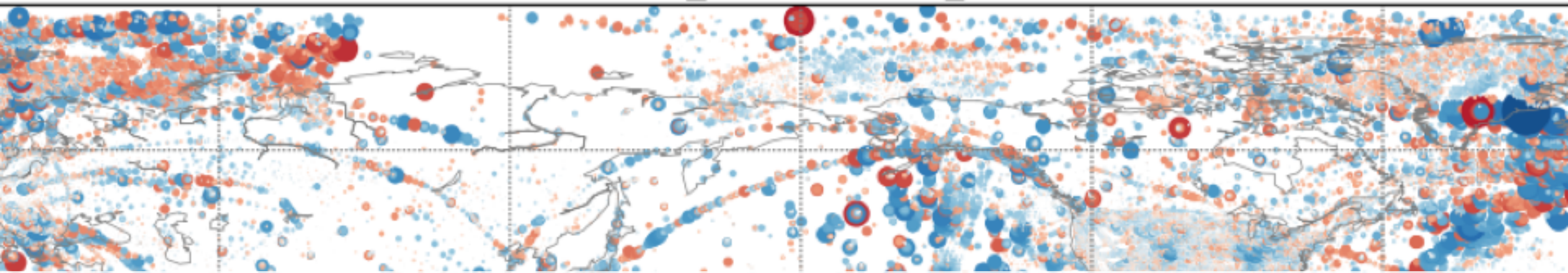
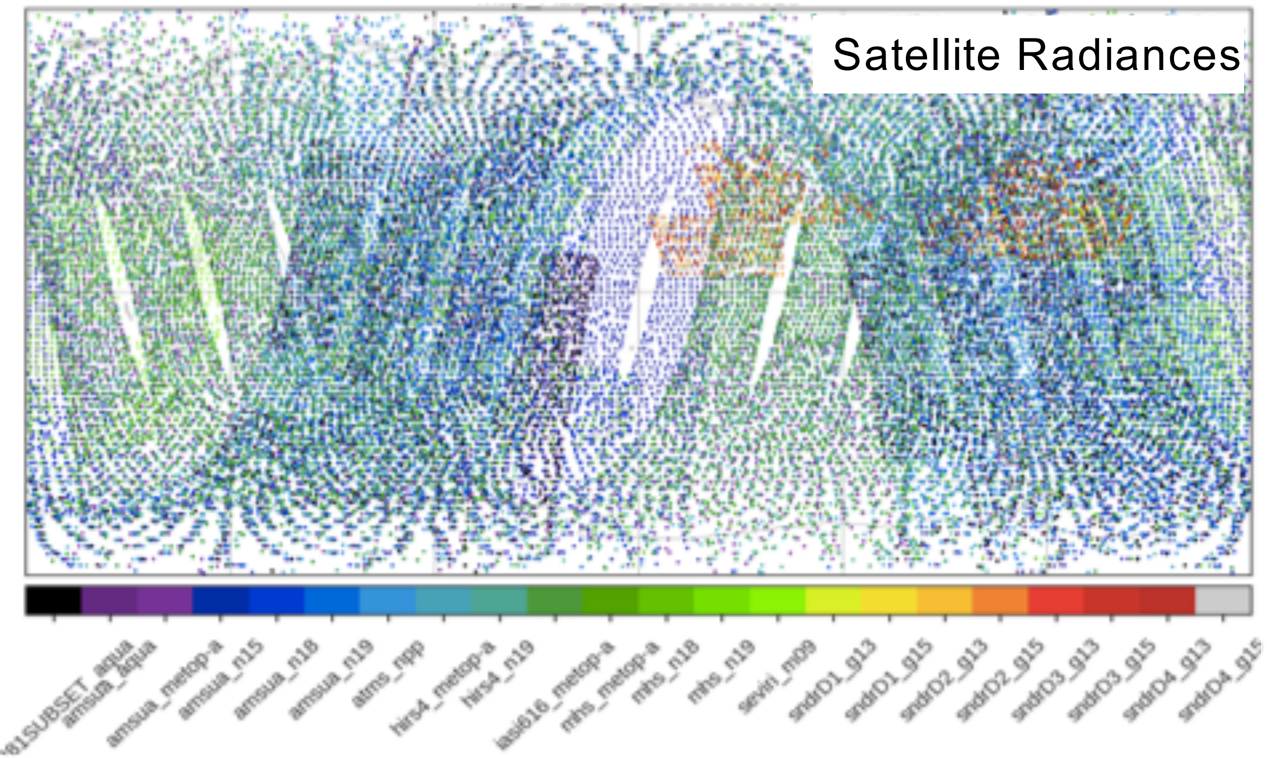


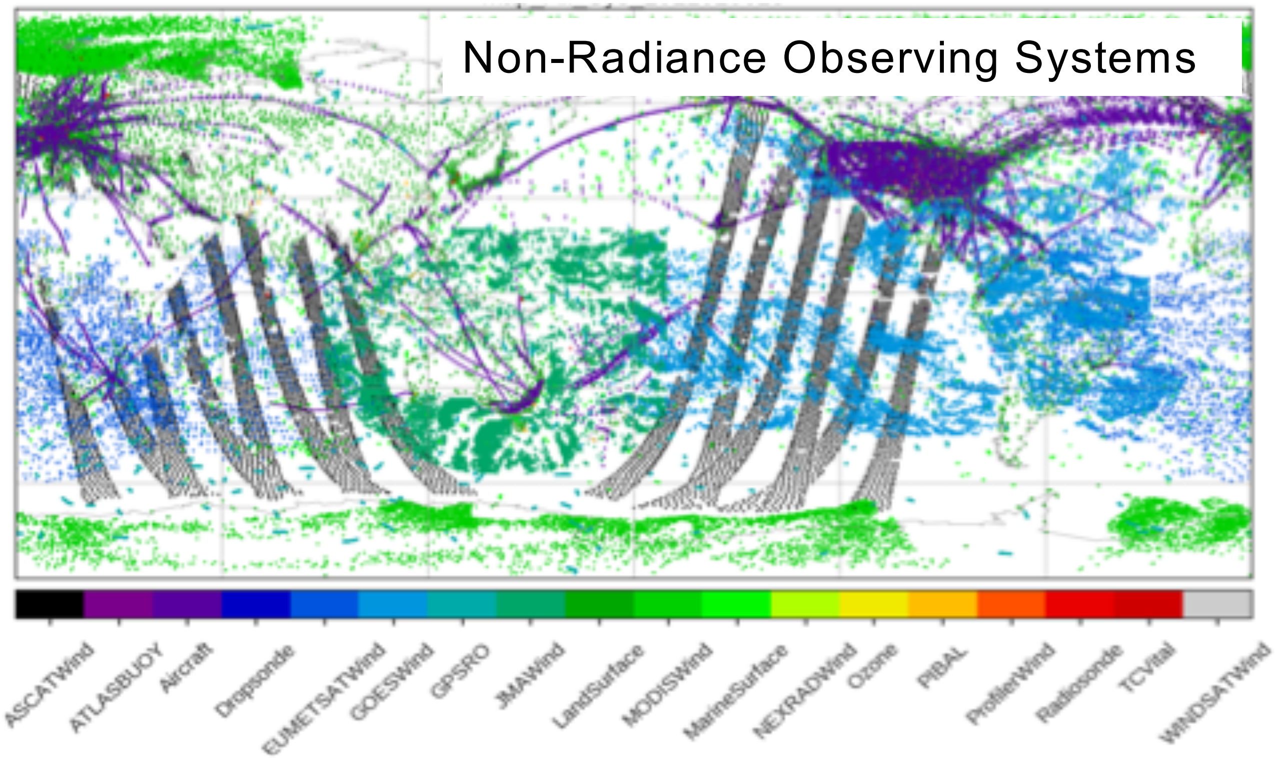
Efficient Data Selection Method for NWP Using EFSO

Tse-Chun Chen
Eugenia Kalnay
University of Maryland

Acknowledgements to:
Daisuke Hotta, Jim Jur
Guo-Yuan Lien, Yoichi
Ota, Krishna Kumar,
Jordan Alpert, and
Cheng Da

Massive Volume of Data





- Millions of observations are assimilated every 6 hours.
- New systems with higher spatial, temporal, and spectral resolution.
  e.g. Next generation GOES, Himawari and Phase Array Weather Radar

How to make use of these data effectively?

Effective use of observations (in NWP)

Data Assimilation

$$\mathbf{x}_0^a = \mathbf{x}_0^b + \boldsymbol{K}\,\delta\mathbf{y}_0^{ob}$$

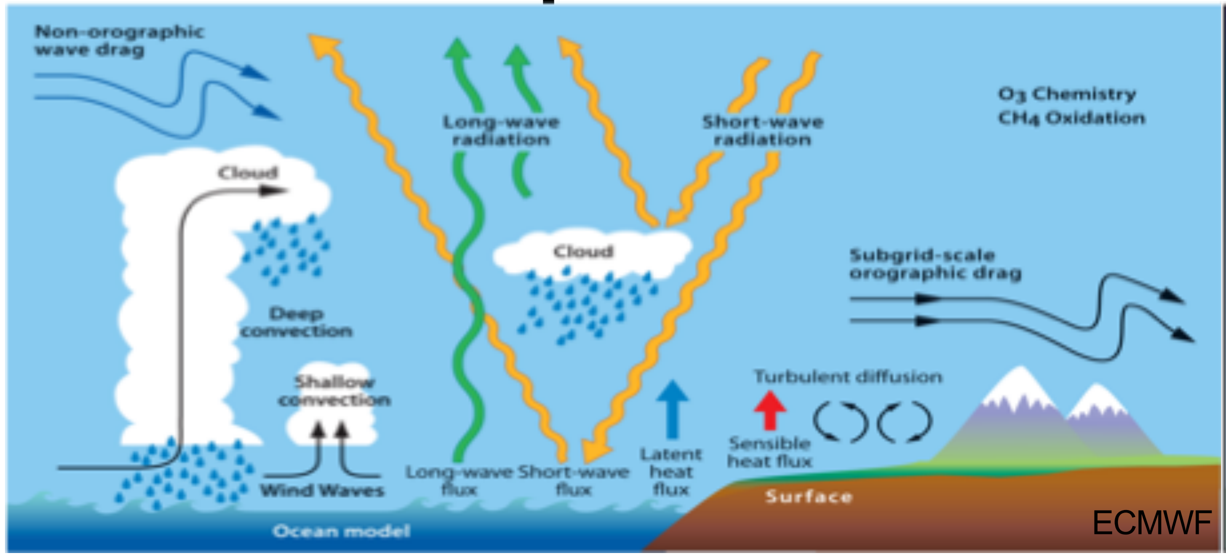
Analysis

Kalman gain

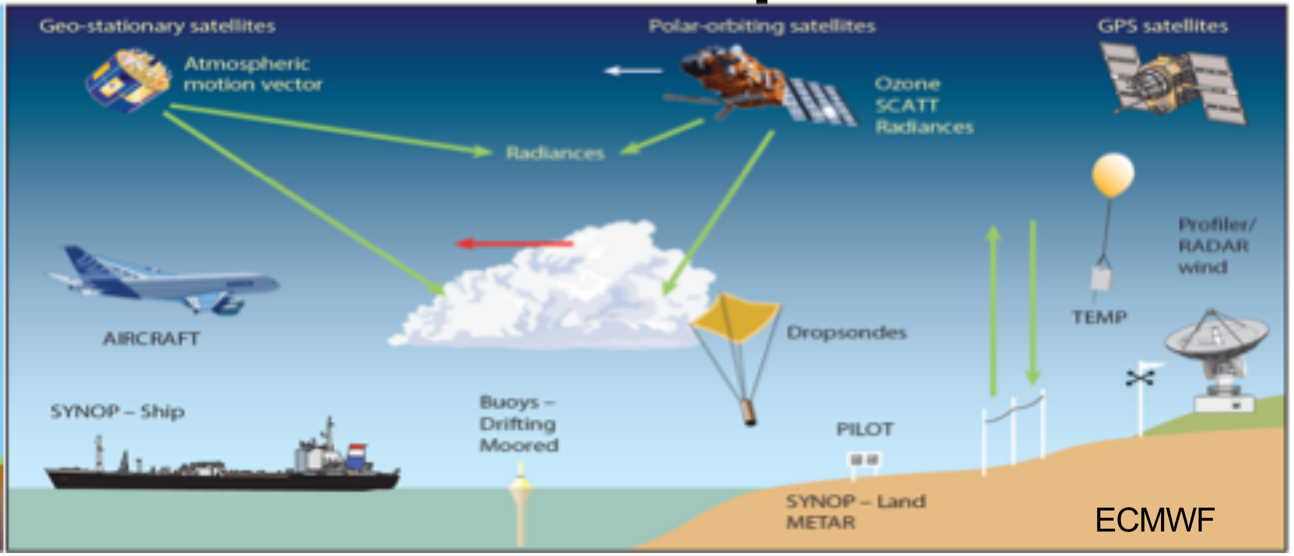
Background (forecast)

Observation-background

Quality Control & Data selection



Numerical Models



Observations

Ensemble Forecast Sensitivity to Observation (EFSO)

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

O-B of the ens. mean

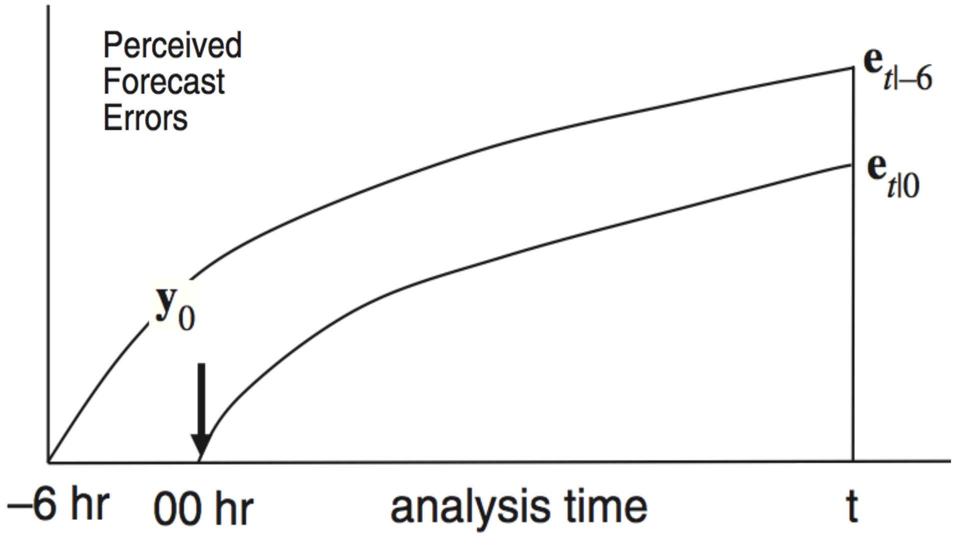
Analysis perturbation in obs. space

Forecast perturbation

Error norm

Forecast errors

Kalnay et al. 2012,
inspired by Langland and Baker (2004)



$$\mathbf{x}_0^a = \mathbf{x}_0^b + \boldsymbol{K} \delta \mathbf{y}_0^{ob}$$

- A **linear mapping** from error changes to each individual observation.
- Most of the components are readily produced by DA process.
- **Economical** and **efficient** for impact evaluation.
- **Negative means Beneficial** observation, Forecast error reduction
- **Positive means Detrimental** observation, Forecast error increase

Efficiently quantifies individual observational impact on forecasts.

EFSO Applications

1. Data Monitoring and Selection

- Efficient EFSO computation allows for **near real-time monitoring** of observational impact on the forecasts.
- Long record (1 month) of EFSO impact reveals detrimental subset of the observations and can **improve existing QC decisions and data selection**.
- **Accelerates** designing QC and selection for new instruments (Lien et al. 2017)

2. Proactive QC

- **Flow dependent QC** to avoid forecast skill dropouts (Hotta et al. 2017)
- Rejects observations at each DA cycle based on **immediate** EFSO impact.

Why perform data selection?

Naively assimilating all observations is expensive and very likely degrades analysis and forecast quality.

Data selection considers:

- Data Quality (bad observations)
- Model Representativeness (imperfect model)
- Redundant Information (overwhelmingly dense observations)
- Information content (insignificant observation)

Data selection ensure assimilating most useful observations.

Data Selection Methods

Physics-based:

- Requires **comprehensive knowledge** of the physical properties of the observation and the corresponding model representation.
(e.g. Gambacorta and Barnet 2013)

OSEs/OSSEs-based:

- **Straight-forward** approach by directly comparing twin experiments with and w/o the targeted observations. Computationally **very expensive**.

Statistics-based:

- **Degrees of Freedom of Signal** (**DFS or Information Content)** measures the expected influence of each observation in DA analysis
(e.g. Rabier et al. 2002, Rodgers 1996).

- These methods are all **complementing each other** and play their own roles in data selection process.

We propose adding (E)FSO-based selection.

EFSO and Degree of Freedom of Signal (DFS)

	Analysis (t=0)	Forecast (t=t')
DFS	$\boldsymbol{S}^a = \frac{\partial \mathbf{y}_0^a}{\partial \mathbf{y}_0^o} = [\boldsymbol{H}\boldsymbol{K}]^T \approx \frac{1}{K-1}\boldsymbol{R}^{-1}\boldsymbol{Y}_0^a\boldsymbol{Y}_0^{aT}$ (Liu et al., 2009)	$\boldsymbol{S}^f = \frac{\partial \mathbf{y}_t^f}{\partial \mathbf{y}_0^o} \approx \boldsymbol{K}^T\boldsymbol{M}^T\boldsymbol{H}^T$ (Chen et al. 2018)
EFSO	$\Delta e^2 \approx \frac{1}{K-1}\delta \mathbf{y}_0^{oT}\boldsymbol{R}^{-1}\boldsymbol{Y}_0^a\boldsymbol{X}_0^{aT}\boldsymbol{M}^T\boldsymbol{C}(\mathbf{e}_{t\|0} + \mathbf{e}_{t\|-6})$	
	$\Delta e^2 \approx -\delta \mathbf{y}_0^{oT}\boldsymbol{S}^a\boldsymbol{S}^{aT}\delta \mathbf{y}_0^o \approx -\delta \mathbf{y}_0^{aT}\delta \mathbf{y}_0^a$ $\delta \mathbf{y}_0^a = \boldsymbol{H}\delta \mathbf{x}_0^a = \boldsymbol{H}\boldsymbol{K}\mathbf{y}_0^o = \boldsymbol{S}^{aT}\delta \mathbf{y}_0^o$	$\Delta e^2 \approx -\delta \mathbf{y}_0^{oT}\boldsymbol{S}^f\boldsymbol{S}^{fT}\delta \mathbf{y}_0^o + \delta \mathbf{y}_0^{oT}\boldsymbol{S}^f\boldsymbol{H}(2\mathbf{e}_{t\|0})$ $\delta \mathbf{y}_t^f \approx \boldsymbol{H}\boldsymbol{M}\delta \mathbf{x}_0^a \approx \boldsymbol{H}\boldsymbol{M}\boldsymbol{K}\mathbf{y}_0^o$

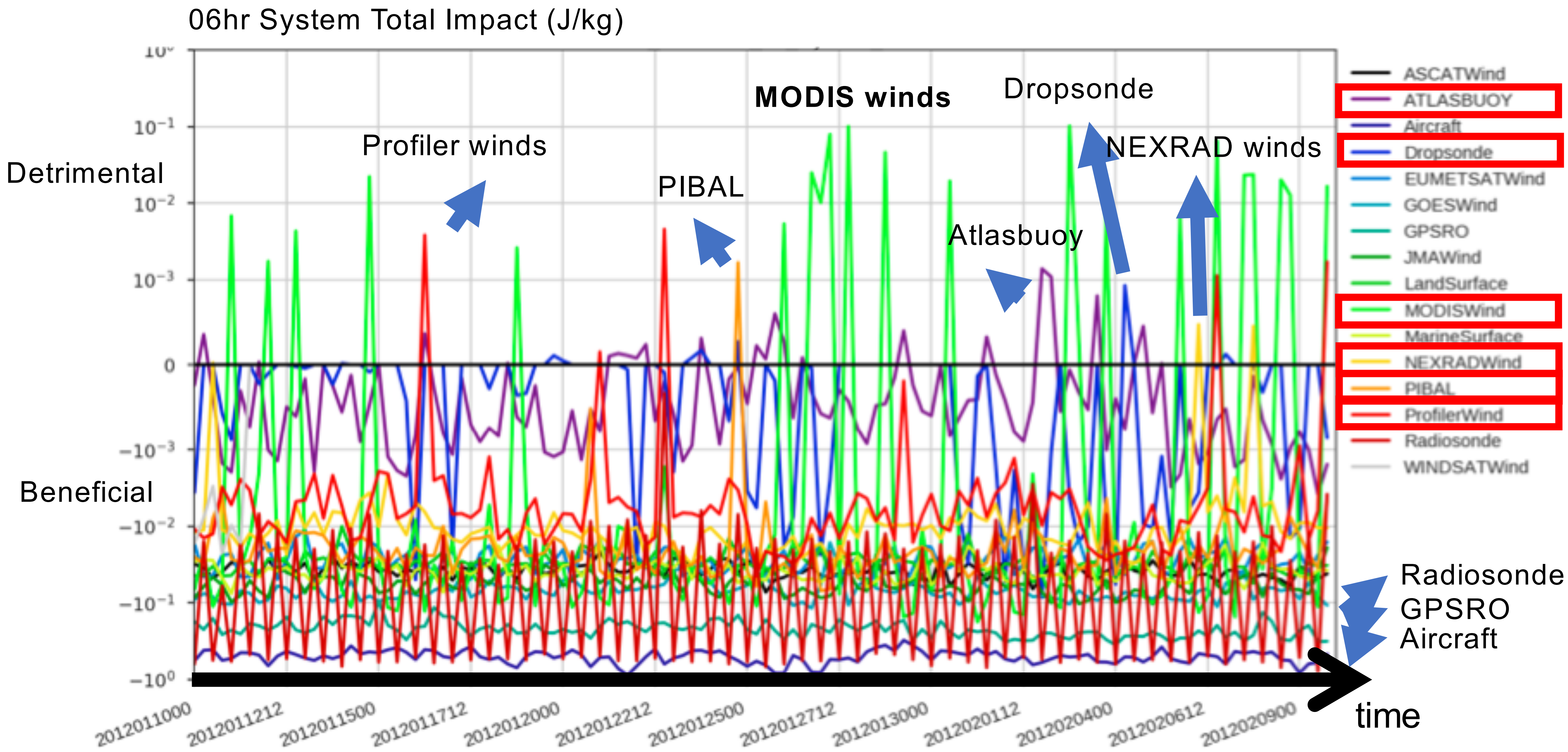
EFSO impact:

- includes **DFS** (gradient / sensitivity; another view of gain matrix)
- computes the **actual impact** on analysis and on forecast
- identify the **sign of impact** (detrimental or beneficial)

Experimental setup for data monitoring & selection

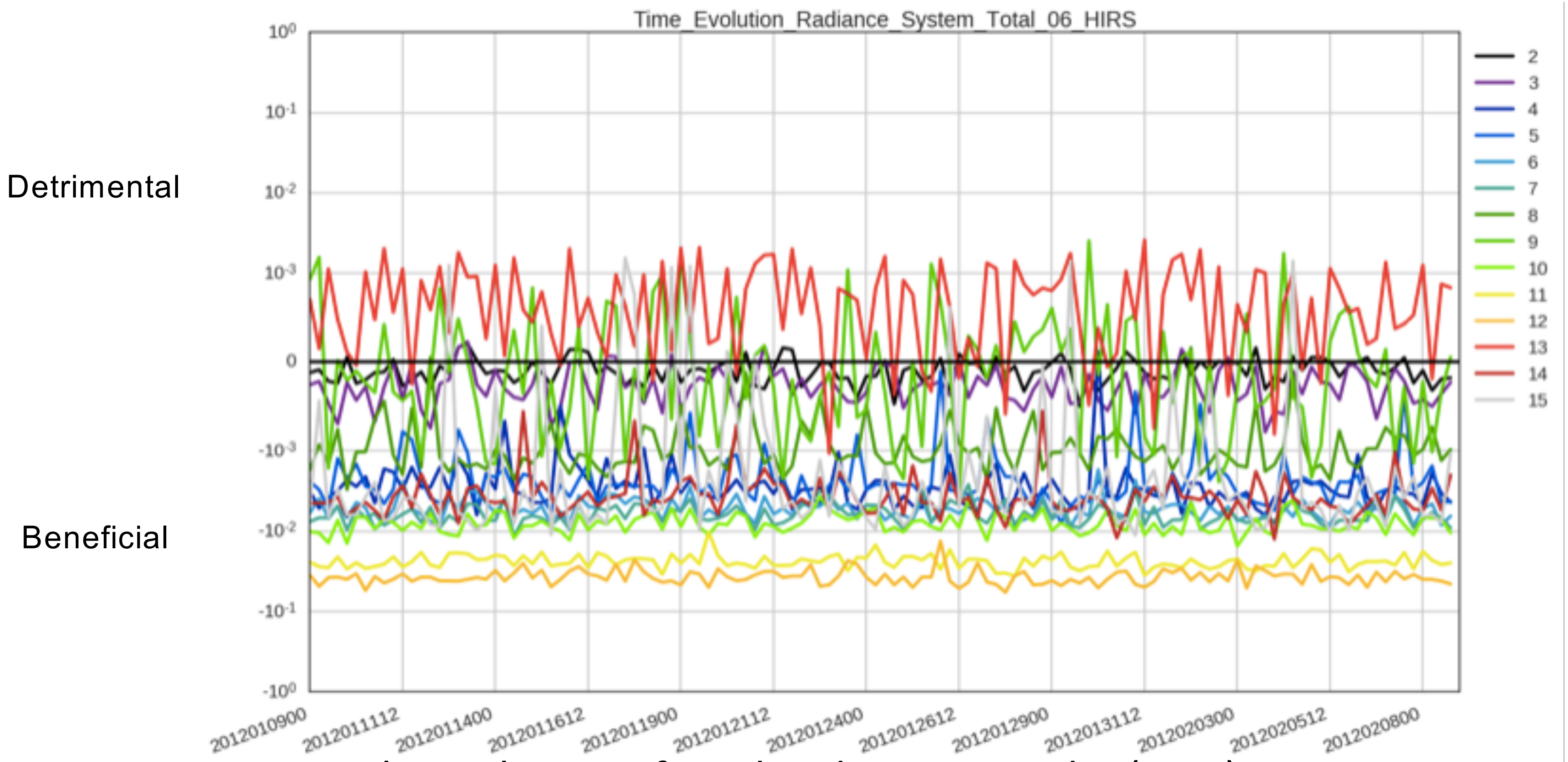
Period (1 month)	Jan/10/2012 00Z – Feb/09/2012 18Z
Model	GFS T254/T126 L64
DA	LETKF/3D-Var Hybrid GSI v2012
Localization cut-off length	2000 km/ 2 scale heights
Error norm	Moist total energy (MTE) $MTE = \frac{1}{2}\frac{1}{\vert S\vert}\int_S\int_0^1\{(u'^2+v'^2)+\frac{C_p}{T_r}T'^2+\frac{R_dT_r}{P_r^2}p_s'^2+w_q\frac{L^2}{C_pT_r}q'^2\}d\sigma dS$

Powerful QC monitoring for every system!



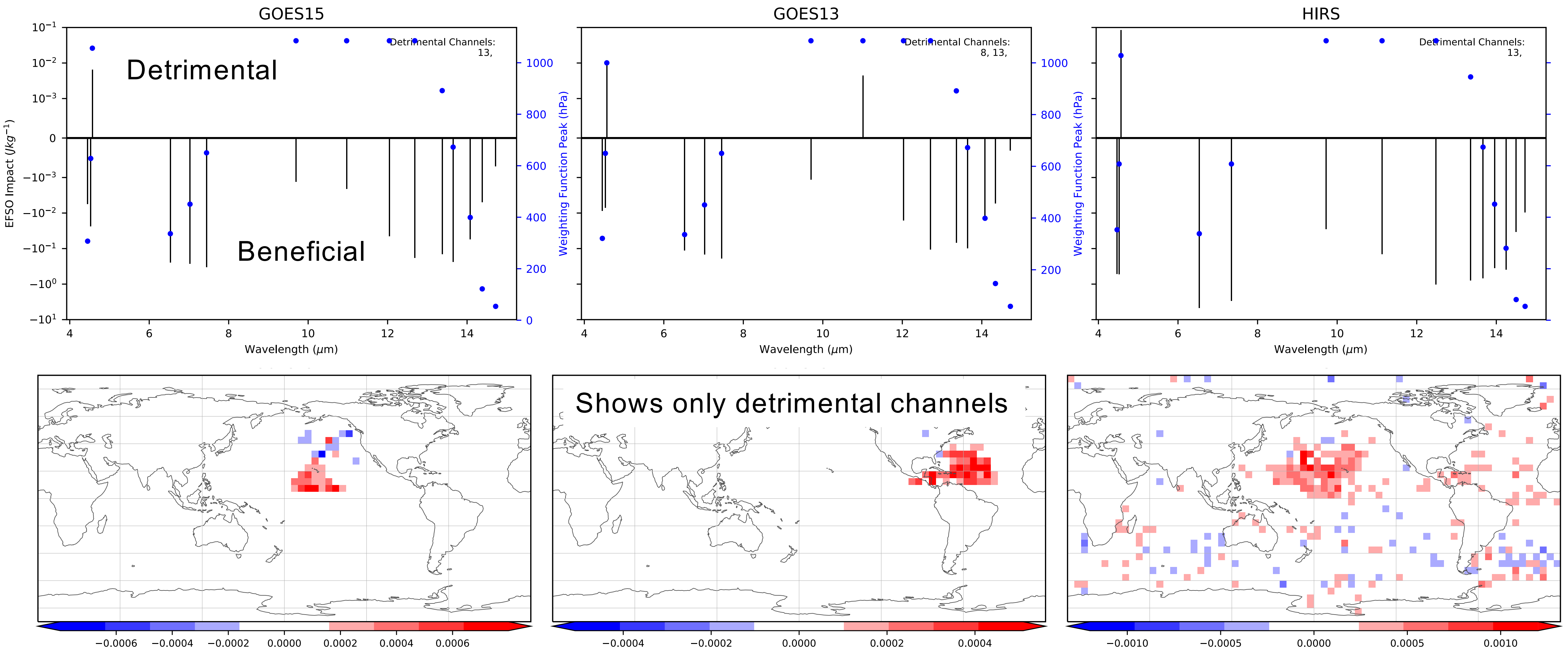
Forecasters can also see which instruments are detrimental

Radiance Channel Selection: **HIRS**



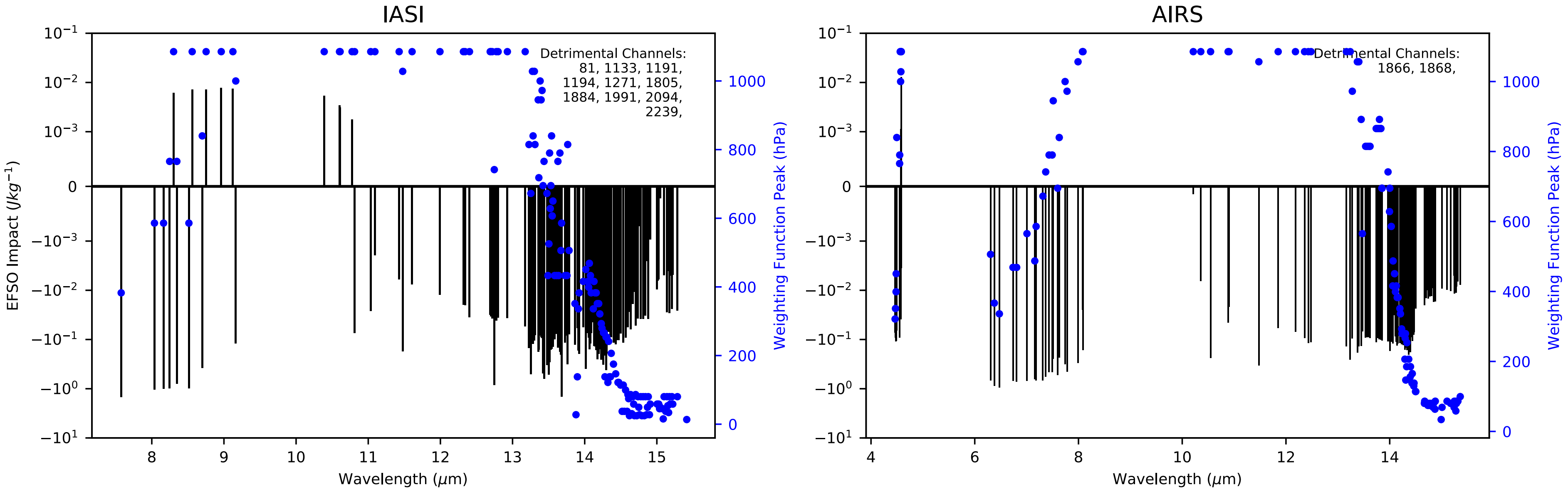
- High-resolution Infrared Radiation Sounder (HIRS)
- Channel **13** of HIRS has always provided detrimental impacts.
- This is easily identified using EFSO (**no OSE required**).

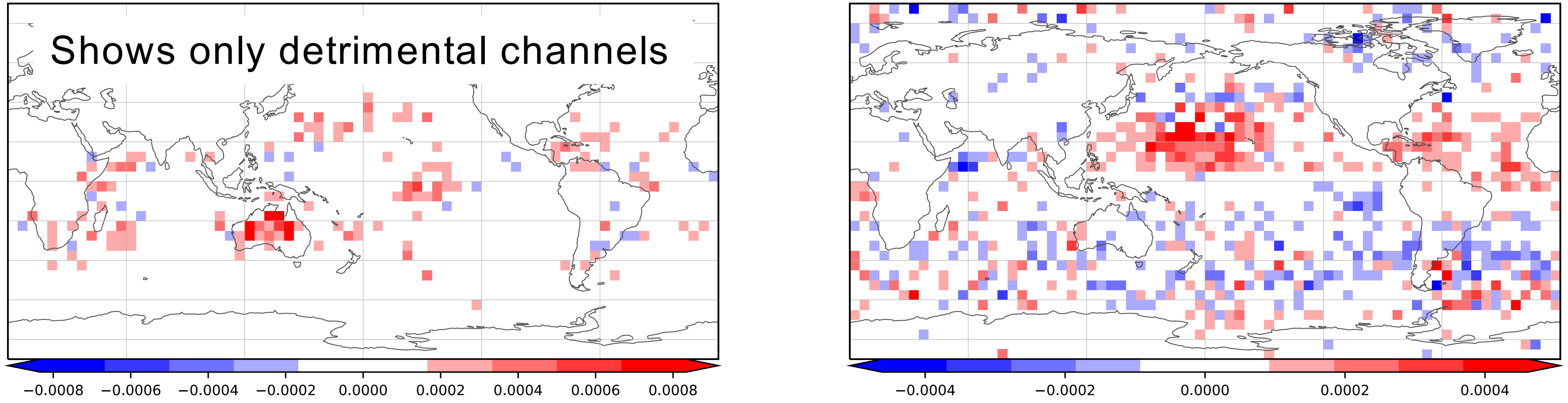
Multi-channel instruments: **GOES sounder, HIRS**



- Channel 8 (11.03 um), 13 (4.57 um): sensitive to surface and low-level temperature.
- Map shows the 2 channels are detrimental in tropical Pacific and Atlantic.

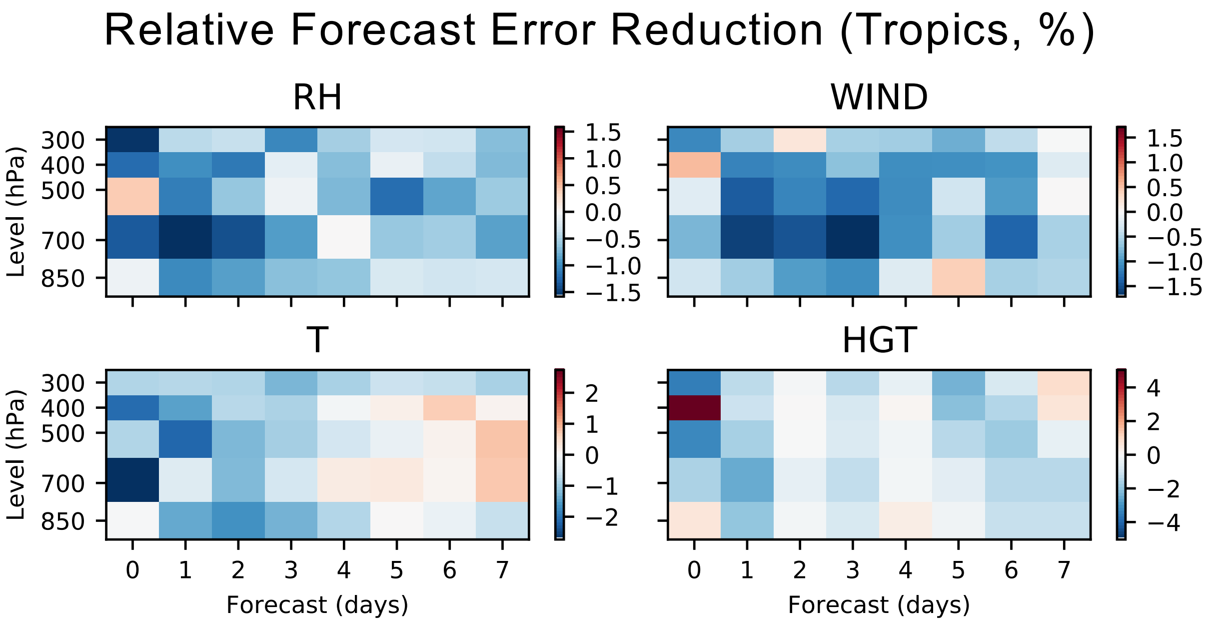
Hyperspectral Instruments: **IASI**, **AIRS**





- Efficient channel-wise impact evaluation even for hyperspectral instruments.
- Detrimental impact from Australia and tropical oceans.

Forecast performance of EFSO-based selection



Instruments:	Rejected channels:
IASI	81, 1133, 1191, 1194, 1271, 1805, 1884, 1991, 2094, 2239
AIRS	1866, 1868
GOES15 sounder	13
GOES13 sounder	8, 13
HIRS	13

- The detrimental impact is mainly from the tropical regions.
- Simply rejecting 16 channels out of hundreds improves the monthly mean tropical forecast by 1%

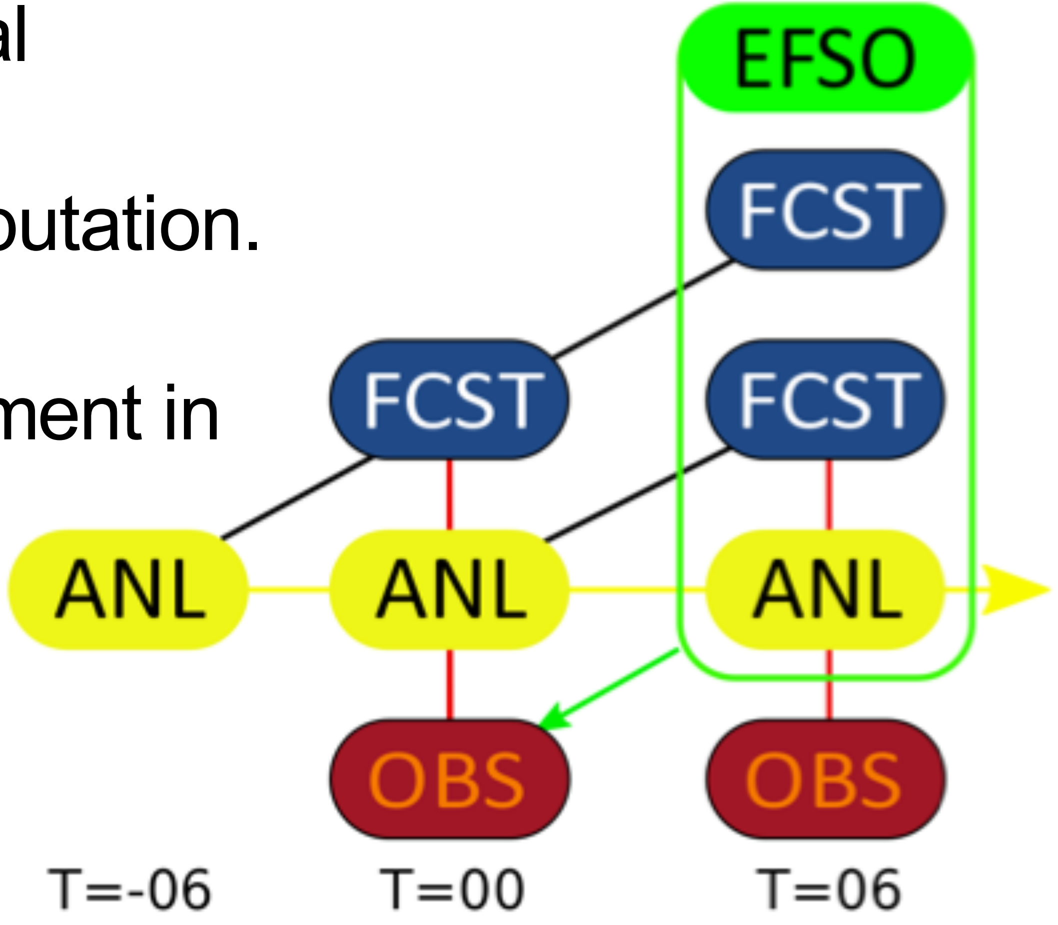
Rejecting the detrimental channels improves tropical forecasts

Proactive Quality Control (PQC)

- Fully **flow-dependent QC** scheme pioneered by Ota et al. (2013) and Hotta et al. (2017) to alleviate forecast skill dropout issue.
- PQC rejects EFSO identified detrimental observations in **each DA cycle**.
- Requires **next analysis** for EFSO computation.
- Long forecast benefits only from the **accumulation** of cycling PQC improvement in operation.
- We test PQC in **GFS**.

PQC rejects observations based on the immediate EFSO impact.

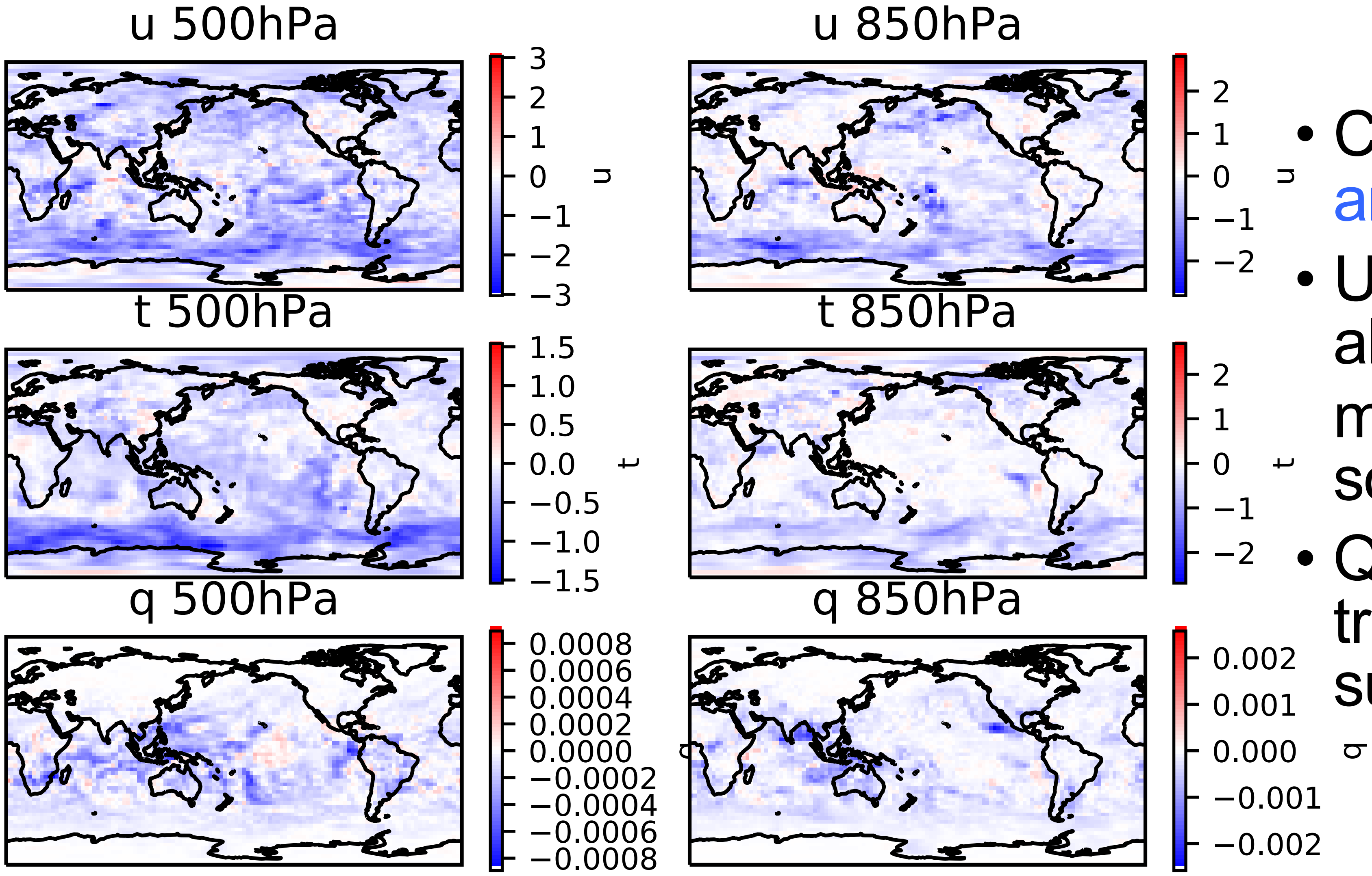
Adapted from Hotta 2017



Experimental setup for GFS-LETKF (Lien, 2015)

Period (~1 month)	Jan/01/2008 00Z – Feb/06/2008 06Z (5 days for DA spinup)
Model	GFS T62 L64
DA	LETKF with 32 ensemble size
Observation	prepBUFR data from NCEP
Localization	Horizontal: 500 km Vertical: 0.4 scale height
Inflation	RTPP (Zhang 2004) + adaptive inflation (Miyoshi 2011)
Verifying truth	NCEP Climate Forecast System Reanalysis (CFSR)

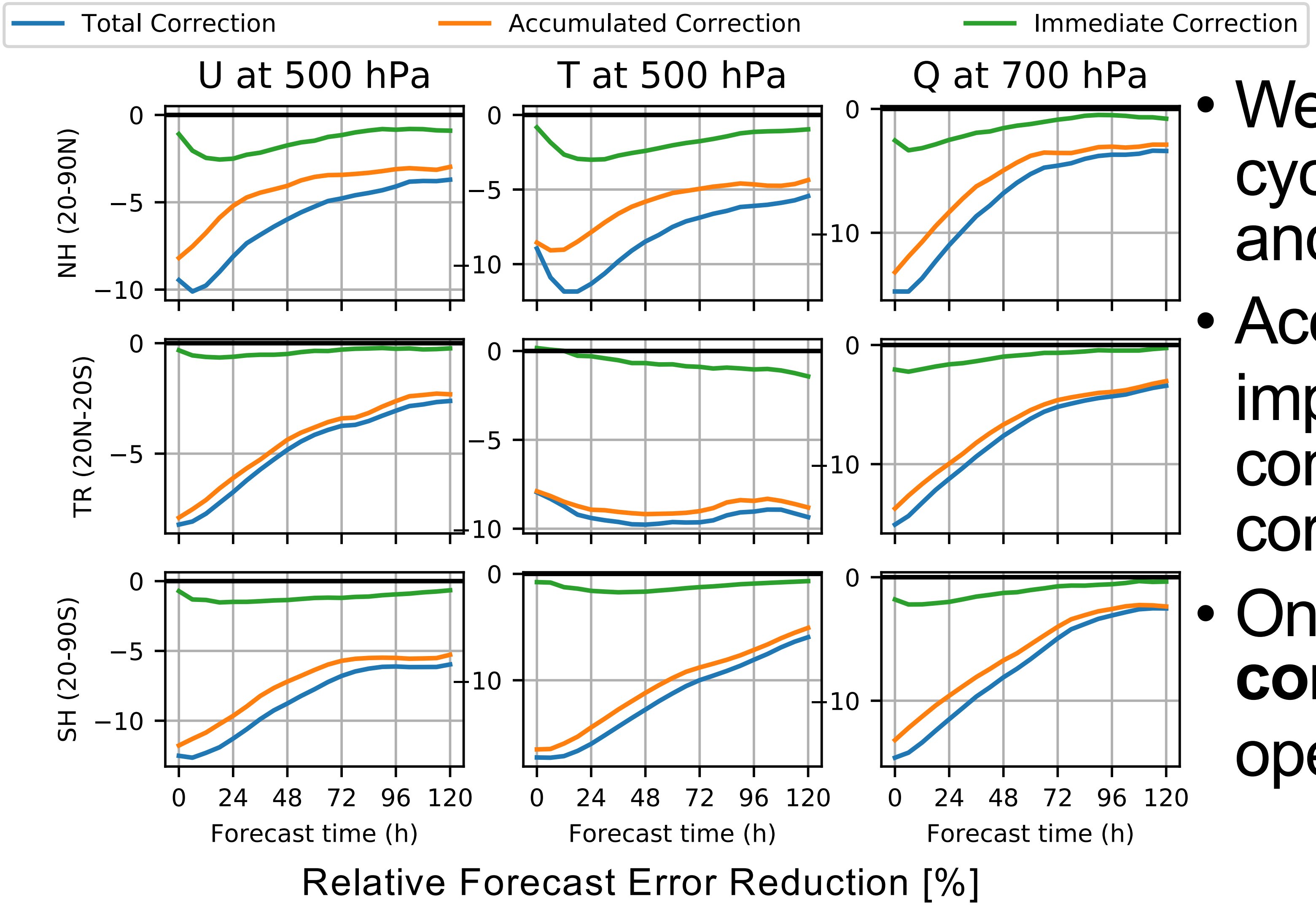
Cycling PQC reduces the analysis error



- Cycling PQC reduces analysis RMSE (blue)
- U and T improves almost globally, but more heavily over the southern ocean
- Q improves over the tropics and the subtropical region.

Analysis is improved globally across every variable!

Immediate and Accumulated impact of cycling PQC



- We separate total correction of cycling PQC into **accumulated** and immediate correction.
- Accumulated correction improves the background and contributes **most** of the total correction.
- Only the **accumulated correction** is **feasible** for GFS operational forecasts.

Most benefit comes from the accumulated correction!

Concluding Remarks

EFSO-based Data Selection

- Agile data monitoring and selection for every observing systems
- Can deal with massive (and increasing) amount of data
- Dropping just few channels improves the forecast by 1%

Proactive QC

- PQC improves the analysis and the forecast across variables over the globe.
- Accumulated corrections by cycling PQC dominate the total correction, indicating that the latest forecast will be improved by cycling PQC.

Future Works

EFSO-based Data Selection

- Perform iterative EFSO-based radiance channel selection.
- Bring back the innocent (beneficial yet discarded) channels to better utilize the data.
- Integrate EFSO into OSSE/OSE in collaboration with Drs. Robert Atlas, Lidia Cucurull, and Sean Casey (QOSAP team)

Proactive QC

- Test PQC with more close-to-operation environment: e.g. 4DEnVar vs. pure LETKF, more realistic model resolution, include radiance data, etc.
- Alleviate GFS forecast skill dropout problem with Jordan Alpert and Krishna Kumar (GFDPT team)

Thank you