mathbf{R}^{-1}\delta y^{oa}\right]_j$$

- It tells us whether the forecast will be improved or degraded by perturbing **R**.

→ **We can optimize R**.

Perfect-model Experiment: Experimental Setup

- **Model:** Lorenz '96 model with N=40 and F=8.0

$$\frac{\mathrm{d}x_j}{\mathrm{d}t} = x_j\left(x_{j+1} - x_{j-2}\right) - x_j + F$$

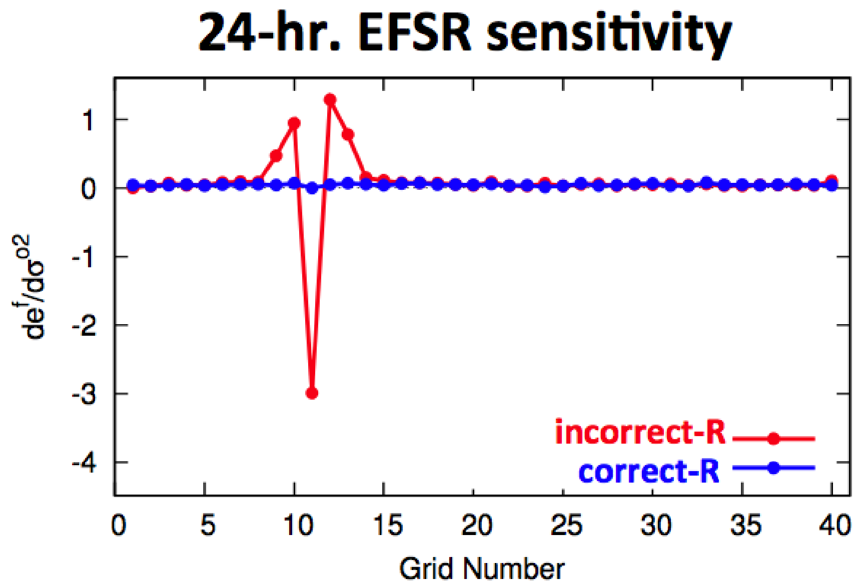
- **DA method:** 40 member LETKF, no localization
- **EFSR:** no localization
- **Observations:** available at every grid point.
- **Specification of R:**

Name	True obs error variance	Prescribed error variance
SPIKE	$\sigma_j^{o,\text{true}2} = \begin{cases} 0.8^2 & j = 11 \\ 0.2^2 & j \neq 11 \end{cases}$	$\sigma_j^{o2} = 0.2^2$ everywhere
STAGGERED	$\sigma_j^{o,\text{true}2} = \begin{cases} 0.1^2 & j\text{: odd} \\ 0.3^2 & j\text{: even} \end{cases}$	$\sigma_j^{o2} = 0.2^2$ everywhere
LAND-OCEAN	$\sigma_j^{o,\text{true}2} = \begin{cases} 0.3^2 & 1 \leq j \leq 20 \ (\text{“land”}) \\ 0.1^2 & 21 \leq j \leq 40 \ (\text{“ocean”}) \end{cases}$	$\sigma_j^{o2} = 0.2^2$ everywhere

- Erroneous obs. variance only at the 11-th grid pt.
- DA system assumes constant **R** for all grid pts.

Design is inspired by Liu and Kalnay (2008)

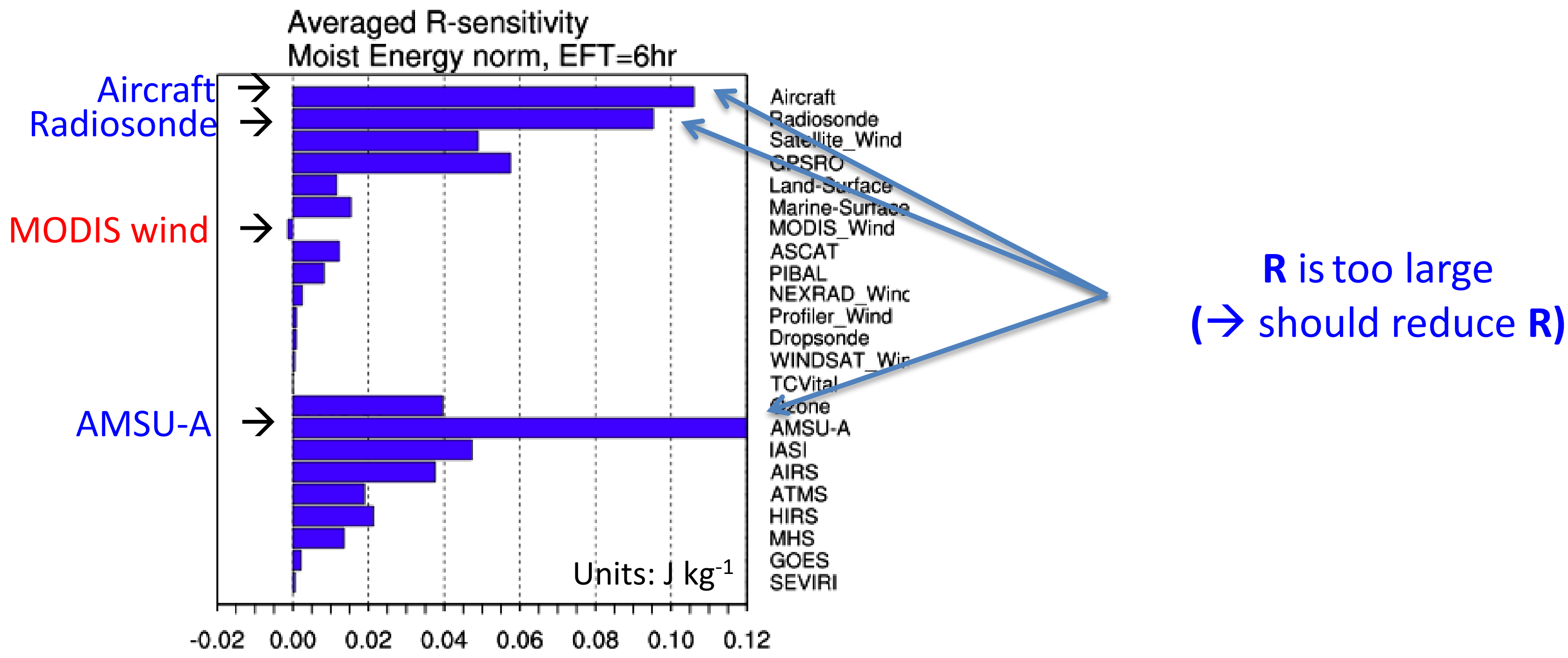
Perfect-model Experiment: Result (SPIKE experiment)



Negative sensitivity: forecast error can be reduced by increasing **R**
→ **R** is too small

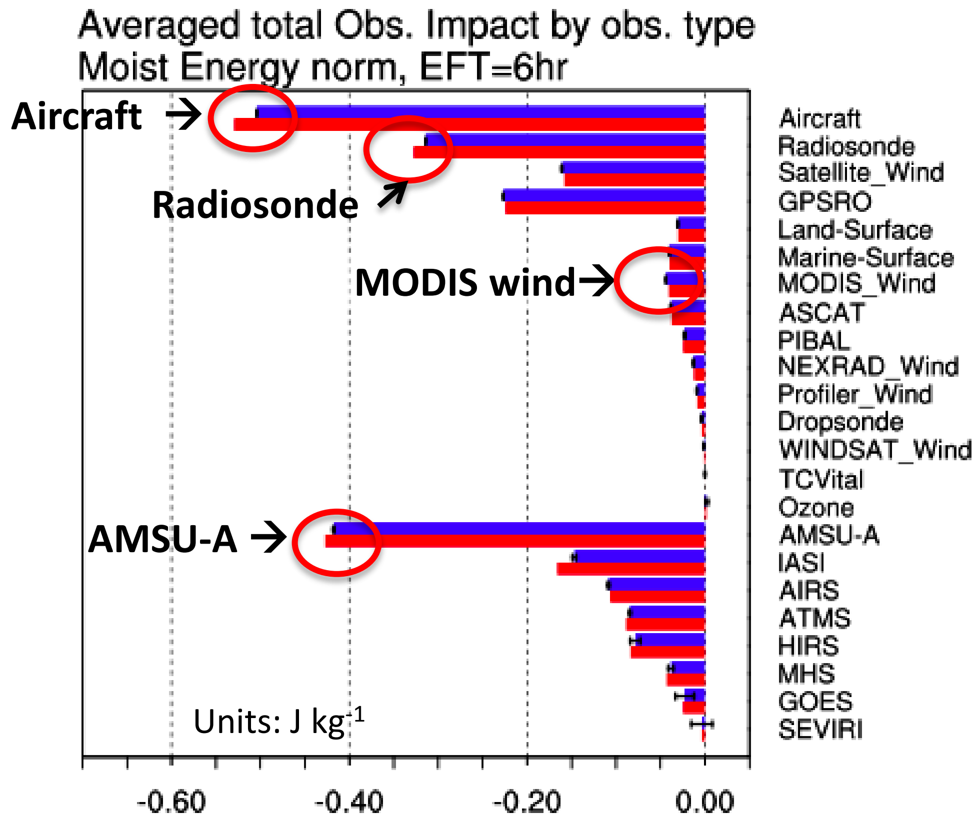
- For "**incorrect-R**," EFSR detects the mis-specification of **R** at the 11th grid point.
  → We can detect mis-specified **R**
- For "**correct-R**," EFSR diagnoses almost-zero sensitivity.
  → No "false alerts"

EFSR for GFS / GSI-LETKF hybrid



- **Aircraft, Radiosonde and AMSU-A**: large positive sensitivity
- **MODIS wind** : negative sensitivity
- → **Tuning experiment**:
    - Aircraft, Radiosonde and AMSU-A: reduce **R** by 0.9
    - MODIS wind: increase **R** by 1.1

Tuning Experiment: Result
EFSO **before**/**after** tuning of **R**



Aircraft, Radiosonde, AMSU-A:

- significant improvement of EFSO-impact

(as expected)

MODIS wind :

- No improvement in EFSO

  (interpretation given in the extra slides)

Summary for EFSR

- EFSR gives information on whether we should increase/reduce prescribed **R**.
- Tuning of **R** based on this diagnostics improves the EFSO.

- → EFSR can be used to *systematically* optimize **R**.

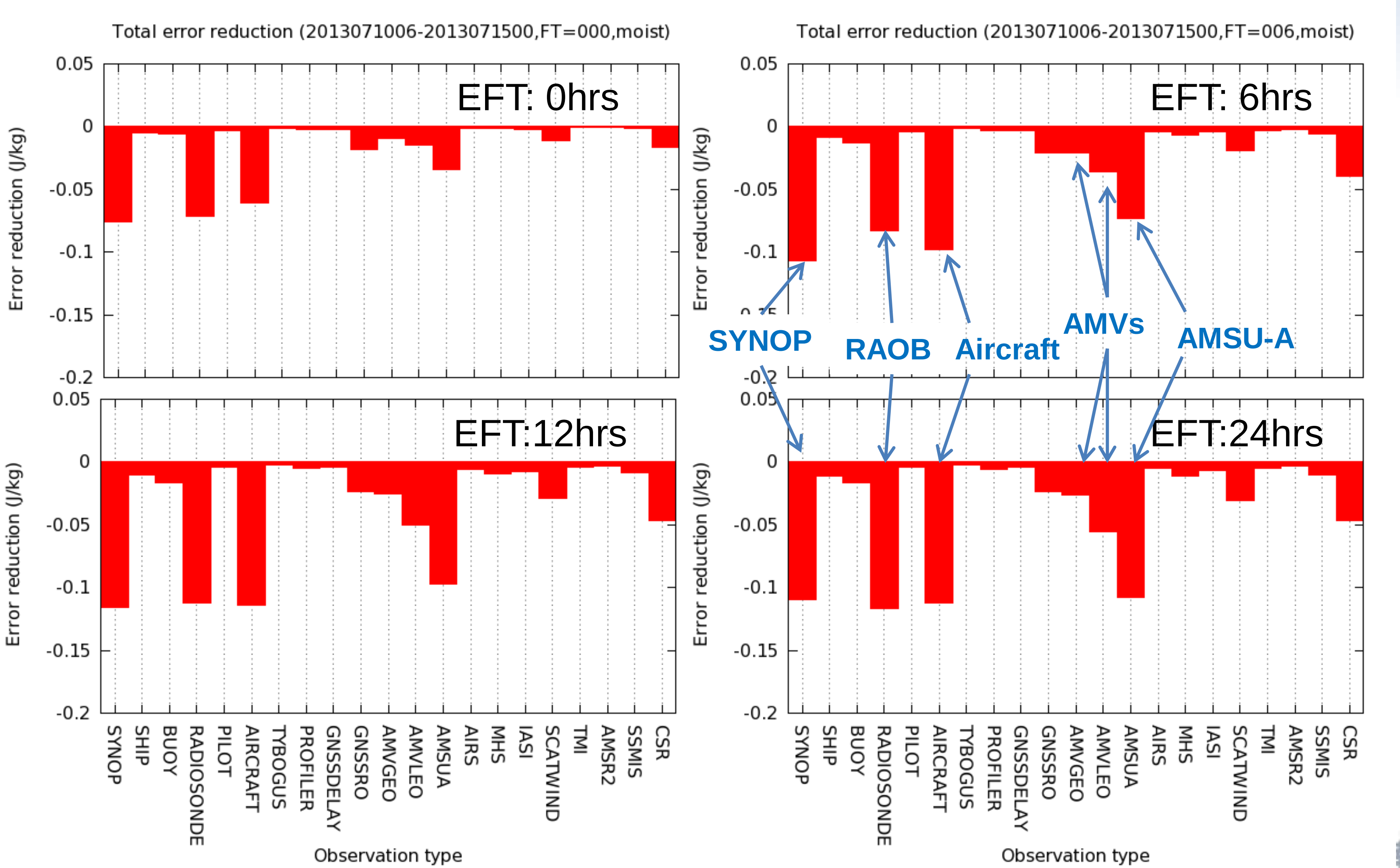
Part II-1:

EFSO at JMA

EFSO implementation at JMA (by Yoichiro Ota)

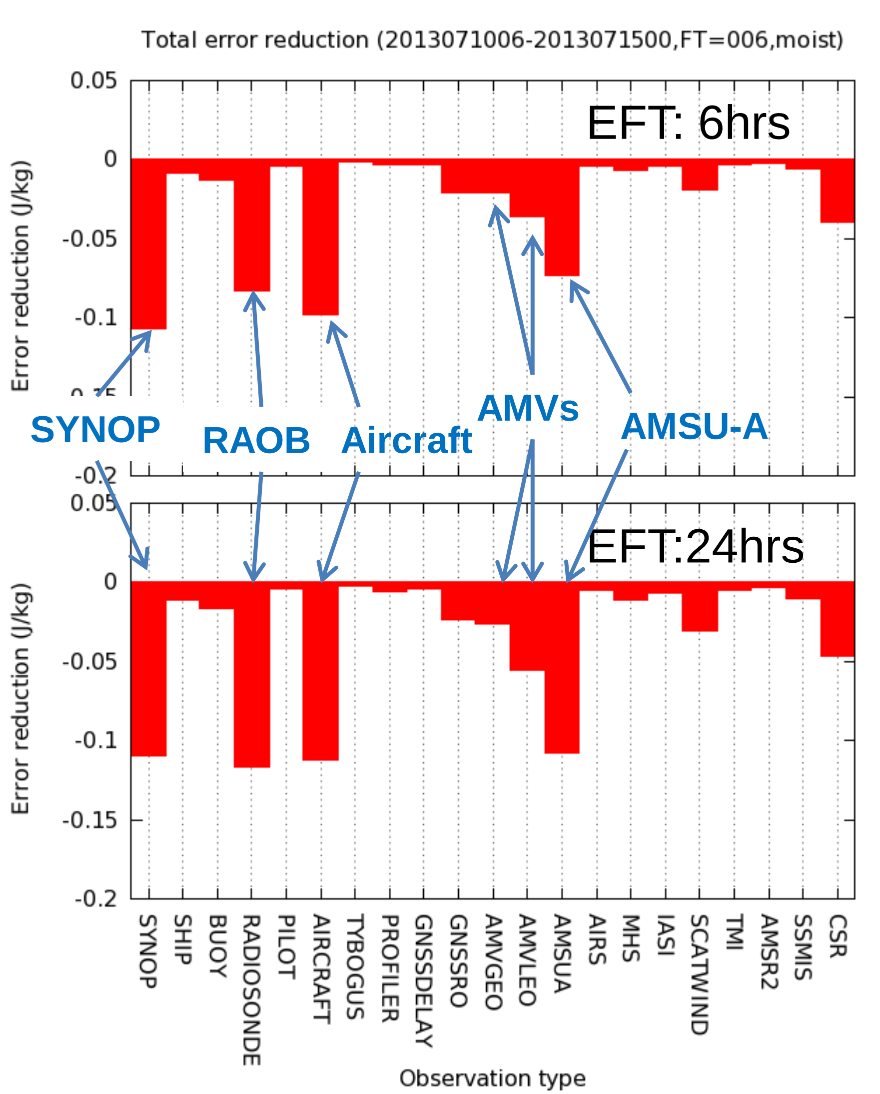
- DA system: hybrid LETKF/4D-Var coupled with JMA GSM
  - Resolution: (outer) TL959L100 ; (inner and ensemble) T319L100
  - Window: 6 hours (analysis time +/- 3 hours)
  - <u>**B** weights: 0.85 for static, 0.25 for ensemble</u>
  - <u>Member size: 50</u>
  - Localization scales (e-folding):
    - LETKF: Horizontal: 400km, Vertical: 0.4 scale heights
    - 4D-Var: Horizontal: 800km, Vertical: 0.8 scale heights
  - Covariance Inflation: Adaptive inflation of Miyoshi (2011)
  - LETKF part initially coded by Dr. T. Miyoshi; maintained and updated by Y. Ota and T. Kadowaki.
- EFSO:
  - Lead-times investigated: FT=0,6,12,24
  - Localization scales: same as LEKTF
    - advection: “moving localization scheme” of Ota et al.(2013) with scaling factor of 0.5 for horizontal wind.
  - Verification: high-resolution analysis from 4D-Var
  - Error norm: KE, Dry TE and Moist TE
- Period: Jul. 10, 2013, 06UTC – Jul. 15, 2013, 18UTC (5days, 20cases)

net EFSO contribution from each observation type (target=globe; norm=moist total energy)



net EFSO contribution from each observation type (target=globe; norm=moist total energy)

- Overall, the results are consistent with other centers:
  - at FT=24, contributions from radiances and conventional data are comparable.
  - AMSU-A, Radiosonde SYNOP and Aircraft are the top contributors to fcst err reduction.
- However, contributions from hyperspectral sounders (AIRS, IASI) are modest compared to ECMWF or NCEP.
- **(important) 6-hour EFSO is surprisingly consistent with 24-hour EFSO!**

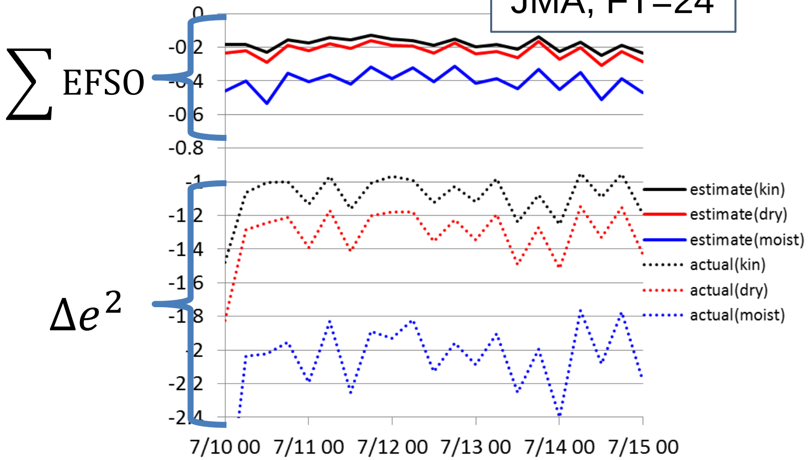


- So far, everything seems working well.
  - EFSO estimation consistent with other FSO studies
  - Plausible relative contributions from different observation types
- However,....

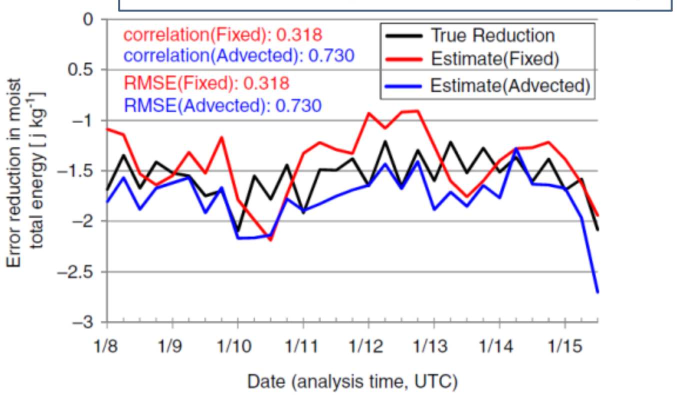
Estimated and actual forecast error reduction

$$\sum \text{EFSO} \qquad \Delta e^2 = \frac{1}{2}\mathbf{e}_{t|0}^{f\mathrm{T}}\mathbf{C}\mathbf{e}_{t|0}^{f} - \frac{1}{2}\mathbf{e}_{t|-6}^{f\mathrm{T}}\mathbf{C}\mathbf{e}_{t|-6}^{f}$$

JMA, FT=24



NCEP, FT=24 (from Ota et al. 2013)



- EFSO successfully reproduces temporal variation of forecast error reductions (correlation coefficient as high as ~ 0.8), but
- Only ~ 20 % of the amplitude explained by EFSO.
  - In contrast to > 100% (overestimation) for NCEP's EnKF (Ota et al. 2013)

A possible reason for impact underestimation

- EFSO implemented for JMA's LETKF underestimates forecast error reduction, whereas, for NCEP's EnKF, EFSO overestimates the actual impact.
- Why?
- Bug? → not found.
- Possible reason: forecast error not well covered by the space spanned by the forecast ensemble

A possible reason for impact underestimation (cont'd)

- EFSO formulation:
  $\Delta e^{f-g} \approx \frac{1}{K-1}\mathbf{d}^T\mathbf{R}^{-1}\left[\rho \circ \mathbf{Y}^a\, \mathbf{X}^{f^T}\right]\mathbf{C}(\mathbf{e}^f_{t|0}+\mathbf{e}^f_{t|-6})$
- In evaluating

$$\mathbf{X}^{f^T}\mathbf{C}(\mathbf{e}^f_{t|0}+\mathbf{e}^f_{t|-6}) = \left(\mathbf{C}^{1/2}\mathbf{X}^f\right)^T[\mathbf{C}^{\frac{1}{2}}(\mathbf{e}^f_{t|0}+\mathbf{e}^f_{t|-6})] =: \widetilde{\mathbf{X}}^{f^T}\tilde{\mathbf{e}}$$

  the portion of $\tilde{\mathbf{e}}$ that lies in the nullspace of $\widetilde{\mathbf{X}}^{f^T}$ does not contribute to $\widetilde{\mathbf{X}}^{f^T}\tilde{\mathbf{e}}$, namely:

- Let

$$\tilde{\mathbf{e}} = \tilde{\mathbf{e}}_{\mathrm{span}} + \tilde{\mathbf{e}}_{\mathrm{null}}, \tilde{\mathbf{e}}_{\mathrm{span}} \in \mathrm{span}\left(\widetilde{\mathbf{X}}^f\right), \tilde{\mathbf{e}}_{\mathrm{null}} \in \mathrm{null}\left(\widetilde{\mathbf{X}}^f\right)$$

  then

$$\widetilde{\mathbf{X}}^{f^T}\tilde{\mathbf{e}} = \widetilde{\mathbf{X}}^{f^T}\left(\tilde{\mathbf{e}}_{\mathrm{span}} + \tilde{\mathbf{e}}_{\mathrm{null}}\right) = \widetilde{\mathbf{X}}^{f^T}\tilde{\mathbf{e}}_{\mathrm{span}}$$

A possible reason for impact underestimation (cont'd)

- Is this really the reason why we had the impact underestimation?

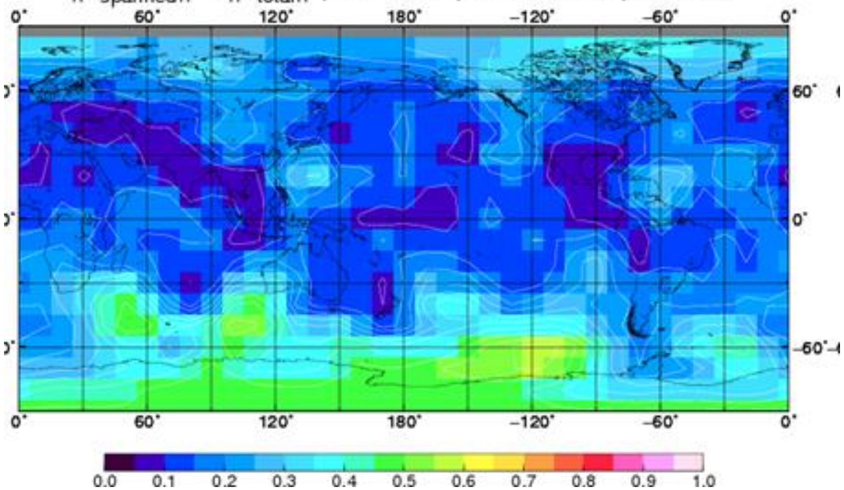
→ Perform diagnostics:

- For each local patch,
  - Decompose $\tilde{\mathbf{e}}$ into $\tilde{\mathbf{e}}_{\mathrm{span}}$ and $\tilde{\mathbf{e}}_{\mathrm{null}}$.
  - Compute the "explained fraction" $\frac{\left\|\tilde{\mathbf{e}}_{\mathrm{span}}\right\|^2}{\|\tilde{\mathbf{e}}\|^2}$.
  - Compare this with the impact underestimation $\frac{\sum \mathrm{EFSO}}{\Delta e^2}$.

Diagnosed "explained fraction" $\frac{\|\tilde{\mathbf{e}}_{\mathrm{span}}\|^2}{\|\tilde{\mathbf{e}}\|^2}$

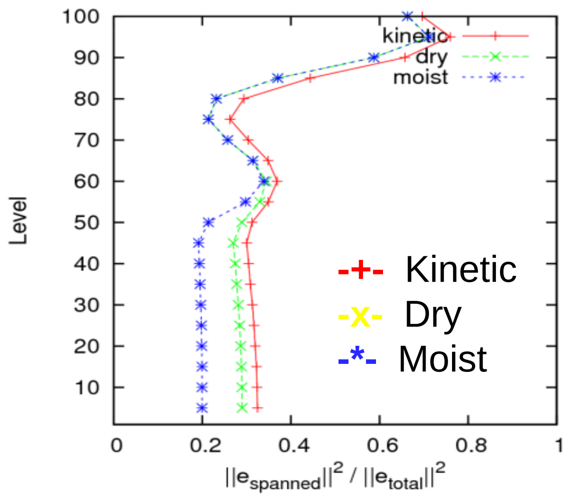
Horizontal distribution

(near tropopause level)



Vertical Profile

(global average)



- Fcst err well-captured by ensemble over the SH ocean, but not over the land.

→ Perhaps related to observation density:

  – Data-sparse area: analysis (verification) and forecast both close to model's free-run → $\mathbf{e}^f_{t|0}$ similar to Bred Vector → covered well by $\mathbf{X}^f$

- Errors in moisture difficult to capture by the ensemble.

Very good agreement between $\frac{\|\tilde{\mathbf{e}}_{\mathrm{span}}\|^2}{\|\tilde{\mathbf{e}}\|^2}$ and $\frac{\sum \mathrm{EFSO}}{\Delta e^2}$! (both ~ 20%)

EFSO at JMA: Summary

- EFSO is successfully implemented on JMA's global DA system, both stand-alone LETKF and LETKF/4D-Var hybrid.
- Plausible impacts from different types of obs. that are consistent with the literature.
- However, EFSO considerably underestimates the actual forecast error reduction Δe^{f-g}.
- Diagnostics that decomposes the fcst err to spanned- and null-spaces of the fcst ensemble $\mathbf{X}^f$ suggests that the underestimate is caused because significant portion of fcst err lies in the null-space of $\mathbf{X}^f$.
- The diagnostics exposes the lack of the ensemble size (currently only 50).
  - → further corroborated by DFS diagnostics (Part II-2)

Part II-2:

Degrees of Freedom for Signals (DFS)

Motivation

- How can we quantify the "value" of each observation?
- One possible quantification:
  - an observation is valuable if it improves the forecast.
- → FSO/EFSO

- Another perspective (inspired from information theory):
  - An observation is valuable if it enhances our "knowledge" about the true state of the atmosphere.
  - Our "knowledge" is enhanced if the uncertainty of the state estimate is reduced by assimilating the observation.
  - → **Degrees of Freedom for Signal** (DFS, or information content).

What is DFS?

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

- Shown to behave similarly to Shannon entropy reduction under loose conditions:

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

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

- First introduced to NWP by 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:

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

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

Ensemble-based DFS diagnostics at JMA Experimental set-up

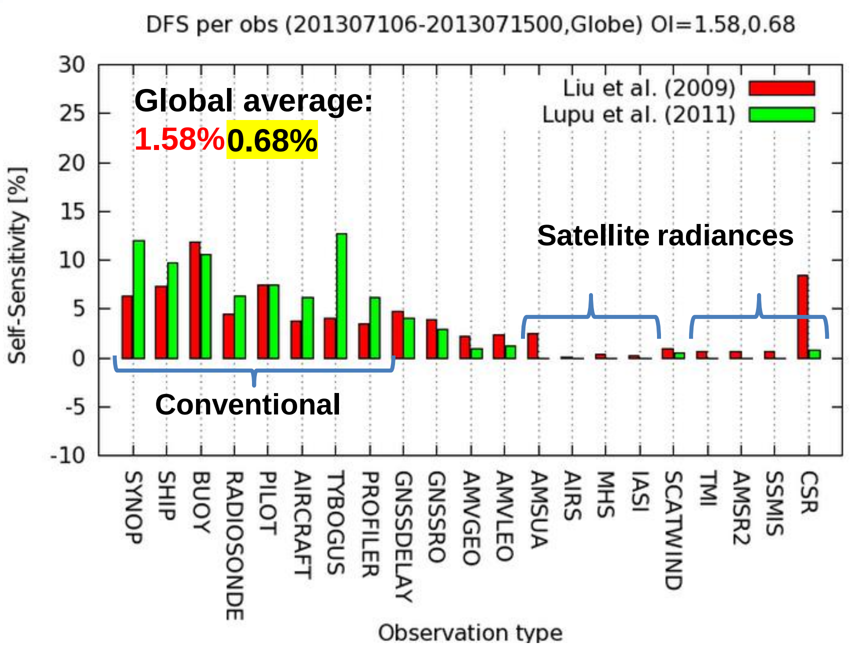
- DA system: hybrid LETKF/4D-Var coupled with JMA GSM
  - Resolution: (outer) TL959L100 ; (inner and ensemble) T319L100
  - Window: 6 hours (analysis time +/- 3 hours)
  - **B** weights: 0.85 for static, 0.25 for ensemble
  - Member size: 50
  - Localization scales (e-folding):
    - LETKF: Horizontal: 400km, Vertical: 0.4 scale heights
    - 4D-Var: Horizontal: 800km, Vertical: 0.8 scale heights
  - Covariance Inflation: Adaptive inflation of Miyoshi (2011)

identical to EFSO in Part II-1

- 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 method of Lupu et al. (2011) as a double check:
    - $\mathrm{tr}(\mathbf{HK}) = \mathrm{tr}\left(\mathbf{R}^{-1}\mathbb{E}\left(\mathbf{d}_{\mathrm{b}}^{\mathrm{a}}(\mathbf{d}_{\mathrm{a}}^{\mathrm{o}})^{\mathrm{T}}\right)\right)$ with the expectation evaluated as the average over a period and samples, assuming ergodicity and homogeneity

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

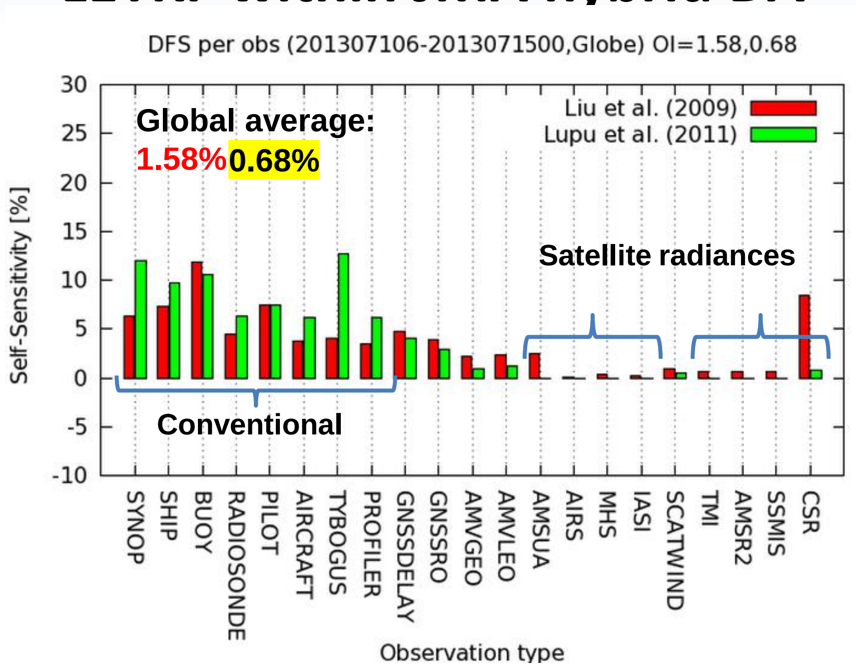
LETKF within JMA hybrid DA



- Reasonable agreement between the two methods (at least for conventional obs).
- **Shockingly small observational impact**:
  - for JMA only about **1%** of information comes from obs.,
  - whereas it is about 20% for ECMWF 4D-Var(Cardinali, 2013; ECMWF lecture notes)

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

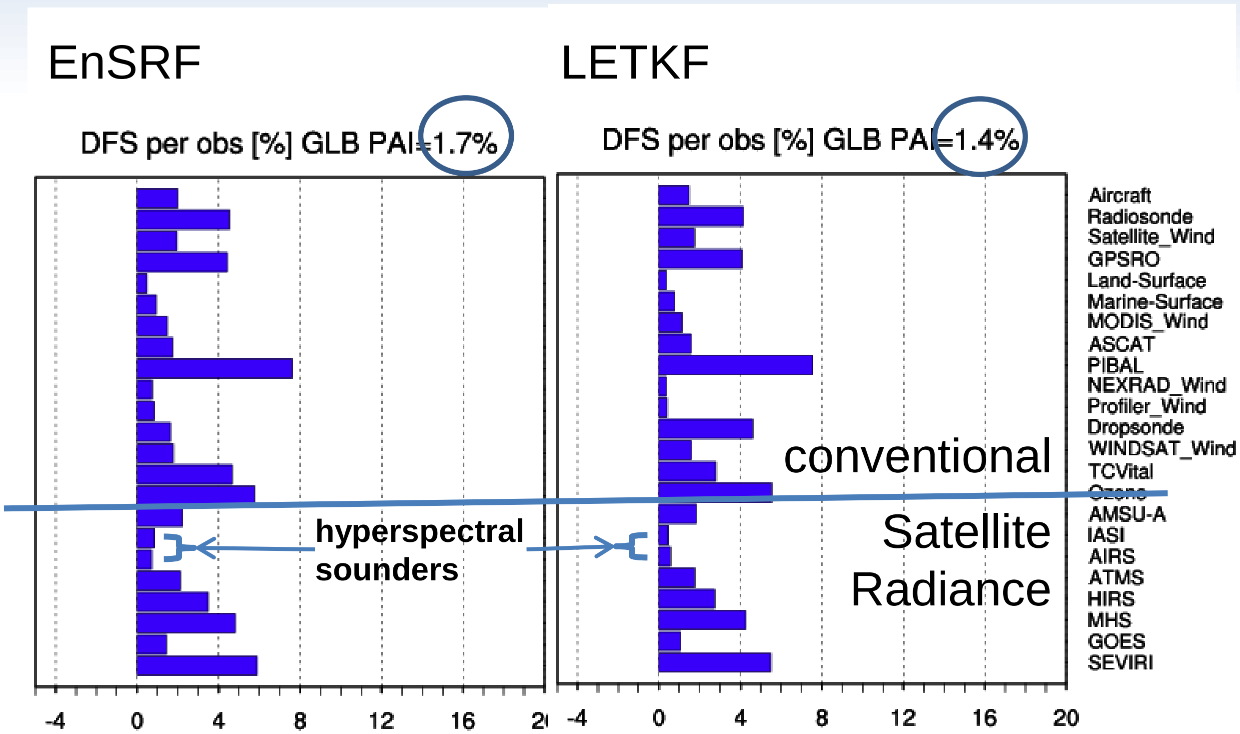
LETKF within JMA hybrid DA



- 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)

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 (from Part I) 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.

Why DFS so small for LETKF?

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

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

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

$$\mathrm{tr}(\boldsymbol{S}_{\mathrm{global}}) = \sum_i (\boldsymbol{S}_{\mathrm{loc}})_{ii} \sim p_{\mathrm{global}} \times \frac{K-1}{p_{\mathrm{loc}}}$$

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

Implications

- We have seen that, for an EnKF with ensemble size much smaller than the number of the locally assimilated observations, DFS is inevitably bounded from above by the member size.
- This means that such a system cannot fully extract information from observations.
- We believe this fact has a lot of important implications, e.g.:
  - why ***covariance inflation*** is necessary,
  - ***what the localization scale should be***, given the ensemble size and observation density,
  - how, ***in serial assimilation,*** the ***order*** of assimilating observations affects the accuracy of the analysis,
  - how ***dense observations*** should be ***thinned***, …etc.

Implication on covariance inflation

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

- Traditionally, nonlinearity and model errors are considered to be the source of necessity for covariance inflation.
  - From this aspect, it is $\mathbf{B}$ rather than $\mathbf{A}$ that is underestimated.
  - This is true for Extended Kalman Filter.
- The inherent underestimation of DFS could be another mechanism that requires covariance inflation for EnKF.
- If so, diagnostics of DFS could give some guidance on how to inflate covariance.

Implication on covariance localization

- Traditionally, it is believed that localization is necessary to filter out spurious correlations in $\mathbf{P}^{\mathbf{b}}$ 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 $\{\lambda_i\}$ be the eigenvalues 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 = $\sum_i \frac{\lambda_i}{1+\lambda_i}$ (see the extra-slide)
- → DFS will not be underestimated if K-th largest eigenvalue λ_K is negligibly small.
- This gives an criterion for the optimal member size K given the observation network and background error covariance.
- Inversely, given the member size K, we can choose localization scale so that DFS is not artificially bounded. For this, we require that the observations within the localized area are few enough such that $\lambda_i \ll 1$ for some $i < K$.

Implication on order of obs. assimilation in serial EnKF

- 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 $\rho := \frac{\mathrm{HAH^T}}{\mathrm{HBH^T}}$ to those with the largest,
  - assigning larger localization scale to obs. with smaller ρ.
- 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. Jeff Whitaker's successful method is equivalent to:
  - assimilating observations from largest DFS to those with smallest,
  - assigning a larger localization scale to obs. with larger DFS
- → DFS argument could provide this method with theoretical justification.

Implication for observation thinning

- Related to the previous argument, in a situation where thinning of observations is necessary (e.g., very dense observation such as satellite hyperspectral sounding, radar data, etc.), it would make sense to assimilate only the observations with largest DFS.

Summary of DFS diagnostics

- Ensemble based DFS estimation of Liu et al. (2009) is implemented for the first time (perhaps) to a quasi-operational DA system with real data.
- It was found that DFS is critically bounded from above by the member size.
- The above result entails a lot of implications for possible improvements to EnKF methodology.

Last comment....

- Our diagnostics, both EFSO and DFS, strongly indicates importance of larger members.

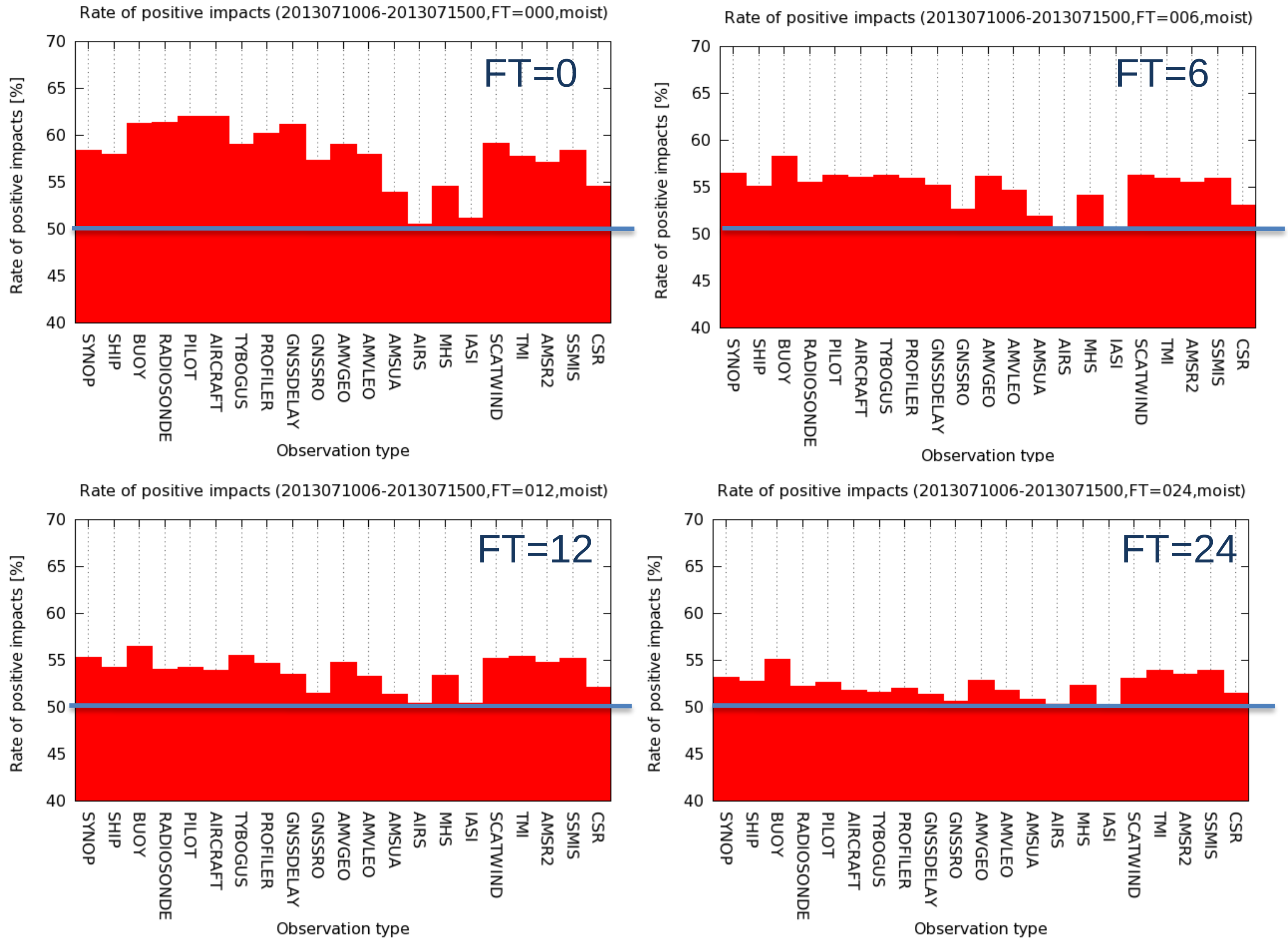
- RIKEN/AICS DA team is leading the world in the area of ensemble DA with massively many members. Your experience/insight will be invaluable for us!

- We would like to collaborate with RIKEN DA team, especially from this aspect!

Thank you very much!

Backup slides for Part I

Percentage of positively-contributing observations (target=globe; norm=moist total energy)



- At FT=24, only slightly more than 50% of the observations contribute to improve forecast (as pointed out by many FSO studies in the literature).
- Percentage of “helpful” observations increases for shorter evaluation lead-time.
- → Consistent with Hotta (2014, PhD. Dissertation).

More applications of EFSO

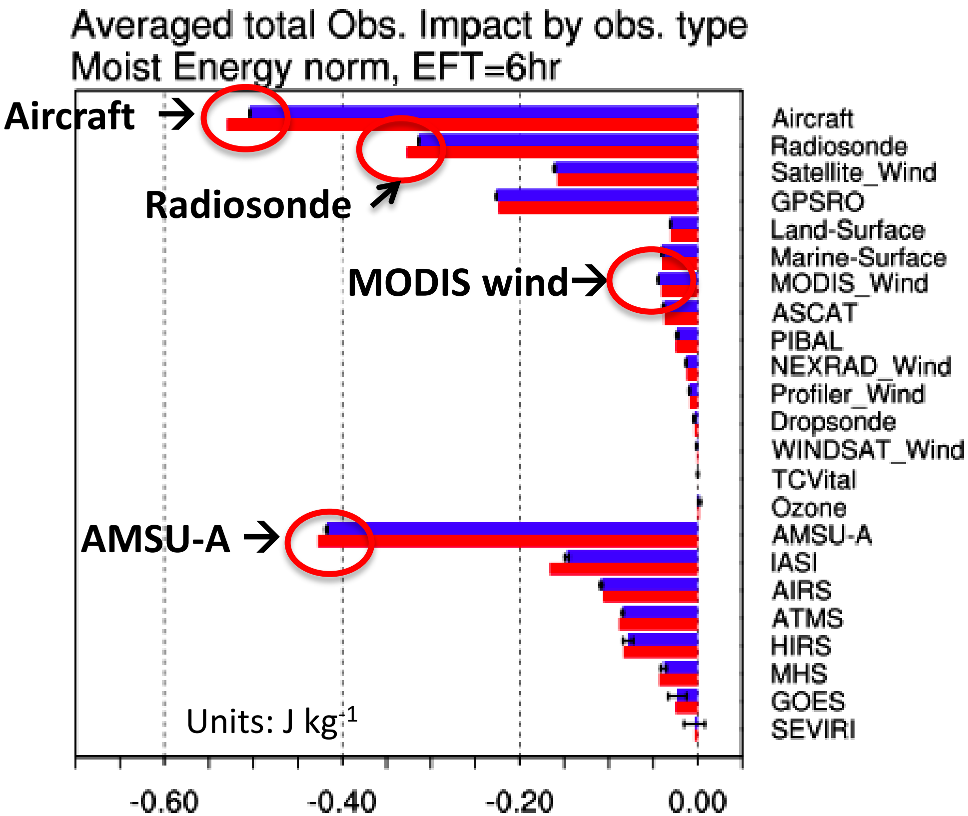
(1) Collaboration with instrument/retrieval developers:

- EFSO can be used to build a database of *flawed* observations along with their relevant metadata.
- → Providing such database to instrument/retrieval developers would help them identify/fix their problems.

(2) Acceleration of development for assimilation of new observing systems:

- Current OSE-based approach:
  - Impact from new obs is obscured by the presence of many obs that are already assimilated.
  - → Difficult to extract statistically significant signals.
  - **EFSO-based data selection will enable efficient determination of an optimal way to assimilate new observing systems.**

Tuning Experiment: Result
EFSO before/after tuning of **R**



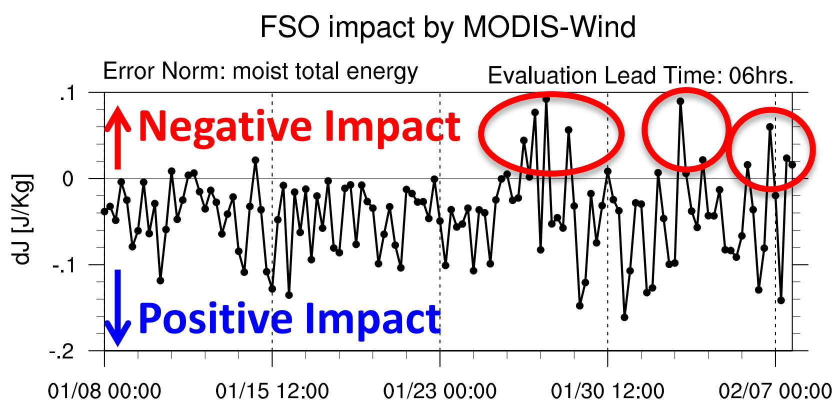
- Aircraft, Radiosonde, AMSU-A:
  - significant improvement of EFSO-impact

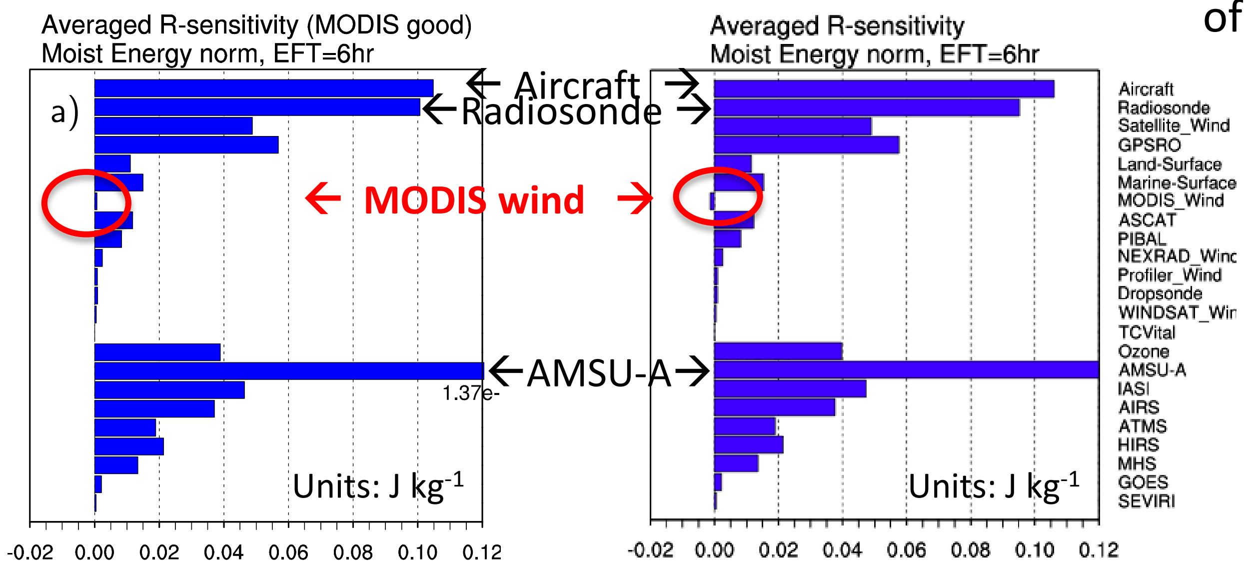
  (as expected)
- MODIS wind :
  - No improvement in EFSO (contrary to expectation)

Why no improvement in MODIS?

- MODIS had “flawed” obs. along with “helpful” obs.
- The “flawed” obs. might have resulted in incorrect estimation of EFSR.

Excluding cases where MODIS wind had negative impact





Excluding "flawed" MODIS case

Including "flawed" MODIS case

- MODIS wind exhibited several negatively-impacting cases.
- **Exclude negative cases**
- **→ EFSR for MODIS becomes neutral**
- → Consistent with the result of tuning experiment

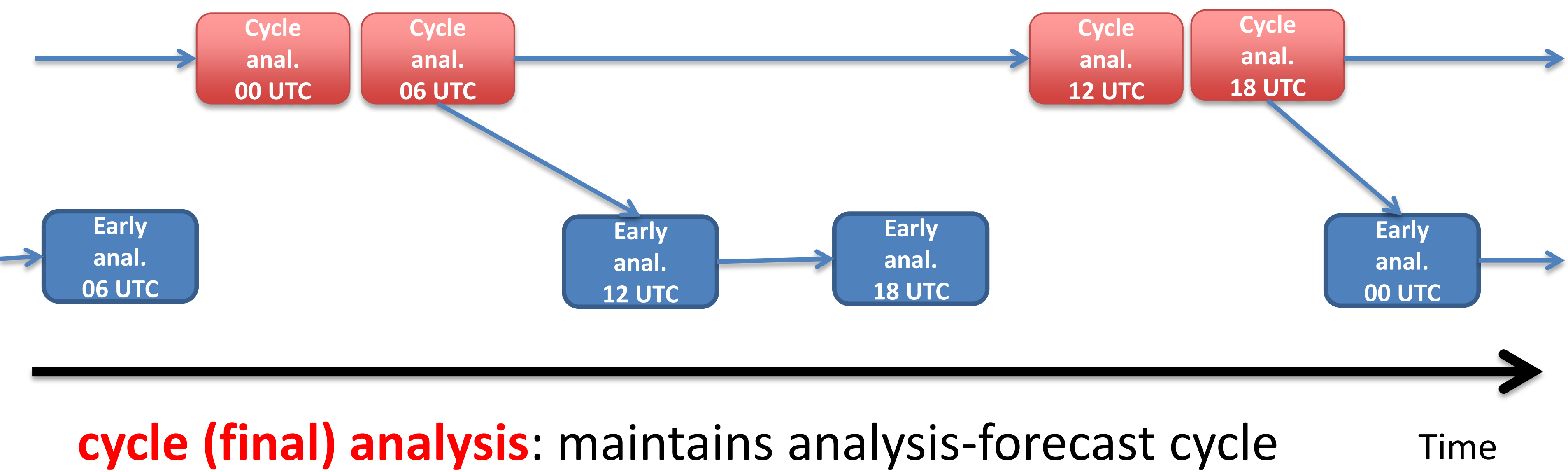
Lesson:

- Before performing EFSR, we should remove "bad" obs.

Implementation to the real operational system
(1) Can we wait for 6 hours?

Idea: Exploit the time lag between "**early analysis**" and "**cycle (final) analysis**"

(suggested by Dr. John Derber, 2013)



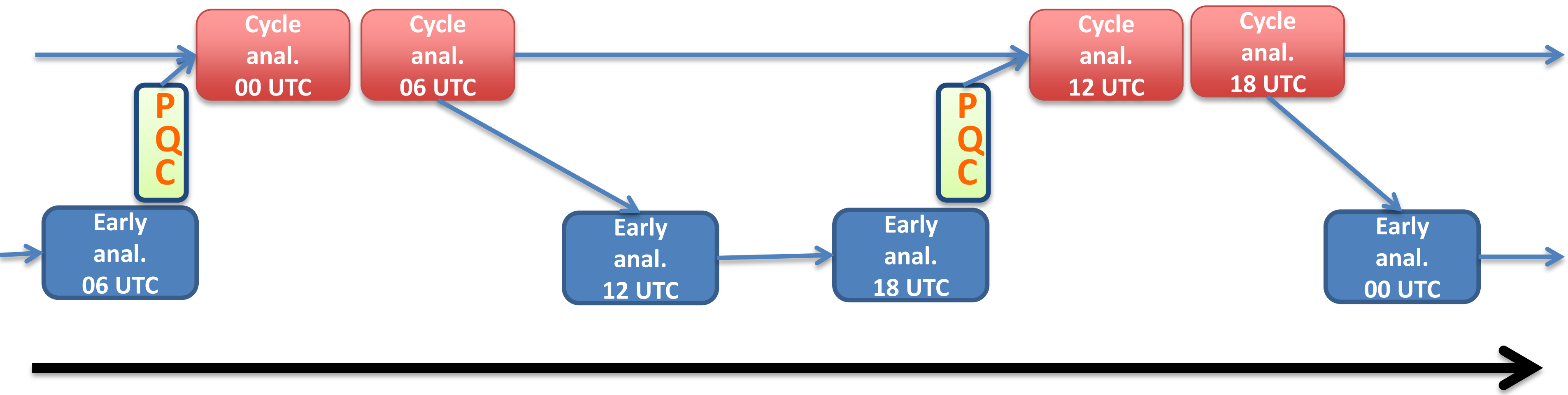
cycle (final) analysis: maintains analysis-forecast cycle

early analysis: provides initial condition for extended forecast

Implementation to the real operational system (1) we don't need to wait 6 hours!

Idea: Exploit the time lag between "**early analysis**" and "**cycle (final) analysis**"

(suggested by Dr. John Derber, 2013)



cycle (final) analysis: maintains analysis-forecast cycle

early analysis: provides initial condition for extended forecast

Backup slides for Part II

Decomposition of fcst error into column- and null- space of fcst ptbs

- Fix a grid and consider a local patch that would be used if an observation was located at the grid point in question. In the derivation below, all vectors/matrices are assumed to be restricted to this local patch.
- In EnKF, the sum of each column of $\mathbf{X}^f$ is zero, so rank($\mathbf{X}^f$)=$K-1$:

  $\mathrm{span}(\tilde{\mathbf{X}}^f) = \mathrm{span}\left([\tilde{\mathbf{X}}^f]_1, \cdots, [\tilde{\mathbf{X}}^f]_{K-1}, [\tilde{\mathbf{X}}^f]_K\right) = \mathrm{span}\left([\tilde{\mathbf{X}}^f]_1, \cdots, [\tilde{\mathbf{X}}^f]_{K-1}\right)$

In light of this, we now denote by $\tilde{\mathbf{X}}^f$ the first $K-1$ columns of the original $\tilde{\mathbf{X}}^f$.

- Now, suppose that $\tilde{\mathbf{e}} := \mathbf{C}^{\frac{1}{2}}(\mathbf{e}^f_{t|0} + \mathbf{e}^f_{t|-6})$ can be decomposed as
- $\tilde{\mathbf{e}} = \tilde{\mathbf{e}}_{\mathrm{span}} + \tilde{\mathbf{e}}_{\mathrm{null}}$, $\tilde{\mathbf{e}}_{\mathrm{span}} = \sum_{\boldsymbol{k=1}}^{\boldsymbol{K-1}} \alpha_k [\tilde{\mathbf{X}}^f]_{\boldsymbol{k}} = \tilde{\mathbf{X}}^f \boldsymbol{\alpha}$,

  $\boldsymbol{\alpha} = (\alpha_1, \quad \cdots \quad \alpha_{K-1})^T$

  Multiplying by $\left(\mathbf{C}^{1/2}\mathbf{X}^f\right)^T =: \tilde{\mathbf{X}}^{f^T}$ from left, $\tilde{\mathbf{e}}_{\mathrm{null}}$ by definition vanishes, giving:

$$\tilde{\mathbf{X}}^{f^T} \tilde{\mathbf{e}} = \tilde{\mathbf{X}}^{f^T}(\tilde{\mathbf{X}}^f \boldsymbol{\alpha} + \tilde{\mathbf{e}}_{\mathrm{null}}) = \tilde{\mathbf{X}}^{f^T} \tilde{\mathbf{X}}^f \boldsymbol{\alpha}$$

$$\therefore \boldsymbol{\alpha} = \left(\tilde{\mathbf{X}}^{f^T} \tilde{\mathbf{X}}^f\right)^{-1} \tilde{\mathbf{X}}^{f^T} \tilde{\mathbf{e}}$$

- Once $\boldsymbol{\alpha}$ is determined, we can obtain $\|\tilde{\mathbf{e}}_{\mathrm{span}}\|^2$ and $\|\tilde{\mathbf{e}}_{\mathrm{null}}\|^2$ by

$$\|\tilde{\mathbf{e}}_{\mathrm{span}}\|^2 = \|\tilde{\mathbf{X}}^f \boldsymbol{\alpha}\|^2 = (\tilde{\mathbf{X}}^f \boldsymbol{\alpha})^T (\tilde{\mathbf{X}}^f \boldsymbol{\alpha}) = \boldsymbol{\alpha}^T \tilde{\mathbf{X}}^{f^T} \tilde{\mathbf{X}}^f \boldsymbol{\alpha} = \boldsymbol{\alpha}^T \tilde{\mathbf{X}}^{f^T} \tilde{\mathbf{e}}$$

$$\|\tilde{\mathbf{e}}_{\mathrm{null}}\|^2 = \|\tilde{\mathbf{e}}\|^2 - \|\tilde{\mathbf{e}}_{\mathrm{span}}\|^2$$

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
- $\mathrm{tr}(\mathbf{S}) \equiv \mathrm{tr}(\mathbf{HK}) = \mathrm{tr}\left(\mathbf{HAH}^{\mathrm{T}}\mathbf{R}^{-1}\right)\ \left(\because \mathbf{K} = \mathbf{AH}^{\mathrm{T}}\mathbf{R}^{-1}\right)$
- LETKF estimates the analysis error covariance by :
- $\mathbf{A} = \mathbf{X^b}\widetilde{\mathbf{A}}\mathbf{X^b}^{\mathrm{T}}, \widetilde{\mathbf{A}} = \left[(\mathrm{K}-1)\mathbf{I} + \mathbf{Y^b}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{Y^b}\right]^{-1} = \frac{1}{\mathrm{K}-1}\left(\mathbf{I}+\mathbf{Z}^{\mathrm{T}}\mathbf{Z}\right)^{-1}$, with $\mathbf{Z} \equiv \frac{1}{\sqrt{\mathrm{K}-1}}\mathbf{R}^{-\frac{1}{2}}\mathbf{Y^b}$
- $\mathbf{Z}^{\mathrm{T}}\mathbf{Z}$ is a $\mathrm{K}\times\mathrm{K}$ positive definite symmetrix matrix. Its eigenvalue decomposition becomes :
- $\mathbf{Z}^{\mathrm{T}}\mathbf{Z} = \mathbf{U\Lambda U}^{-1}, \mathbf{UU}^{-1} = \mathbf{I},\quad \mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_{\mathrm{K}})$
- Since $\mathrm{rank}(\mathbf{Z}) = \mathrm{K}-1$, $\lambda_{\mathrm{K}} = 0$. From positive definiteness of $\mathbf{Z}^{\mathrm{T}}\mathbf{Z}$, $\lambda_{\mathrm{i}} > 0\ (1 \le \mathrm{i} \le \mathrm{K}-1)$.
- Thus:
- $\mathbf{HK} = \mathbf{HAH}^{\mathrm{T}}\mathbf{R}^{-1} = \mathbf{HX^b}\widetilde{\mathbf{A}}\mathbf{X^b}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1} = \mathbf{Y^b}\widetilde{\mathbf{A}}\mathbf{Y^b}^{\mathrm{T}}\mathbf{R}^{-1} = \left(\sqrt{\mathrm{K}-1}\mathbf{R}^{\frac{1}{2}}\mathbf{Z}\right)\widetilde{\mathbf{A}}\left(\sqrt{\mathrm{K}-1}\mathbf{R}^{\frac{1}{2}}\mathbf{Z}\right)^{\mathrm{T}}\mathbf{R}^{-1} =$
  $\mathbf{R}^{\frac{1}{2}}\mathbf{Z}\left(\mathbf{I}+\mathbf{Z}^{\mathrm{T}}\mathbf{Z}\right)^{-1}\left(\mathbf{R}^{\frac{1}{2}}\mathbf{Z}\right)^{\mathrm{T}}\mathbf{R}^{-1}$
- Because trace is invariant under cyclic reordering,
  $\mathrm{tr}(\mathbf{S}) \equiv \mathrm{tr}(\mathbf{HK}) = \mathrm{tr}\left(\mathbf{R}^{\frac{1}{2}}\mathbf{Z}\left(\mathbf{I}+\mathbf{Z}^{\mathrm{T}}\mathbf{Z}\right)^{-1}\left(\mathbf{R}^{\frac{1}{2}}\mathbf{Z}\right)^{\mathrm{T}}\mathbf{R}^{-1}\right) = \mathrm{tr}\left(\mathbf{Z}\left(\mathbf{I}+\mathbf{Z}^{\mathrm{T}}\mathbf{Z}\right)^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{R}^{\frac{1}{2}}\mathbf{R}^{-1}\mathbf{R}^{\frac{1}{2}}\right) =$
  $\mathrm{tr}\left(\mathbf{Z}\left(\mathbf{I}+\mathbf{Z}^{\mathrm{T}}\mathbf{Z}\right)^{-1}\mathbf{Z}^{\mathrm{T}}\right) = \mathrm{tr}\left(\left(\mathbf{I}+\mathbf{Z}^{\mathrm{T}}\mathbf{Z}\right)^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{Z}\right) = \mathrm{tr}\left(\left(\mathbf{I}+\mathbf{U\Lambda U}^{-1}\right)^{-1}\mathbf{U\Lambda U}^{-1}\right) =$
  $\sum_{\mathrm{i}=1}^{\mathrm{K}}\frac{\lambda_{\mathrm{i}}}{1+\lambda_{\mathrm{i}}} = \frac{\lambda_1}{1+\lambda_1} + \cdots + \frac{\lambda_{\mathrm{K}-1}}{1+\lambda_{\mathrm{K}-1}} + \frac{0}{1+0} \le \mathrm{K}-1$