Introduction

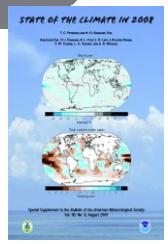
Why should we bother about exchanging reanalysis observation feedback and blacklists?

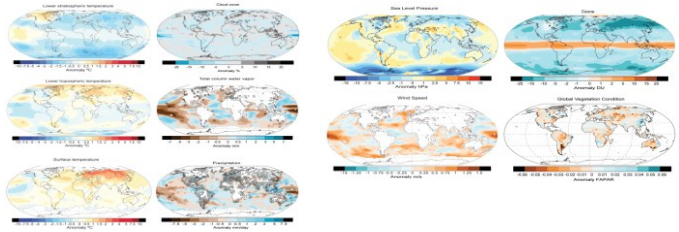
Outline

- Reanalyses becoming highly popular datasets
- With this come greater responsibilities
- ‘Top-down’ approach to explain our observation usage is not sufficient any more
- Other applications of the feedback (historical specialized users)

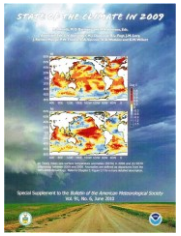
Growing use of reanalyses

BAMS State of the Climate in 2008

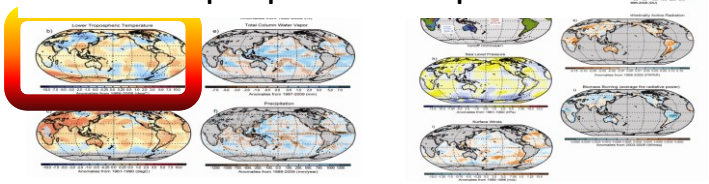




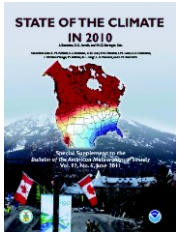
BAMS State of the Climate in 2009

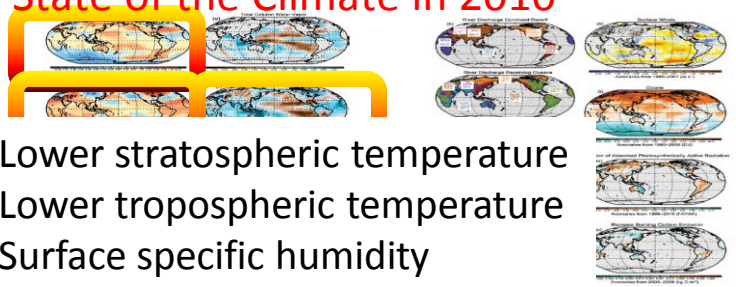


Lower tropospheric temperature



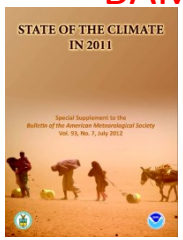
BAMS State of the Climate in 2010

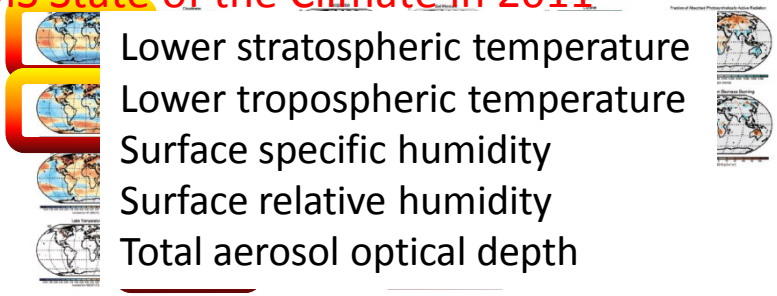




Lower stratospheric temperature
Lower tropospheric temperature
Surface specific humidity

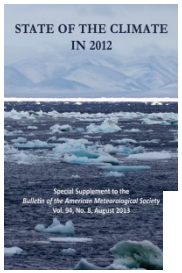
BAMS State of the Climate in 2011

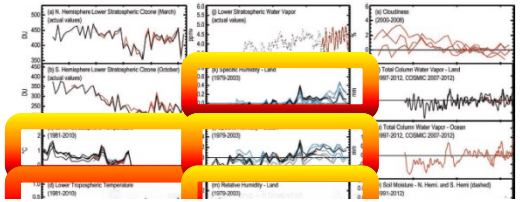




Lower stratospheric temperature
Lower tropospheric temperature
Surface specific humidity
Surface relative humidity
Total aerosol optical depth

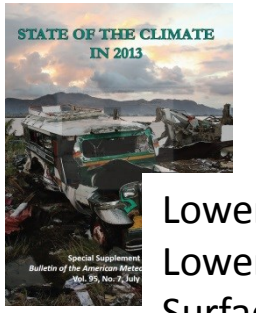
BAMS State of the Climate in 2012

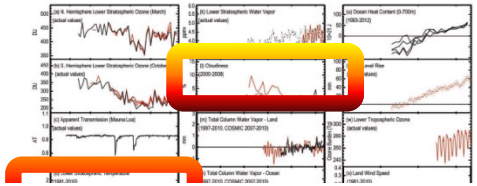




Lower stratospheric temperature
Lower tropospheric temperature
Surface temperature
Ocean wind speed
Surface specific humidity (land, ocean)
Surface specific humidity (land, ocean)

BAMS State of the Climate in 2013





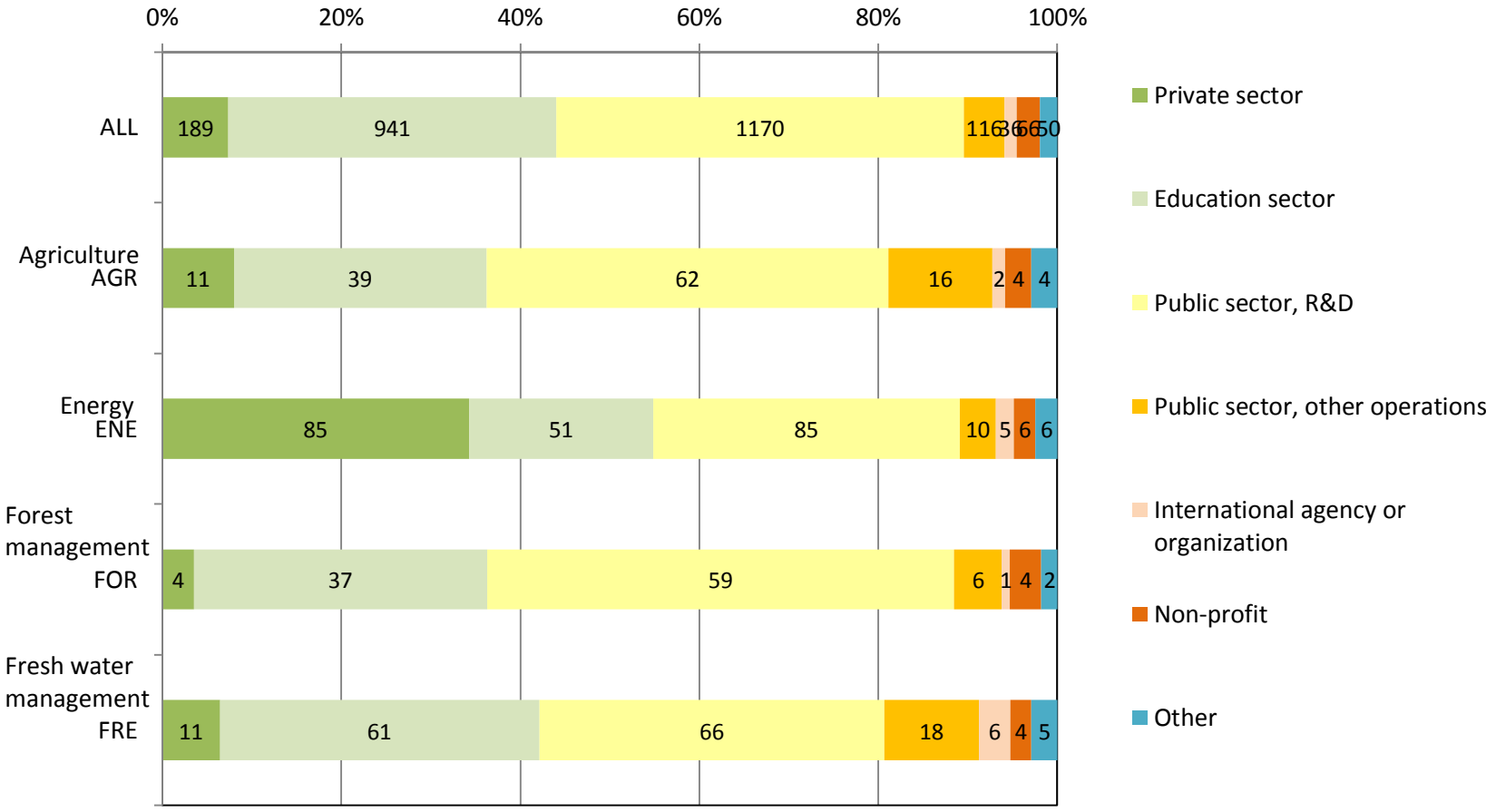
Lower stratospheric temperature
Lower tropospheric temperature
Surface temperature
Ocean wind speed
Surface specific humidity (land, ocean)
Surface specific humidity (land, ocean)
Cloudiness

Reanalyses: high usage

- Among the widest used datasets in Earth sciences, for many applications
  - ERA-40 user survey: 127 responses
  - CC user survey sent to 20,000+ ERA users and advertised on several websites: 2500+ responses
- Reasonable demands and expectations of transparency about what observations data are used, how, how many, where, when…

Reanalyses reach a diverse user base

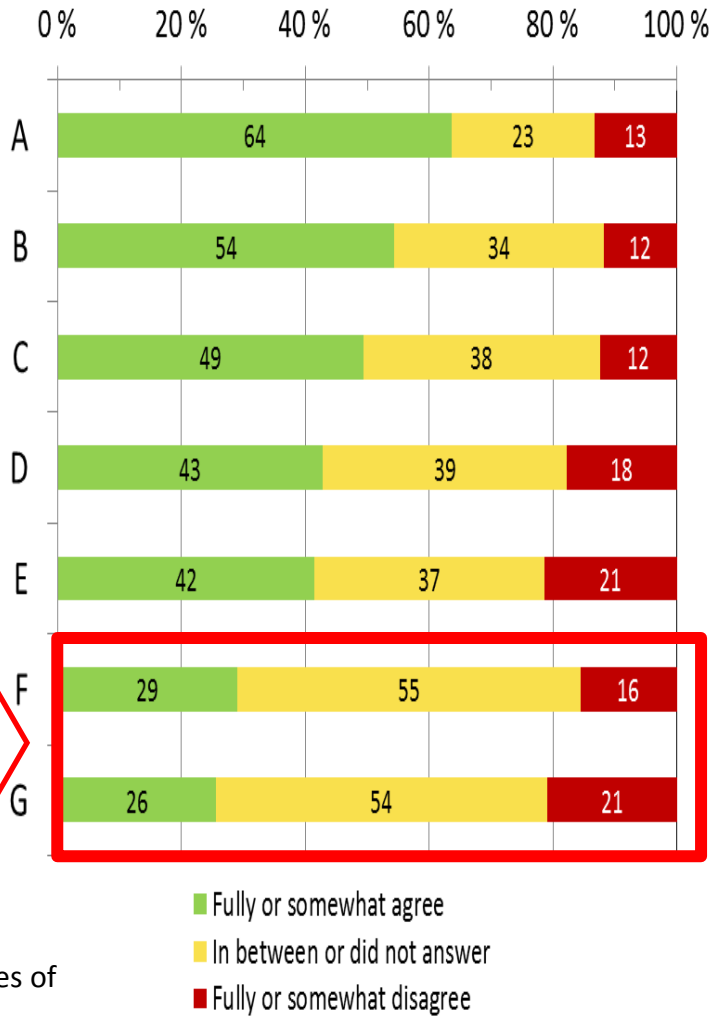
Respondents' sector of work for each study group and sector



Core-Climax user survey, Gregow *et al.*, manuscript in preparation (2014)

Reanalysis users want to know more about observation use, and uncertainties

Class	Statements about reanalysis data
A	• The data is easy to access
B	• The time period covers my interests
	• The general quality is good enough for my needs
	• The data are consistent between the variables
	• The data can be imported easily by my software application
C	• The temporal continuity is adequate
	• The temporal nominal resolution is adequate
	• I know enough to work with the data
	• The vertical nominal (grid cell size) resolution is adequate
D	• Websites provide good information
	• The file sizes are NOT too large to work with
	• The horizontal nominal (grid cell size) resolution is adequate
	• The spatio-temporal scales that I need are well represented
	• The literature provides good information
E	• The data policy is NOT too strict
	• The data DO NOT tend to become available too late for my needs
F	• The biases compared with observations are small enough
	• Time-varying biases DO NOT make the data too instable for my needs
G	• For the climate variables I need, I know how much their spatial true (feature) resolution differs from the nominal resolution
	• I know how much the temporal true (feature) resolution differs from the nominal resolution in time
	• The observation input to reanalysis are clearly explained
	• Plentiful training material is available on the web
	• The uncertainties are well characterized
#	2578



Opinions about characteristics of reanalysis data, grouped in seven homogeneous classes of agreement regarding reanalysis data, based on answers by all the respondents

Core-Climax user survey, Gregow *et al.*, manuscript in preparation (2014)

When people ask about observations: Our first responses

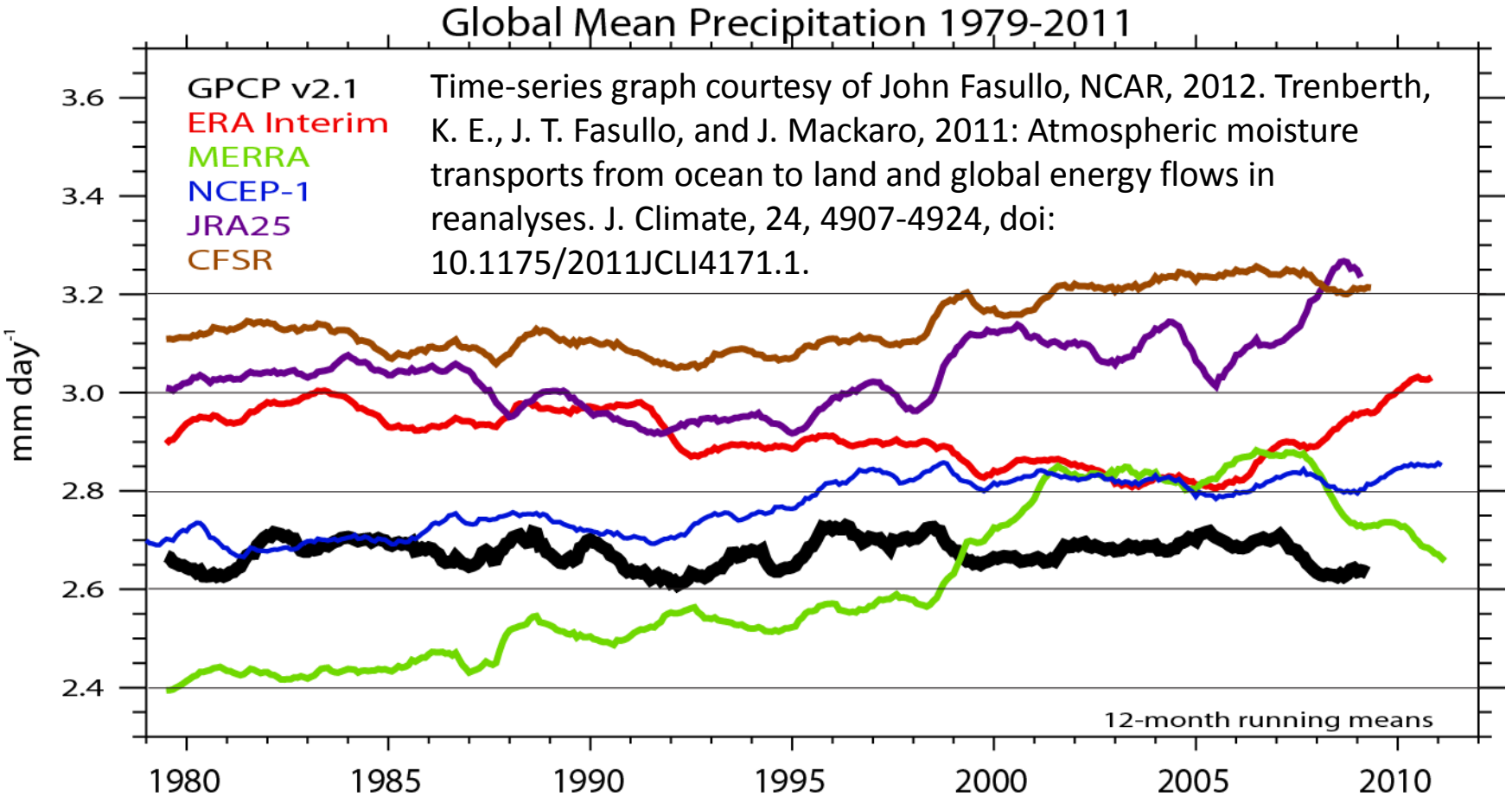
"-- Read the papers!"

- "Everything important is in the papers"
- "Carefully drawn plots show plenty of details"
- "Try and understand what's happening, we cannot do your work and project how observations impact your application"

Applying this recommendation to ourselves…

- …I tried to reconstruct (for the 4th WCRP International Conference on Reanalyses in 2012) the story of what happens to the water cycle as seen from several reanalyses

Precipitation in reanalyses and **GPCP**



Satellite data usage in recent, global reanalyses

Instrument, Observable	CFSR	MERRA	JRA-25	JRA-55	ERA-Interim
MSU Radiances					
AMSU-A Radiances					
SSMI Radiances					
AMSU-B / MHS Radiances					
HIRS Radiances					
SSU Radiances					
GEO Radiances					
SSMI Retrievals					
Imager Upper-air winds					
Scatt Ocean surf. wind					
Ozone Retrievals					

So what, overall, these reanalyses use about the same satellite data?

More details about satellite data usage in reanalyses

Overview of satellite data assimilated in modern reanalyses

Instrument, Observable	CFSR	MERRA	JRA-25	JRA-55	ERA-Interim
MSU Radiances	ch. 1,2,3,4 *For NOAA-10 to -14, NESDIS SNO corrected calibration coefficients; Exclusions: tighten qc limits in tropics and over high [illegible] test ch. 2*	ch. 1,2,3,4 *NESDIS SNO corrected calibration coefficients; Exclusions: snow, ice, mixed surfaces for ch. 1,2*	ch. 2, 3, 4 *Exclusions: land or rain for ch. 2-3*	ch. 2, 3, 4 *Exclusions: land or rain for ch. 2, land for ch. 3*	ch. 2, 3, 4 *Exclusions: land or rain for ch. 2, land for ch. 3*
AMSU-A Radiances	ch. 1-13, 15 *Exclusions: estimated cloud liquid water large for ch. 1-5, 15; Scattering index too large for ch. 1-6, 15; Channel 4 gross check large for ch. 1-5,15; Ch. 6 gross check large for ch. 1-6,15; Over high orography (> 2000m) for ch. 1-5, 15; large fit to emissivity or Ts for ch. 1-5, 15*	ch.1-15 *Exclusions: snow, ice, mixed surfaces for ch. 1-6,15, no offset bias correct for ch.14*	ch. 4-13 *Exclusions: sea-ice or land for ch. 4-5, high terrain for ch. 6-7, rain for ch4-6*	ch. 4-13 *Exclusions: sea-ice or land for ch. 4-5, high terrain for ch.6-7, rain for ch4-8*	ch. 5-14 *Exclusions: high terrain for ch. 5-6, rain for ch. 5-7, no offset bias correct for ch14*
SSMI Radiances		ch. 1-7 *Ocean only*		ch. 1,3,4,6 *Exclusions: land, rain*	ch. 1-7 *Exclusions: land, rain*
AMSU-B / MHS Radiances	ch. 1-5 *Exclusions: scattering index too large, channel 1 fit too large, any channel failing gross check, high orography (>2000m).*	ch. 1-5 *Exclusions: snow, ice, mixed surfaces for ch. 1,2,5*	ch. 3-5 *Exclusions: land, sea-ice, rain*	ch. 3-5 *Exclusions: land, sea-ice, rain*	ch. 3-5 *Exclusions: sea-ice, rain, high terrain for ch. 3-4, land for ch. 5*
HIRS Radiances	ch. 2-15 *Exclusions: Over water wavenumbers > 2400 during day; high orography (> 2000); above model top; only use channels with signal above clouds; surface sensing channels with large difference.*	ch. 2-15 *Exclusions: land for channels 5-8*	ch. 2-7, 11, 12, 14, 15 *Exclusions: clouds or land for ch. 4 and above*	ch. 2-7, 11, 12, 14, 15 *Exclusions: land for ch. 4-7, 11, 14, 15, high terrain for ch. 12, clouds for ch. 3 and above*	ch. 2-7, 11, 12, 14, 15. *Exclusions: clouds, land for ch. 4-7, 11, 14, 15, high terrain for ch. 12*
SSU Radiances	ch. 1,2,3 *All channels bias-corrected.*	ch.1,2,3 *No offset bias correct. for ch.3*	ch. 1,2,3	ch. 1,2,3	ch. 1,2,3 *No offset bias correct. for ch.3*
GEO Radiances	GOES sounder, 5x5 1993-2007, 1x1 2007-	GOES sounder		GOES, METEOSAT, GMS, MTSAT imagers	GOES, METEOSAT, MTSAT imagers
SSMI Retrievals	Ocean surface wind speed	Ocean surface wind speed, Rain rate	Total column water vapor, Snow cover	Snow cover	Total column water vapor in rainy areas
Imager Upper-air winds	GOES, METEOSAT, GMS, MTSAT, MODIS	GOES, METEOSAT, GMS, MODIS	GOES, METEOSAT, GMS, MTSAT, MODIS	GOES, METEOSAT, GMS, MTSAT, MODIS	MTSAT, MODIS
Scatt Ocean surf. wind	ERS, Quikscat, ACATW	ERS, Quikscat	ERS, Quikscat, ASCAT	ERS, Quikscat, ASCAT	ERS, Quikscat
Ozone Retrievals	SBUV V8 Retrievals	SBUV V8 Retrievals	TOMS, OMI (nudging)	TOMS, OMI (nudging)	TOMS, SBUV, GOME, MIPAS, SCIAMACHY, MLS, OMI
Other notable elements	AIRS, IASI, GPS radio occultation, AMSR-E, reprocessed ERS, reprocessed GMS, *AMSU-B NOAA-15*	TMI rain rate, AIRS, *AMSU-B NOAA-15*	Reprocessed GMS winds, AMSR-E Total column water vapor, HIRS NOAA-15 and later not used	Reprocessed winds from GMS/GOES-9/MTSAT (revised) and METEOSAT, Reprocessed radiances from GMS/GOES-9/MTSAT, TMI (from US), AMSR-E (from JAXA), GPS radio occultation, SSMIS, VTPR, HIRS NOAA-15 and later not used	GPS radio occultation, AIRS, SSMI-S, AMSR-E, *HIRS NOAA-18*
Input/checked by	J.Woollen, B.Kistler, D.Kleist, J.Derber	R. Gelaro, M. Bosilovich	S. Kobayashi		P. Poli

AMSU-A channels with surface sensitivity used over land

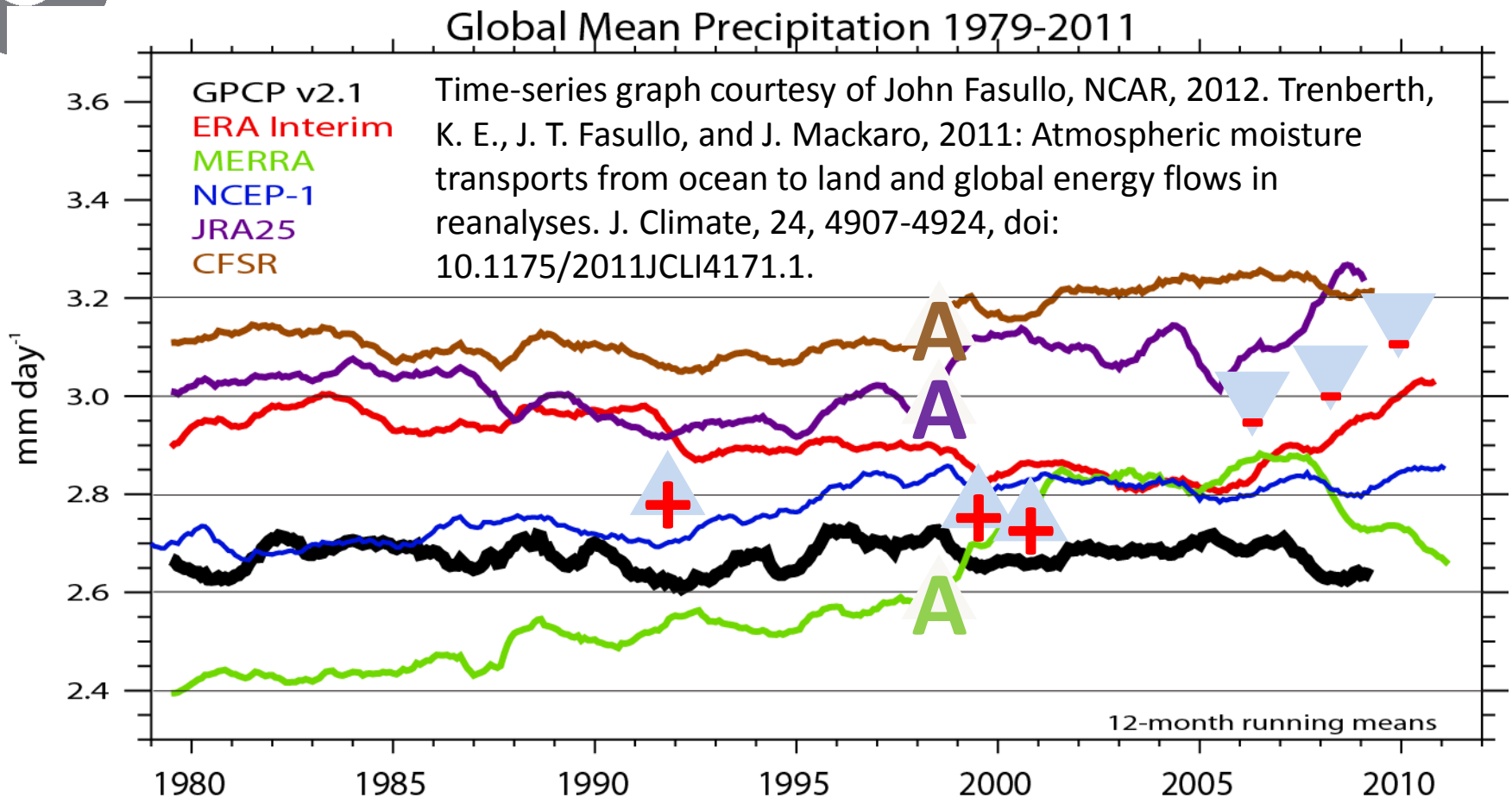
SSM/I 'rain assimilation': Systematic impact in the ECMWF system used in 2006

Material in this table:

1. Gathered first from papers & websites,
2. Then I asked for verification to the experts from each reanalysis – and found I hadn't gotten it always completely right for their reanalyses
3. How difficult can it be for non-expert users to figure out all the nitty gritty details of such differences, and what impact they have?

Table published on https://reanalyses.org/observations/satellite-1

Precipitation in reanalyses and **GPCP**

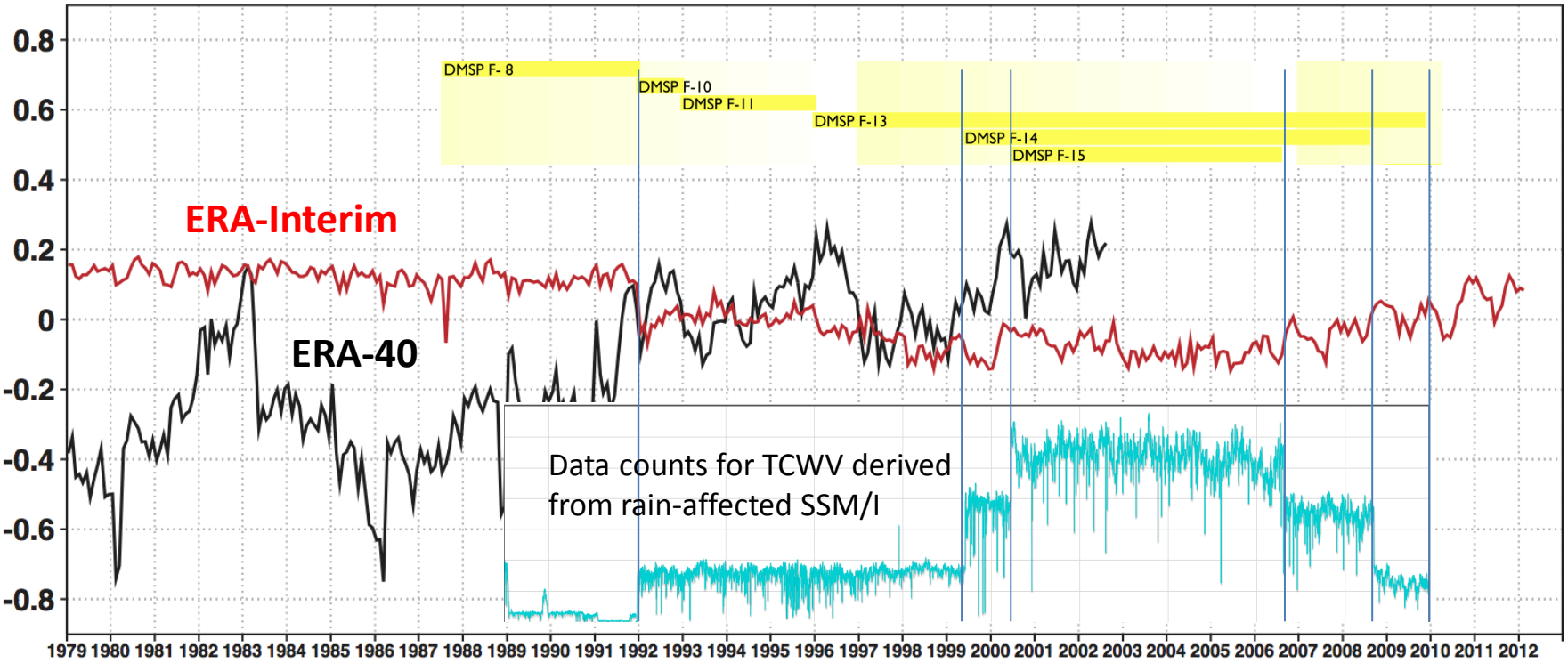


Overlay, selected satellite observing system events:

Addition + or removal - of one SSM/I satellite in ERA-Interim

A Transition from TOVS to ATOVS

Shifts in ERA-Interim precipitation explained



- Due to assimilation of rain-affected radiances from SSM/I over oceans
- Experimentally verified, and now fully understood *(Geer et al. 2008)*
- Effect scales non-linearly with the number of assimilated rain-affected SSM/I data

Differences in observation input and selection…

- Found to explain ‘large’ differences between reanalyses…
- What do we mean by large?
  - Large-scale differences
  - Systematic differences
  - Of a magnitude significant enough to either dominate low-frequency variability or induce a large temporal break
- Such ‘large’ differences raise (appropriately) questions whether reanalyses can be used (‘at all’) to produce a reasonable estimate for the parameters where they disagree
- Increasing the confidence/credibility and use of reanalyses mandates that we open our books about observation usage
- Documentation in the form of tables and plots isn’t enough – especially for non-experts to gather the information from various sources
- Furthermore, an automated exploitation of all this information requires “hard numbers” (with dates)

2 approaches are possible to explain what observations we use

- Top-down approach (high-level summaries)
  - Lists of observations
  - Timelines
  - Plots
    - This is what we have been doing so far
    - Obviously insufficient for automated exploitation – e.g., no ‘diff’ tool…
- Bottom-up approach
  - Starting from each observation, describe where/what/when/used/how much etc…

Our second responses

"-- You can't possibly be serious, the observation feedback represents too much data to exchange"

Size isn't the issue

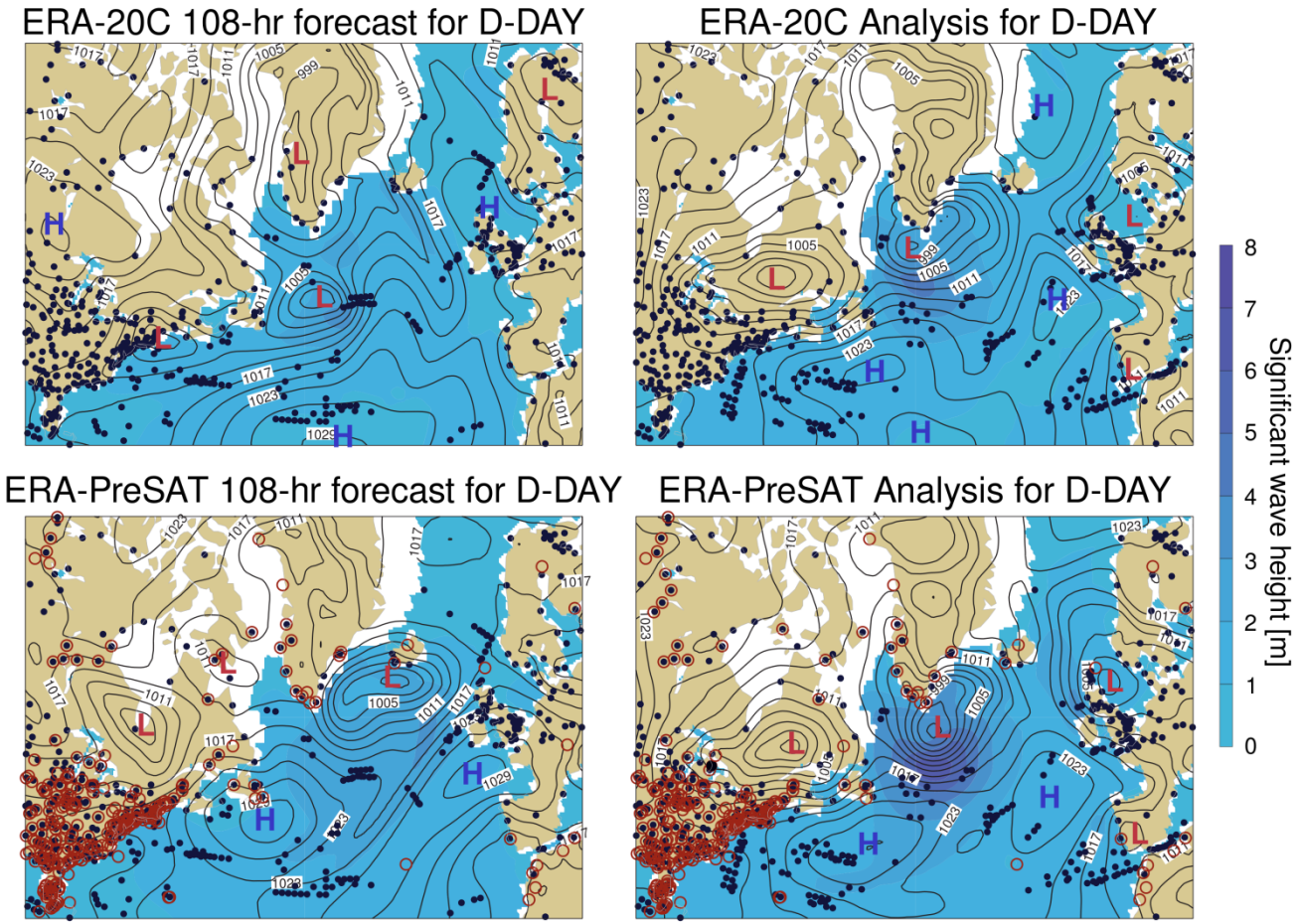
- Observation feedback data for old dates represent a small percentage of reanalysis datasets (fields) volumes
  - The growth in computing allows for always greater model resolution – this applies also (in recent years) to recent observation data, so NWP sees basically a parallel growth in resolution/observation data counts
  - But in a reanalysis we apply state-of-the-art resolution to datasets that increasingly go further back in the past
- The observations feedback is hence only a fraction of the data volume generated by a reanalysis
  - This fraction actually probably decreases for old time periods, as we will not go back and acquire so many new observations for old dates (except for data rescue activities)

Other applications of the reanalysis observation feedback

- Improve the observation record itself
- Diagnose assimilation performance
- Refine observation quality controls
- Identify sources of systematic errors (which obs. system is biased w.r.t. others for a given parameter/location/time period)
- Refine observation error estimates (sigma_o)
- Support reprocessing
- Support historical data rescue

- All 'advanced' users have been motivated enough to find and exploit the feedback; they are hence maybe not the prime motivation for this meeting.
  - However, they will still surely appreciate if we collectively pool our resources to serve the feedback in a ~common way

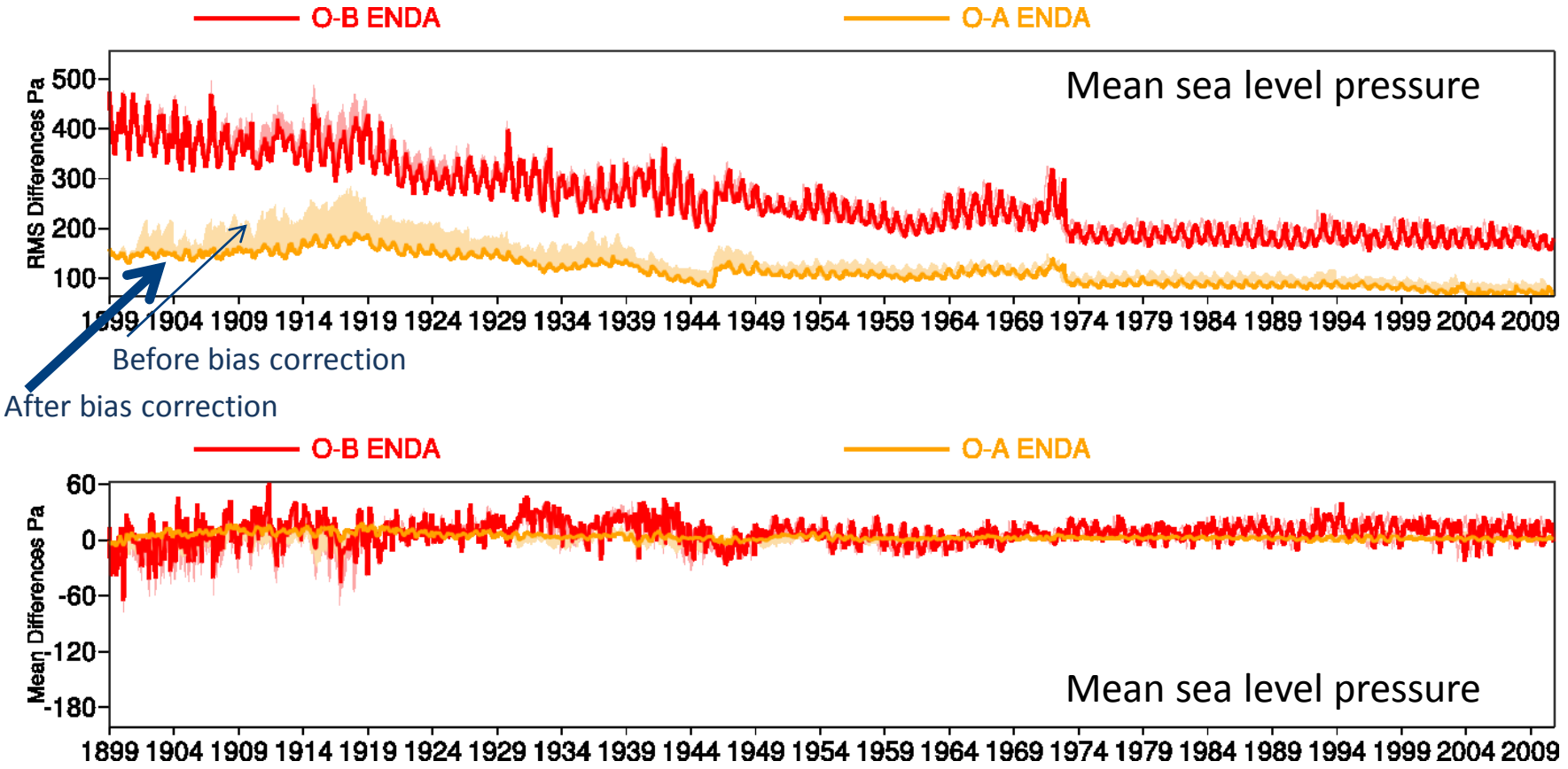
Visualization of observation feedback: where are the data?



Animations of the observation coverage and departures

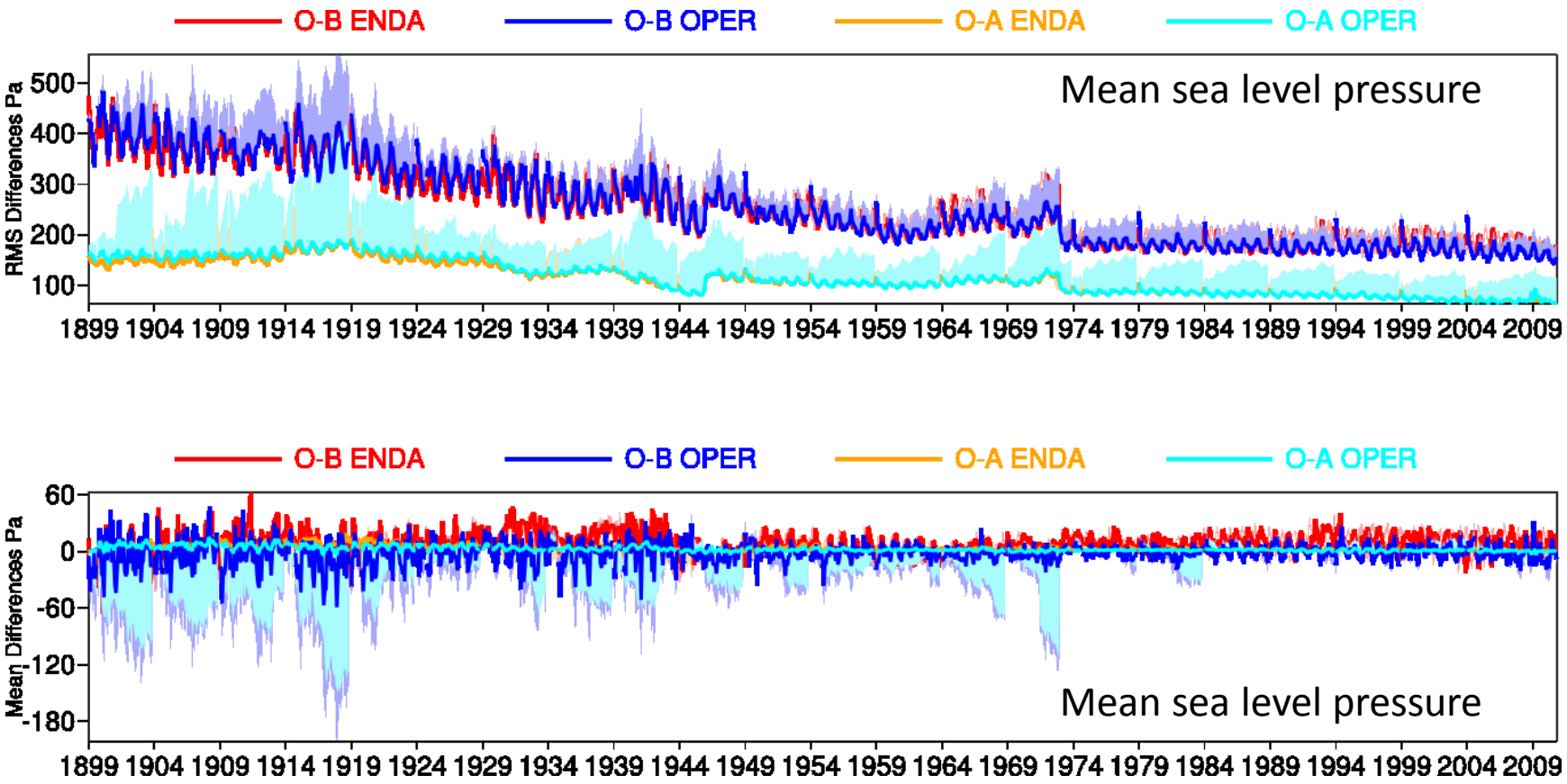
- http://www.youtube.com/playlist?list=PLNgQfkVXCl8gQKXkXEUBEygfSreSwNY1Z

Diagnosing data assimilation performance in ERA-20C ensemble (ENDA)



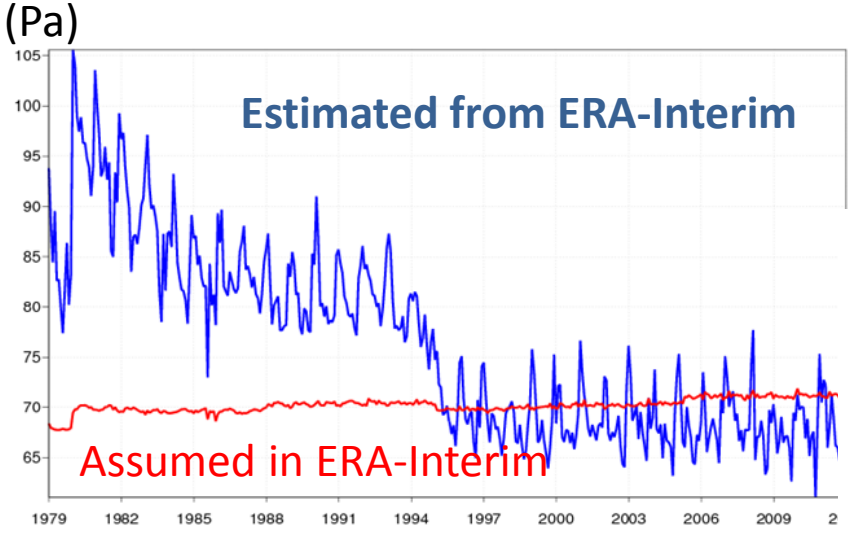
Shading shows work done by observation bias correction instead of the work done by the meteorological analysis to adjust the whole atmospheric state to fit the data

Same but with also ERA-20C re-run (OPER)



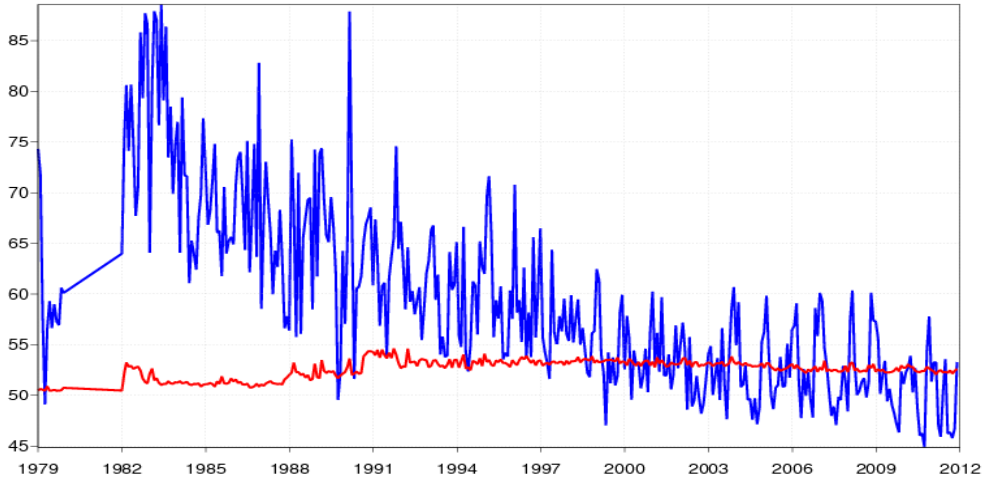
Surface pressure observation error estimates from the ERA-Interim feedback

- Land SYNOP



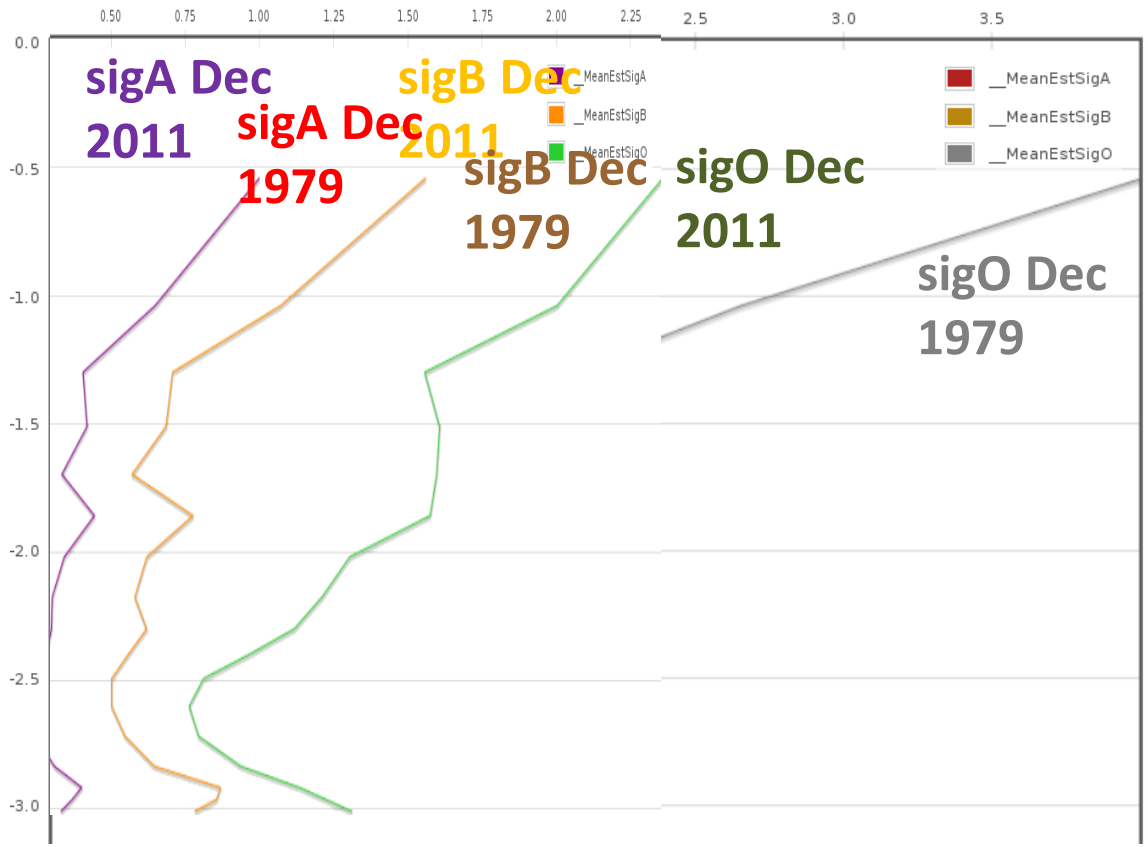
$$\sigma_0 = \sqrt{\frac{1}{N}\sum_{i\in 1\ldots N}\left(y_i^0 - h\left(x^b\right)\right)\left(y^0 - h\left(x^a\right)\right)}$$

- Automatic land SYNOP only



Using the method of Desroziers et al. (2005): Diagnosis of observation, background and analysis-error statistics in observation space. Q.J.R. Meteorol. Soc., 131: 3385–3396. doi: 10.1256/qj.05.108

Desroziers diagnostics applied to Radiosonde temperatures from ERA-Interim

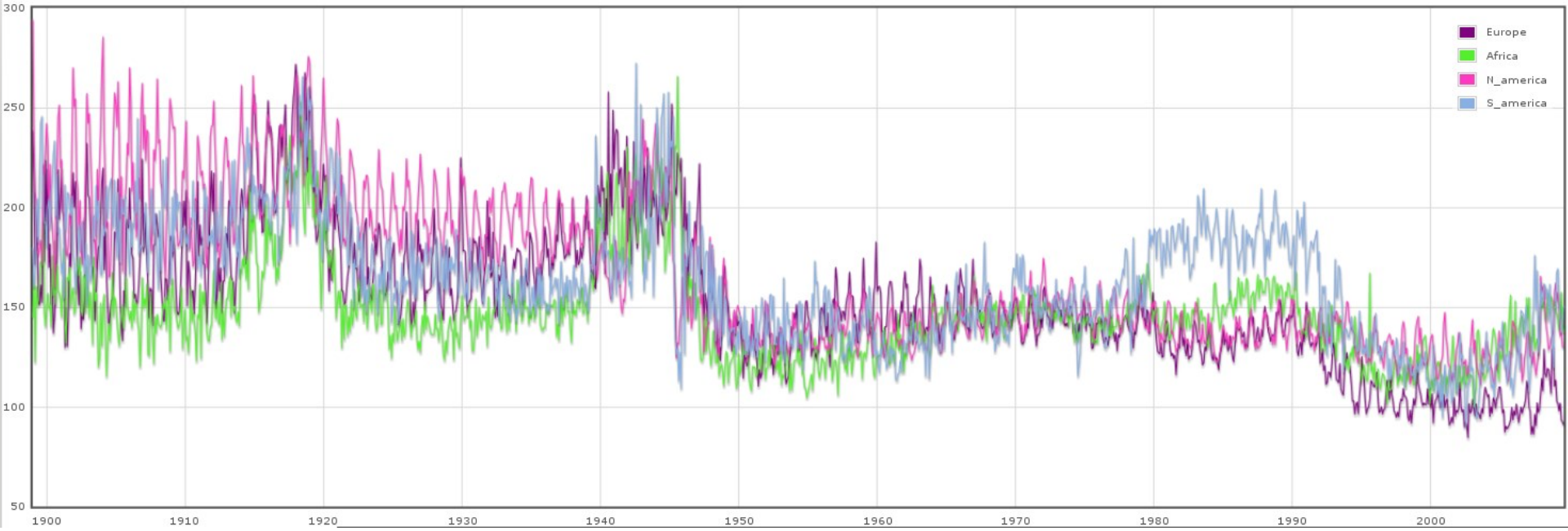


Radiosonde equipment has changed quite a bit in 30 years…

Probably shouldn't have assumed that obs. errors are constant in time or across RS types

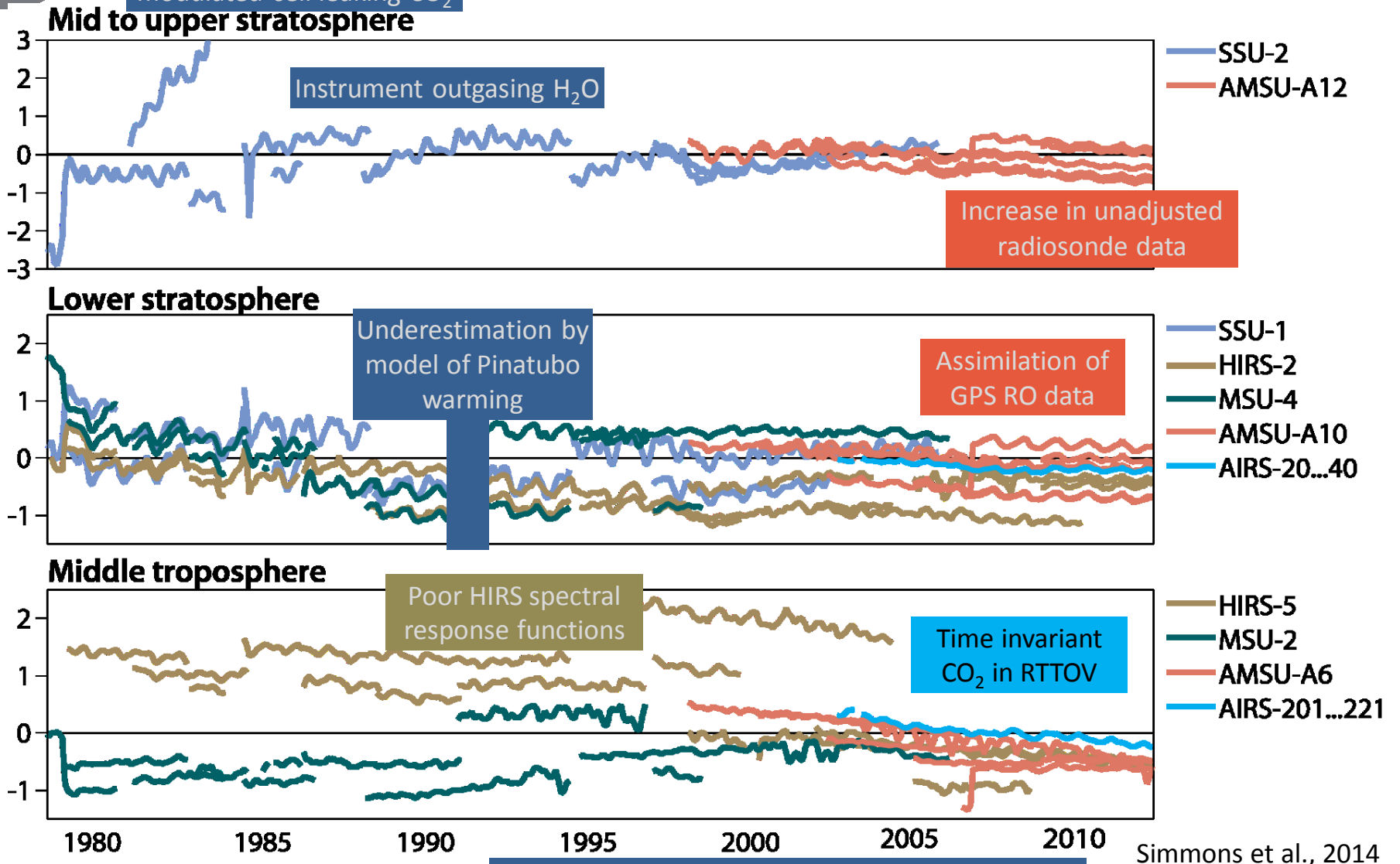
How reliable are such observation error estimates?

- Quick answer: We can't really tell right away!
- Regional disparities larger for older dates
- Some common features beg for explanations



For ship reporting observations of mean sea level pressure, from ICOADS 2.5.1, after assimilation in ERA-20C ensemble reanalysis

Example of feedback interpretation: adjustments in ERA-Interim



Simmons et al., 2014

Frequency shifts of local oscillator w.r.t pre-launch specifications

Solar heating change caused by orbital drift (insufficient non-linearity calibration correction)

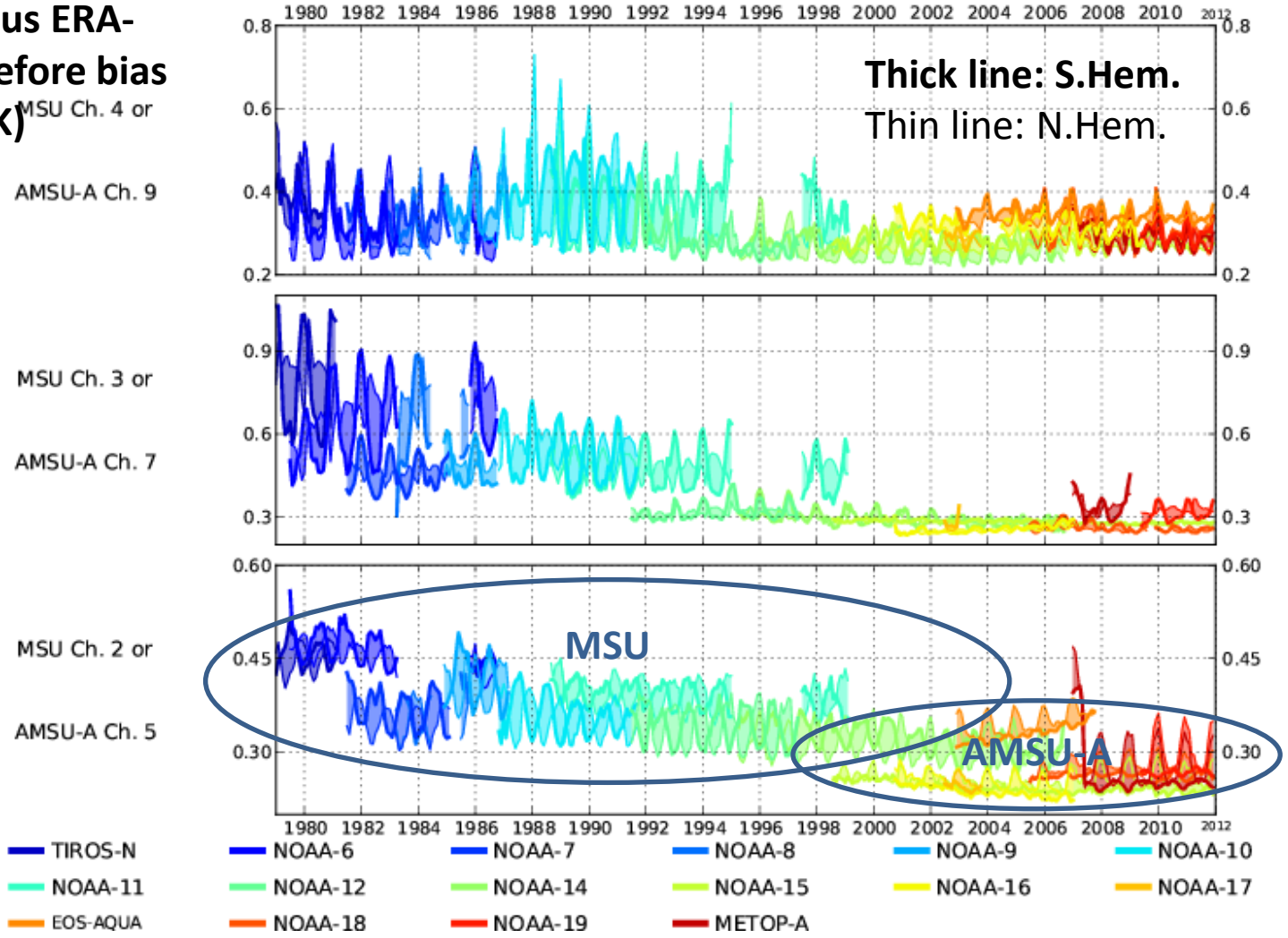
Observation feedback: witness of satellite instrument improvement

Stdev(Obs minus ERA-Interim bg.), before bias correction (in K)

a.k.a TLS
~90 hPa

a.k.a TTS
~300 hPa

a.k.a TMT
~600 hPa



Conclusions

- Reanalysis observation feedback = a mine of information
- Is it fair to say that right now we pretty much sit on it and only let few others use it? I would hope not…
  - Disclaimer: This is not because we all run away to write lots of papers with this feedback!
  - NOAA/CIRES 20CR is an exception there, having opted early on from the start for an open-access feedback
  - ERA-20C follows this example
    - ERA-20C feedback to come out any time 'soon'
- We all have limited time to look at this feedback in details
- Surely making it available should help us:
  - More eyes to look at it and tell us about missing QC, etc…
  - Promote investment to improve the past observation record where & when it is most needed