

# Using Data Assimilation to improve the models and the observations

---

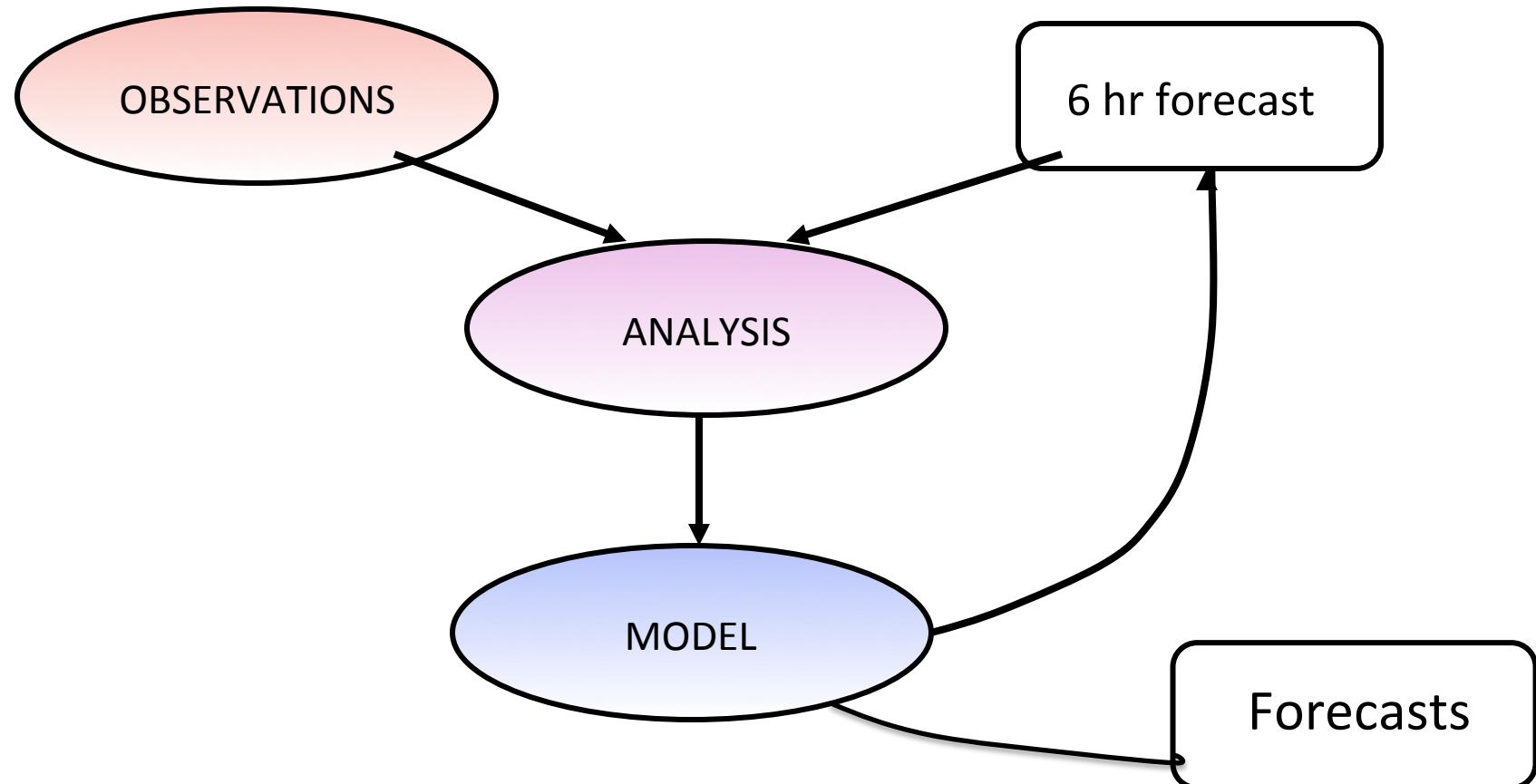
Tse-Chun Chen<sup>1</sup>, Kriti Bhargava<sup>1</sup>, Daisuke Hotta<sup>2</sup>,  
Ji-Sun Kang<sup>3</sup>

Jim Carton<sup>1</sup>, Eugenia Kalnay<sup>1</sup>

<sup>1</sup>AOSC <sup>2</sup>JMA <sup>1,3</sup>KISTI

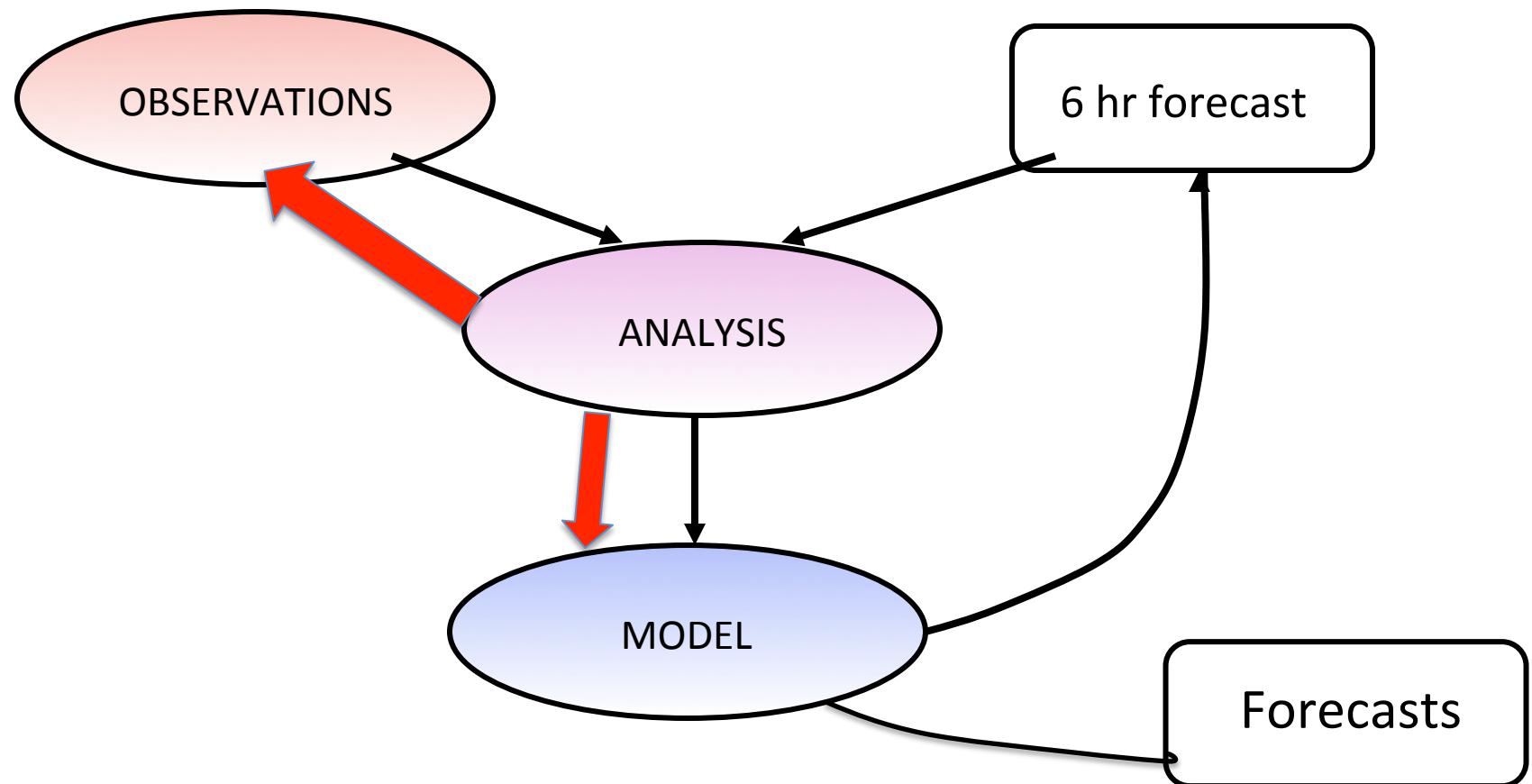
**CICS 22 and 23 November 2015**

**Classic Data Assimilation:** For NWP we need to improve **observations**, **analysis scheme** and **model**



**New Data Assimilation:** We can also use DA  
to improve **observations** and **model**

---



The simplicity and power of EnKF should encourage the use of DA for improvements beyond its main goal

---

## **Combine optimally observations and model forecasts (mostly done! ☺)**

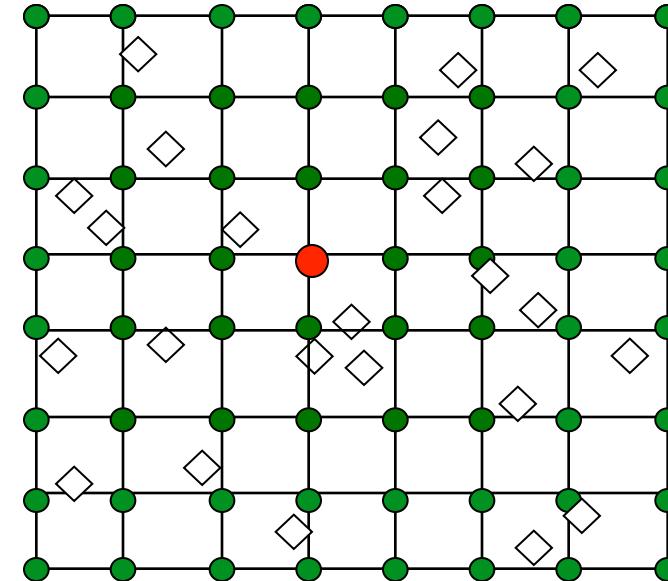
- We should also use DA to:
  - 1) Improve the observations**
  - 2) Improve the model**

# LETKF: Localization based on observations

---

Perform data assimilation in a local volume, choosing observations

**The state estimate is updated at the central grid red dot**



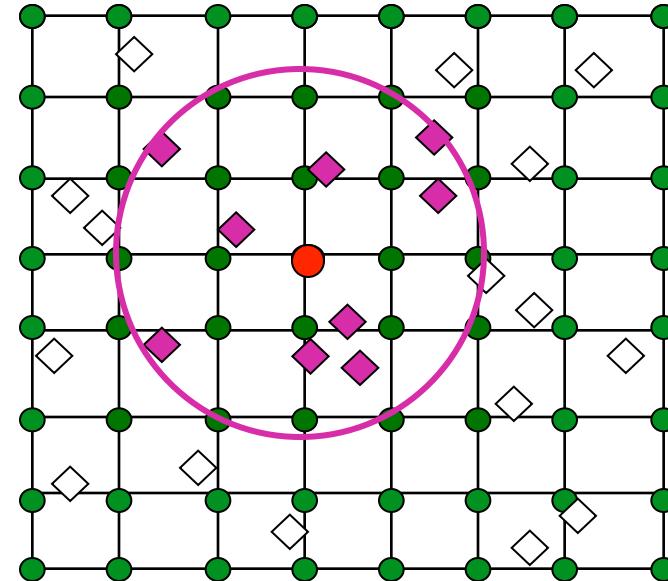
# LETKF: Localization based on observations

---

Perform data assimilation in a local volume, choosing observations

**The state estimate is updated at the central grid red dot**

**All observations (purple diamonds) within the local region are assimilated**



The LETKF algorithm can be described in a single slide!

# Local Ensemble Transform Kalman Filter (Hunt et al, 2007)

---

**Globally:**

**Forecast step:**

**Analysis step: construct**

$$\mathbf{x}_{n,k}^b = M_n \left( \mathbf{x}_{n-1,k}^a \right)$$

$$\mathbf{X}^b = \left[ \mathbf{x}_1^b - \bar{\mathbf{x}}^b \mid \dots \mid \mathbf{x}_K^b - \bar{\mathbf{x}}^b \right];$$

$$\mathbf{y}_i^b = H(\mathbf{x}_i^b); \mathbf{Y}_n^b = \left[ \mathbf{y}_1^b - \bar{\mathbf{y}}^b \mid \dots \mid \mathbf{y}_K^b - \bar{\mathbf{y}}^b \right]$$

---

**Locally:** Choose for each grid point the observations to be used, and compute the local analysis error covariance and perturbations in ensemble space:

$$\tilde{\mathbf{P}}^a = \left[ (K-1) \mathbf{I} + \mathbf{Y}^T \mathbf{R}^{-1} \mathbf{Y} \right]^{-1}; \mathbf{W}^a = [(K-1) \tilde{\mathbf{P}}^a]^{1/2}$$

**Analysis mean in ensemble space:**  $\bar{\mathbf{w}}^a = \tilde{\mathbf{P}}^a \mathbf{Y}^{bT} \mathbf{R}^{-1} (\mathbf{y}^o - \bar{\mathbf{y}}^b)$

and add to  $\mathbf{W}^a$  to get the analysis ensemble in ensemble space.

The new ensemble analyses in model space are the columns of  $\mathbf{X}_n^a = \mathbf{X}_n^b \mathbf{W}^a + \bar{\mathbf{x}}^b$ . Gathering the grid point analyses forms the new global analyses. Note that the the output of the LETKF are analysis weights  $\bar{\mathbf{w}}^a$  and perturbation analysis matrices of weights  $\mathbf{W}^a$ . These weights multiply the ensemble forecasts.

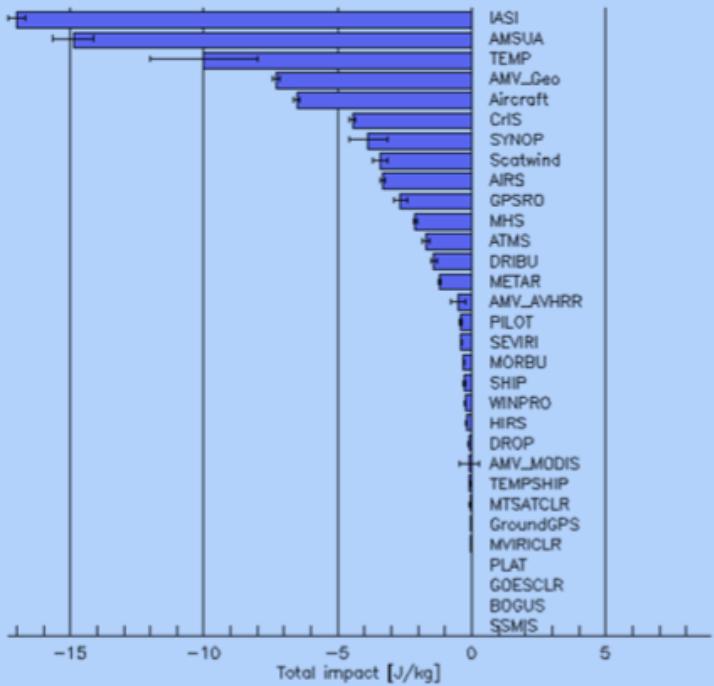
# Forecast Sensitivity to Observations (Langland and Baker, 2004)



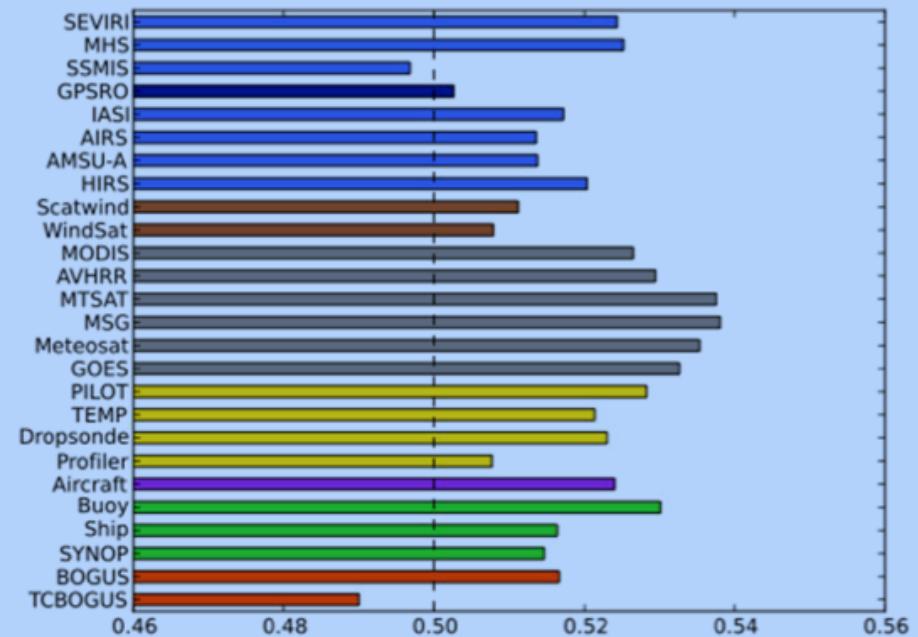
Met Office

## FSOI in Global NWP

Total Observation Impact (Aug 2014)



Fraction Obs that Improve Forecast



- Infra-Red (IASI) and microwave (AMSSUA) radiances now biggest impact.
- Note only ~50% of observations reduce forecast error(!).
- Estimate: need 6 months time series to assess impact for single observing site.
- EFSO methodology now being considered when no adjoint available

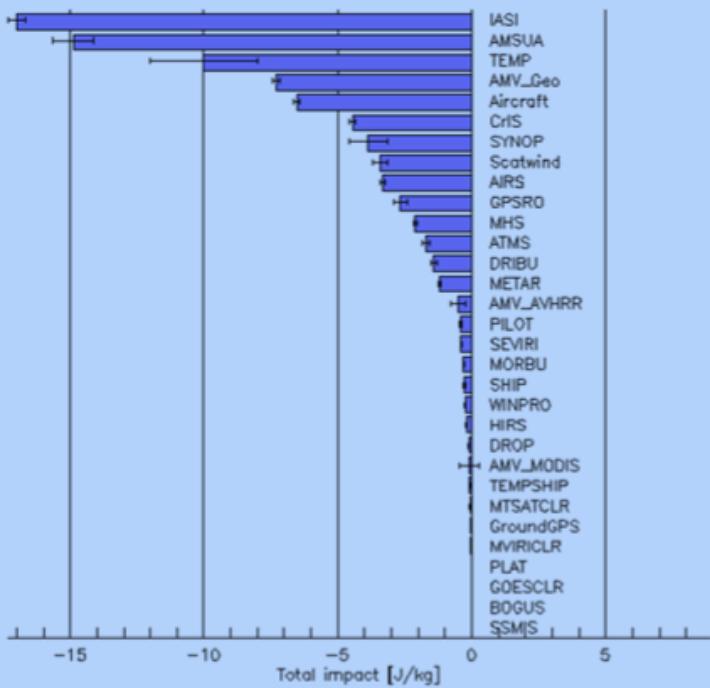
# Can we identify bad observations?



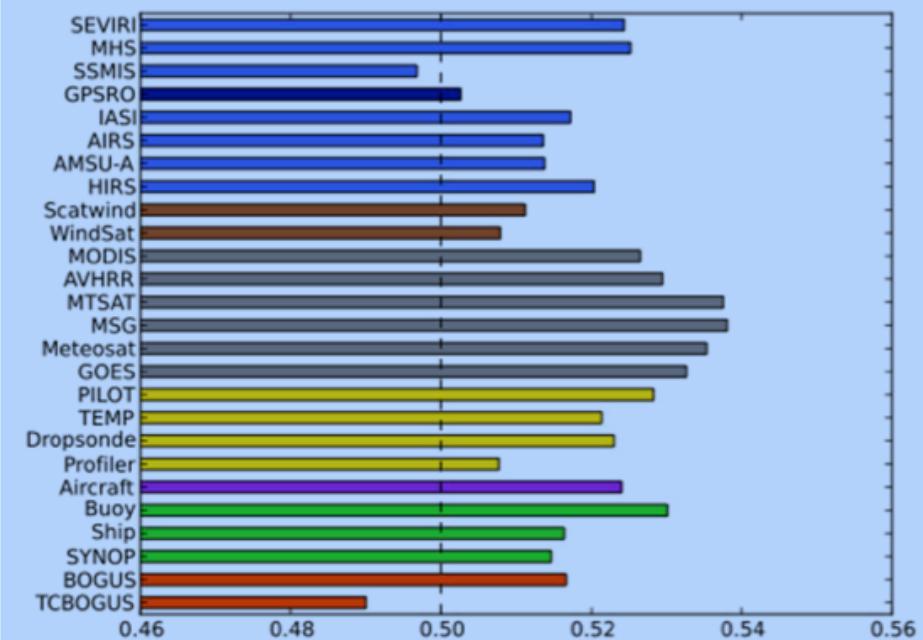
## FSOI in Global NWP

Met Office

Total Observation Impact (Aug 2014)



Fraction Obs that Improve Forecast



- Infra-Red (IASI) and microwave (AMSAU) radiances now biggest impact.
- Note only ~50% of observations reduce forecast error(!).
- Estimate: need 6 months time series to assess impact for single observing site.
- EFSO methodology now being considered when no adjoint available

# 1) Improve the observations: Ensemble Forecast Sensitivity to Observations and Proactive QC

---

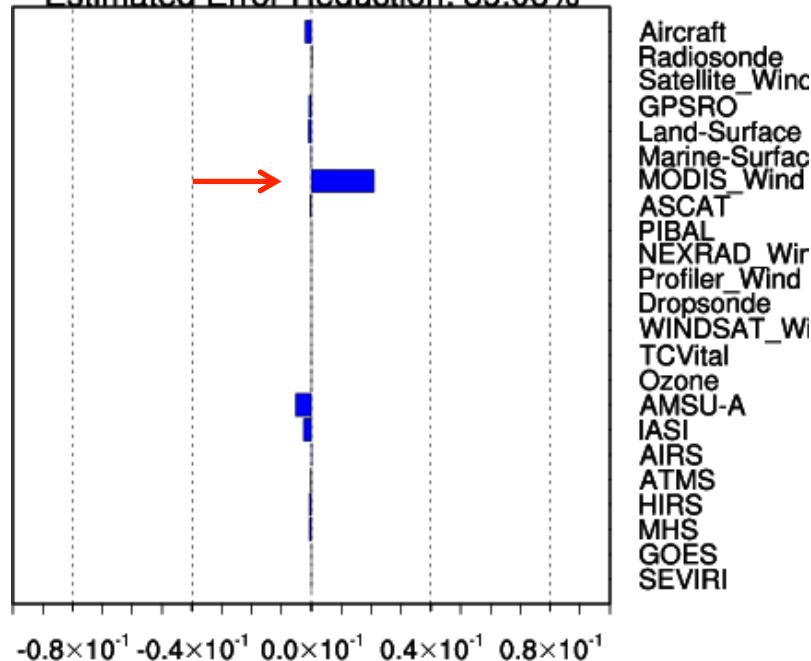
- Kalnay et al. (2012) derived EFSO.
- Ota et al. (2013) tested 24hr GFS forecasts and showed EFSO could be used to identify bad obs.
- D. Hotta (2014): EFSO can be used after only 6 hours: the bad obs. can be collected and withdrawn, with all needed useful metadata, so **the algorithm can be improved**.
- The analysis is corrected with EFSO.
- We call this **Proactive QC**, much stronger than QC.
- Hotta also showed EFSO **can be used to tune R**
- Tse-Chun Chen (2015) tested impact of EFSO/PQC over 5 day forecasts: **PROMISING RESULTS**

# Hotta (2014)

Feb. 18 06UTC, near the North Pole  
(Ota et al. 2013 case). Bad obs: MODIS WINDS

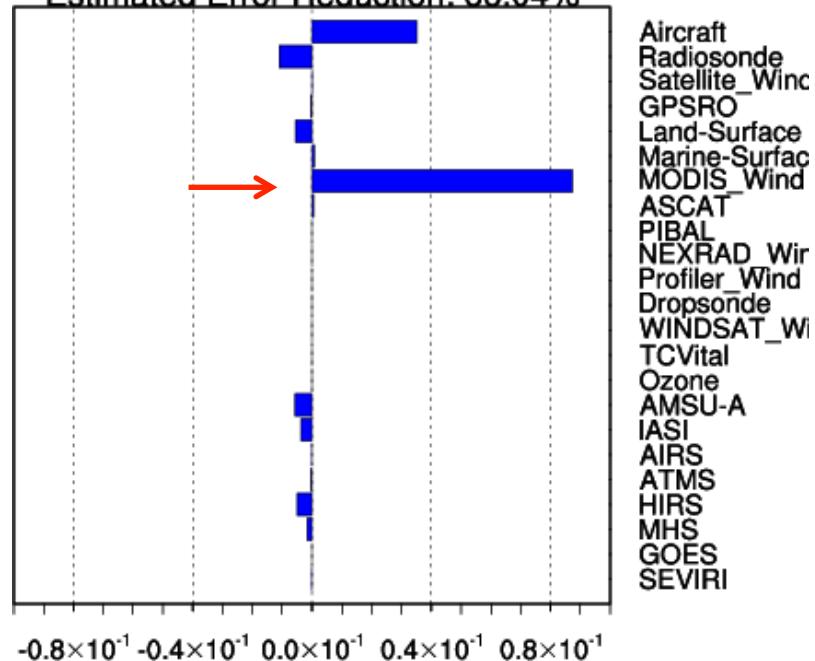
**FT=06 hr.**

2012020618  
Total Obs. Impact by obs. type  
Moist Energy norm, EFT=6hr  
[60°N,40°E,70°E]  
Estimated Error Reduction: 39.06%



**FT=24 hr.**

2012020618  
Total Obs. Impact by obs. type  
Moist Energy norm, EFT=24hr  
[60°N,40°E,70°E]  
Estimated Error Reduction: 66.04%

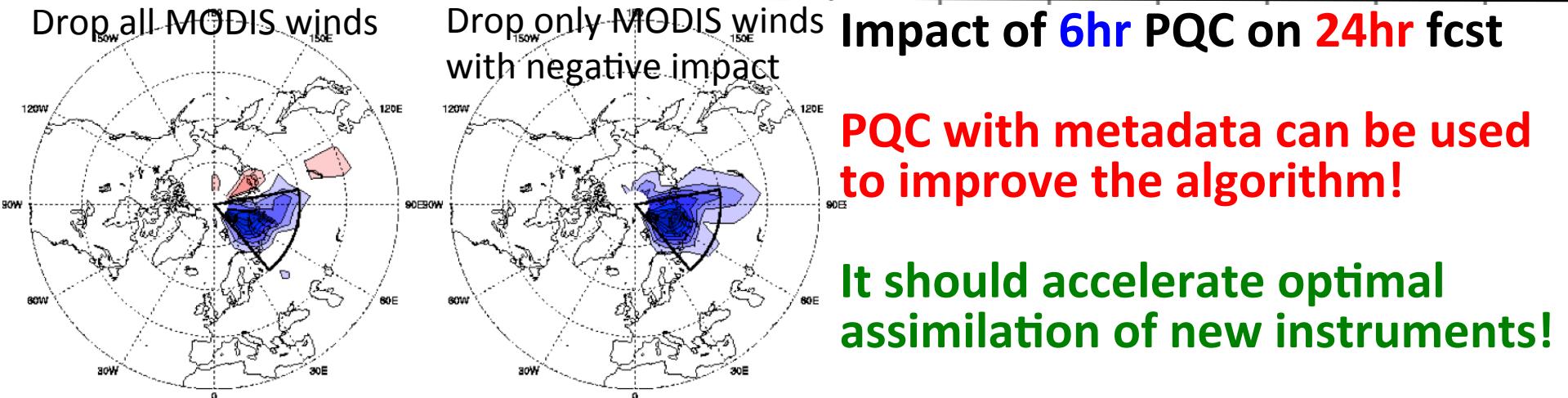
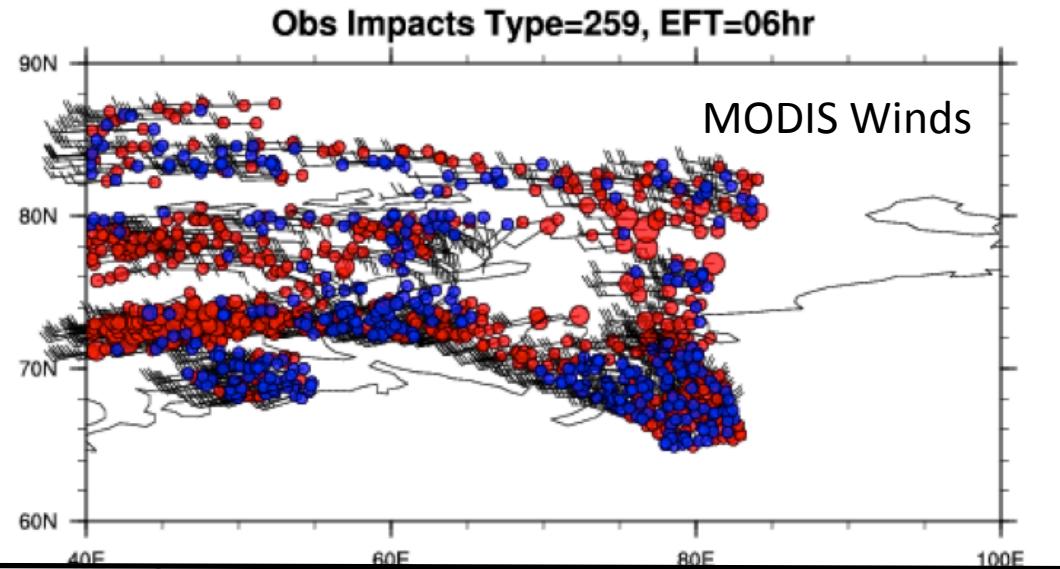


Can identify the bad observations after only 6 hours!

# Improve observations:

## Proactive QC: Find and delete the obs. that make the 6hr forecast worse using EFSO

**Dr. Daisuke Hotta (2014):**  
EFSO is able to find whether each observation improves (blue) or makes the 6hr forecast **worse** (red)



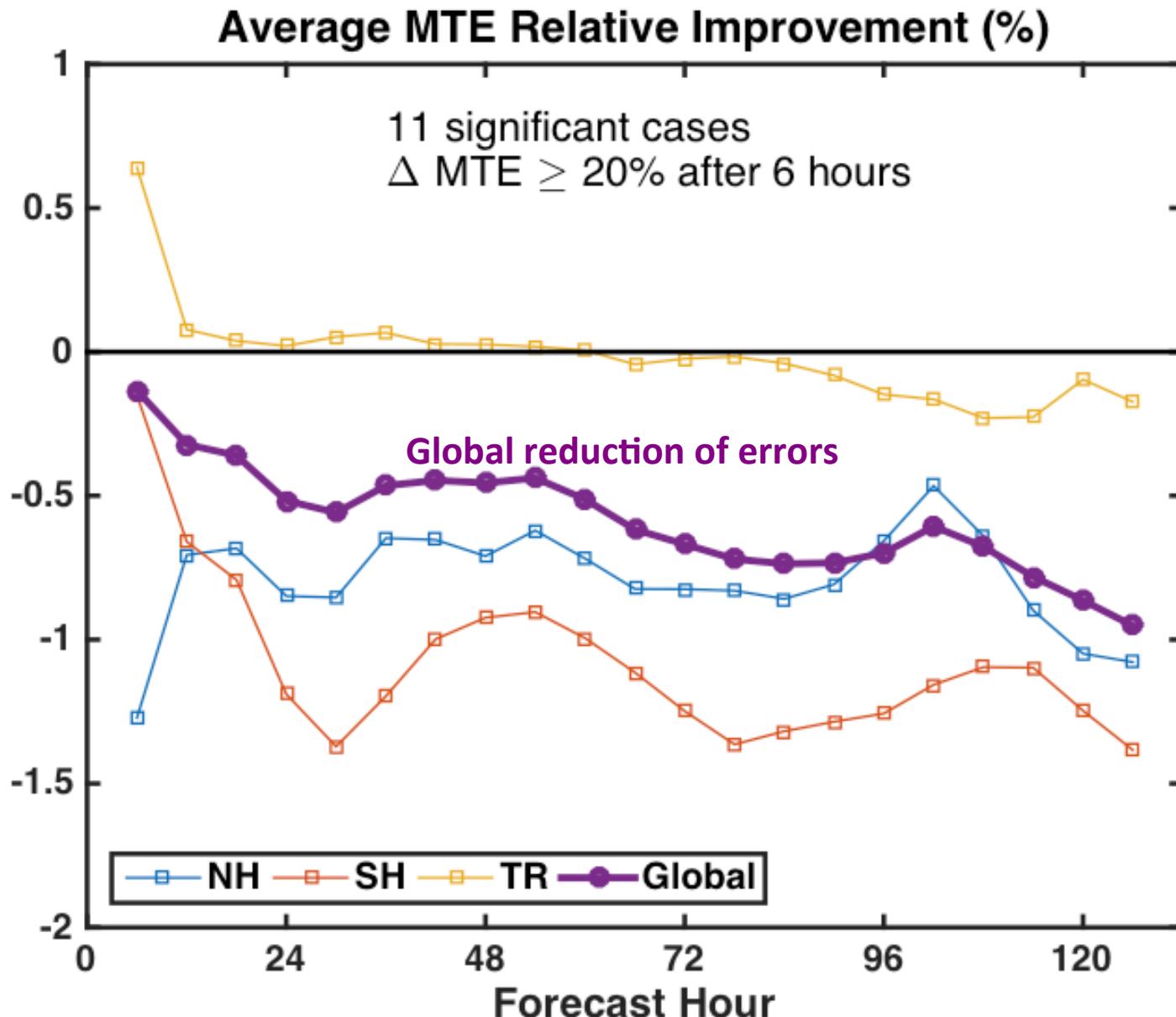
**In one month: 20 cases of skill dropout due to flawed observations that passed the operational QC**

**Tse-Chun Chen:** classify the 20 cases into

- 11 **SIGNIFICANT** cases, where EFSO estimates that withdrawing the flawed observations reduces the 6hour forecast error by more than 20% (in Total Moist Energy)
  - $\Delta \text{TME} > 20\%$
- 9 **NON-SIGNIFICANT CASES**, with  $\Delta \text{TME} < 20\%$

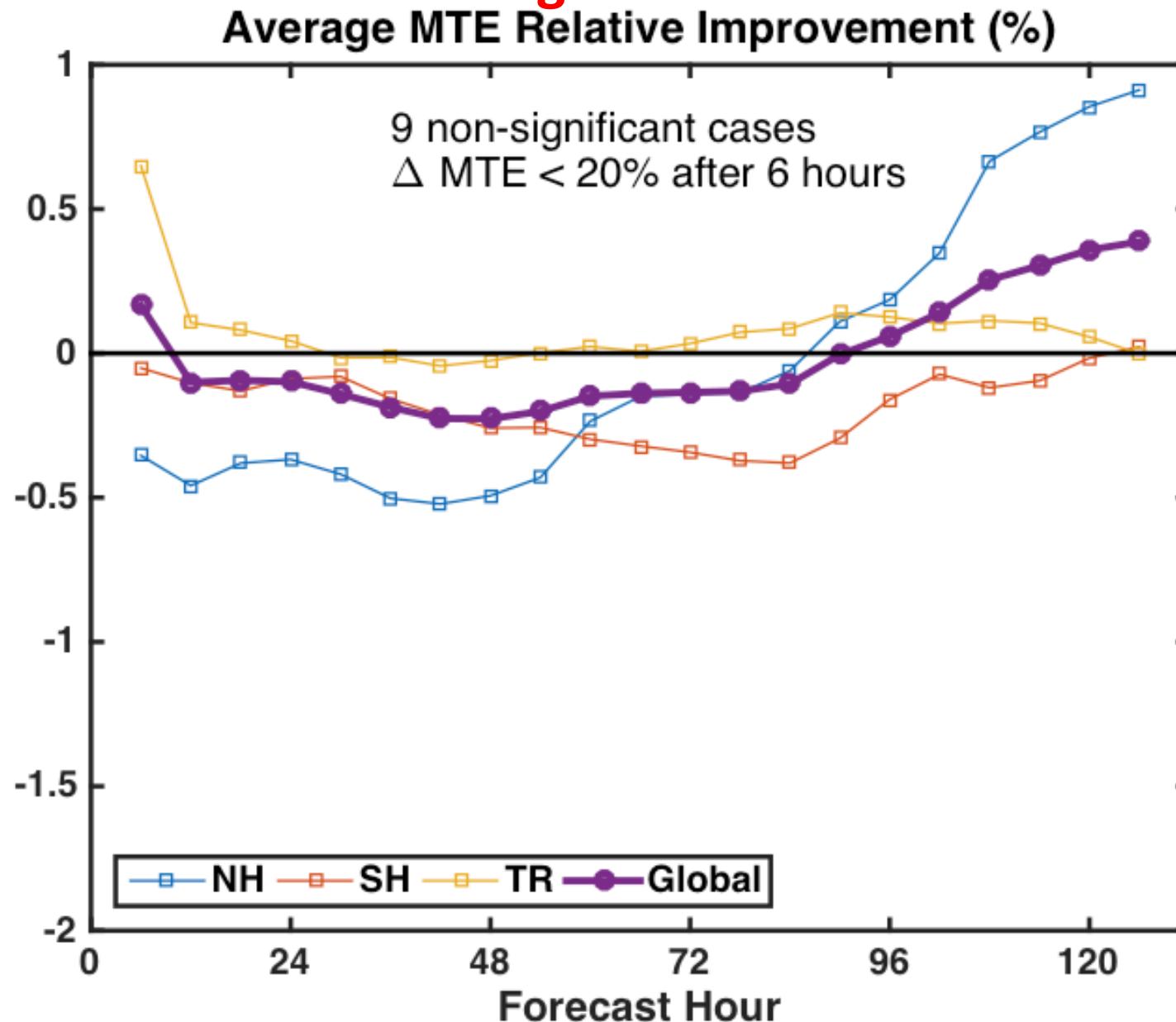
# 5-day reduction of total moist energy of the forecast error

## 11 significant cases



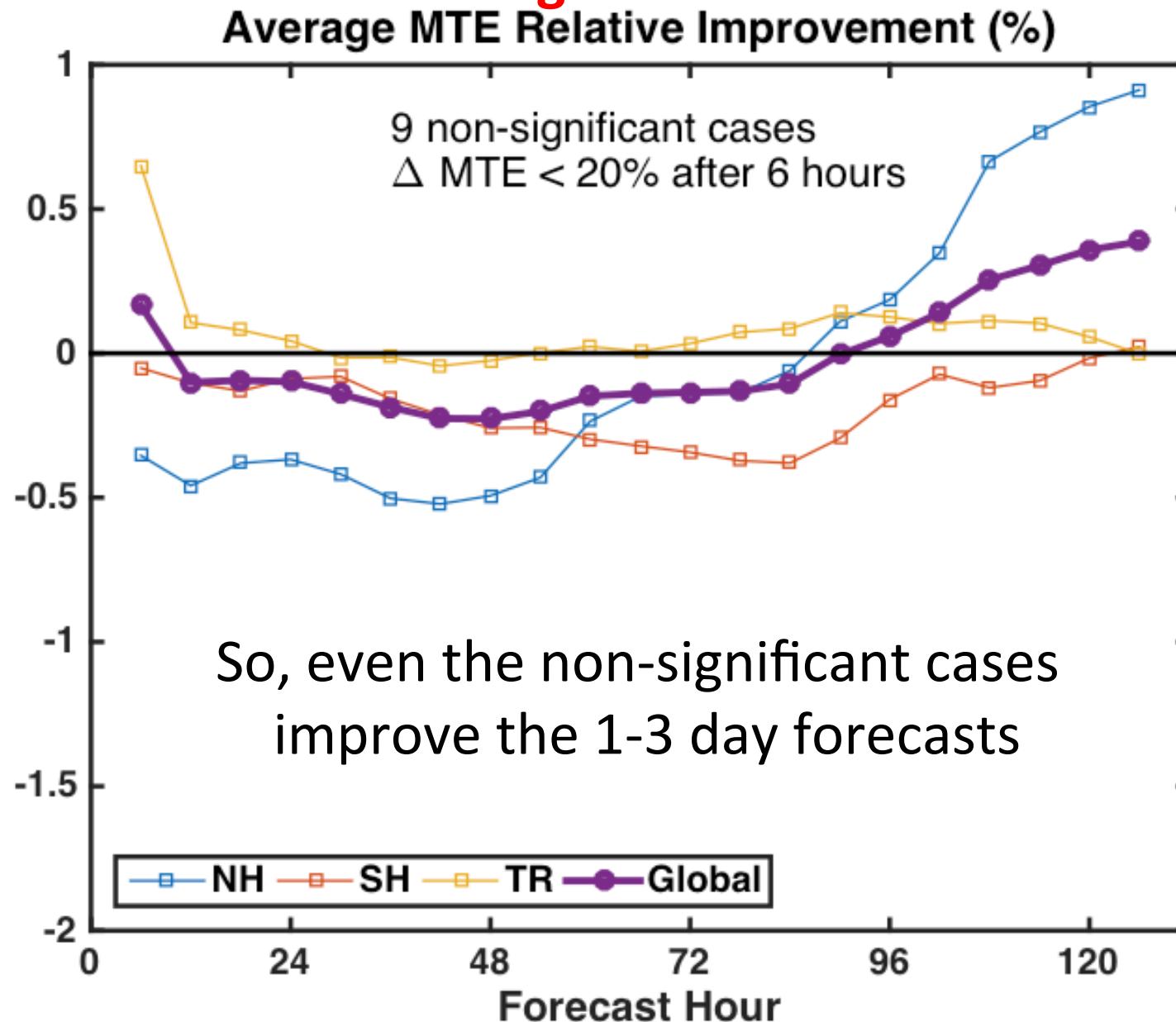
# 5-day reduction of total moist energy of the forecast error

## 9 non-significant cases



## 5-day reduction of total moist energy of the forecast error

### 9 non-significant cases



## 2) Ensemble Forecast Sensitivity to Error Covariances

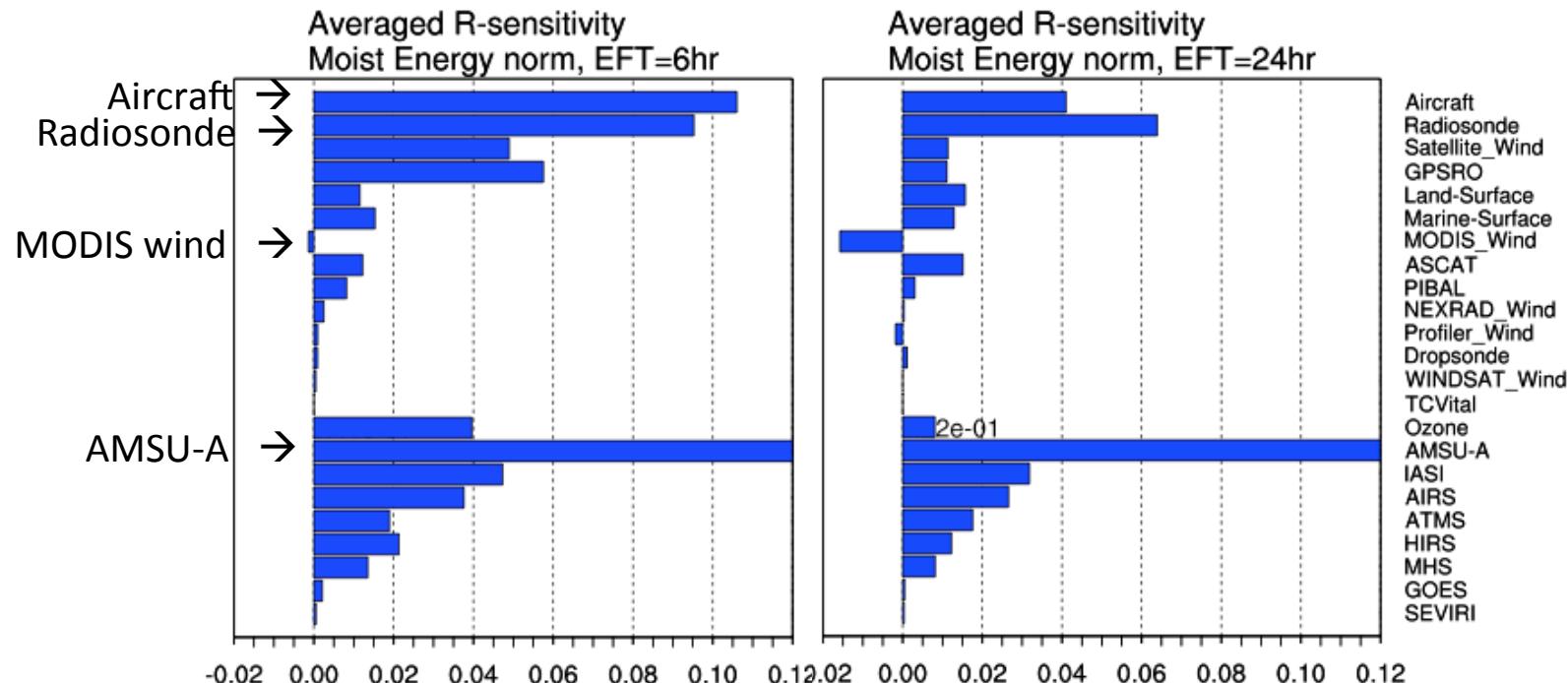
Hotta (2014)

- Daescu and Langland (2013, *QJRMS*) proposed an adjoint-based formulation of forecast sensitivity to **B** and **R** matrix.
- Daisuke Hotta formulated its ensemble equivalent for **R** using **EFSO** by Kalnay et al. (2012) :

$$\left[ \frac{\partial e}{\partial \mathbf{R}} \right]_{ij} \approx \frac{\partial e}{\partial y_i} z_j \approx -\frac{1}{K-1} \left[ \mathbf{R}^{-1} \mathbf{Y}_0^a \mathbf{X}_{t|0}^{fT} \mathbf{C} (\mathbf{e}_{t|0} + \mathbf{e}_{t|-6}) \right]_i [\mathbf{R}^{-1} \delta y^{oa}]_j$$

where **z** is an "intermediate analysis increment" in observation space

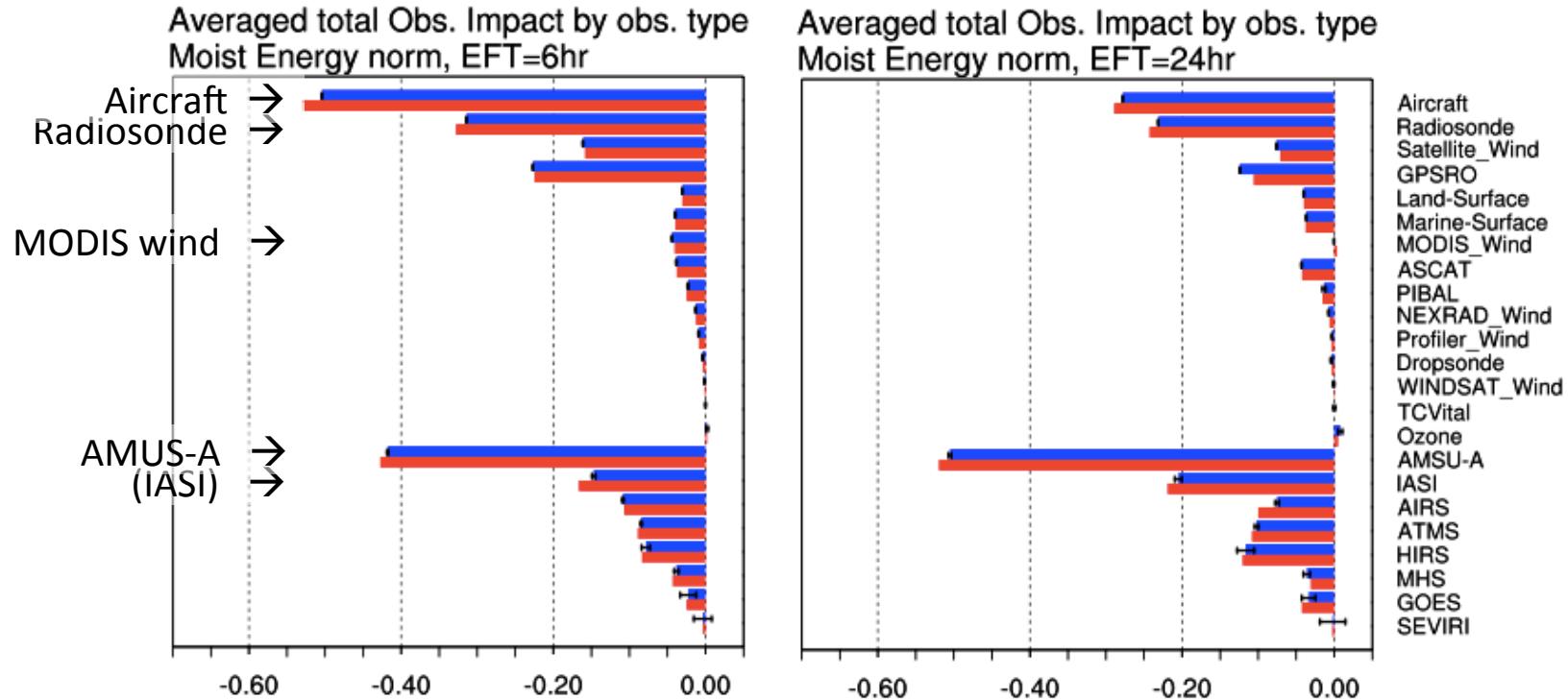
# R-sensitivity results from GFS / GSI-LETKF hybrid



- Positive value: error increases as  $s_o^2$  increases → should decrease  $s_o^2$
- Aircraft, Radiosonde and AMSU-A: large positive sensitivity
- MODIS wind : negative sensitivity
- **Tuning experiment:**
  - Aircraft, Radiosonde and AMSU-A: scale  $s_o^2$  by 0.9
  - MODIS wind: scale  $s_o^2$  by 1.1

# Tuning Experiment: Result

## EFSO before/after tuning of R



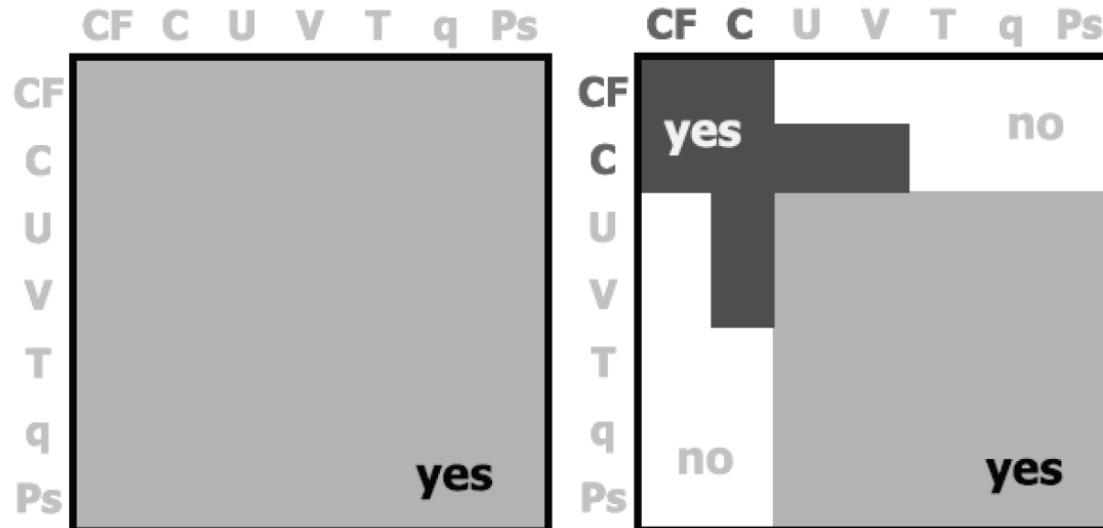
- Aircraft, Radiosonde and AMSU-A: significant improvement of EFSO-impact
- IASI: Significant improvement in EFSO although its error covariance is untouched!
- Very promising results for quick testing of new observing systems!

## 2) Improve the models: Parameter estimation and Estimation of model Bias using DA

---

- Model tuning on long time scales should be done with EnKF parameter estimation.
- Kang et al., JGR, 2011, 2012 showed that evolving surface carbon fluxes can be estimated accurately at the model grid resolution from simulated atmospheric CO<sub>2</sub> observations (OCO-2) as **evolving parameters**.
- Another approach is the use of **analysis increments** to estimate model bias (Greybush et al., 2012, Mars) and even state-dependent model bias (e.g., El Niño bias), as in **Danforth et al. 2007**.

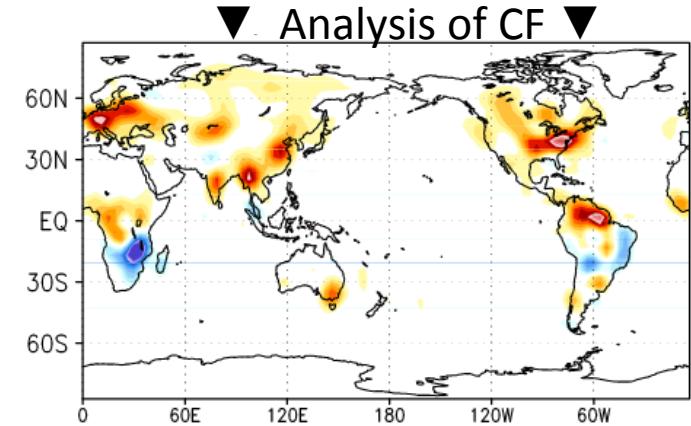
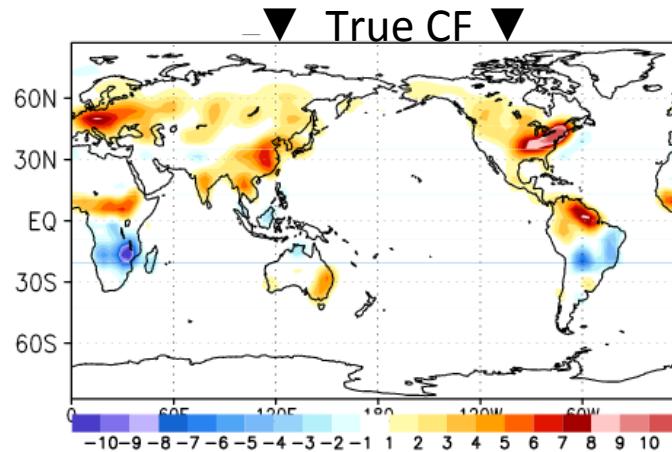
Surface carbon fluxes CF from atmospheric assimilation of meteorological variables and CO<sub>2</sub> obtained as **evolving parameters** (OSSE). Kang et al., JGR, 2011, 2012



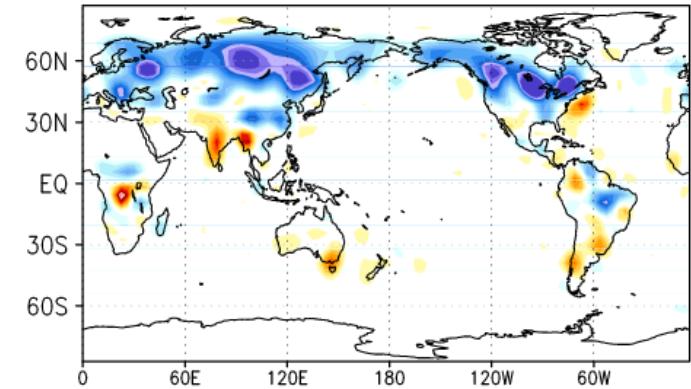
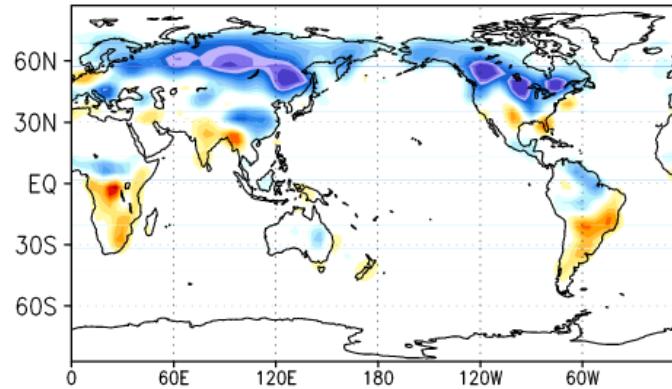
**“Variable  
Localization”  
in the B  
matrix**

# OSSE Results

00Z01APR ►  
After three months of DA

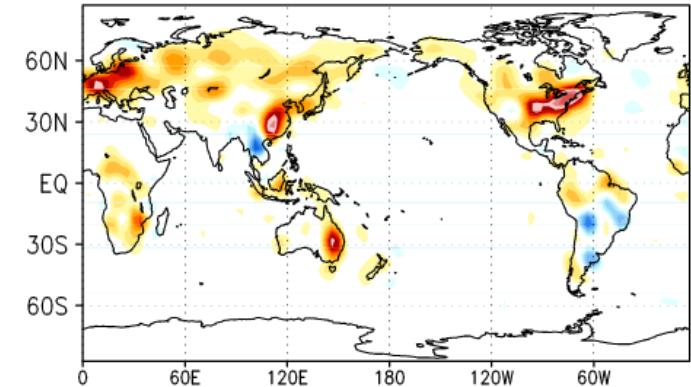
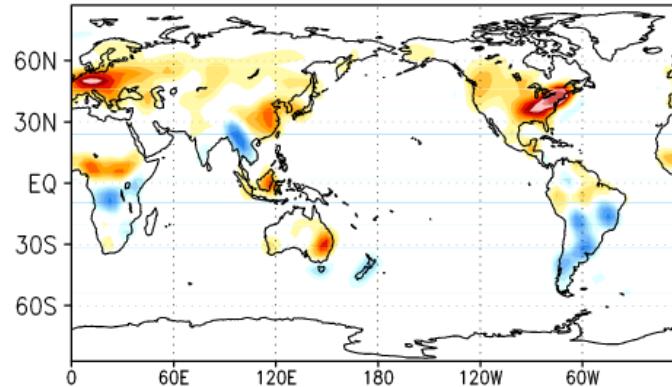


00Z01AUG ►  
After seven months of DA



*We succeeded in estimating time-evolving CF at model-grid scale*

00Z01JAN ►  
After one year of DA



### 3) How can we estimate and correct Big Model bias?

---

- The best current estimate of nature is the Analysis.
- The First Guess (6hr forecast) contains the initial forecast errors (**before they grow nonlinearly**).
- Analysis - First Guess = Analysis Increments (**AI**) =  
- Initial (linear) model errors.
- **The time average of AI is the best estimate of the error growth due to model bias in 6 hr.**
- Danforth, Kalnay and Miyoshi (DKM-2007) estimated the 6hr errors of the SPEEDY model.
- Estimated the average SPEEDY model error (bias) by averaging over several years the 6 hour forecast (started from reanalysis R1) minus the reanalysis.

## DKM-2007 results

---

- Estimated the monthly mean 6hr forecast bias
- Corrected the model by adding  $(-\text{bias}/6\text{hr})$  to each variable time derivative, at each grid point.

### Results

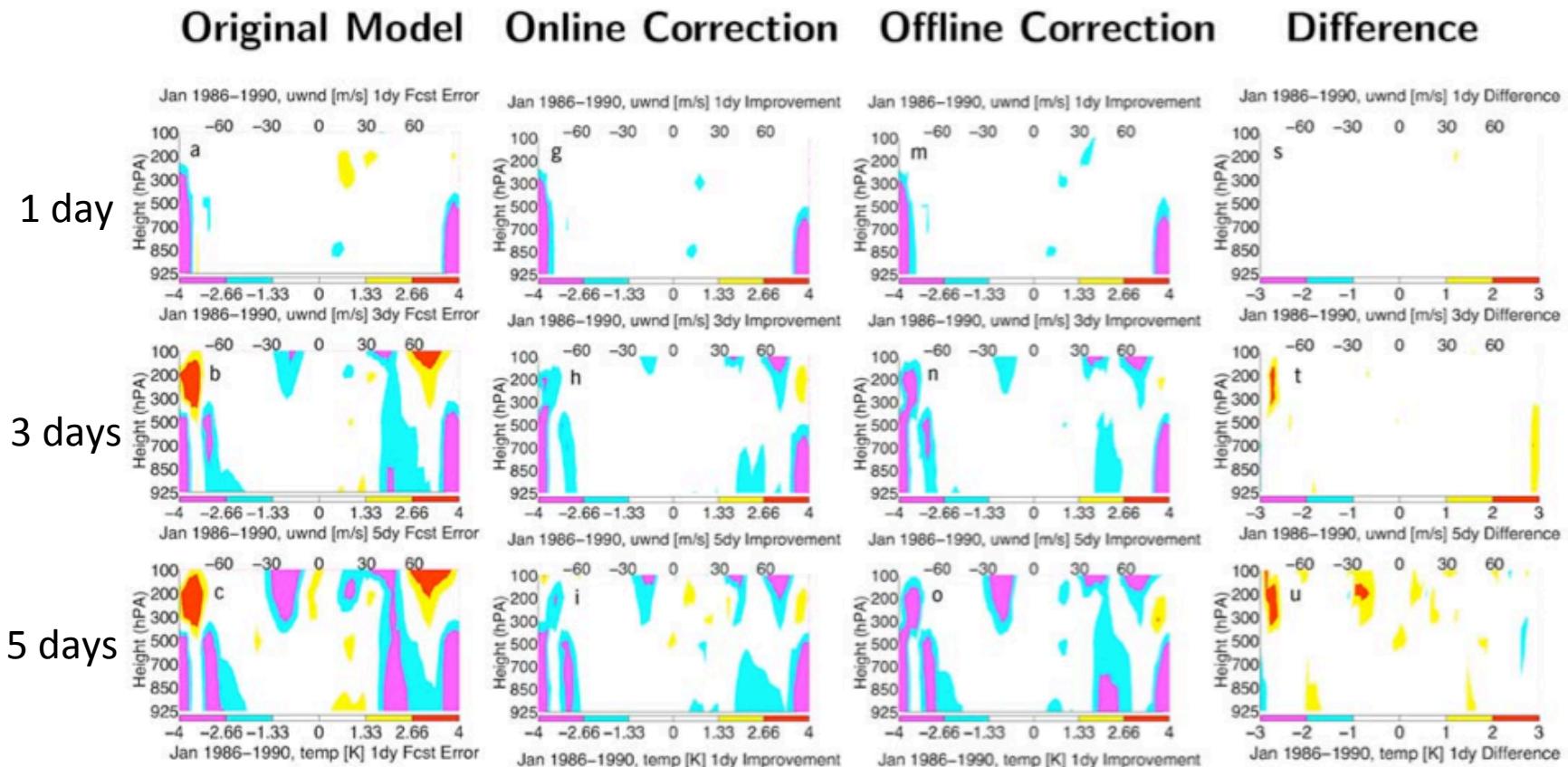
- The bias correction after 3 or 5 days was the same as the best *a posteriori* bias correction.
- But the random errors were **smaller**.
- The dominant EOFs of the 6hr debiased forecast errors were the errors in the diurnal cycle.
- It was possible to estimate the **systematic errors for anomalies** (e.g., ENSO, lows over land or over ocean)

The model corrected online did same or better than the model statistically corrected off-line

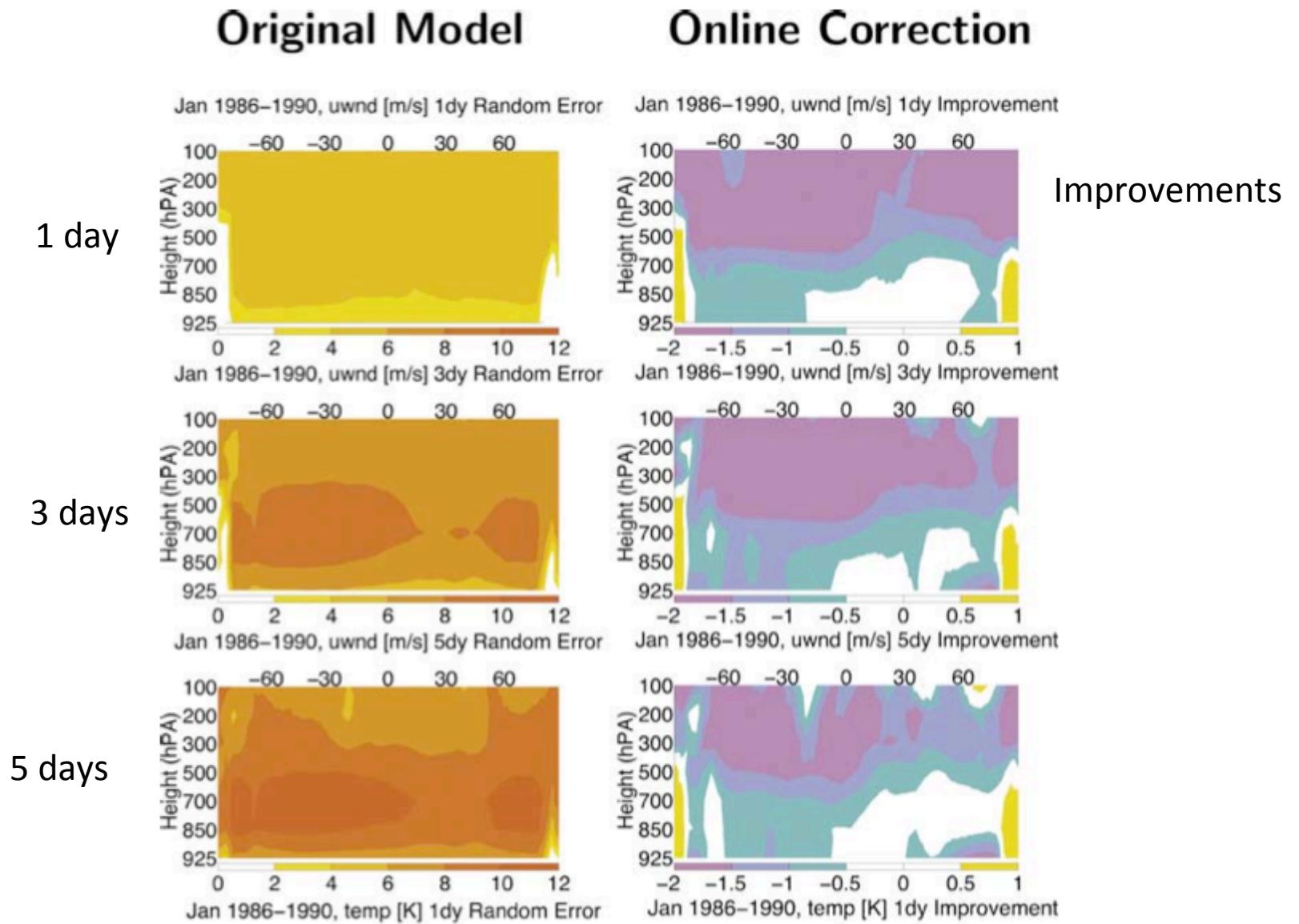
L24805

DANFORTH AND KALNAY: NONLINEAR ERROR GROWTH

L24805



# And the random errors were significantly smaller!

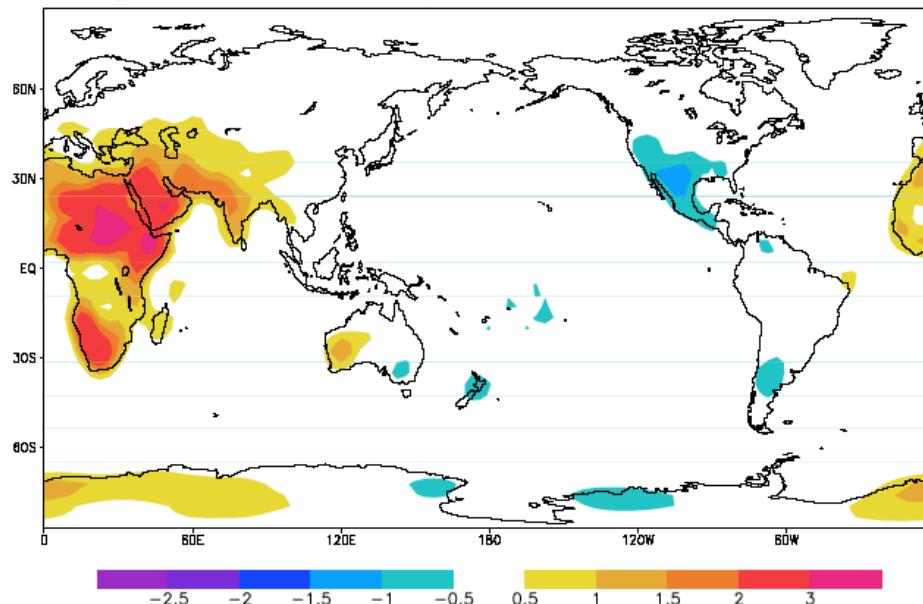


# How to find the **diurnal cycle** model errors using EOFs from a Reanalysis (Danforth et al., 2007)

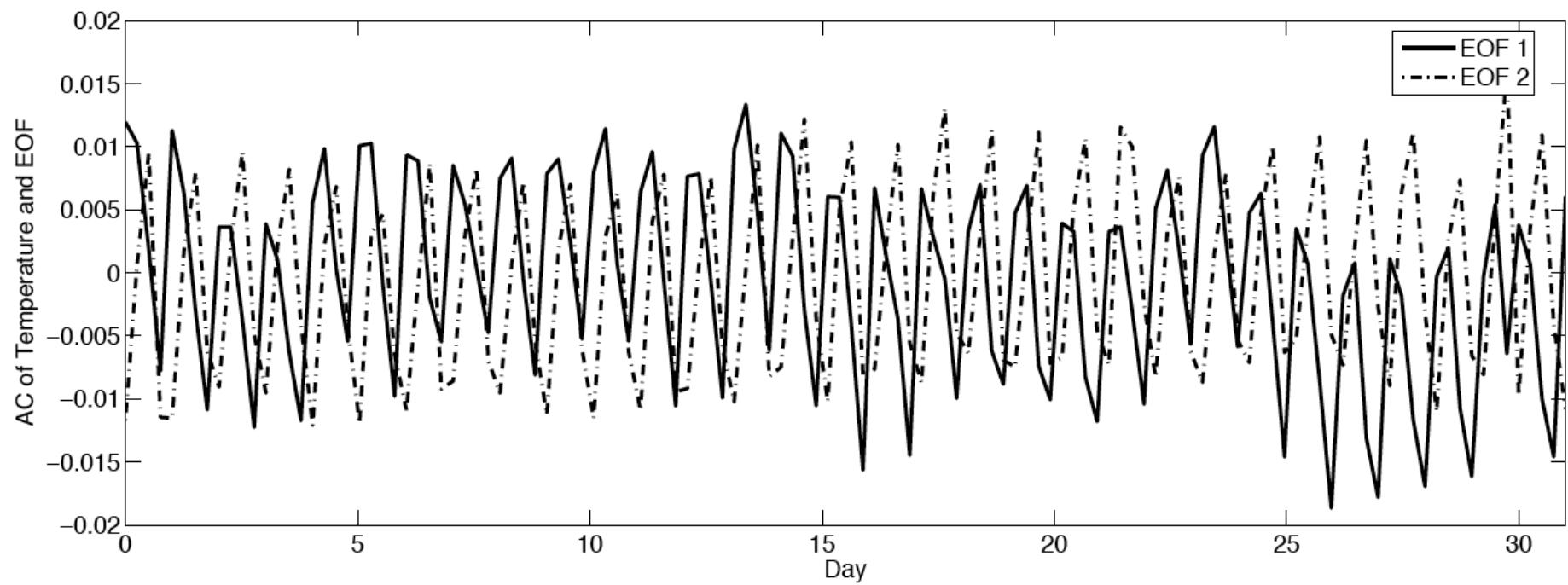
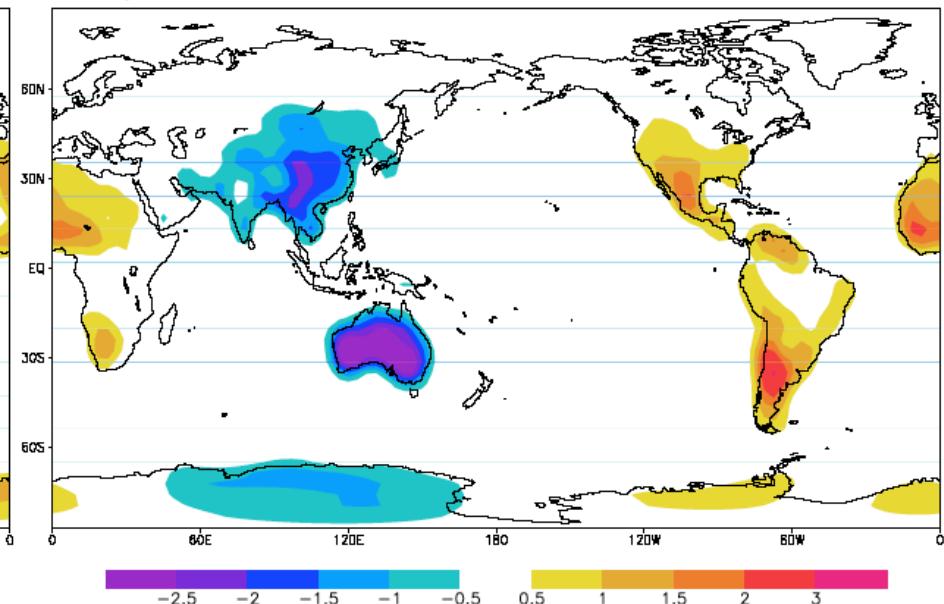
Estimated the average SPEEDY model error (bias) by averaging over several years the 6 hour forecast (started from reanalysis) minus the reanalysis.

Then they computed the EOFs of the anomaly in the model error, and found two dominant EOFs representing the model error in representing the diurnal cycle:

sig=0.95 debiased Temp Jan 1982–86 Increment EOF1



sig=0.95 debiased Temp Jan 1982–86 Increment EOF2

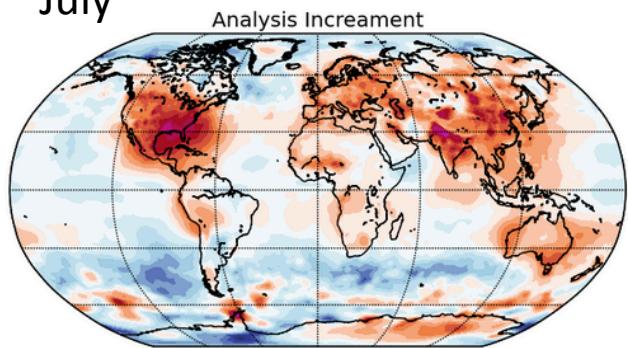
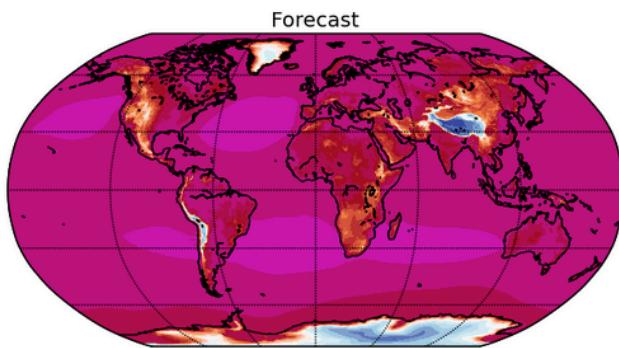
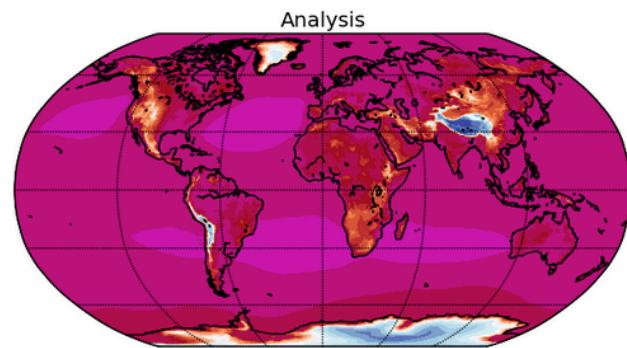
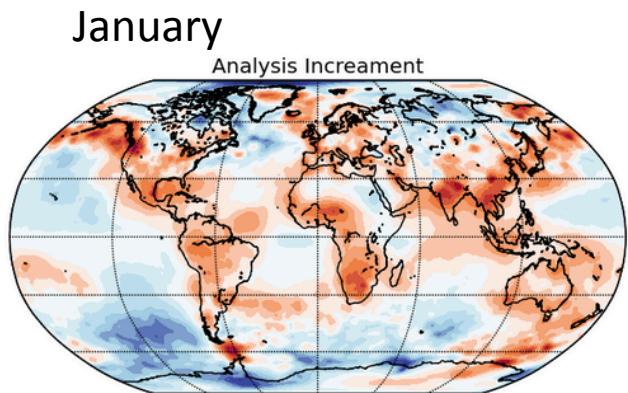
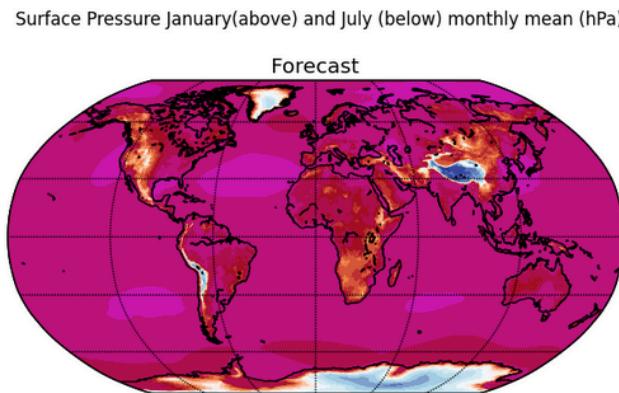
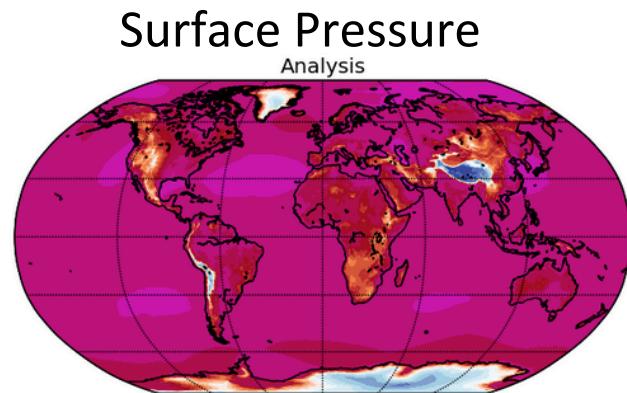


# Implications for improving the model bias

---

- The DKM2007 method gave very good results with the SPEEDY model, using R1 as an approximation of the true atmosphere.
- The -bias/6hr was added to the SPEEDY time derivatives ( $u, v, T, p_s$ ).
- This corrected the bias, **getting similar or better results than an *a posteriori* bias correction!** In addition, random forecast errors were also reduced.
- It was also used to improve the diurnal cycle and to find the state dependent systematic errors (e.g., during an El Niño).
- **It can be tried on the GFS (or the CFS!) taking advantage of the Analysis Increments, i.e., the difference between the Analysis and the Forecast.**
- Dr. Fanglin Yang (NCEP) very kindly provided us with 2012, 2013, and 2014 Analyses and Forecasts.

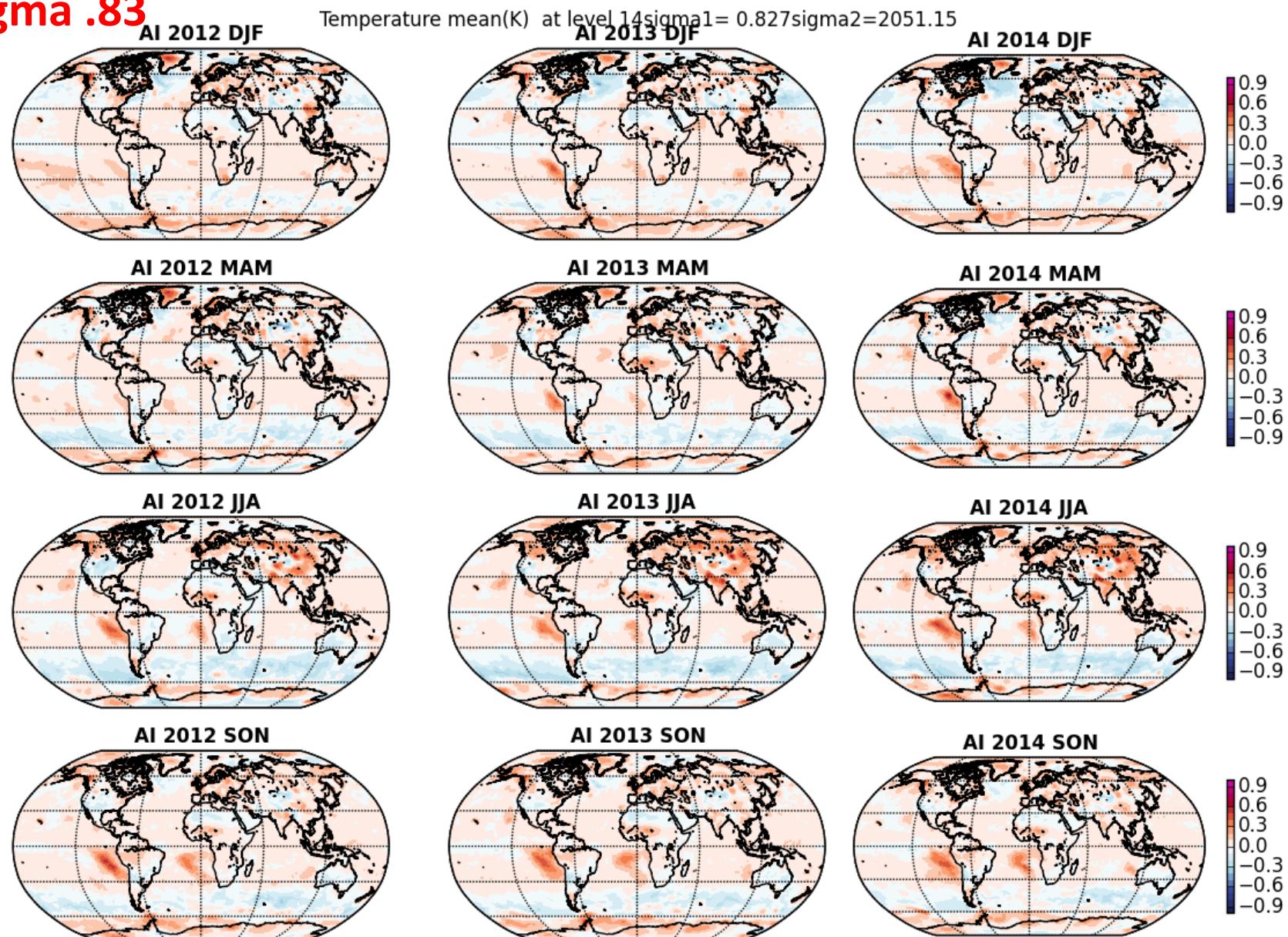
# First results: 2014 Analyses, Forecasts and AIs



$P_s$  is too low over continents, too high over oceans in both winter and summer.

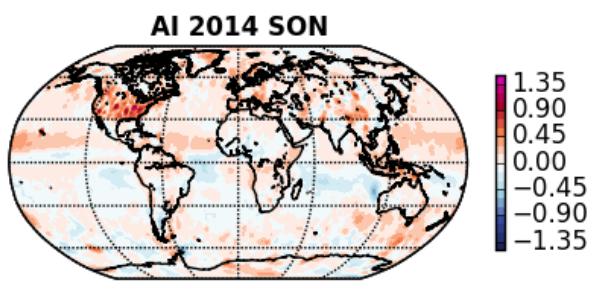
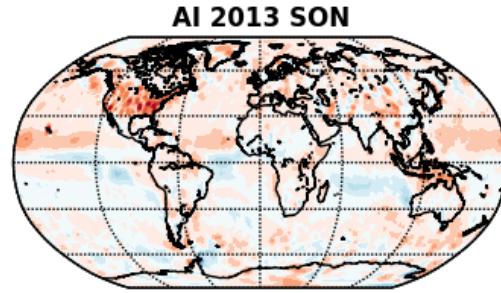
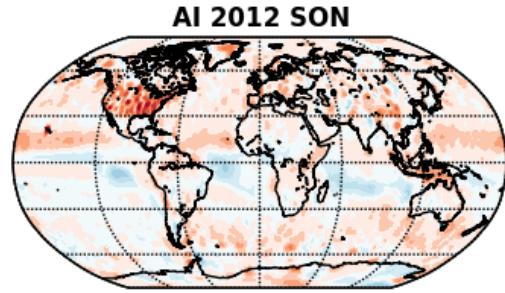
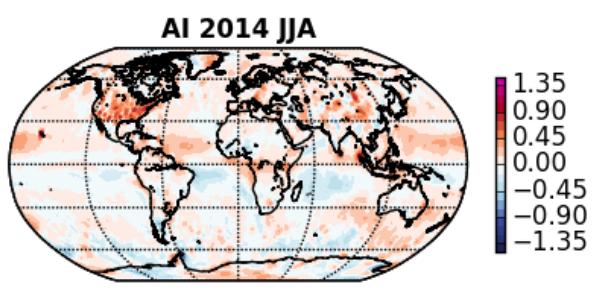
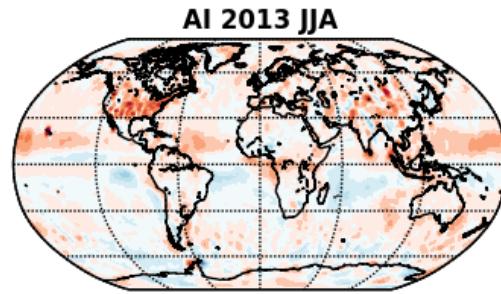
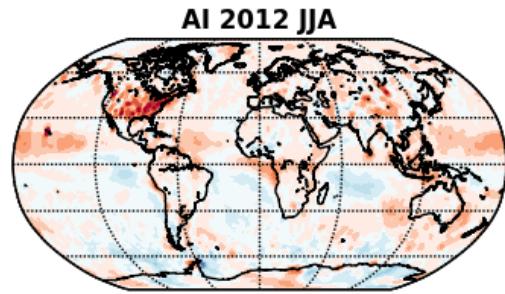
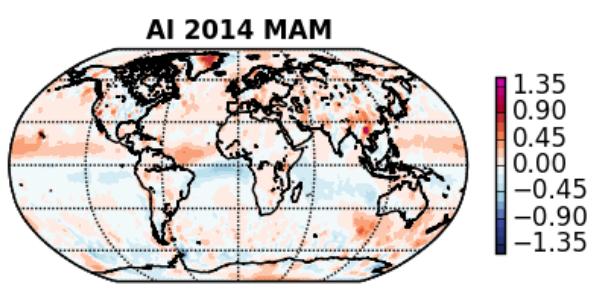
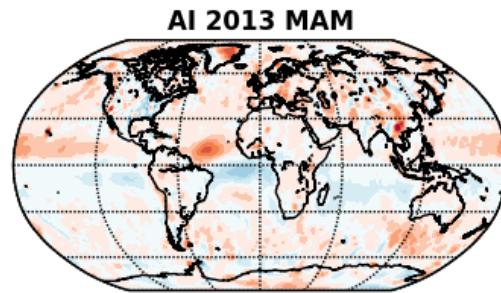
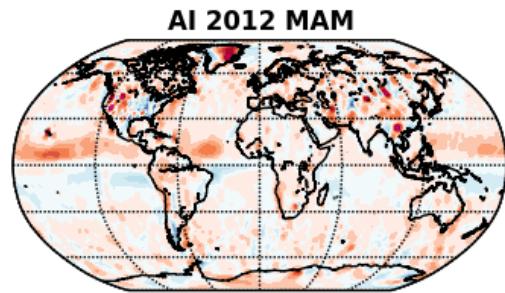
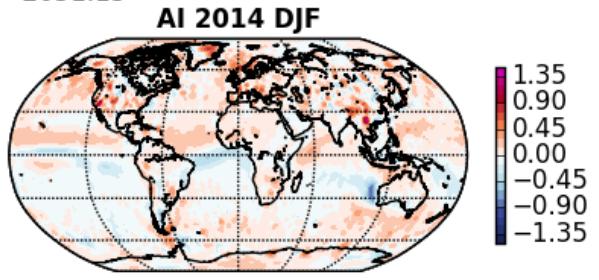
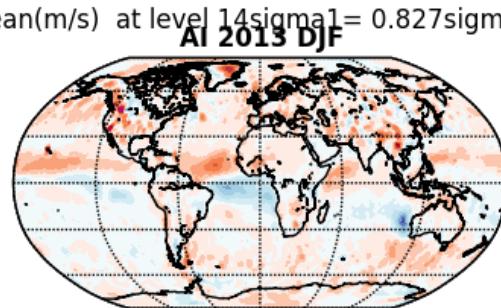
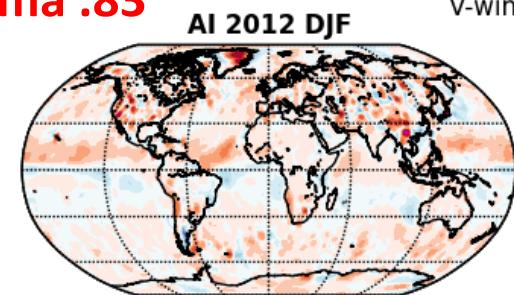
# Seasonal Analysis Increment 2012-2013-2014

T sigma .83



# Seasonal Analysis Increment 2012-2013-2014

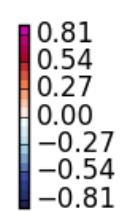
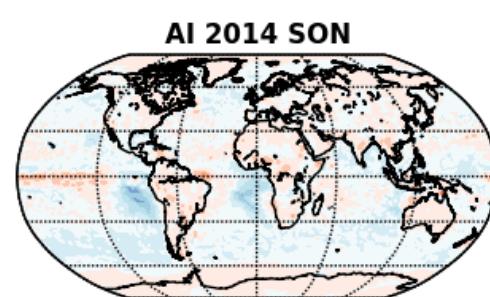
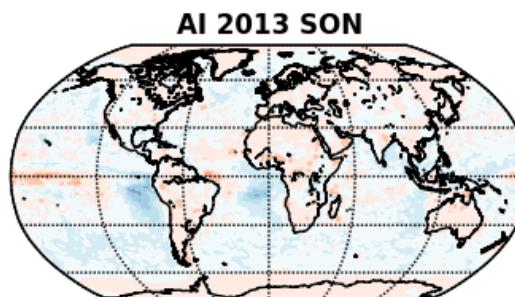
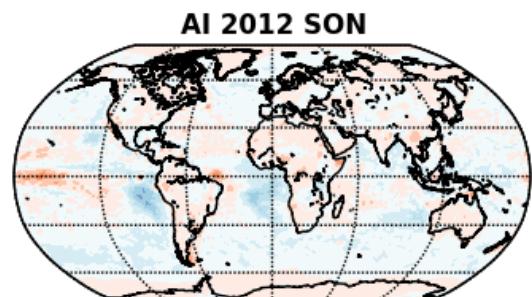
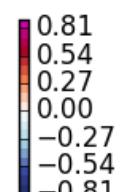
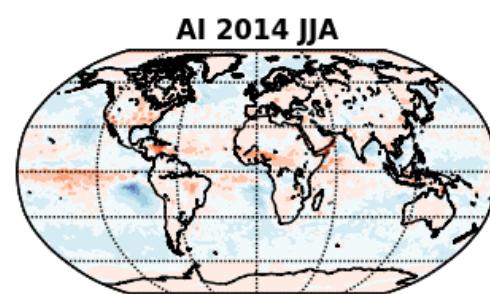
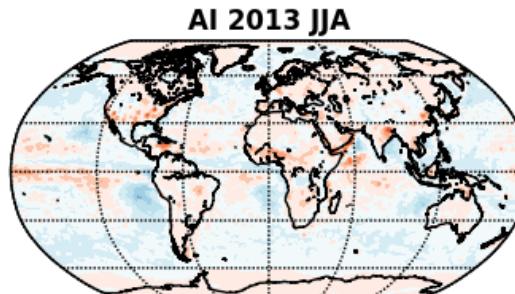
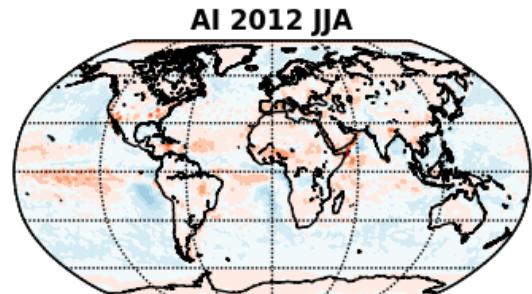
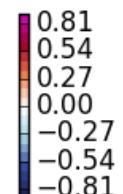
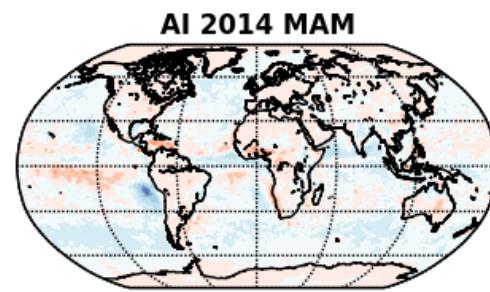
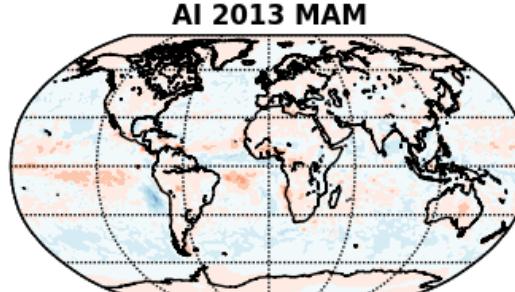
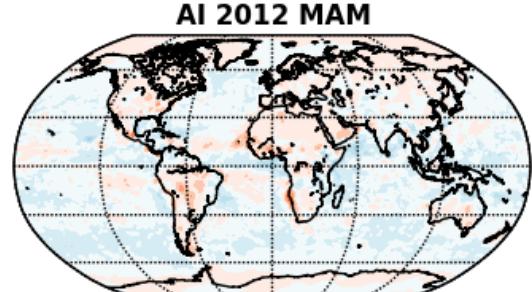
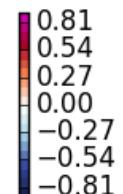
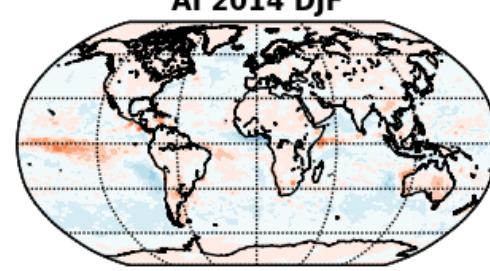
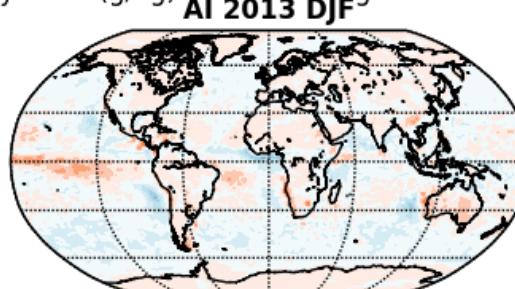
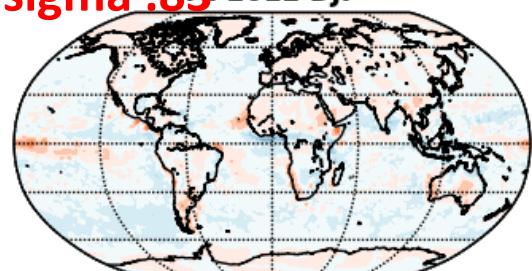
V sigma .83



# Seasonal Analysis Increment 2012-2013-2014

Q sigma .83

Specific Humidity mean(g/kg) at level 14sigma1= 0.827sigma2=2051.15



# How do we plan to reduce model bias?

---

- Check the robustness of the monthly or seasonal averaged AI (2014 vs. 2013 vs. 2012) ✓
- Perform exploratory low resolution (T254) experiments correcting the perceived model bias by adding AI/6hr to each variable time derivative.
- Test the impact on the forecast skill.
- Explore the diurnal cycle of the AI. Test if the diurnal cycle errors can be reduced.
- **If successful, the AI bias correction will also guide the development of the physical parameterizations.**

# SUMMARY

---

- Applications of EnKF-based data assimilation can go beyond providing the best initial conditions:
- They can improve both **observations** and **models**:
- Improve observations with EFSO and PQC.
- Improve the Obs Error Covariance R
- Improve the model parameters
- Improve the models by using the Analysis Increments to correct the models' bias