5th WMO workshop on the impact of various observing systems on NWP

Observation impact estimates using the NCEP GFS/EnKF

Yoichiro Ota (NCEP/EMC visiting from JMA)

Acknowledgement to the collaboration with
Eugenia Kalnay, Takemasa Miyoshi and Weather/Chaos group (University of Maryland)
Daryl Kleist, Russ Treadon, John Derber (NCEP/EMC), Ricardo Todling (NASA/GMAO), Jeff Whitaker (ESRL), and Advanced Data Assimilation team

Outline

- Estimating the impact of observations in the forecast
  - Nonlinear approach: data denial/addition experiment like OSE, OSSE
    - Impact with and without certain sets of observations
  - Linear approach: estimating the sensitivity to the observations within the assimilated datasets
    - Relative impact within the assimilated observations
    - Adjoint-based method (Langland and Baker 2004)
    - EnKF-based method (Liu and Kalnay 2008)

Adjoint-based and EnKF-based

Adjoint-based method: Langland and Baker (2004)
EnKF-based method: Liu and Kalnay (2008), Kalnay et al. (2012, Tellus, submitted)

	Adjoint	EnKF
Adjoint model	Yes	No
Applicable assimilation scheme	All type	Ensemble based
Deterministic forecast from first guess and analysis	Yes	Yes
Analysis value at evaluation time	Yes	Yes
Ensemble forecast from analysis	No	Yes
Analysis ensemble perturbation projected on observation space	No	Yes
Observation departure	Yes	Yes
Observation error	Yes	Yes
Back ground error covariance	Yes	No

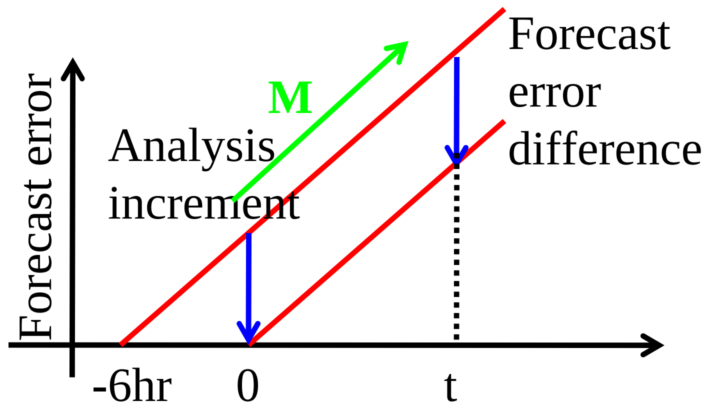
EnKF-based method uses ensemble perturbations instead of the adjoint model.

Adjoint-based formulation

Forecast error reduction is defined as

$$\delta J = \frac{1}{2}\left(\mathbf{e}_{t|0}^{\mathbf{T}}\mathbf{C}\mathbf{e}_{t|0} - \mathbf{e}_{t|-6}^{\mathbf{T}}\mathbf{C}\mathbf{e}_{t|-6}\right) = \left(\mathbf{e}_{t|0} - \mathbf{e}_{t|-6}\right)^{\mathbf{T}}\mathbf{C}\left(\mathbf{e}_{t|0} + \mathbf{e}_{t|-6}\right) = \left(\overline{\mathbf{x}}_{t|0}^{f} - \overline{\mathbf{x}}_{t|-6}^{f}\right)^{\mathbf{T}}\mathbf{C}\left(\mathbf{e}_{t|0} + \mathbf{e}_{t|-6}\right)$$

where **C**: norm operator, $\mathbf{e}_{t|-6}$,$\mathbf{e}_{t|0}$: error of forecast from the first guess and analysis at FT=*t*, $\mathbf{x}^{f}_{t|0}$, $\mathbf{x}^{f}_{t|-6}$: forecast from the first guess and analysis at FT=*t*.



Forecast error difference can be approximated with the tangent linear evolution of the analysis increment.

$$\overline{\mathbf{x}}_{t|0}^{f} - \overline{\mathbf{x}}_{t|-6}^{f} \approx \mathbf{M}\left(\overline{\mathbf{x}}_{0}^{a} - \overline{\mathbf{x}}_{0|-6}^{f}\right)$$

Also, analysis increment can be described as

$$\overline{\mathbf{x}}_{0}^{a} - \overline{\mathbf{x}}_{0|-6}^{f} = \mathbf{K}\mathbf{d}$$

where **K**: Kalman gain, **d**: observational increment from the first guess.

Then,

$$\delta J \approx \frac{1}{2}\mathbf{d}^{\mathbf{T}}\mathbf{K}^{\mathbf{T}}\mathbf{M}^{\mathbf{T}}\mathbf{C}\left(\mathbf{e}_{t|0} + \mathbf{e}_{t|-6}\right)$$

EnKF-based formulation

Kalman gain in the EnKF can be expressed as

$$\mathbf{K} = \frac{1}{K-1}\mathbf{X}_0^a \mathbf{X}_0^{a\mathbf{T}} \mathbf{H}^{\mathbf{T}} \mathbf{R}^{-1}$$

where K: number of member, $\mathbf{R}$: observation error covariance, $\mathbf{H}$: tangent linear observation operator, $\mathbf{X}^a$: analysis ensemble perturbations.

And, using tangent linear approximation, $\mathbf{M}\mathbf{X}_0^a \approx \mathbf{X}_{t|0}^f$

where $\mathbf{X}^f$: perturbations of forecast ensemble.

So, Kalnay et al. (2012) proposed a following formulation.

$$\delta J = \frac{1}{2}\left(\mathbf{e}_{t|0} - \mathbf{e}_{t|-6}\right)^{\mathbf{T}} \mathbf{C}\left(\mathbf{e}_{t|0} + \mathbf{e}_{t|-6}\right) \approx \frac{1}{2}\mathbf{d}^{\mathbf{T}} \frac{1}{K-1}\mathbf{R}^{-1}\left(\mathbf{H}\mathbf{X}_0^a\right)\mathbf{X}_{t|0}^{f\mathbf{T}} \mathbf{C}\left(\mathbf{e}_{t|0} + \mathbf{e}_{t|-6}\right)$$

This formulation uses analysis and forecast ensembles instead of the adjoint model.

Experiments with GFS/EnKF

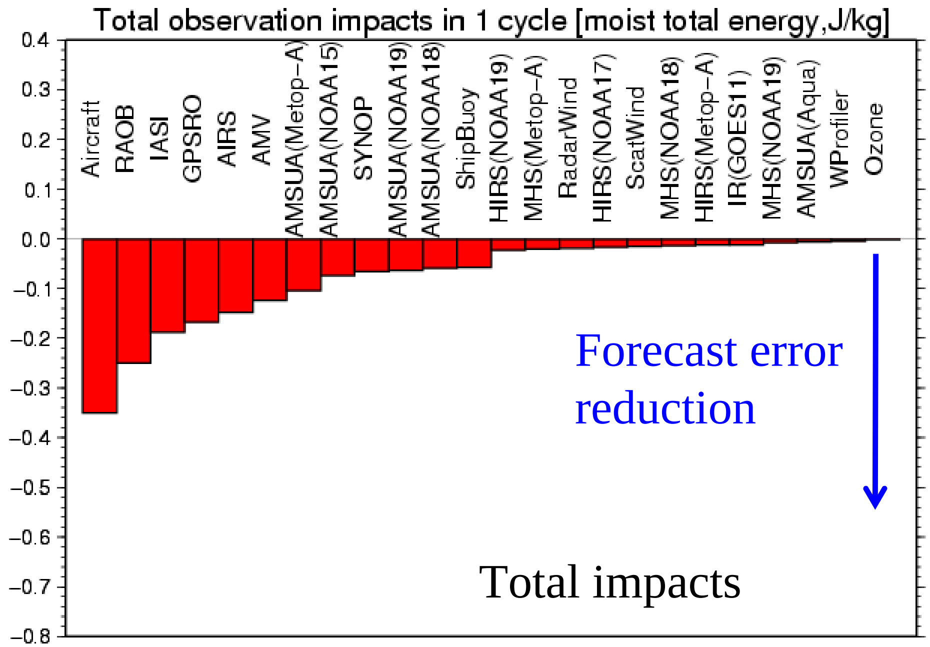
- LETKF and serial EnSRF experiments with 80 member GFS at T62 and 64 vertical levels
- Almost all observation types assimilated on the operational GDAS are used (N~1.7 million).
- Localization scale (cut-off length): 1500km in horizontal, 3.3, 2.2 and 1.5 scale heights in vertical for satellite radiance, Ps and others
- Period: from 12Z Oct. 21 to 06Z Oct. 28 2010
- Impacts are measured with the moist total energy (and dry total energy).
- The evaluation is on FT=24 with global domain.

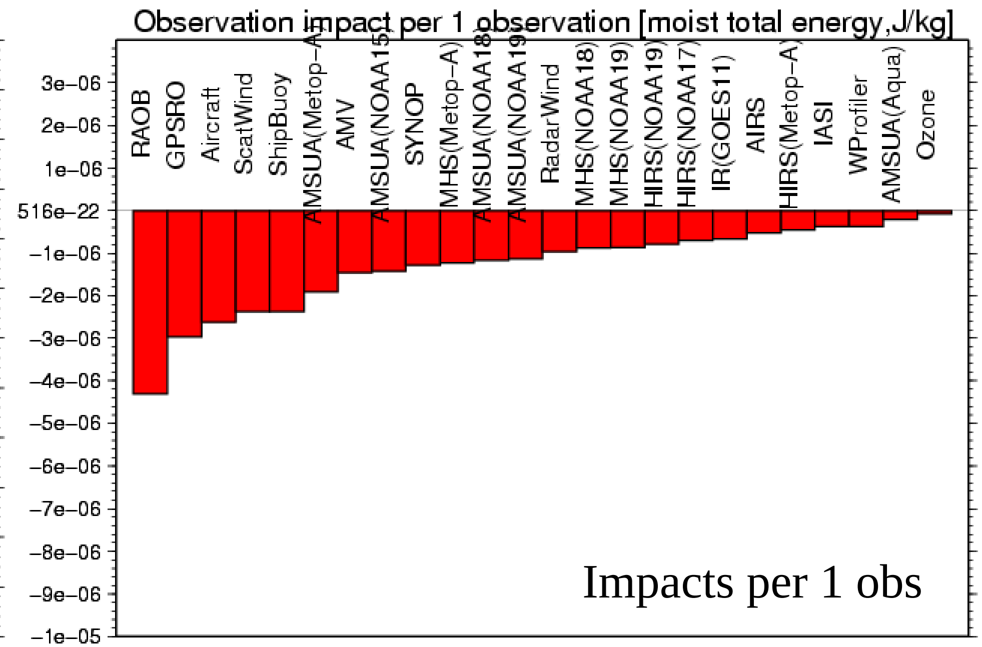
Only the result for the EnSRF is shown in this presentation.

Observation list

Type of data	Description	Number	Thinning
Aircraft	T, u, and v observations from the aircraft	134,000	
AMV	Atmospheric Motion Vectors	85,000	
Ozone	Ozone retrievals from satellite radiances	12,000	
RadarWind	u and v observations from the NEXRAD (radar)	19,000	
RAOB	Radiosonde observations (u, v, T, q and Ps) including PIBAL, flight level reconnaissance and dropsonde	58,000	
ScatWind	u and v observations from satellite microwave scatterometer	6,000	
ShipBuoy	Surface u, v, T, q and Ps observations from the buoys and ships	24,000	
SYNOP	Ps observations from land surface stations	51,000	
WProfiler	u and v observations from wind profiler and PILOT	12,000	
AIRS	Satellite infrared hyper spectral sounder radiances	283,000	180 km
AMSUA	Satellite microwave sounder radiances (from 5 satellites)	244,000	145 km
GPSRO	GPS radio occultation (reflectivity)	57,000	
HIRS	Satellite infrared radiances (from 3 satellites)	79,000	180 km
IASI	Satellite infrared hyper spectral sounder radiances	501,000	180 km
IR	Geostationary satellite infrared sounder radiances	18,000	180 km
MHS	Satellite microwave sounder radiances (from 3 satellites)	41,000	240 km

Impact summary



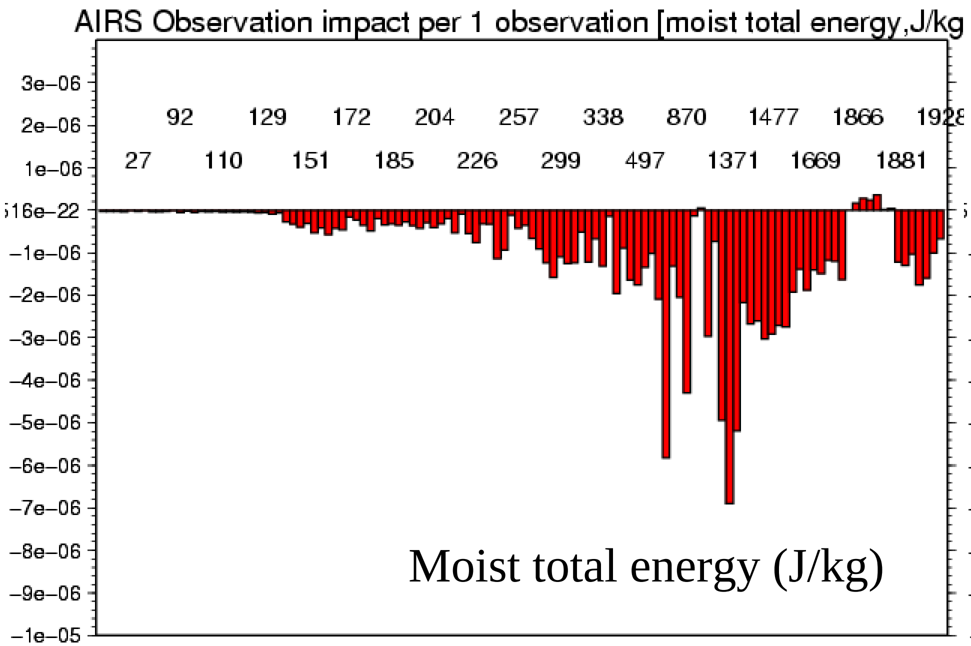


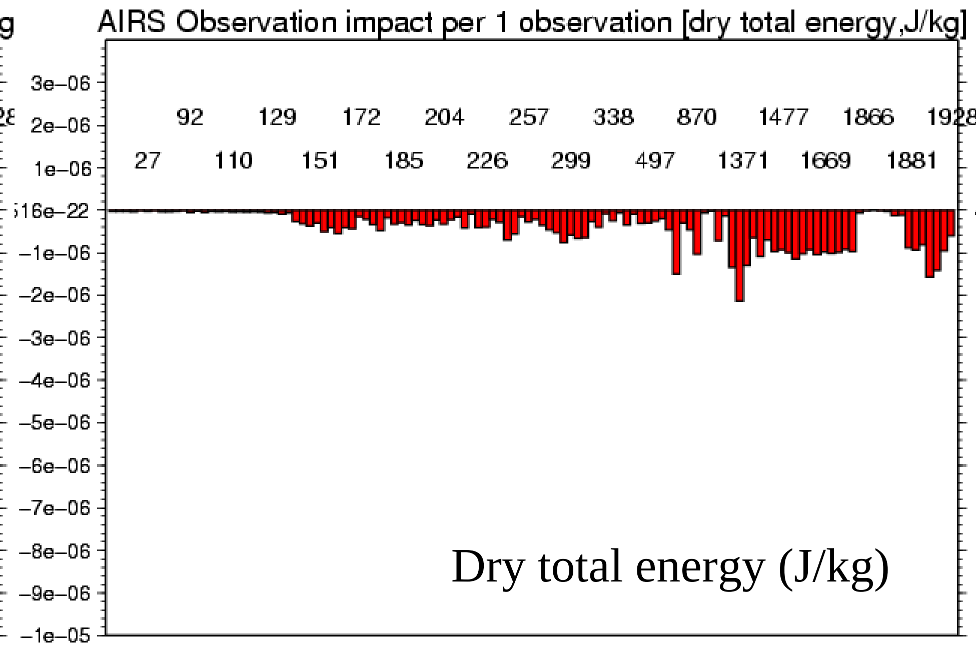
All observation types have positive impacts on average.

For the total impact, 1: aircraft, 2: AMSU-A, 3: radiosonde, 4: IASI, 5: GPSRO

For impact per 1 obs., 1: radiosonde, 2: GPSRO, 3: aircraft, 4: Scattrometer wind, 5: marine surface observation

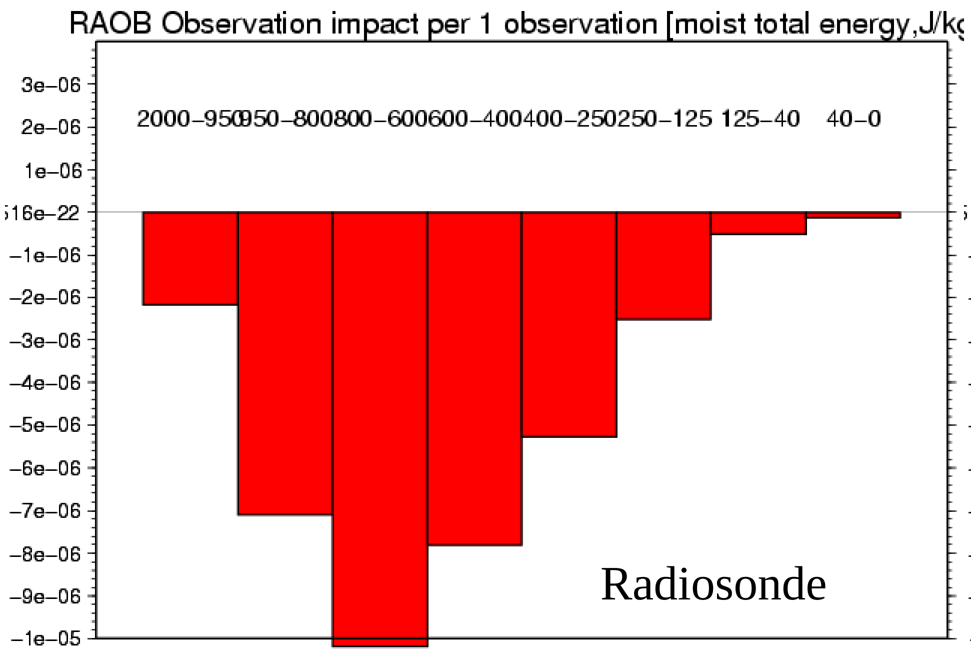
AIRS

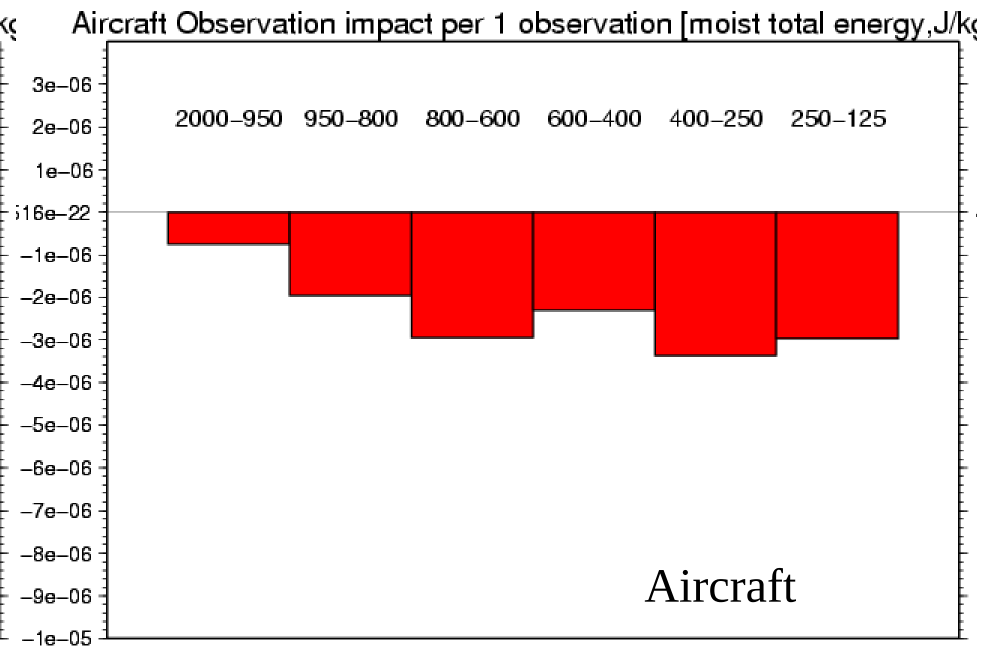




Some channels have large impacts and others does not.
Larger peaks in the moist total energy than in the dry total energy means these channels are sensitive to the forecast of moisture variable.

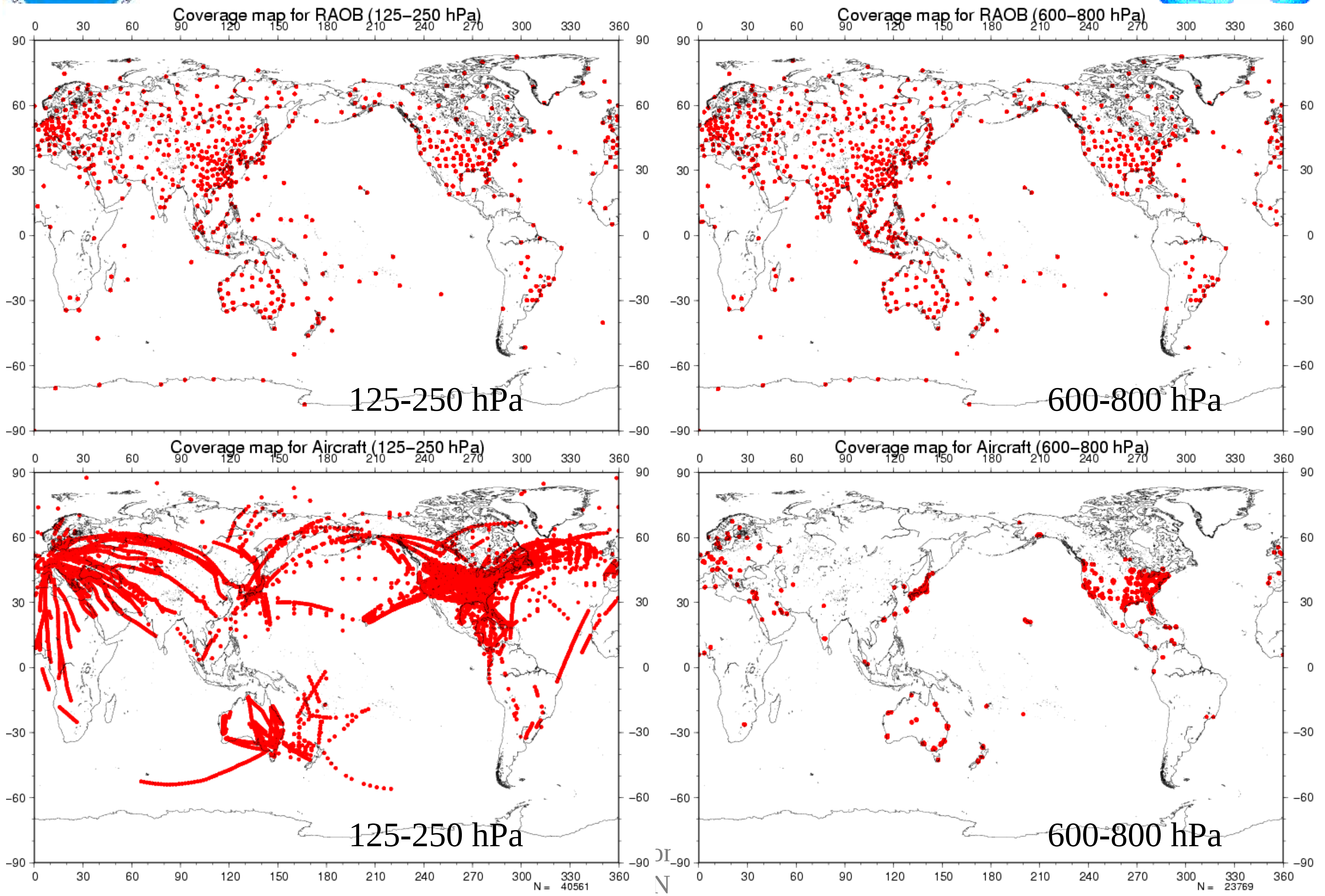
Radiosonde and aircraft



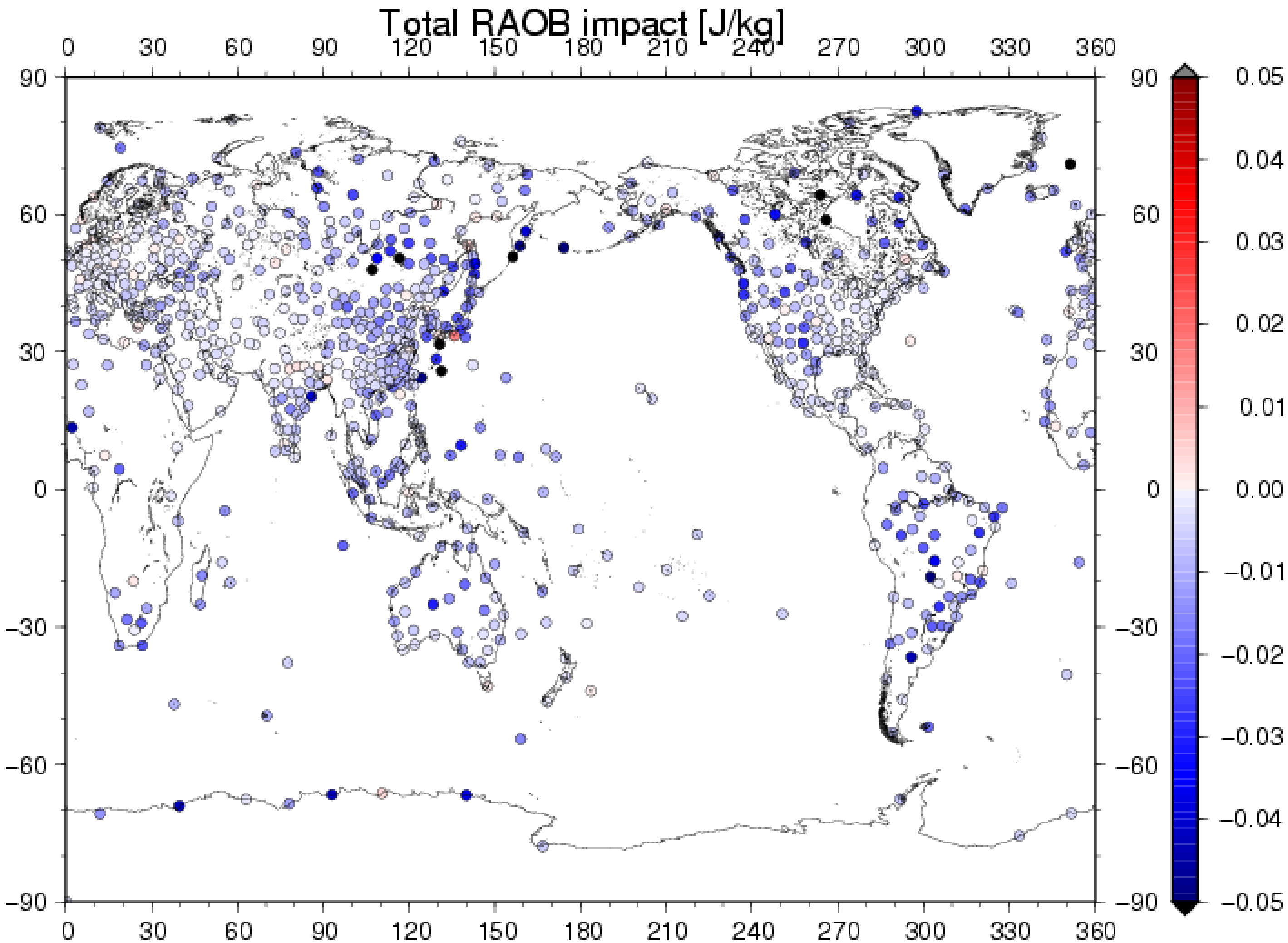


Radiosonde observations on mid- to lower troposphere have larger impacts compared to the aircraft observations.

Radiosonde and aircraft



Radiosonde impacts

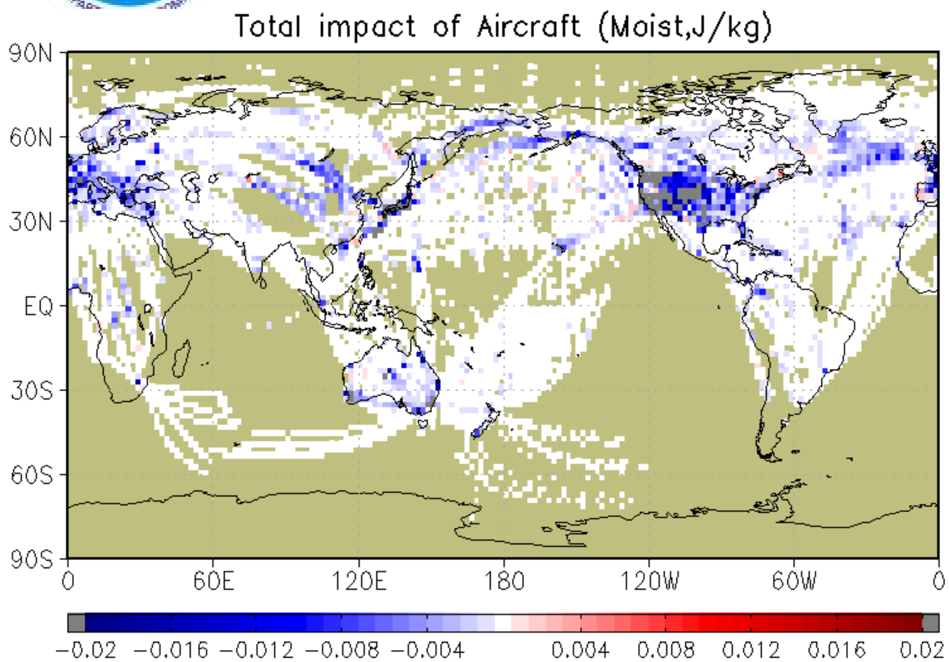


Total impacts of radiosonde (12UTC October 21 to 06UTC October 28)

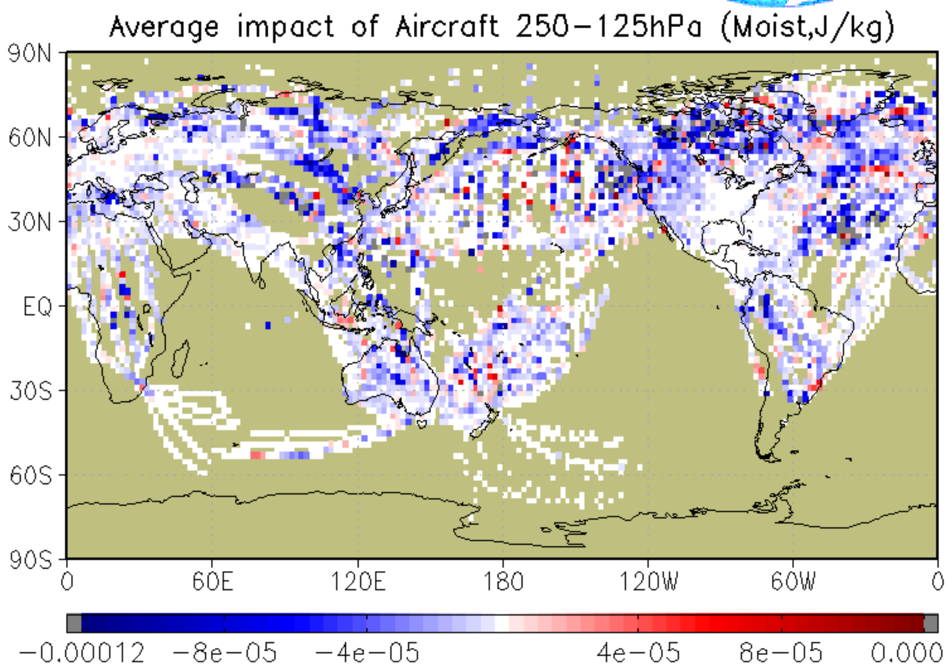
Most observations have positive impacts on average

Relatively large impacts for East Asia, Western US, Canada, and South America.

Aircraft impacts



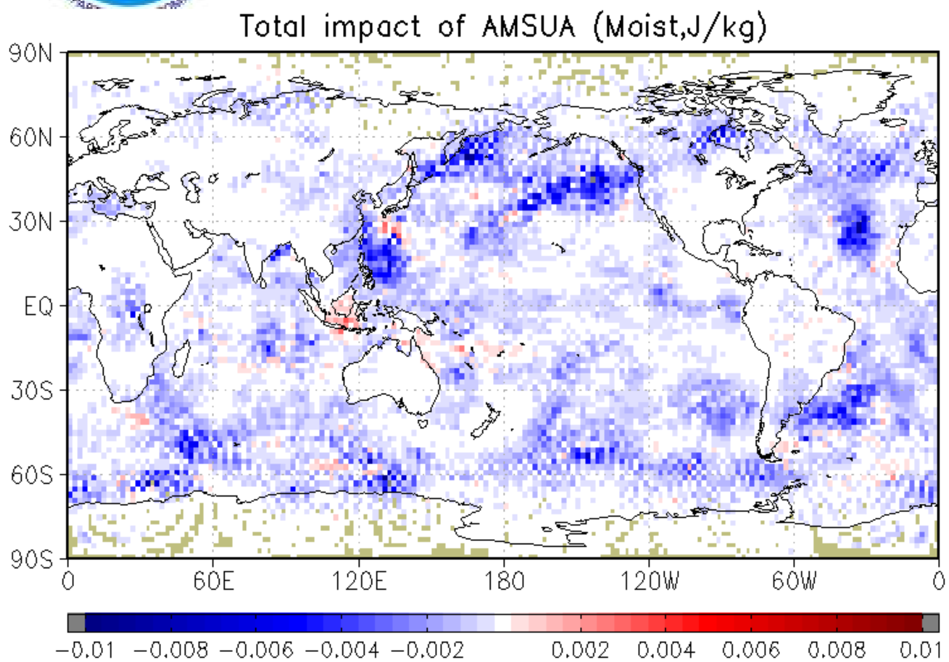
Total impacts, moist total energy



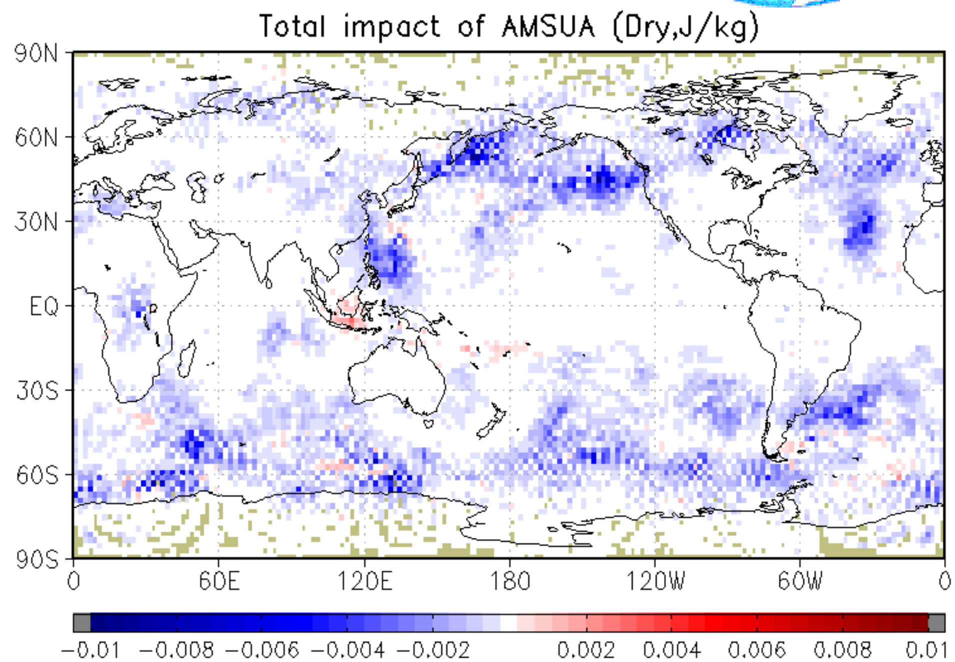
Impacts per 1 obs (250-125 hPa)

Aircraft observations over US, Europe and East Asia have large positive impacts. The impact of aircraft observations is extremely large over US, however impact per 1 observation is small.

AMSUA impacts



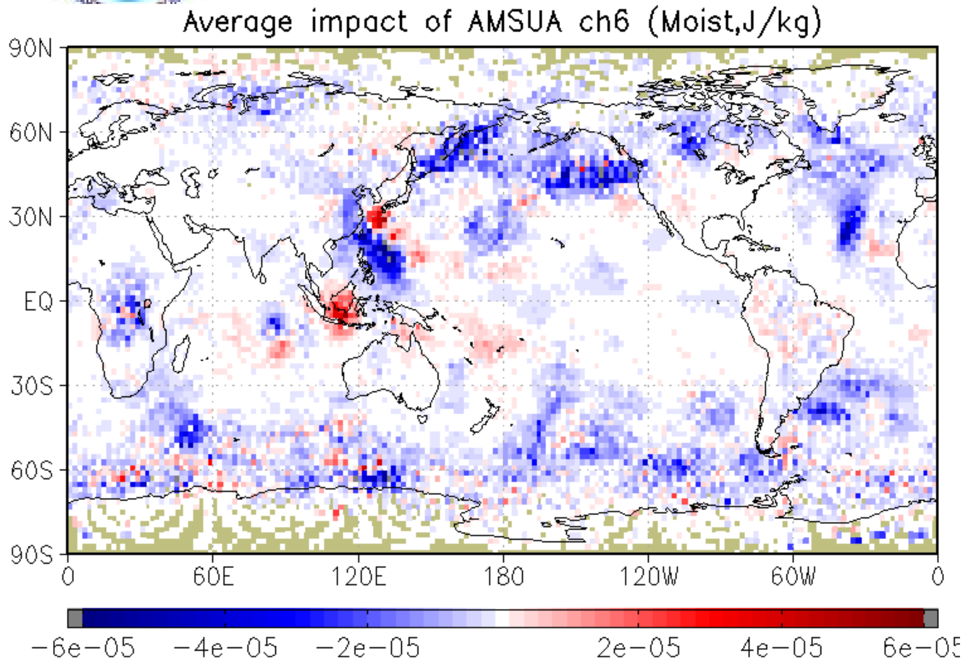
Total impacts, moist total energy



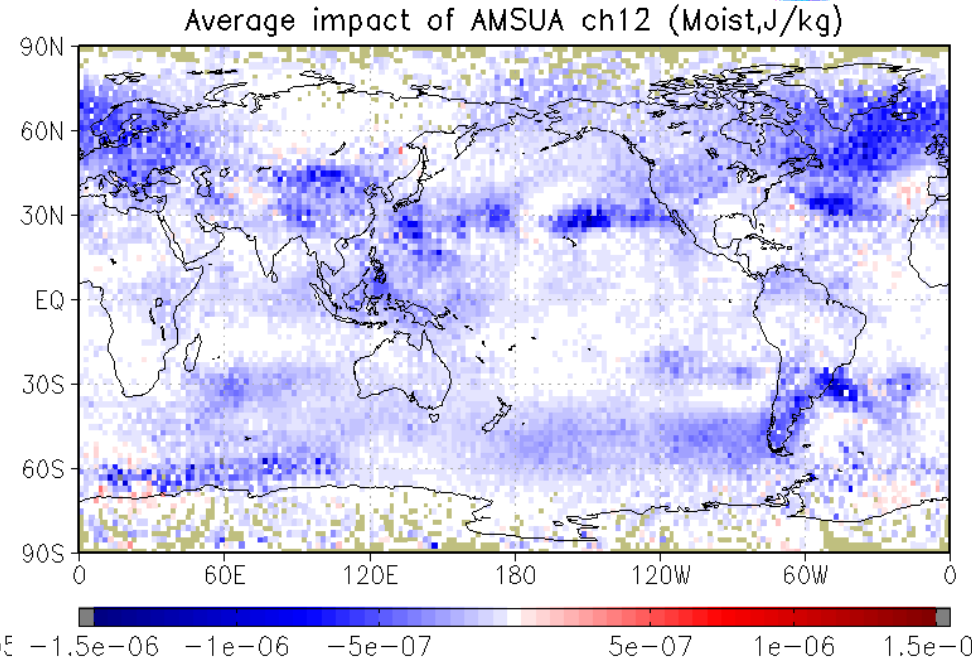
Total impacts, dry total energy

In most locations, observation impacts are positive.
Observations over mid-latitude ocean have large impacts.
Similar impacts are derived with dry and moist total energy norm (slightly large impacts for moist total energy in tropics).

AMSUA impacts



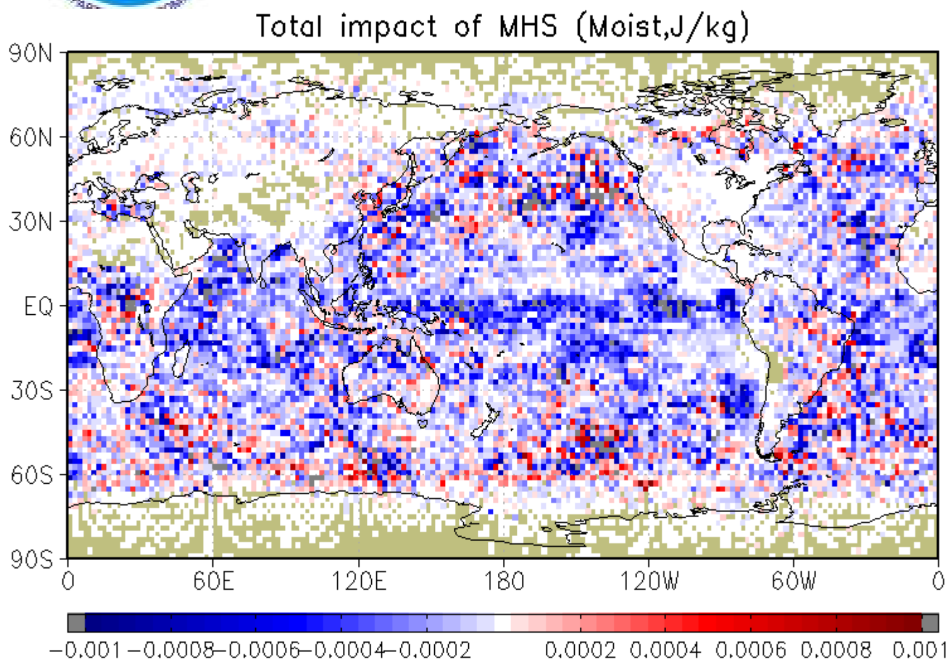
Average impact, ch6, moist total energy



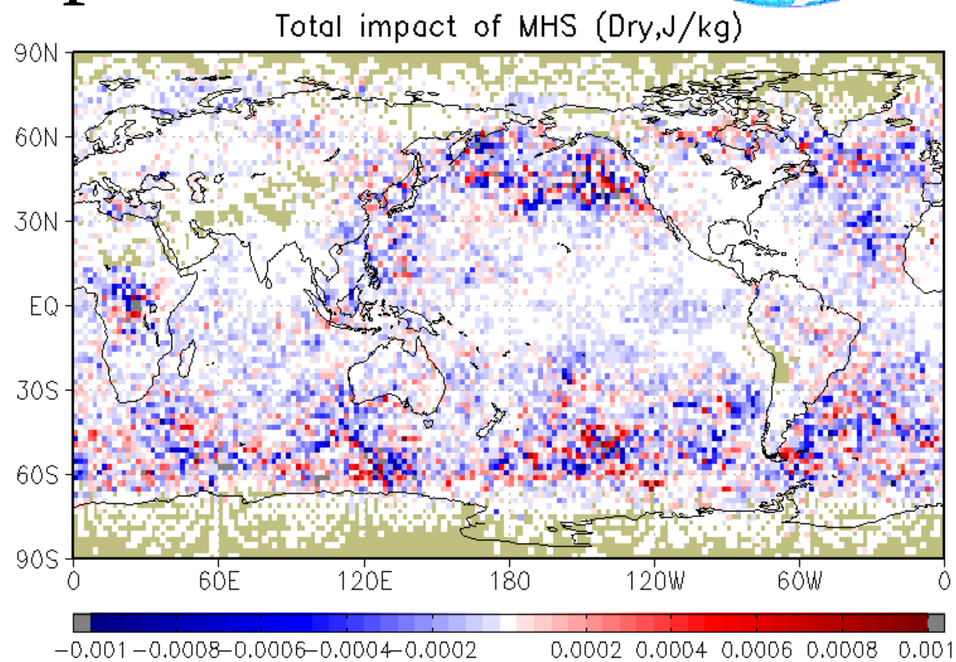
Average impact, ch12, moist total energy

In most locations, observations have positive impacts on average.
However, on some channels sensitive to lower to mid troposphere have negative impacts on some areas especially in tropics.
For the channels sensitive to the stratosphere, observations in almost all areas have positive impacts and the impacts are more uniform.

MHS impacts



Total impacts, moist total energy



Total impacts, dry total energy

Measuring with dry total energy, the impacts of MHS are mixed with positive and negative.

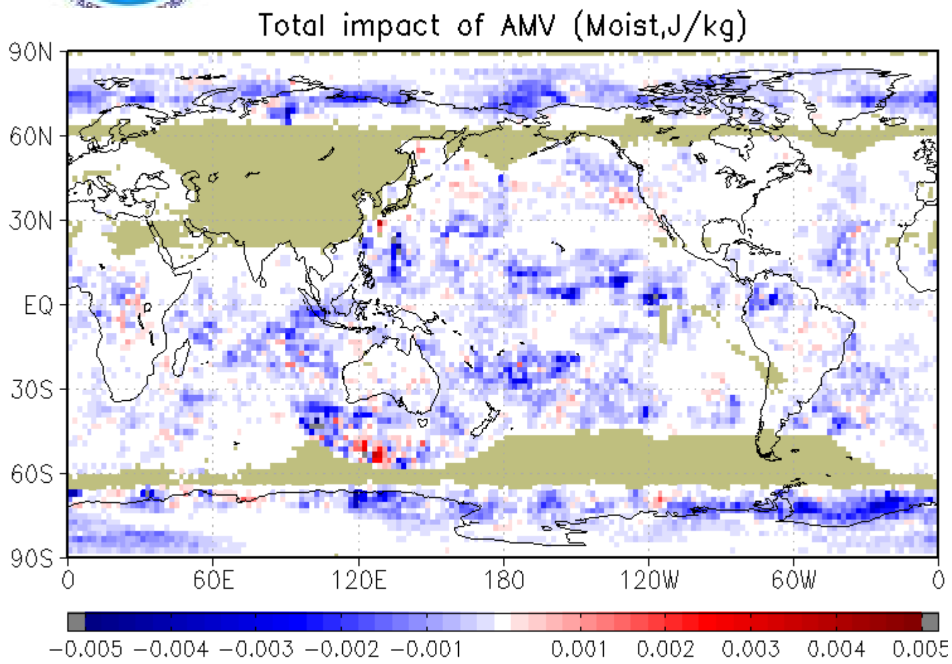
Most impacts are coming from observations over the mid-latitude ocean.

With moist total energy, the impact of the MHS is much larger especially in tropics.

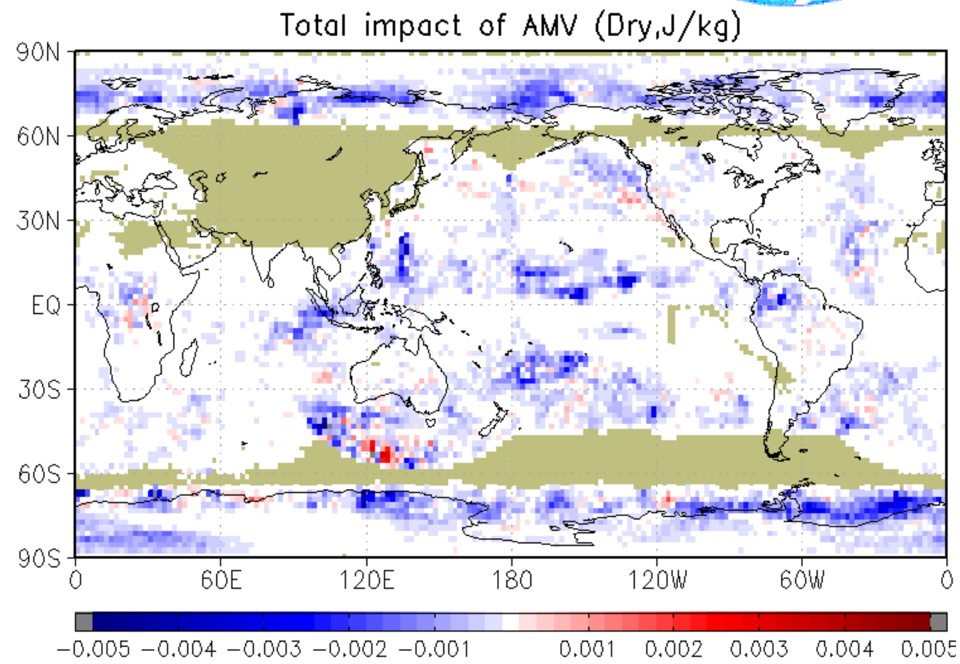
Thus, assimilation of MHS is beneficial to the moisture forecasts in tropics.

* Note: scales are 0.1 times smaller than AMSUA

AMV impacts



Total impacts, moist total energy



Total impacts, dry total energy

Both AMV from polar orbiting satellites and geostationary satellites have large positive impacts overall.
Impact with moist total energy is slightly larger than that with dry total energy especially for tropics.

Application to the hybrid EnKF/Var

- There are something we need to consider
  - Dual resolution
  - Assimilation of different observation datasets in EnKF and 3/4DVAR
- One possible alternative is to solve only $\mathbf{M}^T$ part with the ensemble-based sensitivity method (Ancell and Hakim 2007).
- Need more work to apply it to the observation impact study.

Summary

- Observation impacts are estimated within the NCEP GFS/EnKF using Kalnay et al. (2012).
- All observation types contribute to reduce the short-range forecast error on average.
- This method would provide diagnostics of the data assimilation system and observing system in EnKF.
- Future work
  - Application to the EnKF/Var framework
  - Improvements on the EnKF-based impact estimates (localization issue)
  - Robustness of the estimate (horizontal resolution, number of ensemble members)
  - Verification in observation space

Thank you very much !

P.S. Global hybrid 3DVAR/EnKF become operational today at NCEP.

Two EnKF-based formulation

From Liu and Kalnay (2008) and Li et al. (2010),

[1] $$\delta J = \mathbf{d}^{\mathbf{T}}\tilde{\mathbf{K}}_0^{\mathbf{T}}\mathbf{X}_{t|-6}^{f\mathbf{T}}\mathbf{C}\left(\mathbf{e}_{t|-6} + \frac{1}{2}\mathbf{X}_{t|-6}^{f}\tilde{\mathbf{K}}_0\mathbf{d}\right)$$

(1) and (2) (1) and (2)

From Kalnay et al. (2012),

[2] $$\delta J = \frac{1}{2}\frac{1}{K-1}\mathbf{d}^{\mathbf{T}}\mathbf{R}^{-1}\left(\mathbf{H}\mathbf{X}_0^a\right)\mathbf{X}_{t|0}^{f\mathbf{T}}\mathbf{C}\left(\mathbf{e}_{t|0} + \mathbf{e}_{t|-6}\right)$$

(1) and (3)

NOTE

$$\tilde{\mathbf{K}}_0 \equiv \tilde{\mathbf{P}}_0^a\left(\mathbf{H}\mathbf{X}_{0|-6}^f\right)^{\mathbf{T}}\mathbf{R}^{-1}$$

$$\tilde{\mathbf{P}}_0^a \equiv \left[(K-1)\mathbf{I} + \left(\mathbf{H}\mathbf{X}_{0|-6}^f\right)^{\mathbf{T}}\mathbf{R}^{-1}\left(\mathbf{H}\mathbf{X}_{0|-6}^f\right)\right]^{-1}$$

$$\mathbf{X}_0^a = \mathbf{X}_{0|-6}^f\left[(K-1)\tilde{\mathbf{P}}_0^a\right]^{1/2}$$

And, here is a definition,

$$\delta J \equiv \frac{1}{2}\left(\mathbf{e}_{t|0}\mathbf{C}\mathbf{e}_{t|0}^{\mathbf{T}} - \mathbf{e}_{t|-6}\mathbf{C}\mathbf{e}_{t|-6}^{\mathbf{T}}\right) = \frac{1}{2}\left(\mathbf{e}_{t|0} - \mathbf{e}_{t|-6}\right)^{\mathbf{T}}\mathbf{C}\left(\mathbf{e}_{t|0} + \mathbf{e}_{t|-6}\right)$$

Basically, the difference of formulations is coming from how to approximate $\mathbf{e}_{t|0}$-$\mathbf{e}_{t|-6}$ and how to use that.

(1): Evolution of the analysis increment

(2): Evolution of the first guess perturbation

(3): Evolution of the analysis perturbation

For the formulation [1],

$$\mathbf{e}_{t|0} - \mathbf{e}_{t|-6} = \overline{\mathbf{x}_{t|0}^f} - \overline{\mathbf{x}_{t|-6}^f} \approx \mathbf{M}_{0\to t}\left(\overline{\mathbf{x}_0^a} - \overline{\mathbf{x}_{0|-6}^f}\right) = \mathbf{M}_{0\to t}\mathbf{X}_{0|-6}^f\tilde{\mathbf{K}}_0\mathbf{d} \approx \mathbf{X}_{t|-6}^f\tilde{\mathbf{K}}_0\mathbf{d}$$

(1) (2)

For the formulation [2],

$$\mathbf{e}_{t|0} - \mathbf{e}_{t|-6} = \overline{\mathbf{x}_{t|0}^f} - \overline{\mathbf{x}_{t|-6}^f} \approx \mathbf{M}_{0\to t}\left(\overline{\mathbf{x}_0^a} - \overline{\mathbf{x}_{0|-6}^f}\right) = \mathbf{M}_{0\to t}\mathbf{X}_{0|-6}^f\tilde{\mathbf{K}}_0\mathbf{d}$$

(1)

$$= \mathbf{M}_{0\to t}\mathbf{X}_{0|-6}^f\tilde{\mathbf{P}}_0^a\left(\mathbf{H}\mathbf{X}_{0|-6}^f\right)^{\mathbf{T}}\mathbf{R}^{-1}\mathbf{d} = \mathbf{M}_{0\to t}\frac{1}{K-1}\mathbf{X}_0^a\mathbf{X}_0^{a\mathbf{T}}\mathbf{H}^{\mathbf{T}}\mathbf{R}^{-1}\mathbf{d}$$

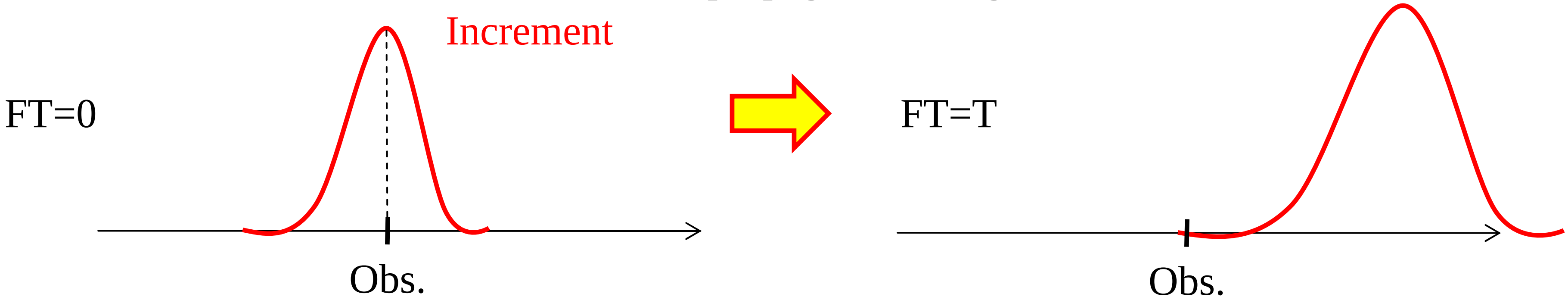
(3)

$$\approx \frac{1}{K-1}\mathbf{X}_{t|0}^f\left(\mathbf{H}\mathbf{X}_0^a\right)^{\mathbf{T}}\mathbf{R}^{-1}\mathbf{d}$$

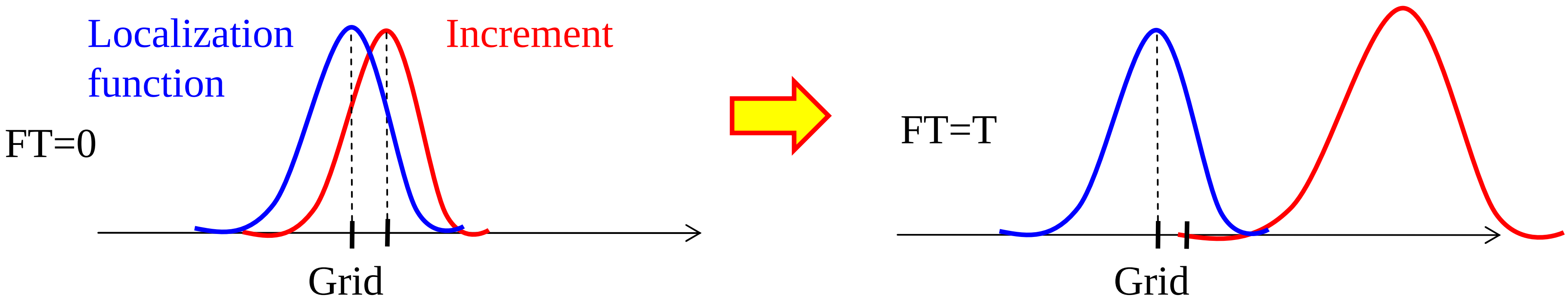
(3) may be better than (2), because perturbation is smaller for the analysis ensemble.

Problem of localization

Observation information moves and propagates during the forecast.



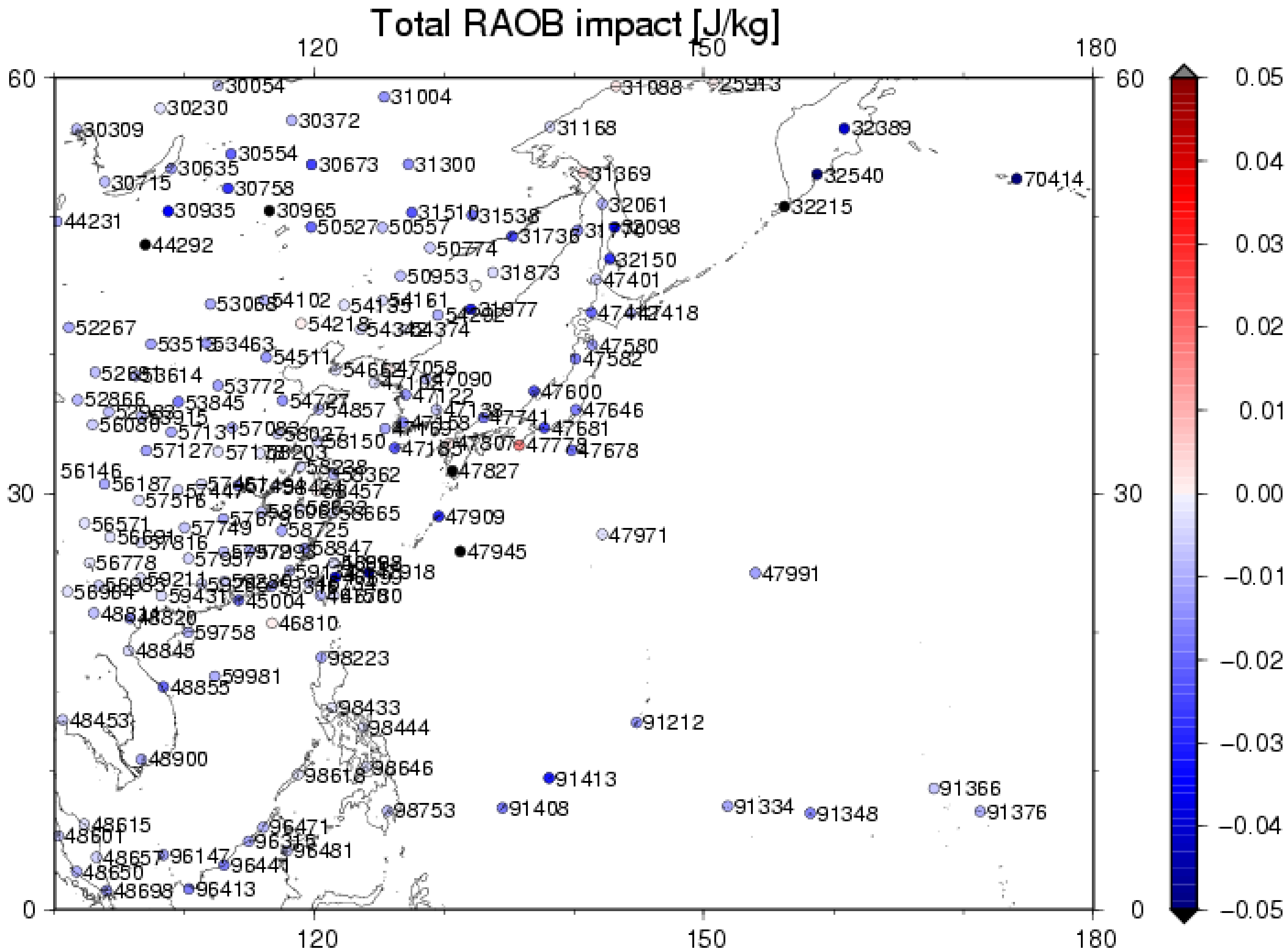
In FSO using EnKF, we apply the localization on perturbations at the evaluation time.



For longer FTs, most of the information moves away from the localized area. For limited number of members and fixed localization, we cannot capture enough signals.

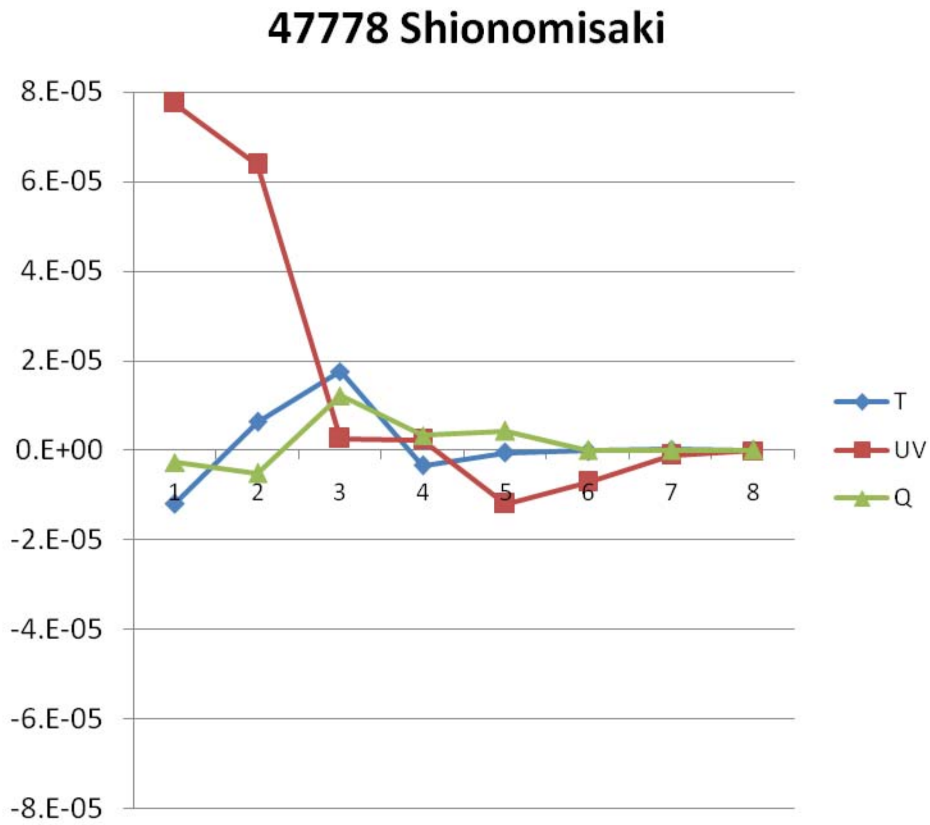
Currently, localization center point for observation impact estimate is moved with the horizontal wind.

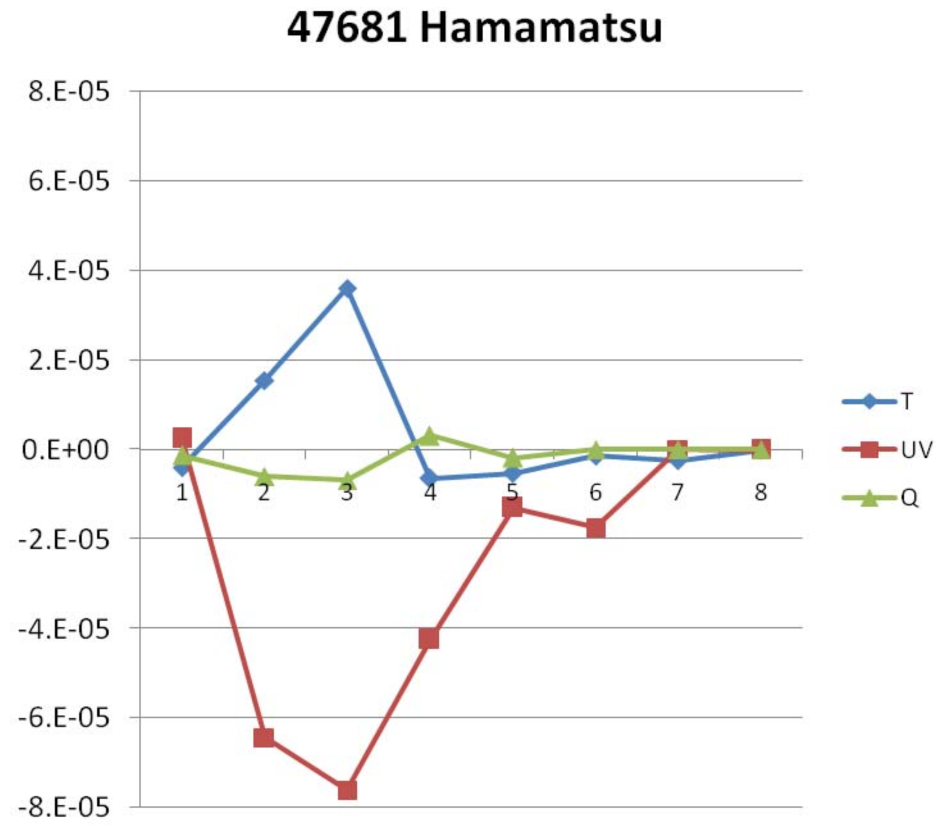
Radiosonde impacts



47778 (Shionomisaki) shows relatively large negative impact.

Radiosonde impacts



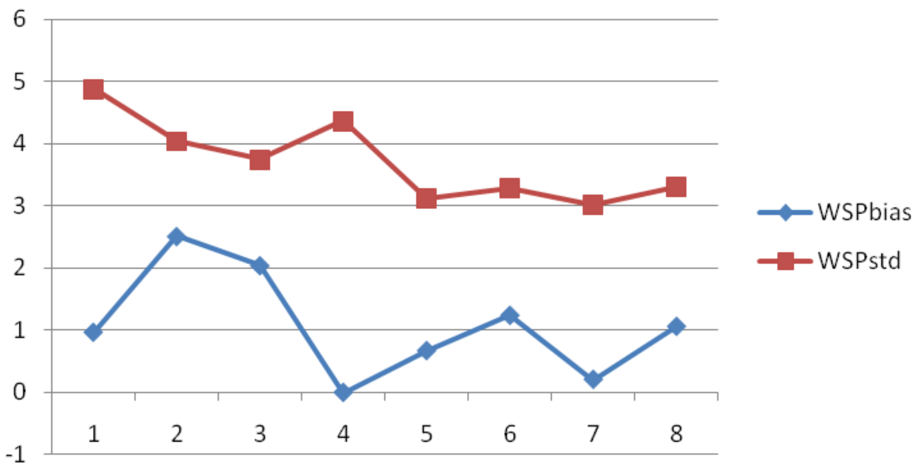


1: -950hPa, 2: 800-950hPa, 3: 600-800hPa, 4: 400-600hPa, 5: 250-400hPa, 6: 125-250hPa, 7: 40-125hPa, 8: 0-40hPa

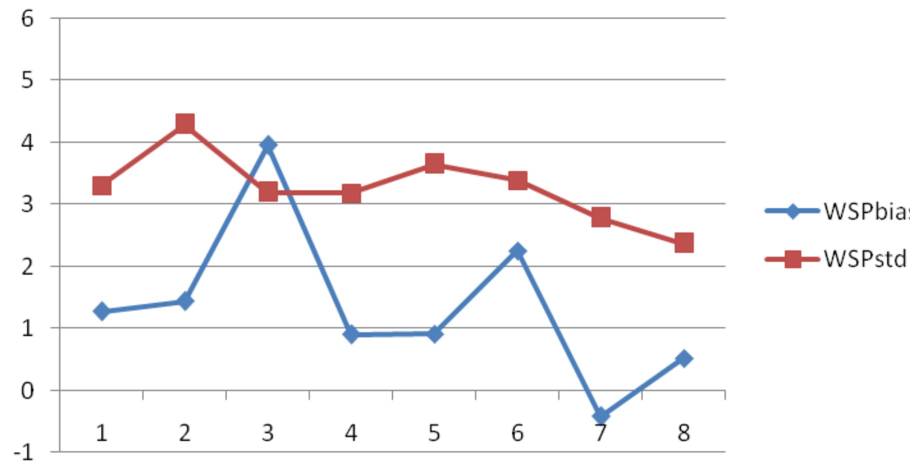
Large negative impacts from wind observations on lower troposphere

Wind observation innovations

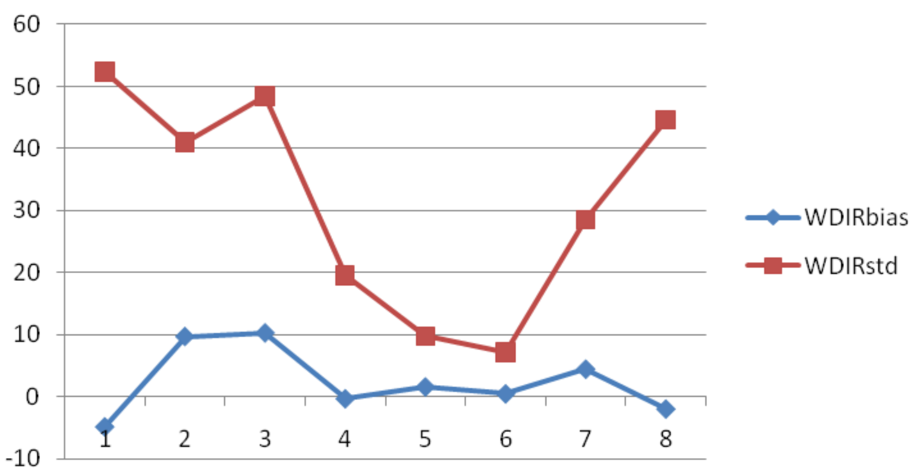
Wind Speed (47778, m/s)



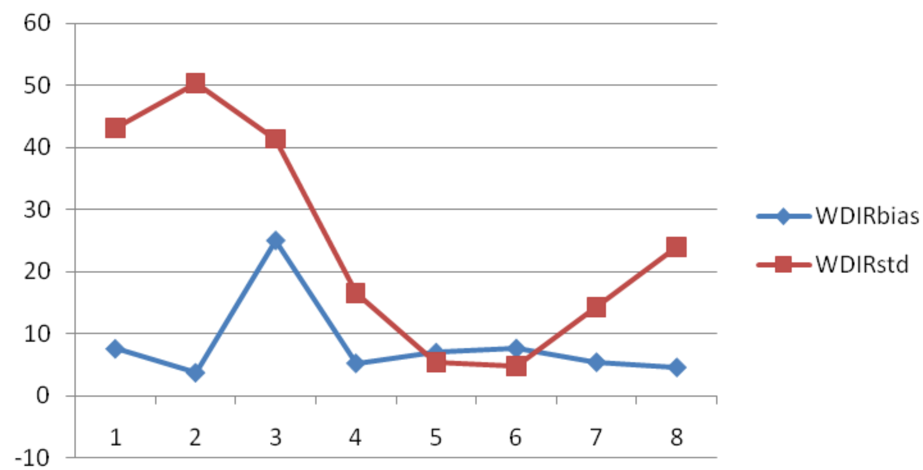
Wind Speed (47681, m/s)



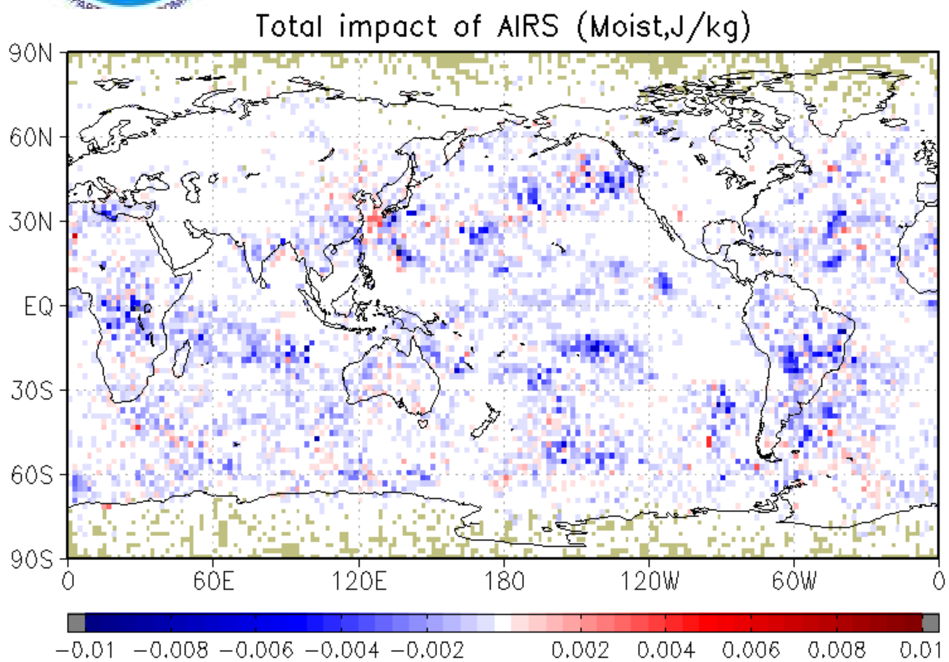
Wind Direction (47778, deg)



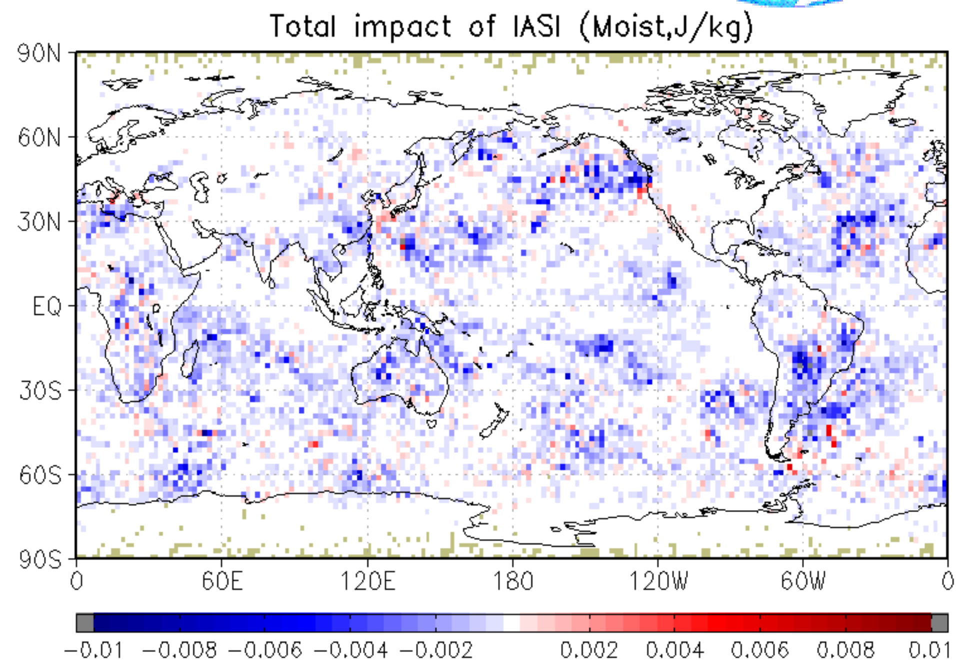
Wind Direction (47681, deg)



AIRS and IASI impacts



AIRS total impacts, moist total energy

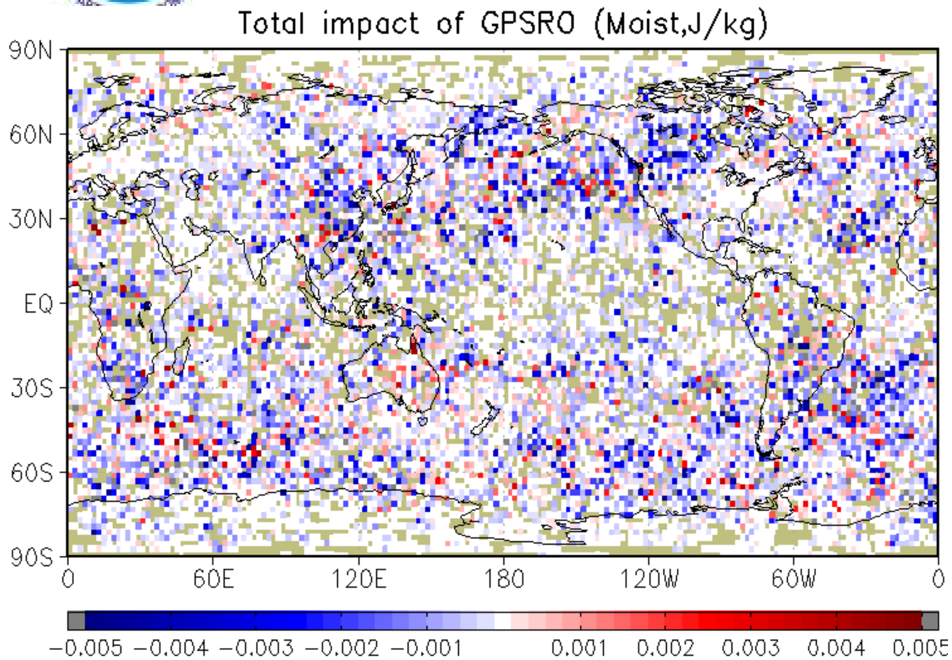


IASI total impacts, moist total energy

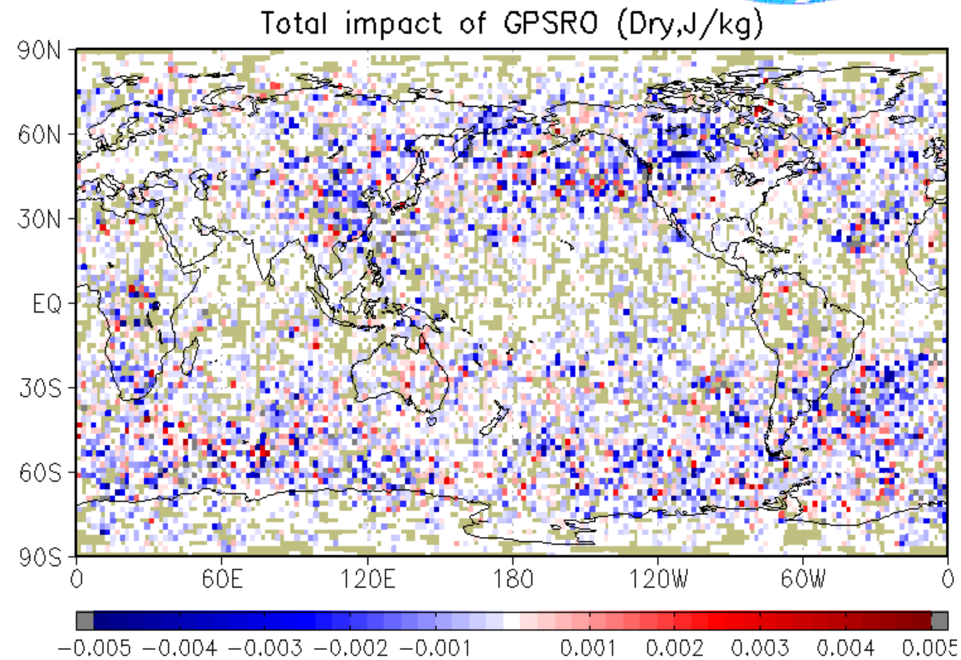
In most areas, both AIRS and IASI have positive impacts especially over NH ocean and SH.

Total impact distributions are very similar for AIRS and IASI.

GPSRO impacts



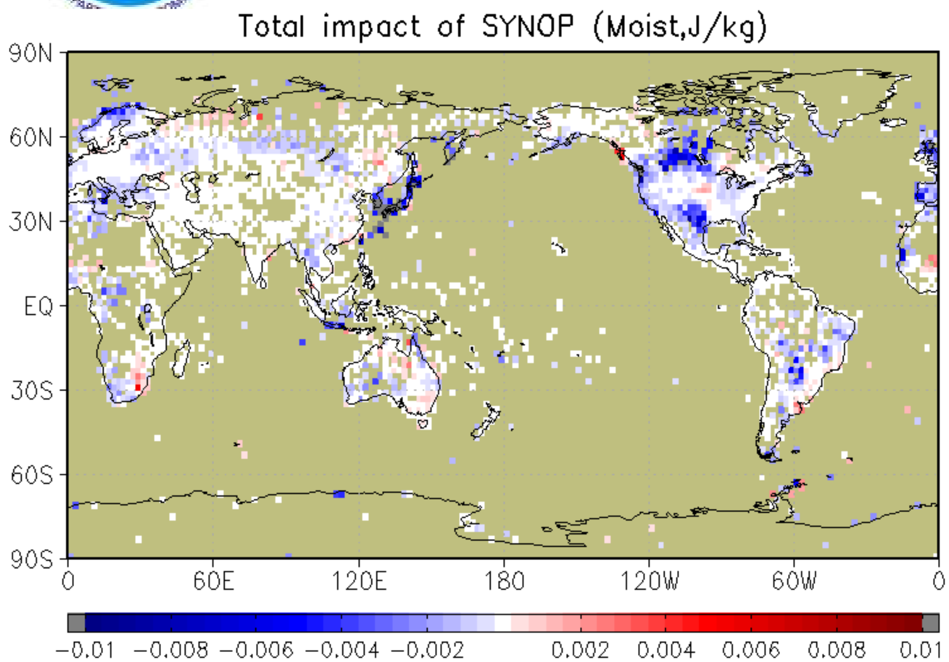
Total impacts, moist total energy



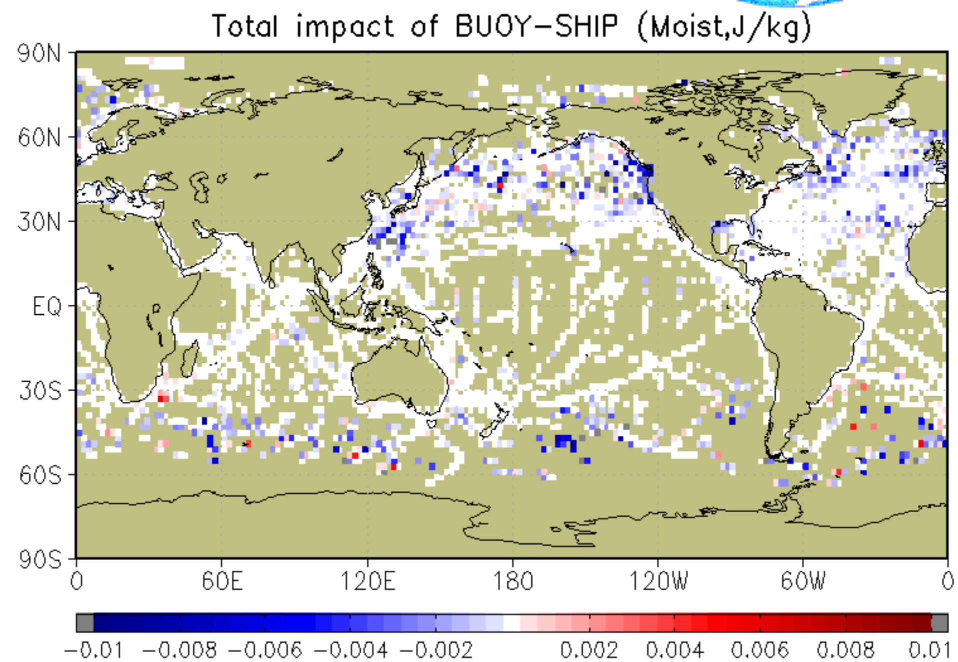
Total impacts, dry total energy

Impacts of GPSRO are mixed with positive and negative.
Maybe the number of sample is not enough to get average impacts.

SYNOP and SHIP-BUOY impacts

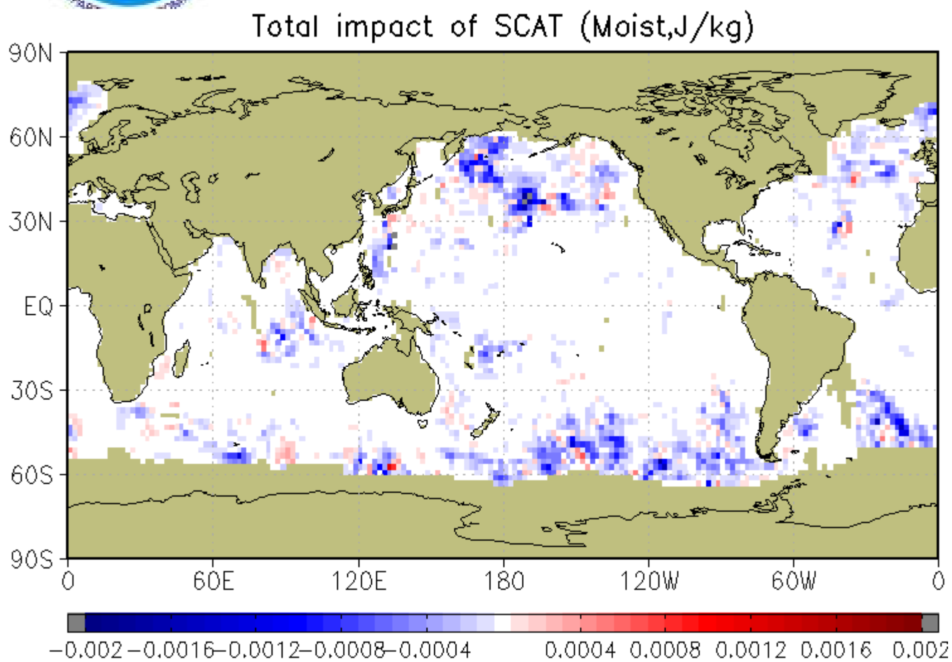


SYNOP total impacts, moist total energy

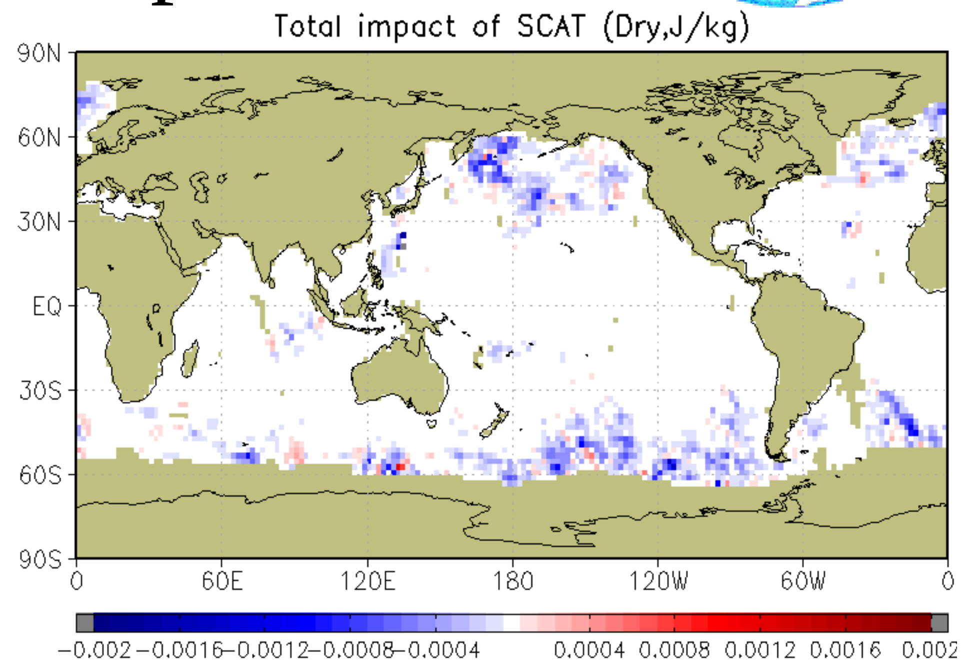


BUOY,SHIP total impacts, moist total energy

ScatWind impacts



Total impacts, moist total energy



Total impacts, dry total energy

Large impacts are coming from the observations over mid-latitude ocean where strong extratropical cyclones are common.