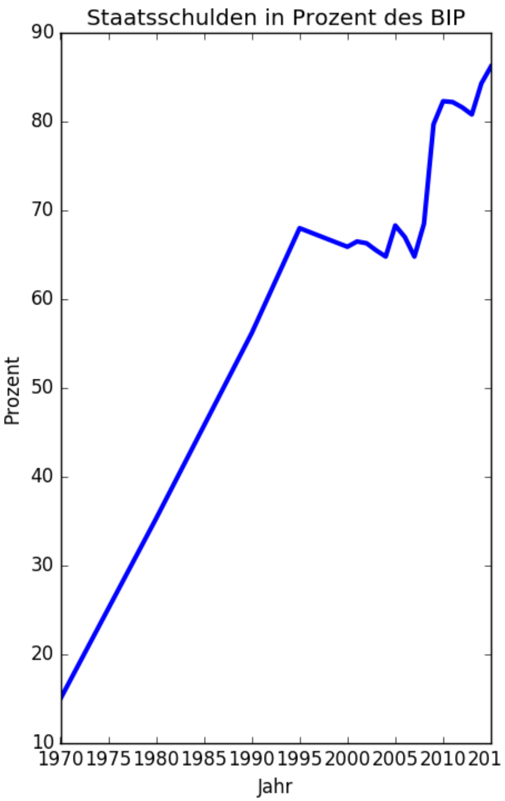
Reducing uncertainties in reanalysis input and products

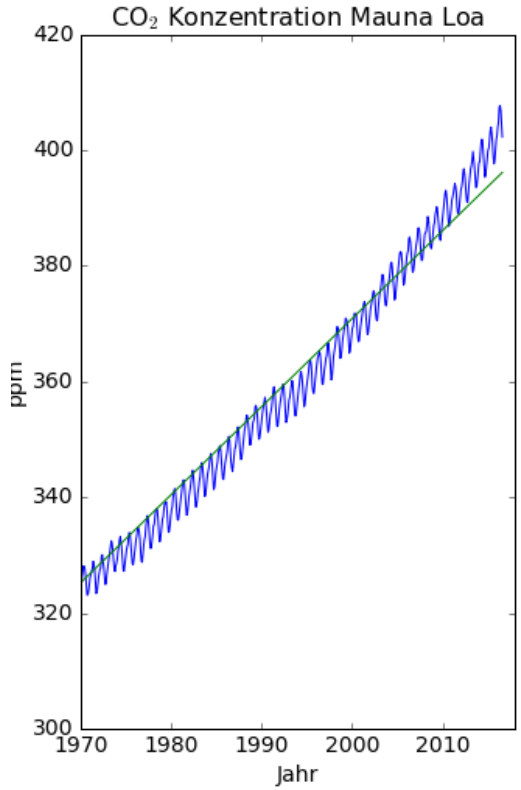
L Haimberger
University of Vienna, Austria

With thanks to many collegues in ERA-CLIM2 consortium

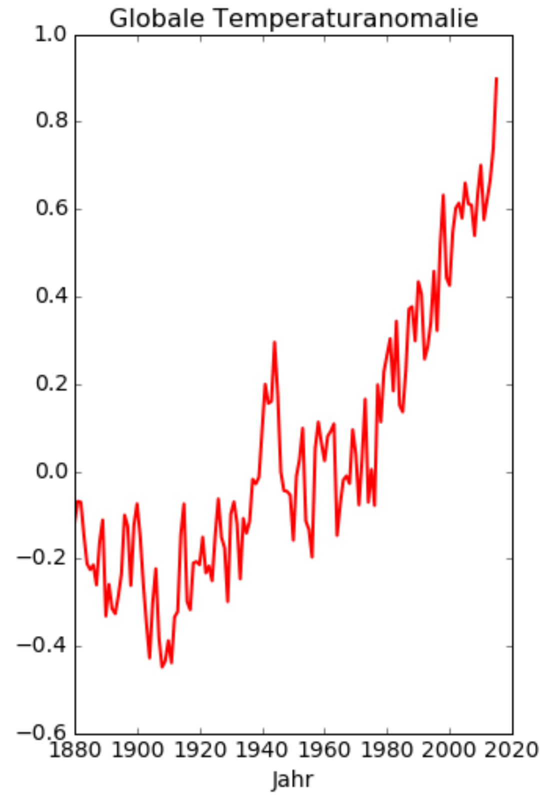
How large is the deficit (surplus)?



Debt problem



CO_2 Problem



Energy/Climate problem

Long term loss/accumulation causes problems ...

Climate Reanalyses

- Help assessing magnitude and cause of climate change
- Combine climate model with observations
  - Dynamic Data assimilation
  - Role of systematic errors
- Comparison with independent data

Fundamental Climate Budget

$$C\frac{\mathrm{d}T}{\mathrm{d}t}+\overbrace{-\frac{S}{4}(1-\alpha)}^{F_{\mathrm{Sun}}}+\overbrace{\varepsilon\sigma T^4}^{F_{\mathrm{Earth}}}=0$$

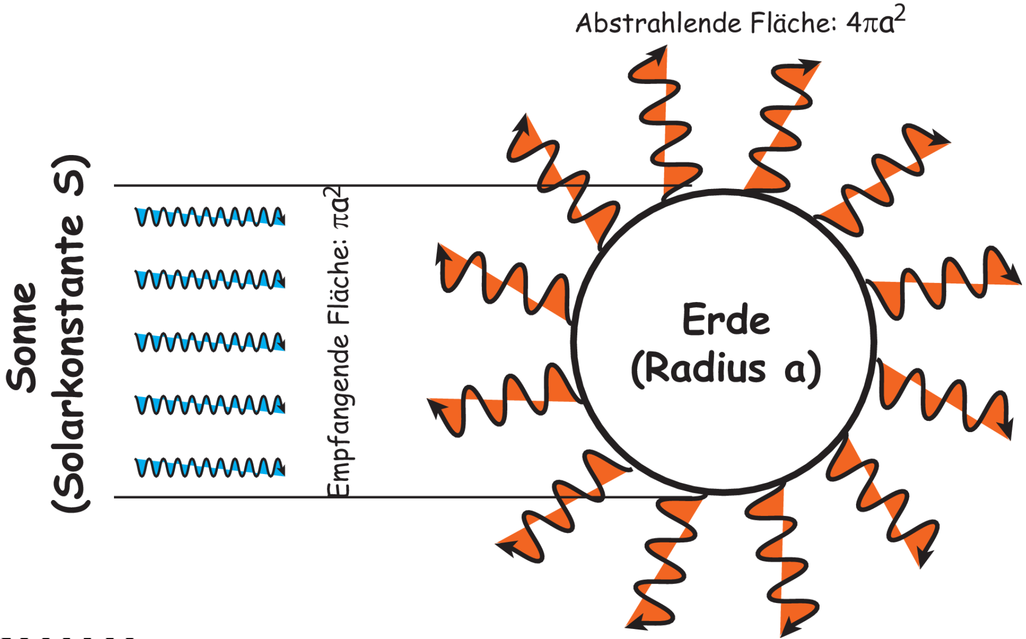
C = Heat Capacity,

S = Solar Constant,

α = Planetary Albedo,

ε = Emissivity (<1),

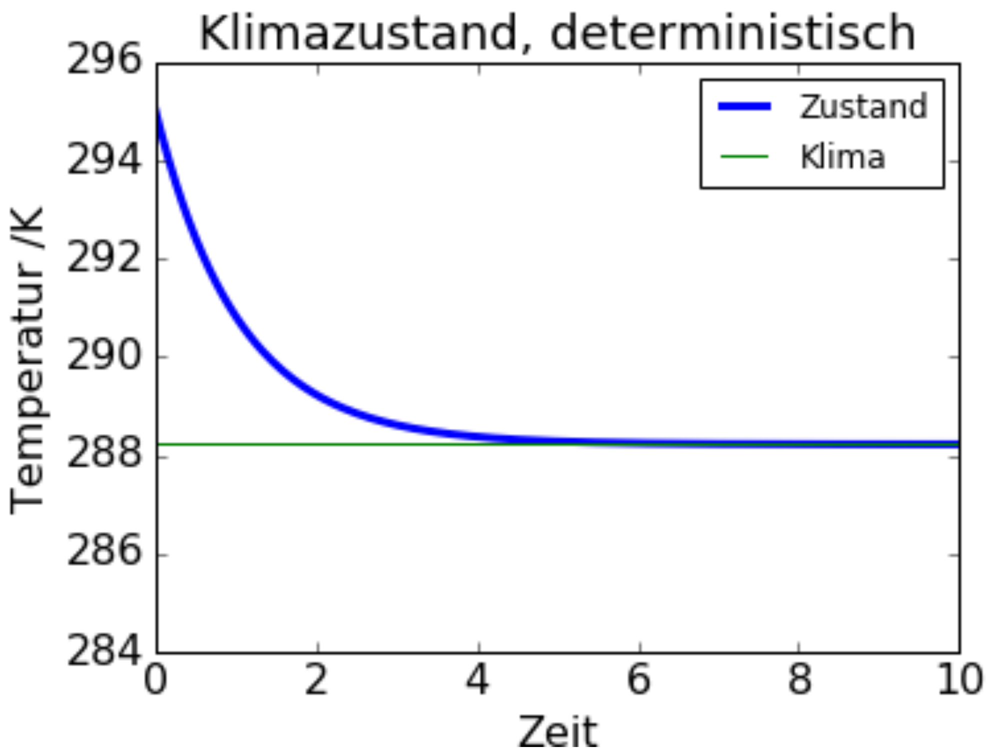
σ = S. Boltzmann cons



Deterministic Energy Balance Model (EBM)

$$C\frac{\mathrm{d}T}{\mathrm{d}t} = \frac{S}{4}(1-\alpha) - \varepsilon\sigma T^4$$

- "State" $T(t)$
- Constant Climate
- Departures quickly damped
- C determines damping rate



EBM with random component

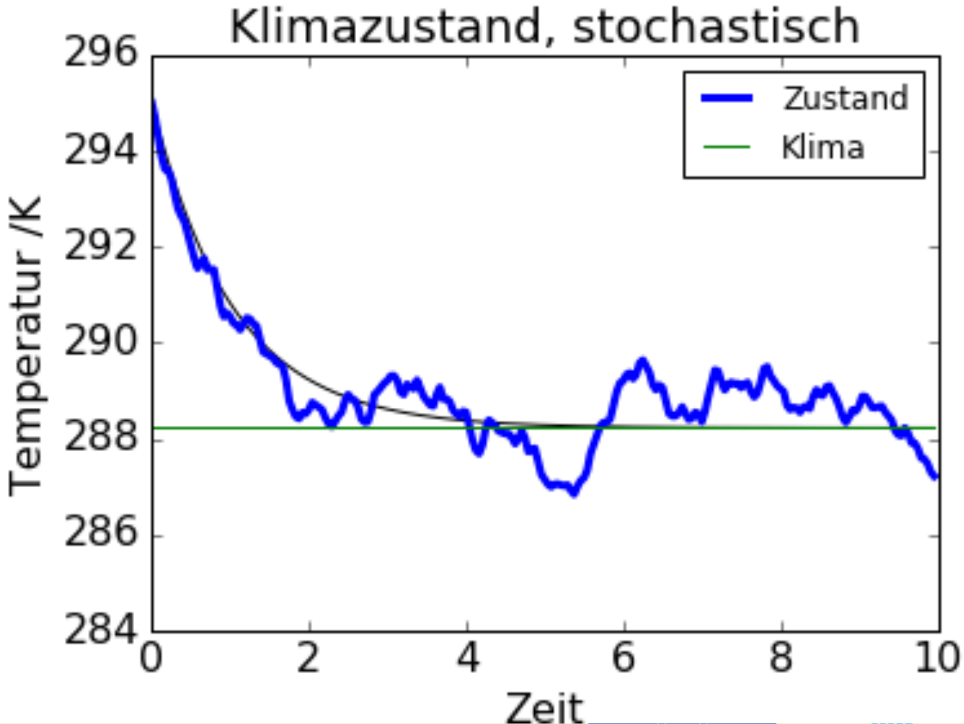
$$C\frac{\mathrm{d}T}{\mathrm{d}t} = \frac{S}{4}[1 - \underbrace{(\alpha_o + \boxed{Z\Delta\alpha)}}_{\alpha}] - \sigma T^4$$

Z=Random Variable – Clouds?

Modell is now stochastic...

... as is climate!

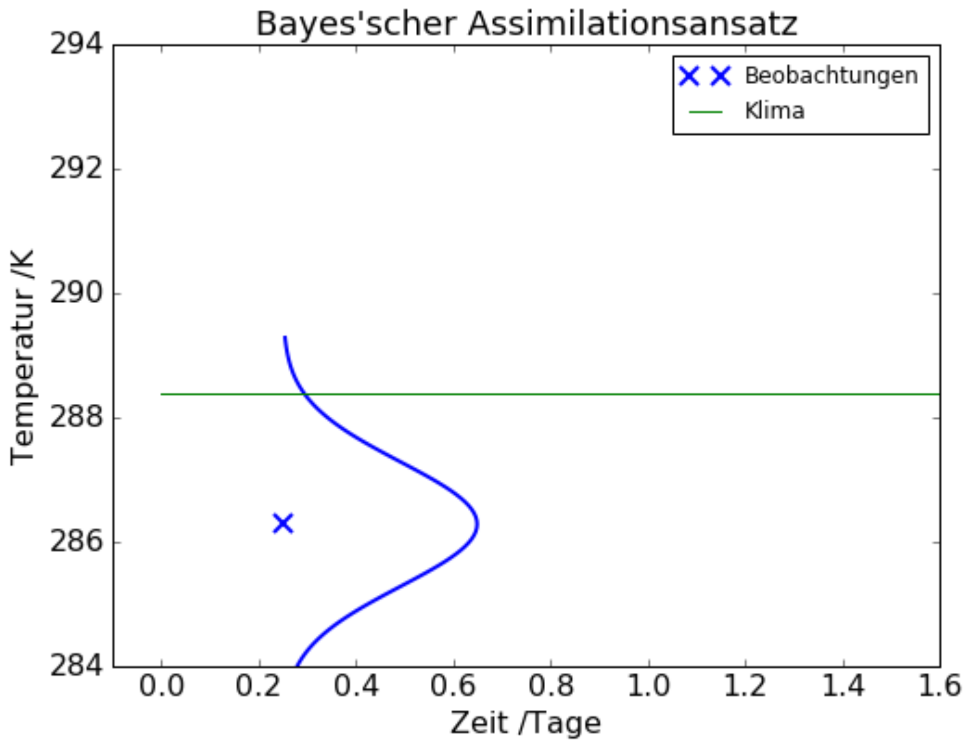
A main cause of natural climate variability



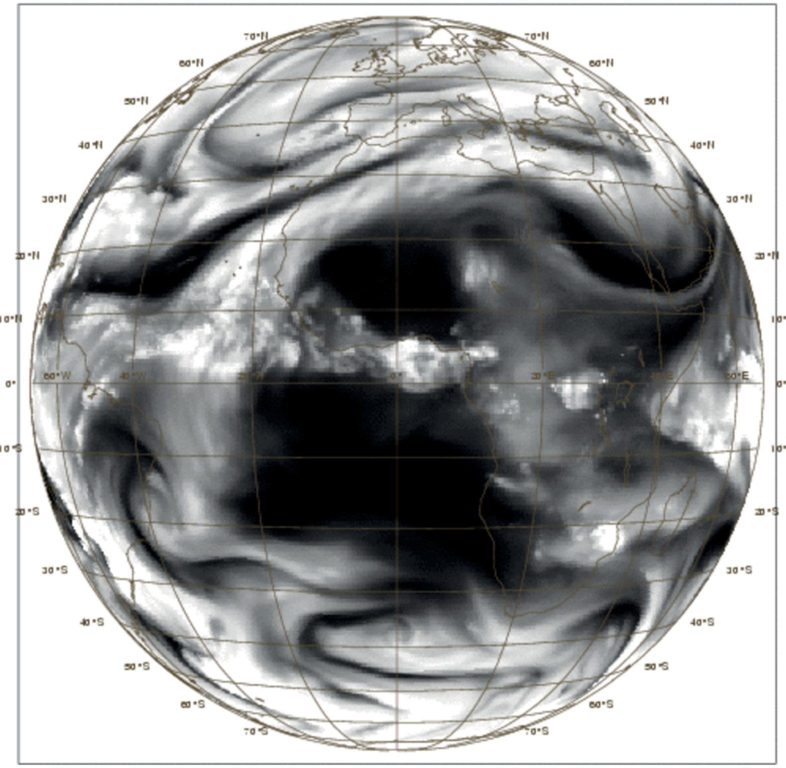
Climate Observations

$$Y = H(T) + \boxed{Z\Delta Y}$$

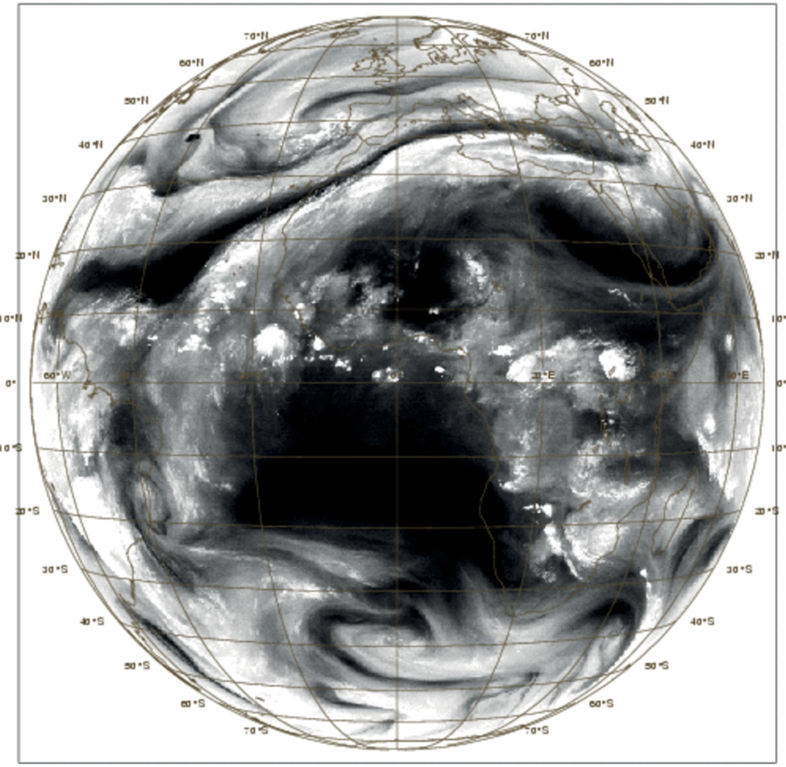
- Observations Y can be similar to state variables but also very different e.g. radiance data
- An observation operator H estimates what should be measured given the current state
- There are also random errors $Z\Delta Y$
- $\rightarrow$ p.d.f. of observations
- How do I get the best estimate for climate state T?



Observation operator allows comparison of model state with satellite picture



Calculated from climate state **H(X)**

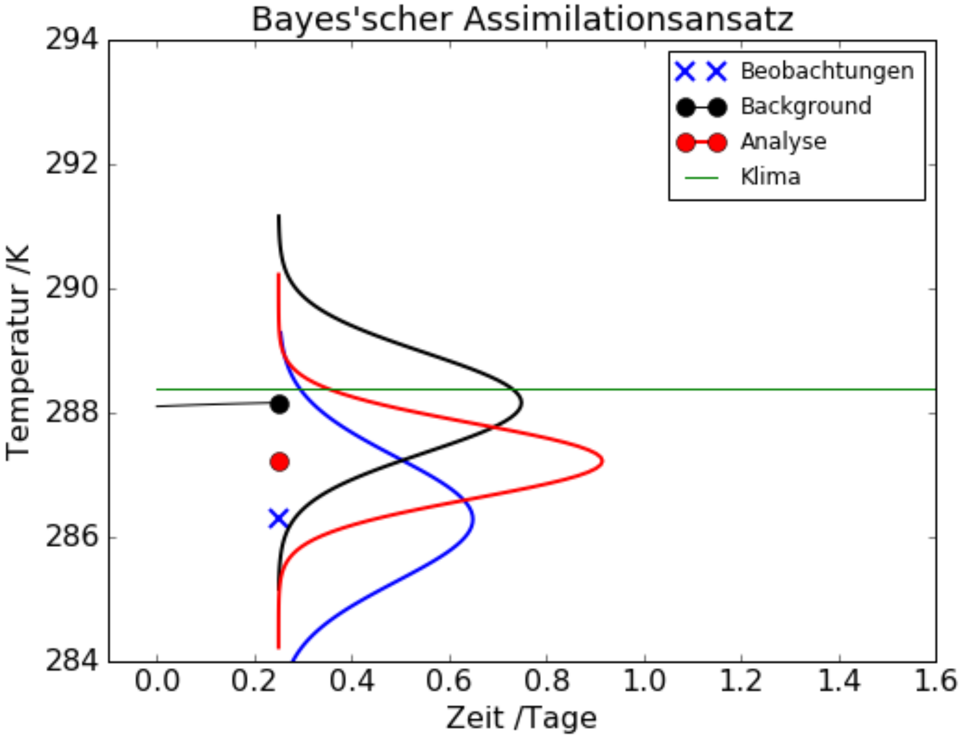


Observed by satellite **Y**

Include prior knowledge - Bayes' theorem

- Climate state vaguely known (e.g. climatology) - prior knowledge)
- or from a forecast
- Optimal combination of prior knowledge ("Background" T_b), observation (Y) yields analysis
- Bayes' theorem is the basis:

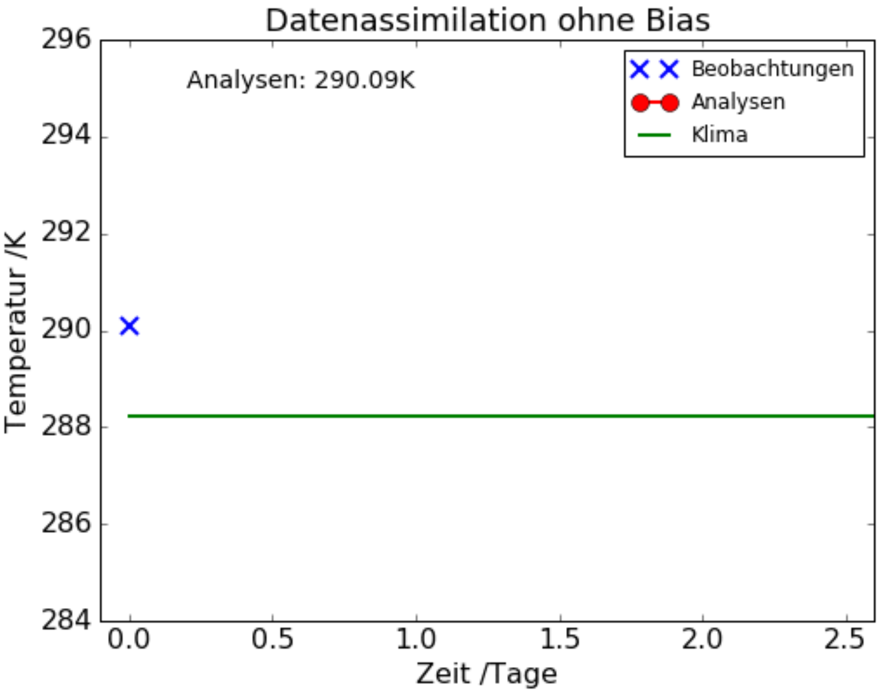
$$\text{p.d.f}(T|Y) = \frac{\text{p.d.f}(T) * \text{p.d.f}(Y}{\text{p.d.f(Y)}}$$



Data assimilation with a deterministic EBM

$$C\frac{\mathrm{d}T}{\mathrm{d}t} = \left[\frac{S}{4}(1-\alpha) - \varepsilon\sigma T^4\right] + \boxed{K[Y - H(T_b)]}$$

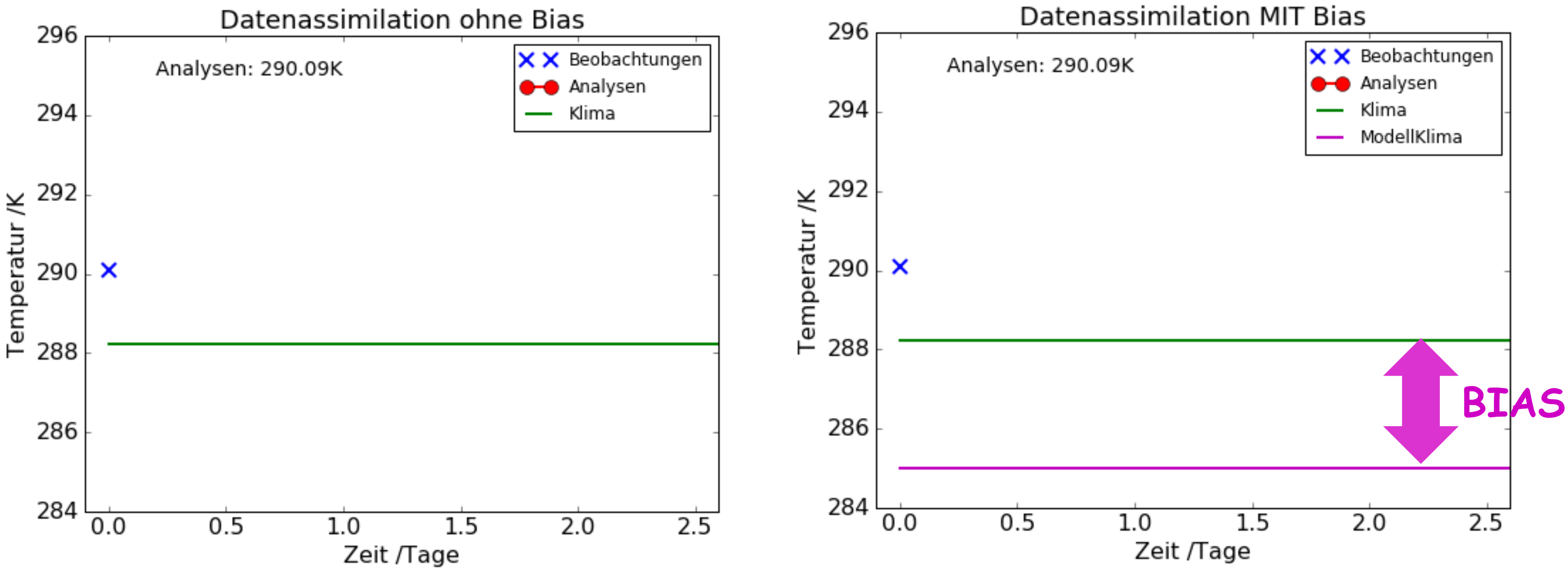
- T_b=Background, K ist a filtering operator.
- Model "assimilates" Observations
- $\rightarrow$ gets stochastic!!
- Best estimates: "Analyses"



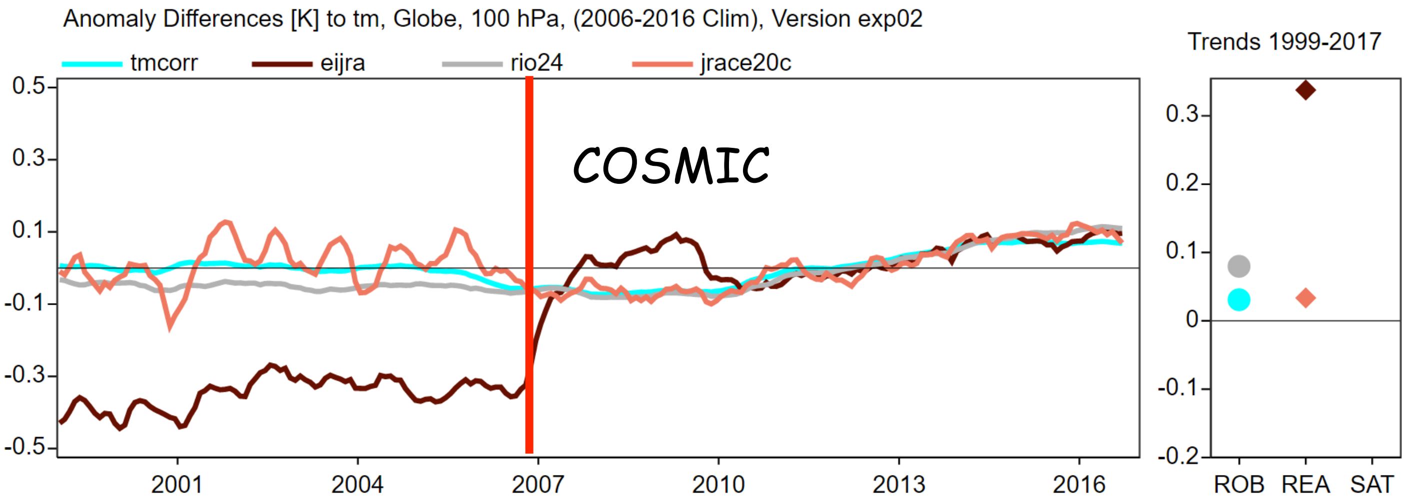
Datenassimilation with Bias

$$Y - H(T_b) \neq 0 \text{ in the mean}$$

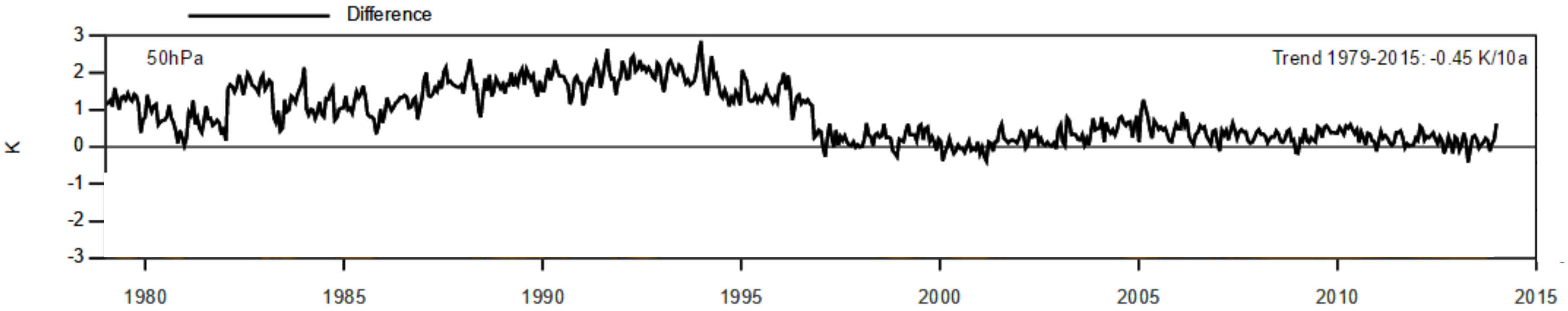
- "Bias": inconsistency between "Model Climate" und "Observation Climate".
- Y values to high?? ε in Model too large??

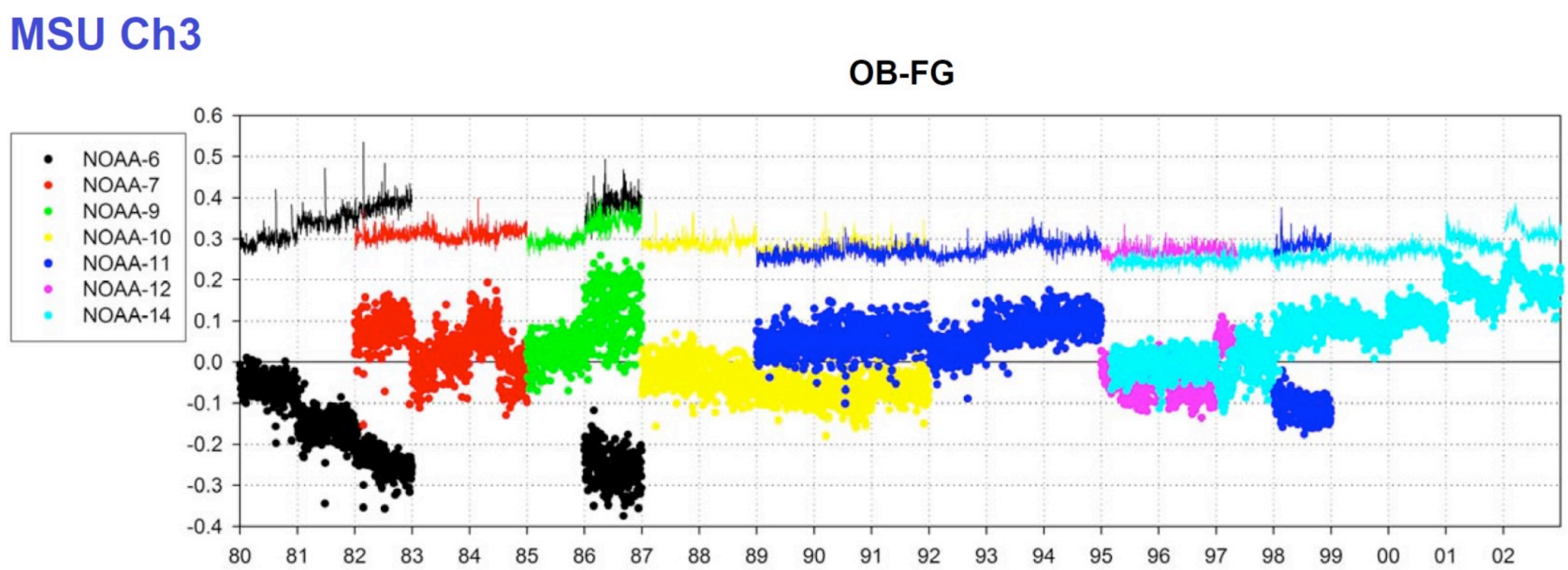


A very real problem: Uncertainty due to model bias



Biases in observations relative to reanalysis background

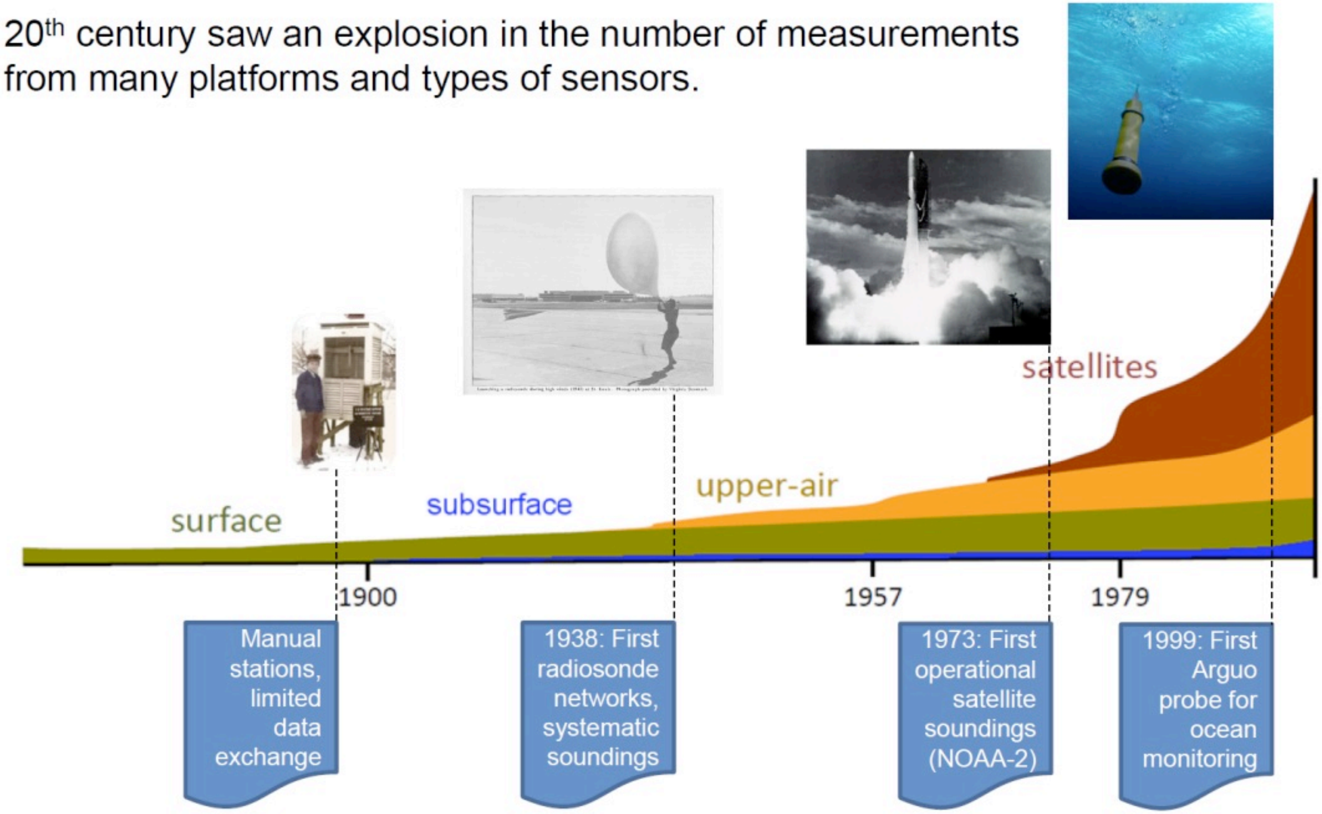




Dee and Uppala 2009

Evolution of the observing system

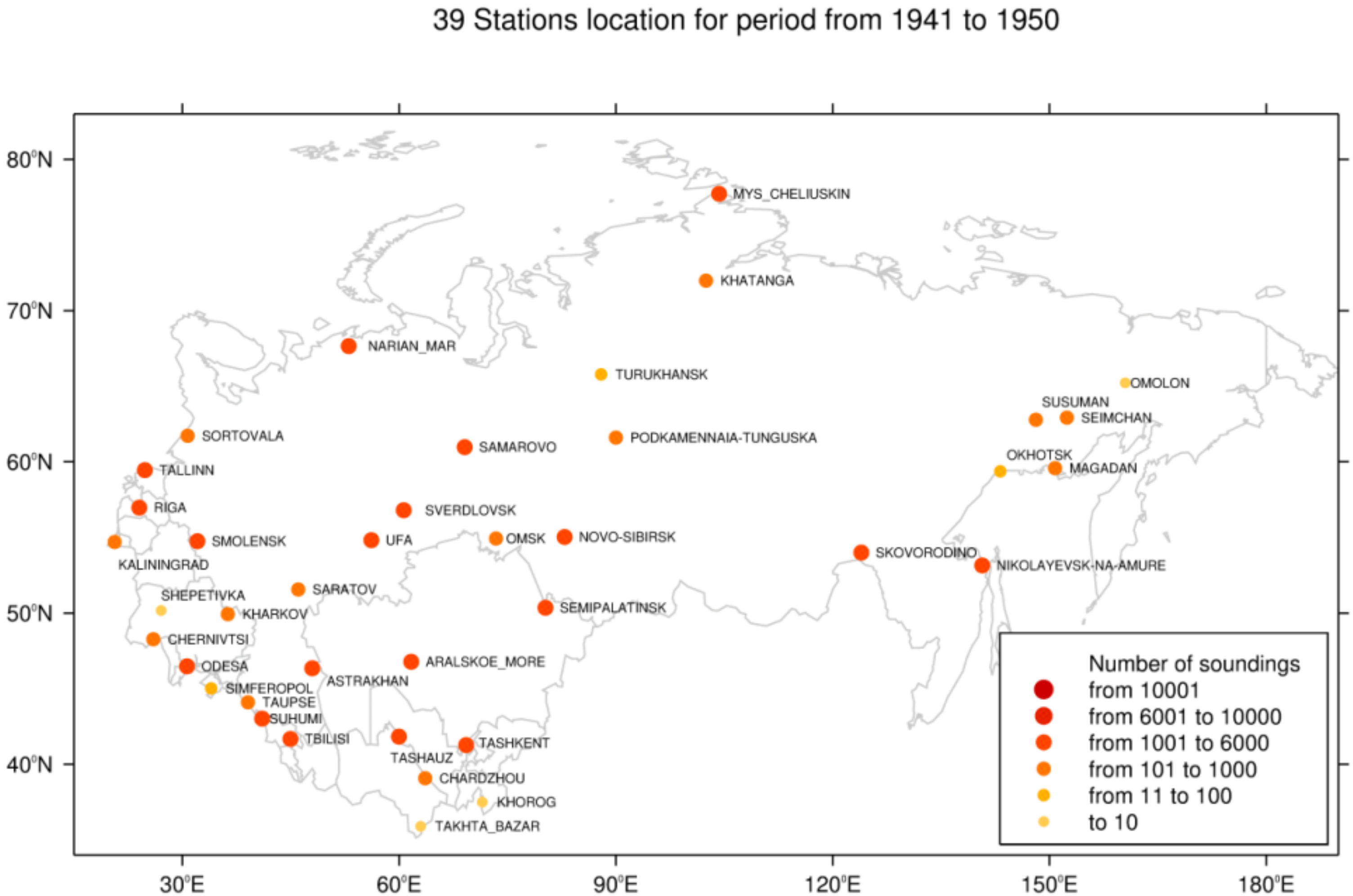
20th century saw an explosion in the number of measurements from many platforms and types of sensors.



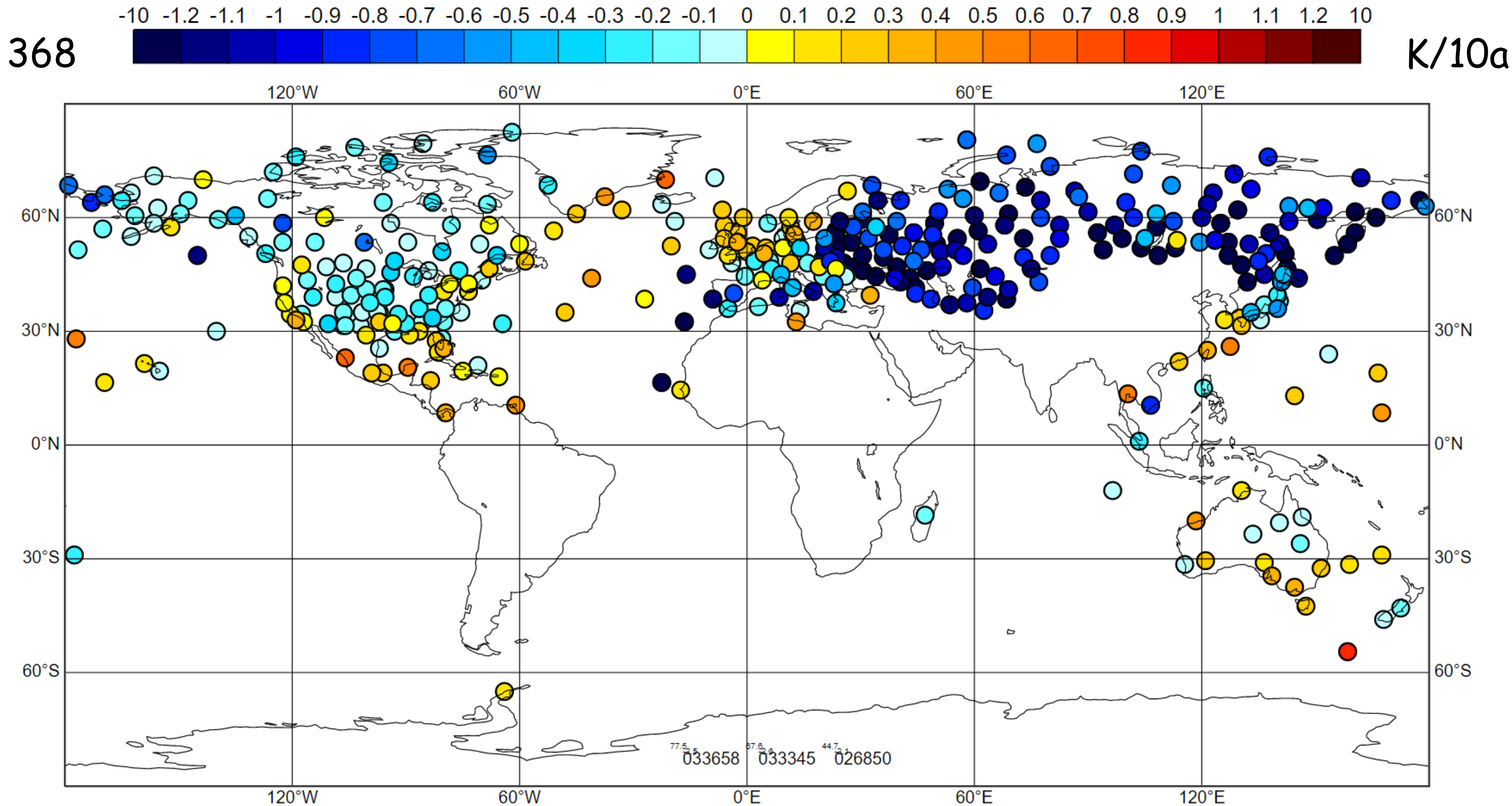
$I(x)=-_a\log(p(x))$ C. Shannon

Events are most valuable if they are unlikely.

Rescuing early upper air data

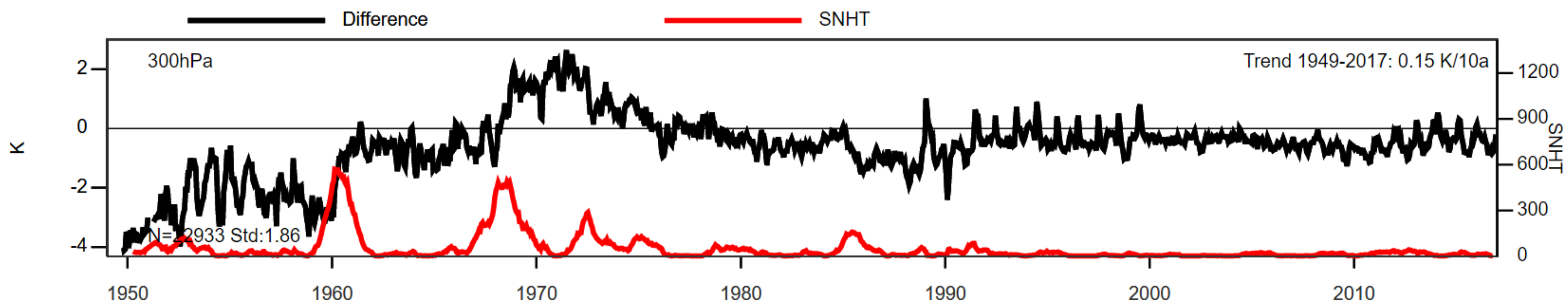


T- Trends from radiosondes, 1954-1974, 300 hPa

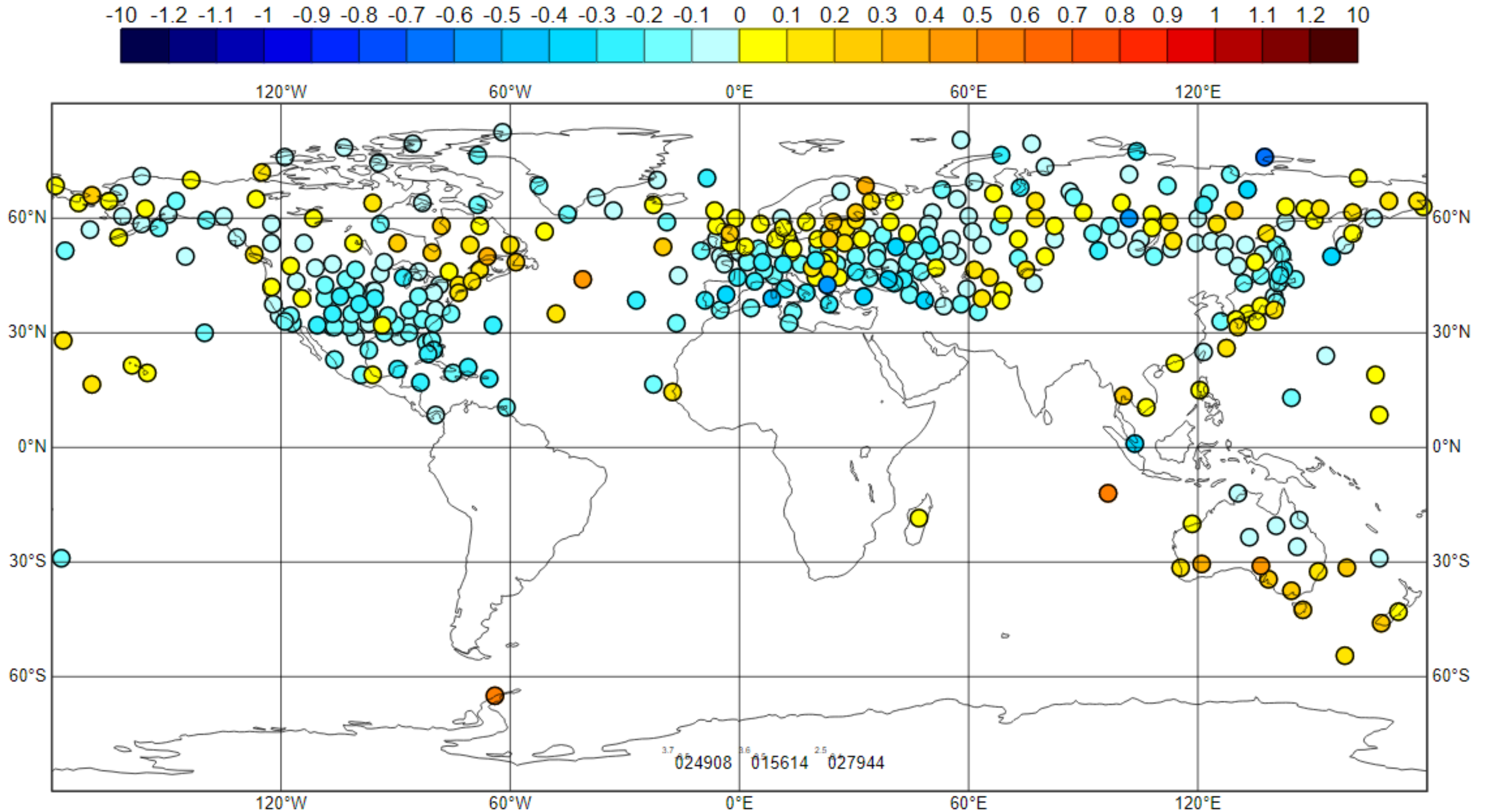


Radiosonde bias correction

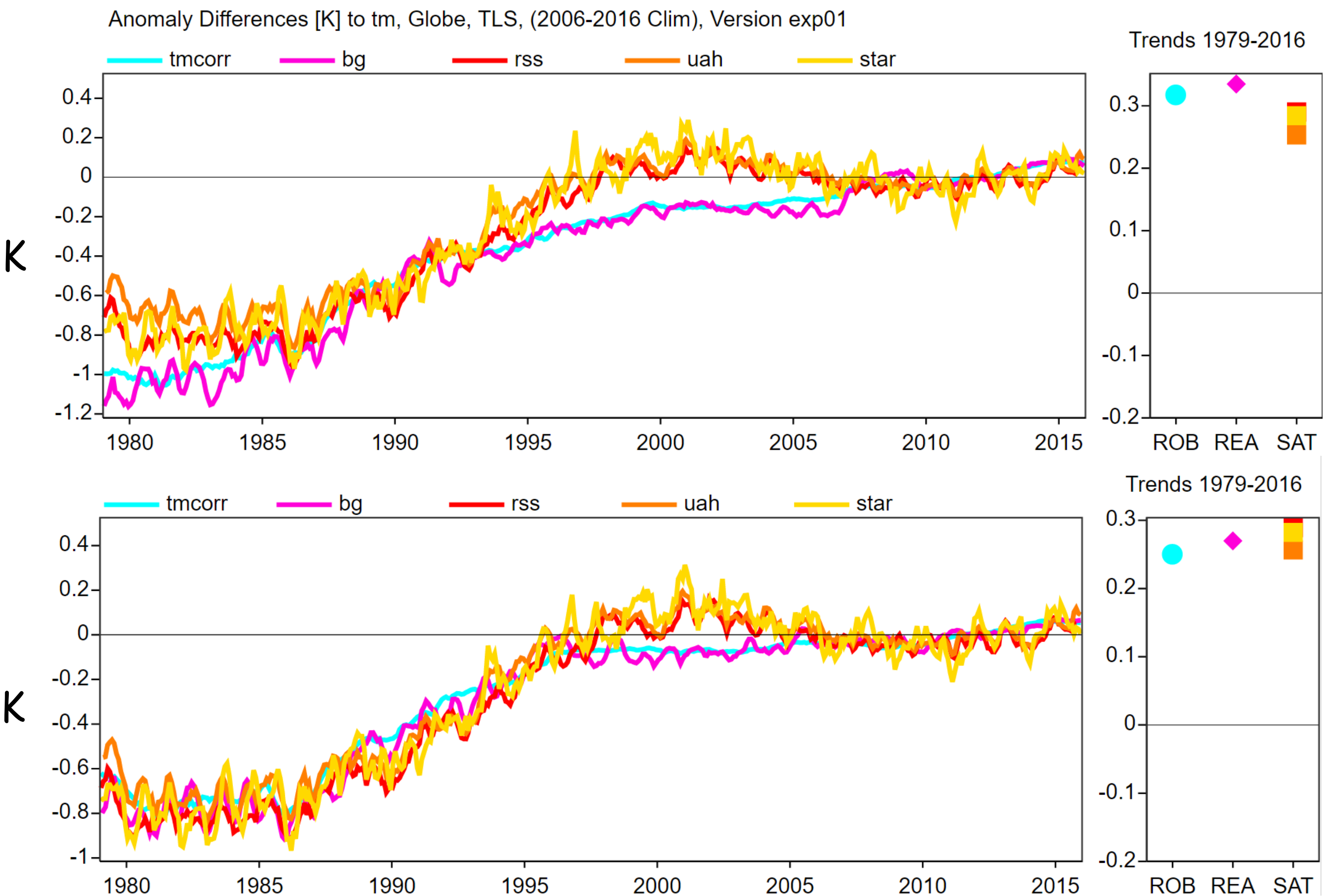
- Detect jumps in radiosonde temperature time series by comparing with reanalysis reference time series
- Automatic break detection with Statistical Test
- Adjustment to reanalysis reference or reference from neighbouring radiosondes



Temperature Trends from adjusted data 1954-1974



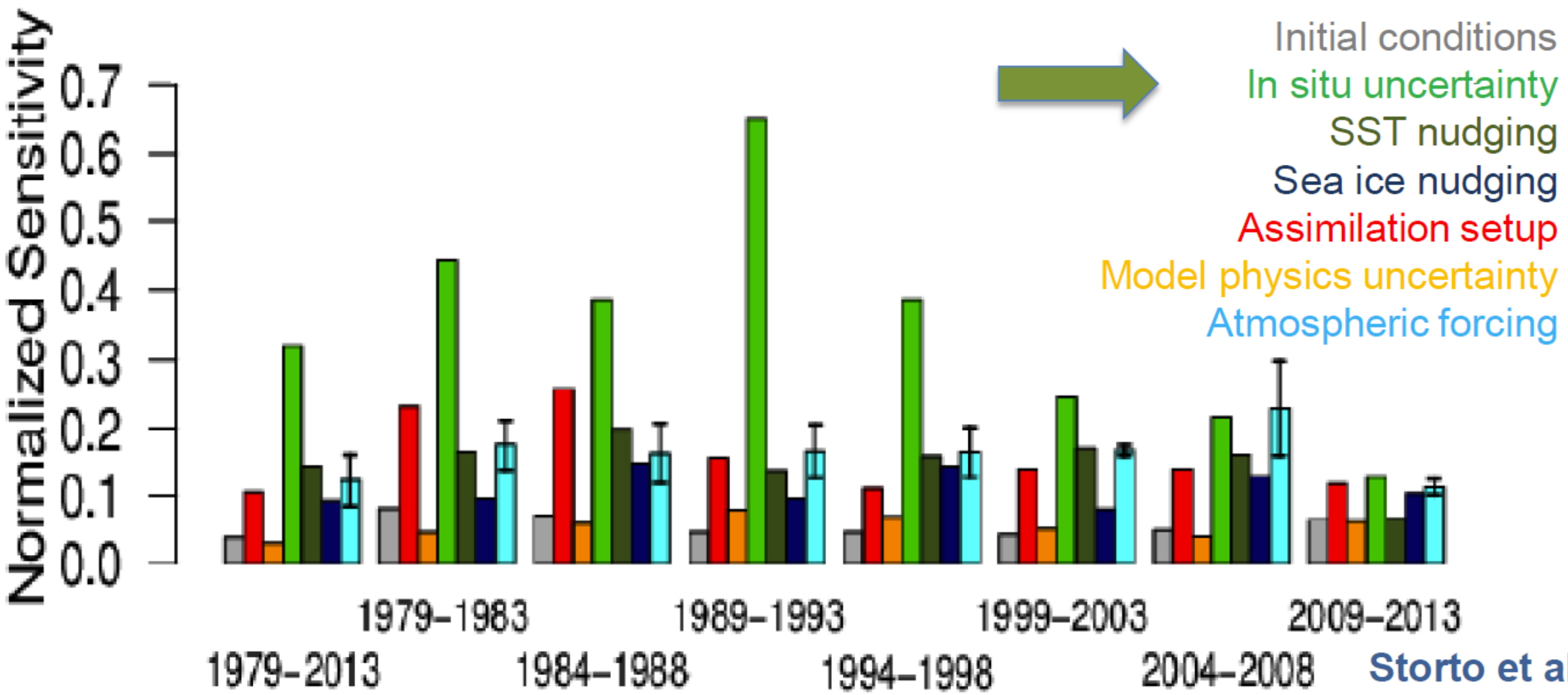
Comparisons for Satellite Era: MSU TLS



ERAI

JRA55

GOHC anomaly sensitivity to reanalysis components & atmos. forcing

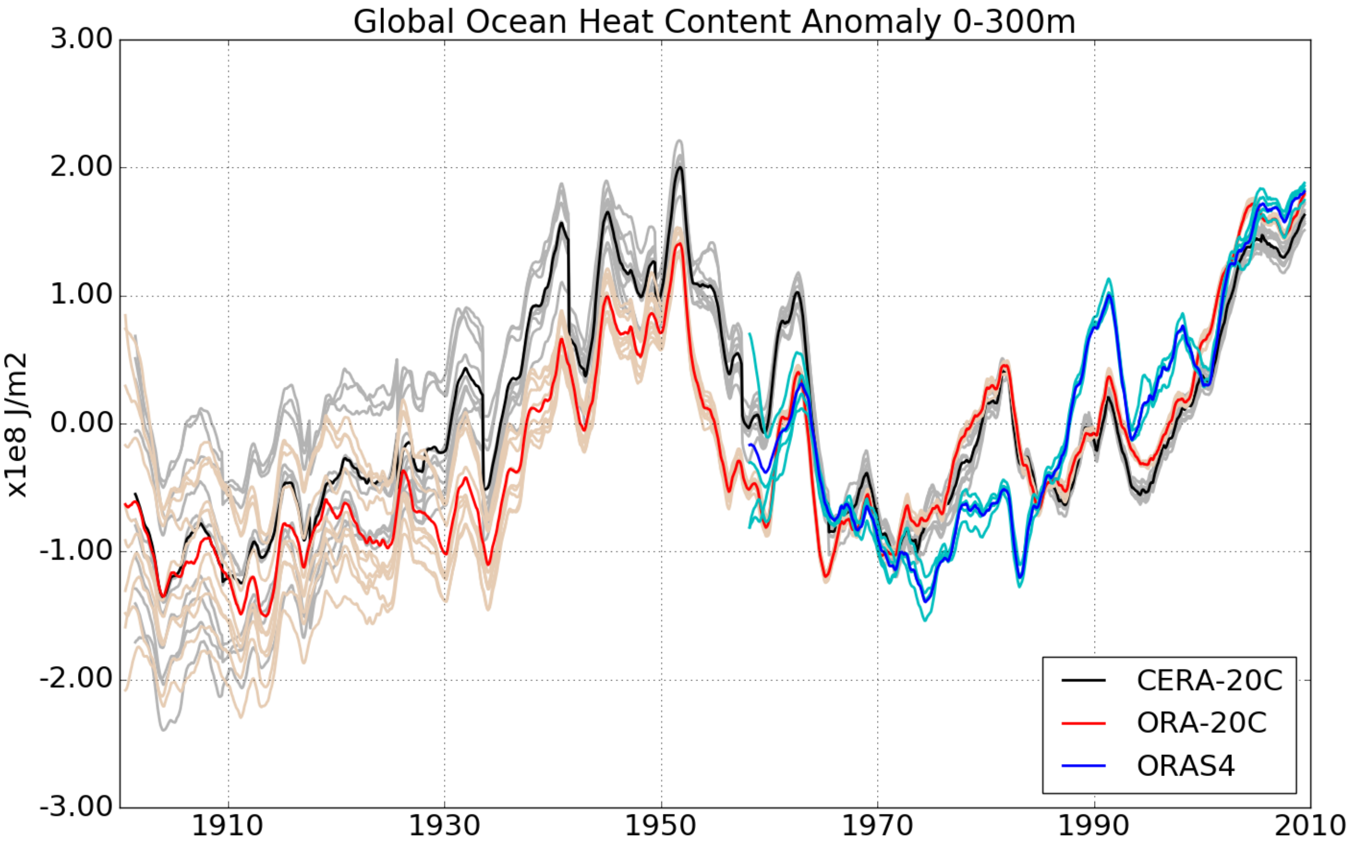


Storto et al., 2016

Bias correction and preprocessing of in situ observations represent the most crucial component of the reanalysis, whose perturbation accounts for up to 60% of the ocean heat content anomaly variability in the pre-Argo period

v. Schuckmann, ICR5

Oceanic heat content estimates



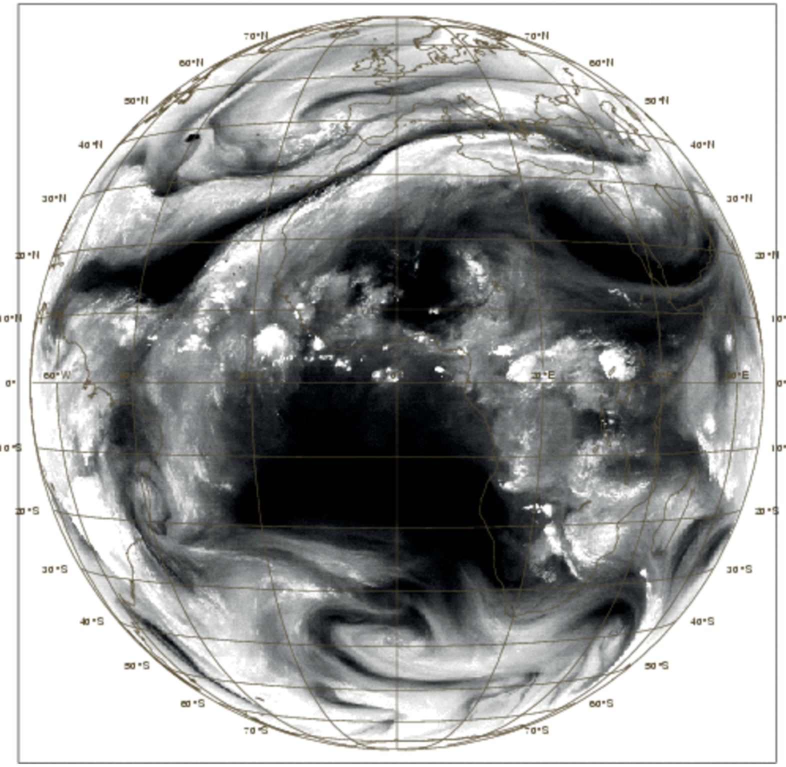
$$C\frac{\mathrm{d}T}{\mathrm{d}t}$$

Few observations before 1950 – large ensemble spread

Conclusion

- Uncertainty constrained by observations
  - Information content per observation much higher in early period – justifies data rescue efforts in ERA-CLIM2
  - Enormous observation data increase in past decades, global mean temperature error at 100hPa now below 0.1K
  - Model and observation biases major limitation but decreasing
- More and more efforts to express uncertainty through ensembles
  - Differences larger than ensemble spread valuable indicator of still unresolved biases
- Reduction of uncertainties in both models, observations is an iterative process
  - Continued research on reanalyses essential

Observation operator allows comparison of model state with satellite picture

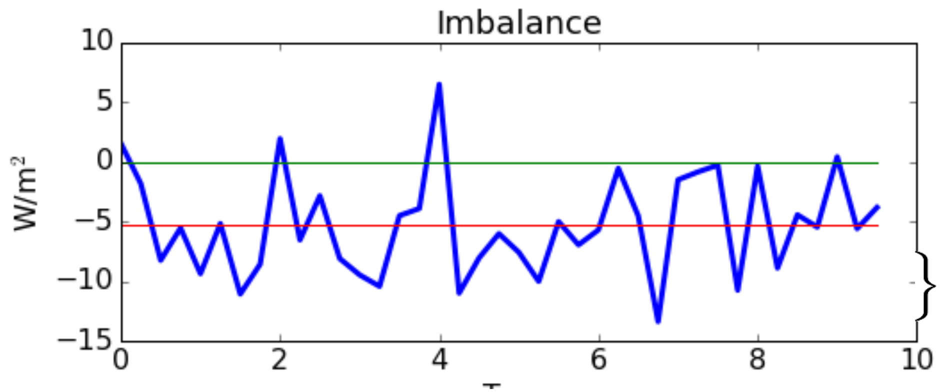


Observed by satellite **y**

Overspecification and imbalance

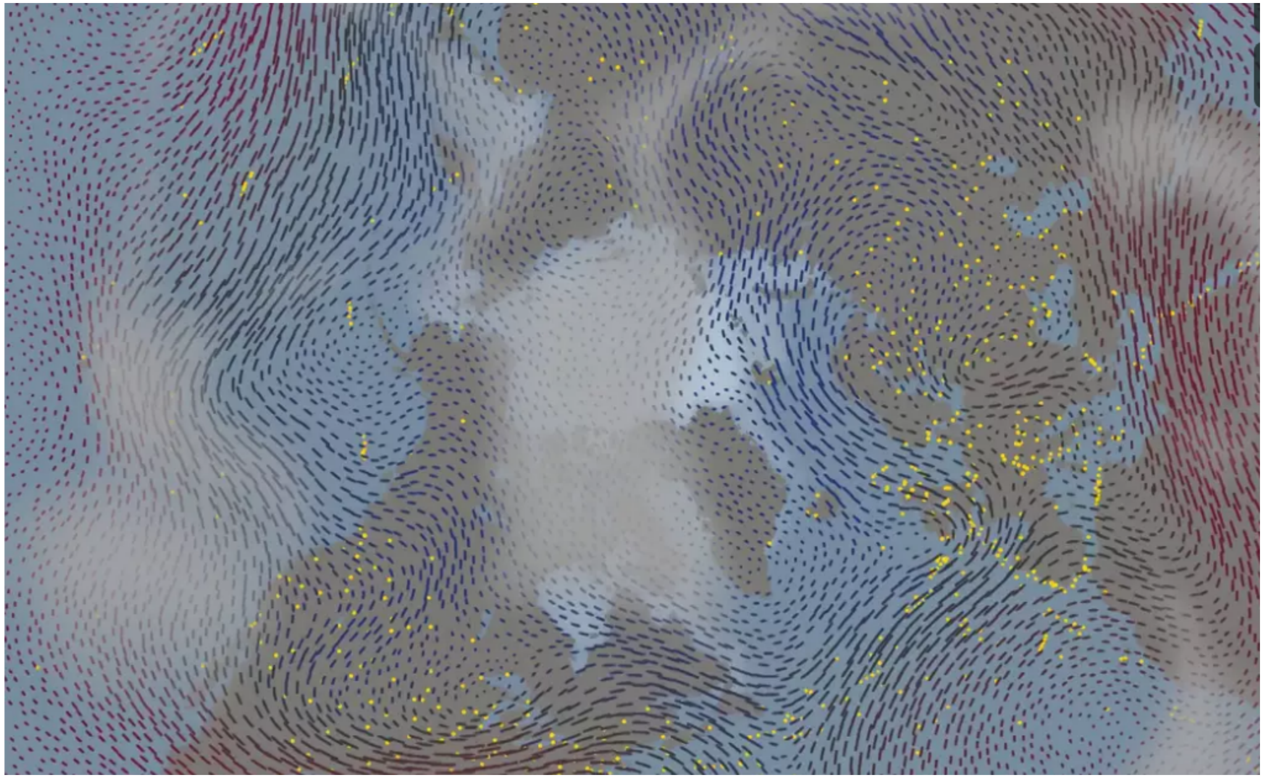
$$\underbrace{C\frac{\mathrm{d}T}{\mathrm{d}t}}_{\text{aus } T_{\mathrm{end}}-T_{\mathrm{begin}}} - \underbrace{\frac{S}{4}(1-\alpha)}_{\text{gemessen}} + \underbrace{\varepsilon\sigma T^4}_{\text{from } T} = \underbrace{\mathbf{??}}_{I}$$

- Storage can be calculated from differences of T-Analyses.
- Fluxes evaluated from independent data or from model state.
- So left side is available but perhaps not zero.
- Averaged over a longer time period, the imbalance is much smaller than the fluxes(240W/m^2) but often larger than storage. Indication of bias.



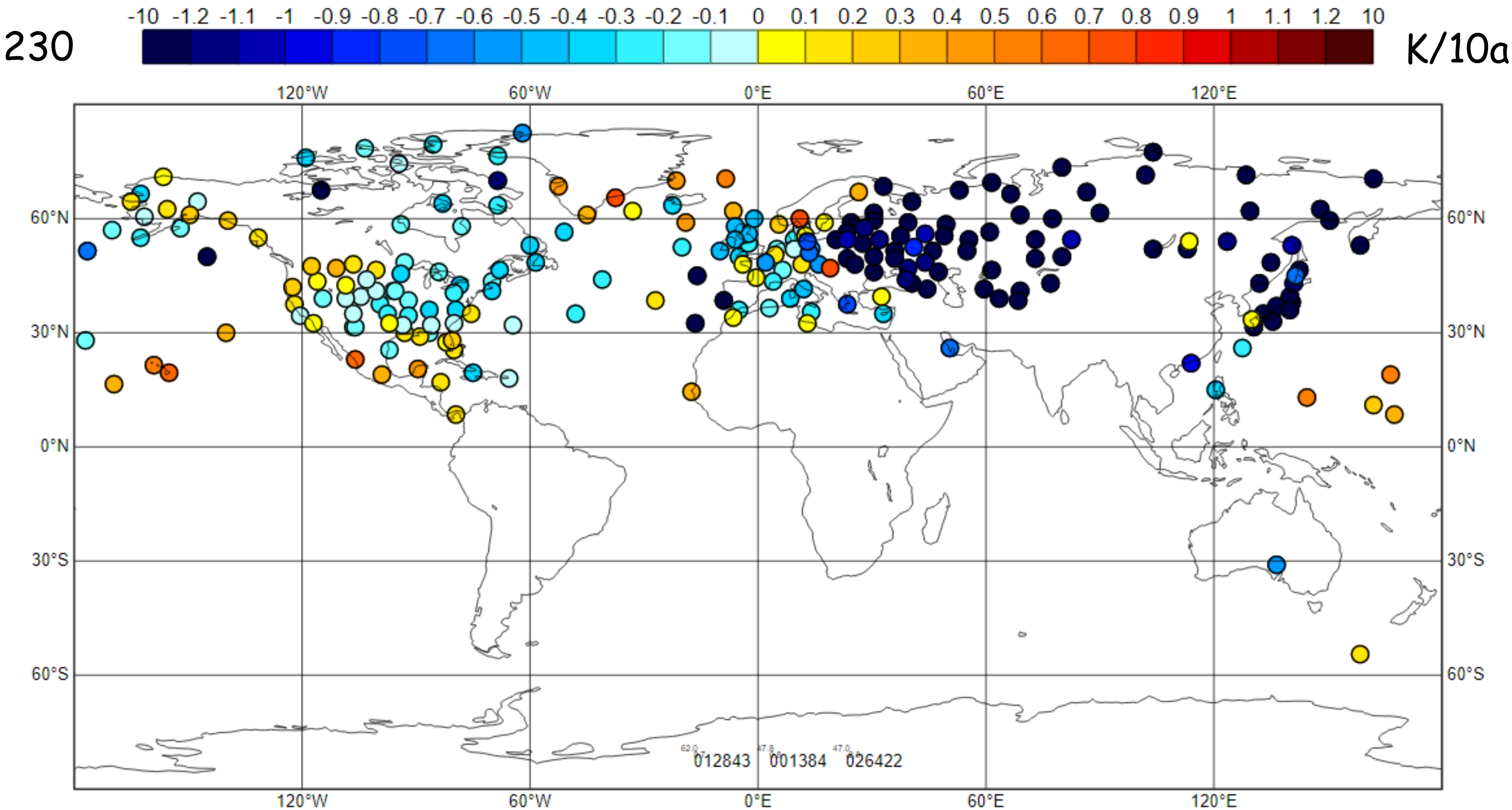
} Mittlere Imbalance

More observations help reducing fog of uncertainty

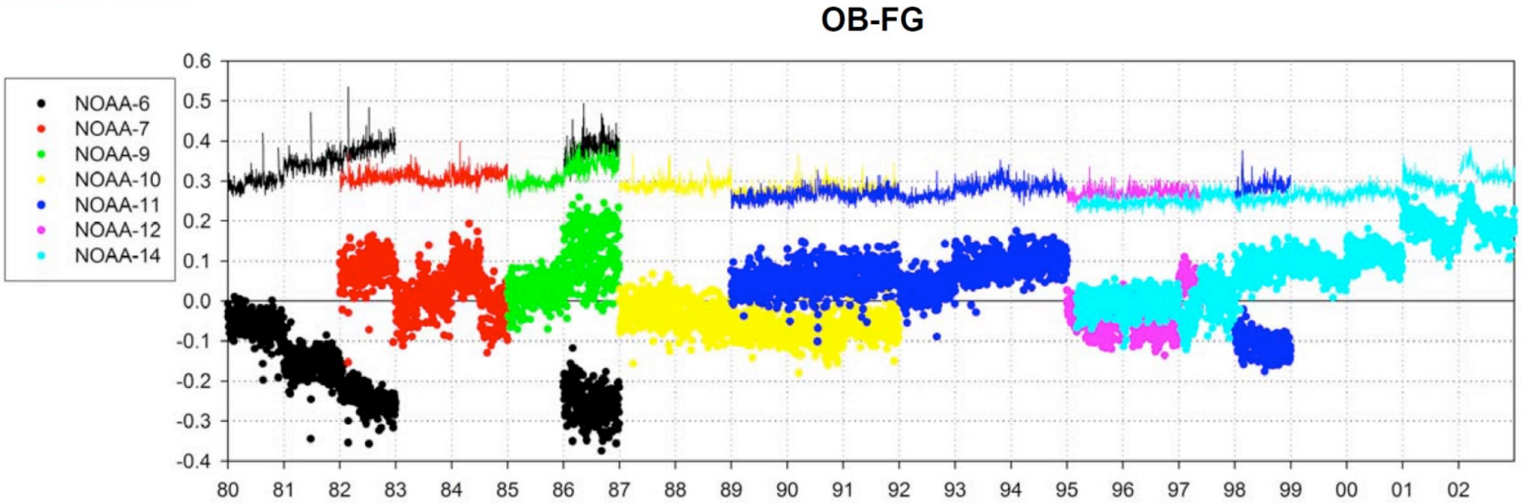


P. Brohan

Trends 1949-1969, 300 hPa



MSU Ch3


OB-FG
NOAA-6
NOAA-7
NOAA-9
NOAA-10
NOAA-11
NOAA-12
NOAA-14
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.4
80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 00 01 02

Definition of uncertainty

- The reanalysis input vector y and the reanalysis state vector x are multivariate random variables
- Components of y highly versatile, length of y changes by several orders of magnitude from say 1900 to today
- Typically one realization of y in a time interval
  - Distribution of y, or at least mean and covariance, are estimated using other components of y or from temporal variability
- Some components of y may be biased.

Simplified view of data assimilation

- Let's assume variational assimilation context

$$J(\mathrm{x}) = (\mathrm{x} - \mathrm{x}_b)^T \mathrm{B}^{-1} (\mathrm{x} - \mathrm{x}_b) + (\mathrm{y} - \mathrm{H}\mathrm{x})^T \mathrm{R}^{-1} (\mathrm{y} - \mathrm{H}\mathrm{x})$$

- We observe measurements y but we want to estimate state x
- J should be sharp, minimum should be at the right place.