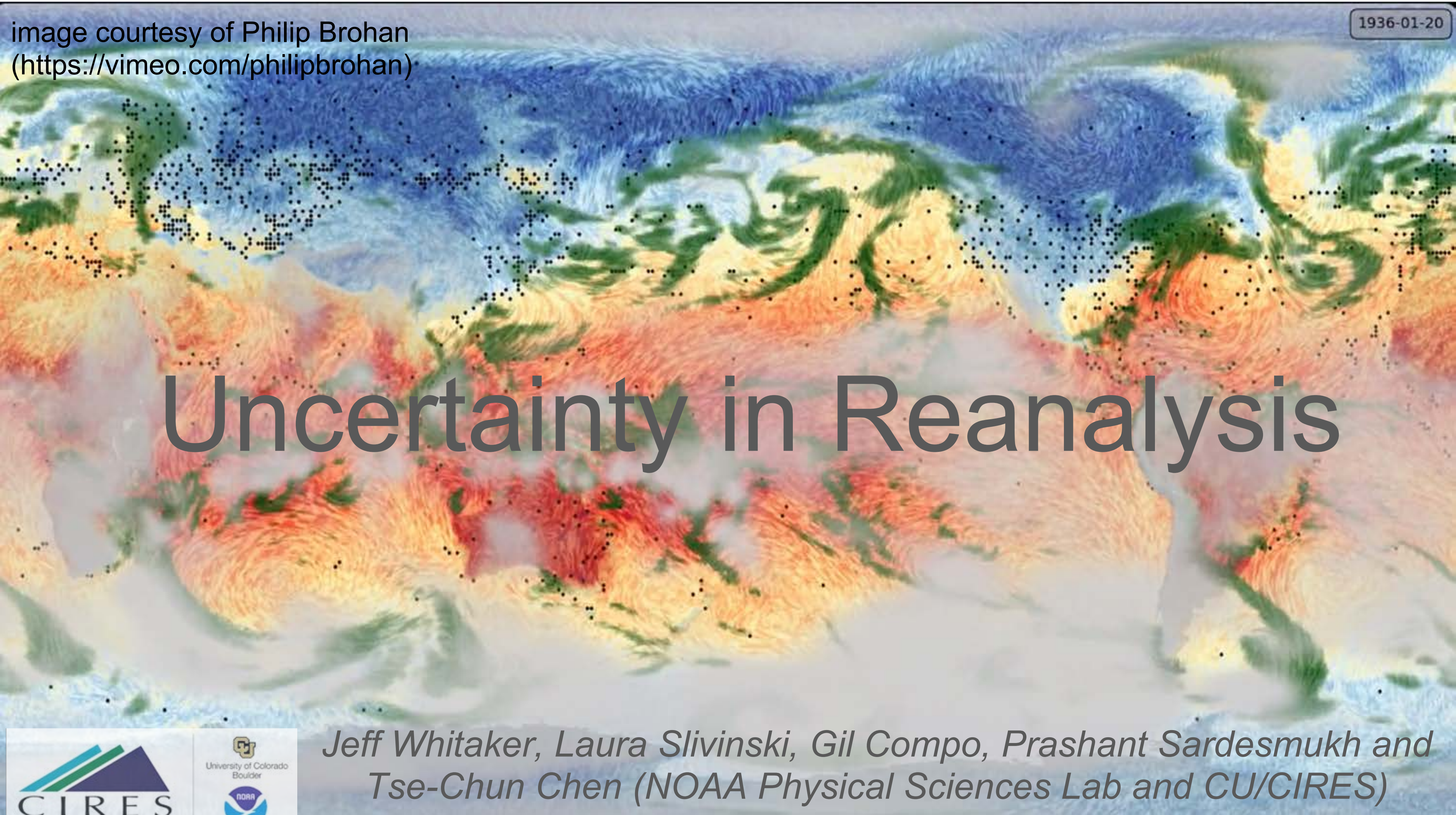
image courtesy of Philip Brohan
(https://vimeo.com/philipbrohan)



Uncertainty in Reanalysis

Jeff Whitaker, Laura Slivinski, Gil Compo, Prashant Sardesmukh and Tse-Chun Chen (NOAA Physical Sciences Lab and CU/CIRES)

Types of uncertainty and how to estimate them

- Random uncertainties associated with observation errors, chaotic error growth, model parameterizations, representation error, sub-optimal data assimilation algorithms, etc (not including systematic bias)
  - Estimated using ensemble spread in an ensemble data assimilation system
    - Only works if ensemble data assimilation (DA) system is well-tuned so that ensemble spread matches innovation statistics (inflation, localization)
- Systematic biases associated with model errors
  - Estimated via slowly varying component of analysis increments
    - Only works when observations are available to correct for these biases.

Estimating random component of analysis uncertainty with ensemble spread - experience with 20CR

- Multiplicative inflation used to represent missing or under-represented sources of random error in the ensemble DA system.. Values need to be observation-network dependent!
  - No inflation where there are no obs, more inflation where obs are dense.
- 20CRv2 used preset constant values that varied with time

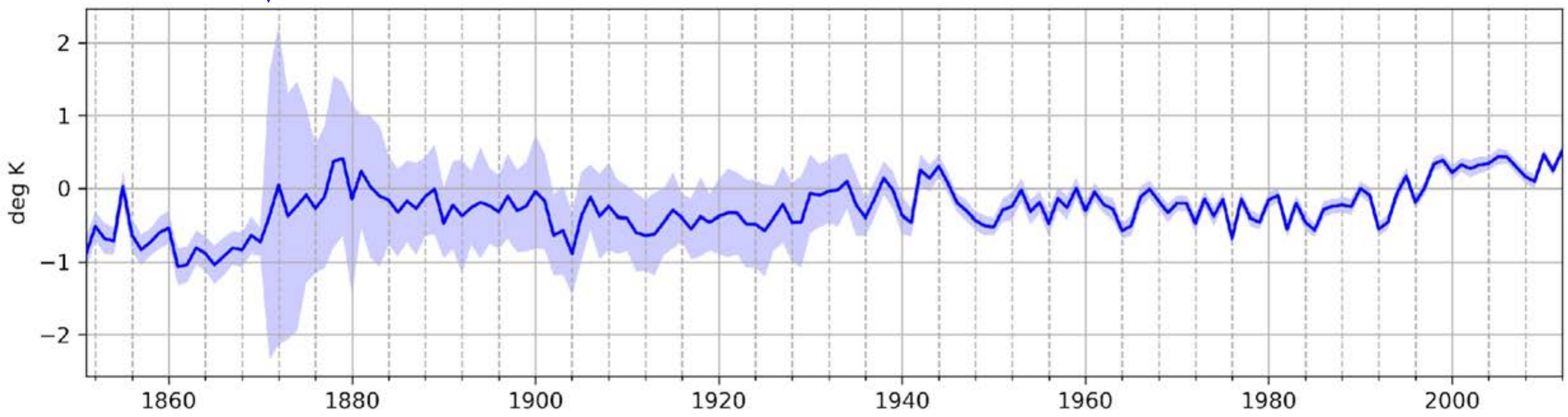
	Northern Hemisphere	Tropics	Southern Hemisphere
1851 – 1870	1.01	1.01	1.01
1871 – 1890	1.05	1.01	1.01
1891 – 1920	1.09	1.02	1.01
1921 – 1950	1.12	1.03	1.02
1951 – 2012	1.12	1.07	1.07

20CRv2

	Northern Hemisphere	Tropics	Southern Hemisphere
1851 – 1870	1.01	1.01	1.01
1871 – 1890	1.05	1.01	1.01
1891 – 1920	1.09	1.02	1.01
1921 – 1950	1.12	1.03	1.02
1951 – 2012	1.12	1.07	1.07

- Unrealistic signals in uncertainty
- Inhibits accurate studies of significance of long-term trends

Atmospheric layer temperature anomalies, Northern Hemisphere



‘Relaxation to prior spread’ posterior inflation

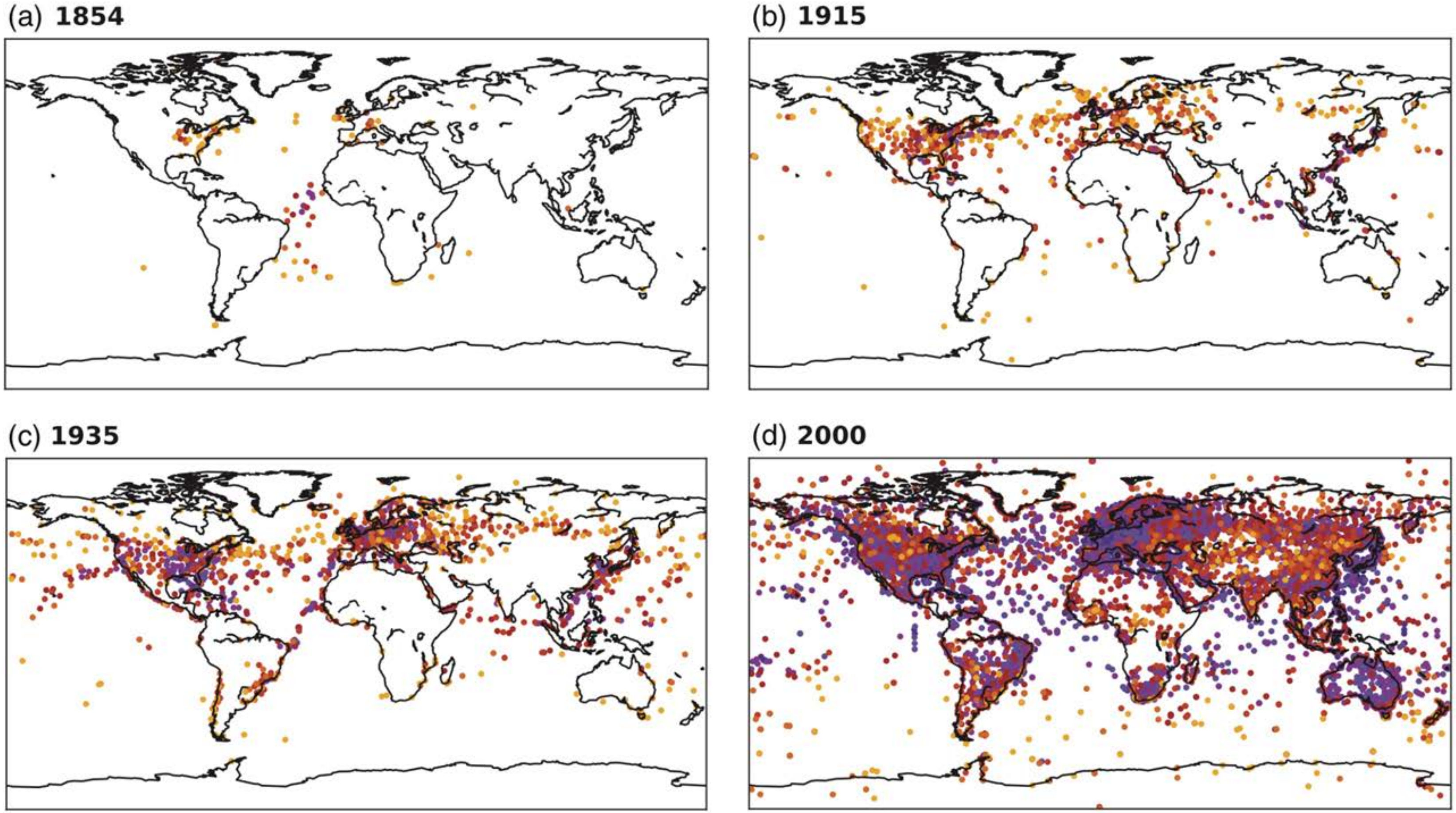
(RTPS, Whitaker and Hamill 2012)

$\sigma^a \leftarrow (1-\alpha)\sigma^a + \alpha\sigma^b$ which implies $\mathbf{x}_i'^a \leftarrow \mathbf{x}_i'^a \sqrt{\alpha \frac{\sigma^b - \sigma^a}{\sigma^a} + 1}$

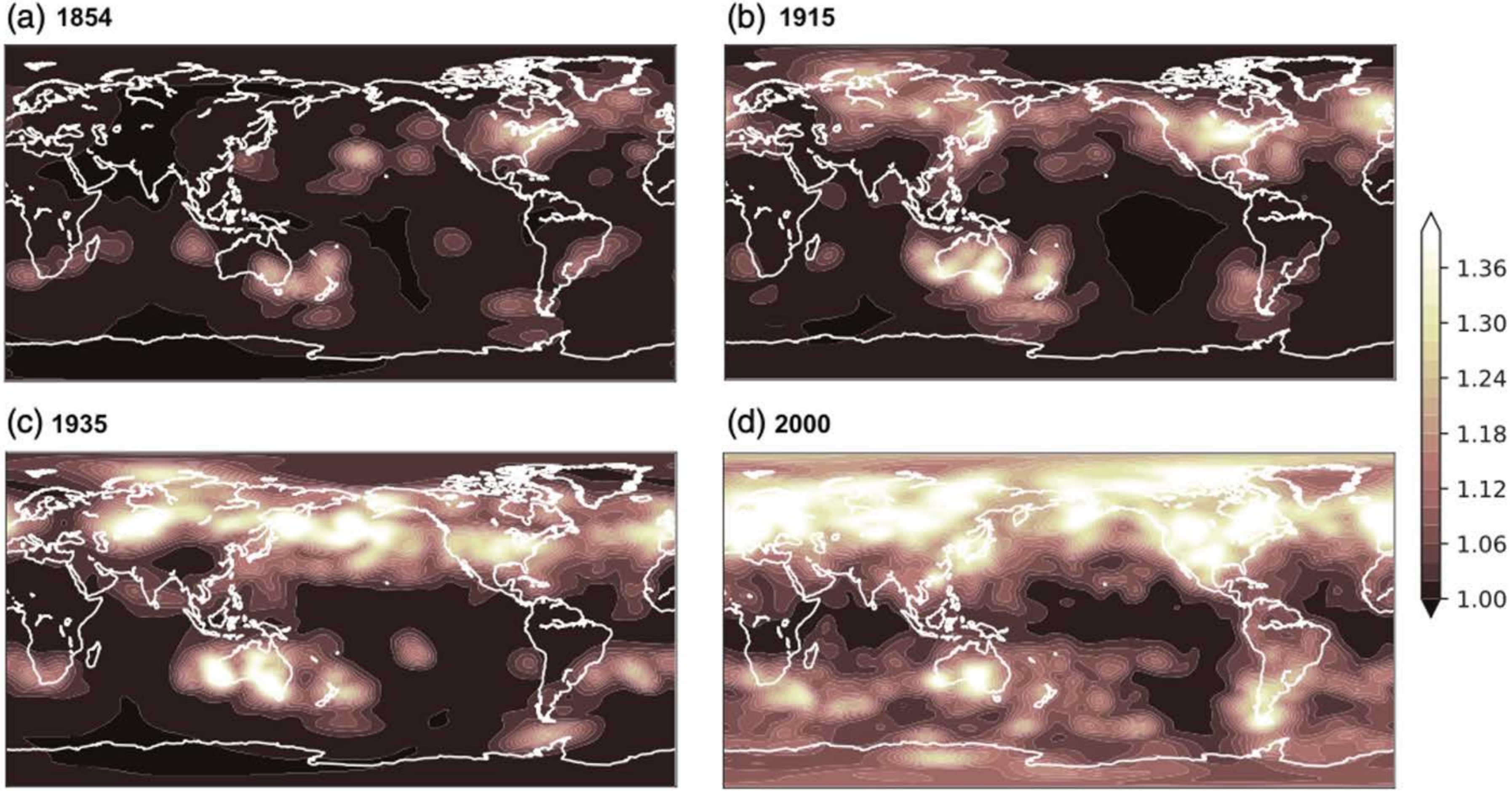
Adapts to observing network: inflation factor = 1 if no obs are assimilated, increases as reduction of ensemble spread by assimilation of obs increases.

α set to 0.9 for entire period in 20CRv3

Observing network



20CRv3 - Inflation factor using RTPS with α=0.9

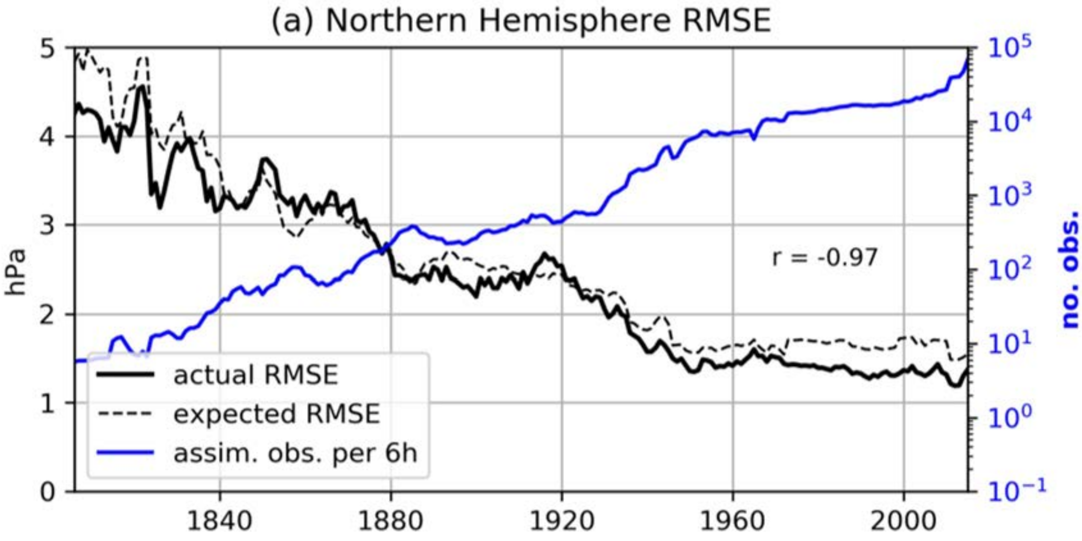


Innovation statistics for 20CRv3

(Fig 4 from Slivinski et al 2021

https://doi.org/10.1175/JCLI-D-20-0505.1)

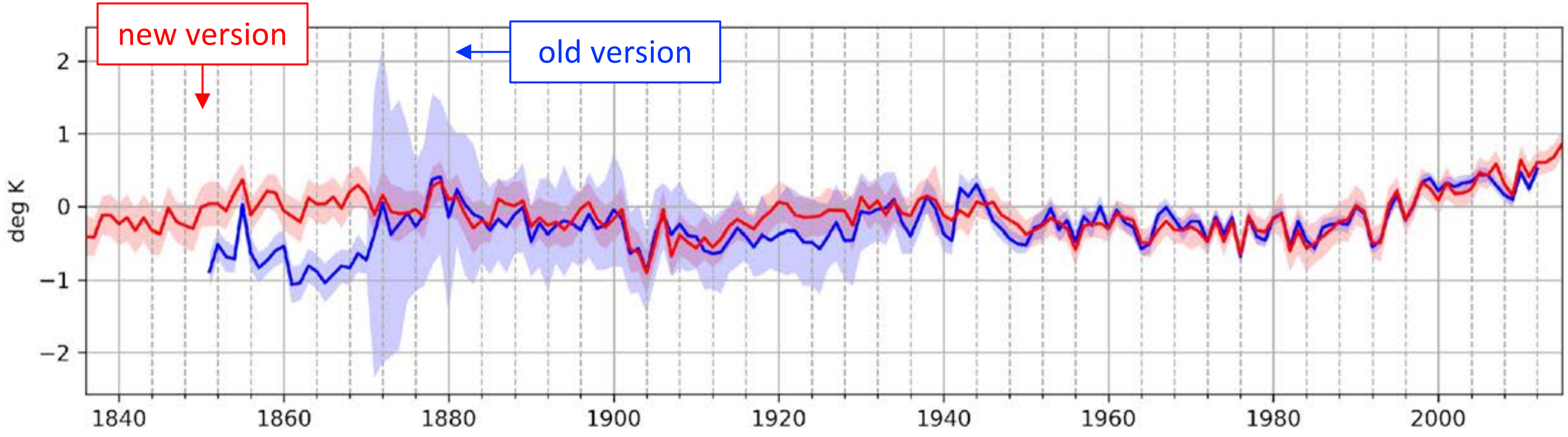
- if the observation and background errc are uncorrelated and unbiased, then RMSDactual should be equivalent to RMSDexp = sqrt(spread + oberror)
- As obs density increases (blue curve), both actual and expected RMSE go down, correlation is > 0.9 (0.97 in NH)



20CRv2 vs 20CRv3 time series

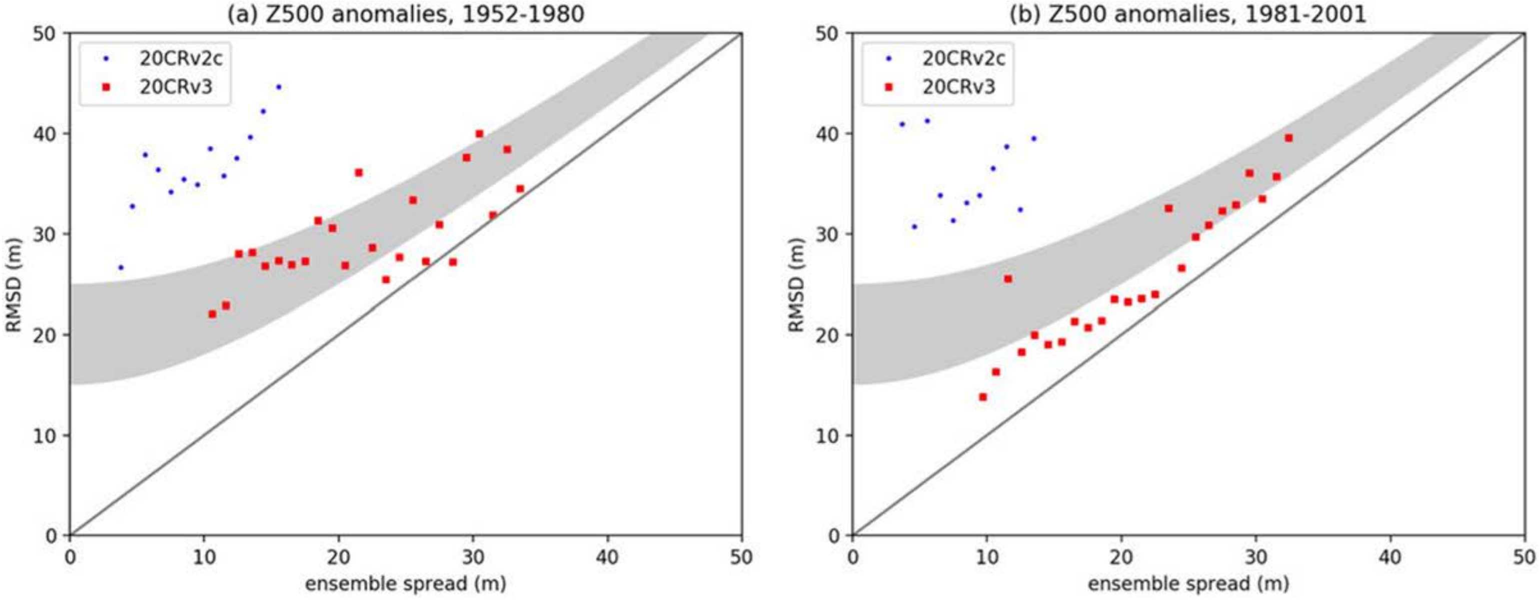
- More accurate, consistent estimates of uncertainty
- Can make stronger statements about trends
- Other relevant changes in 20CRv3 - stochastic physics, adaptive localization length scale

Atmospheric layer temperature anomalies, Northern Hemisphere



Validation against independent upper air obs at Lindenberg, Germany (Fig 8 from Slivinski et al 2021)

- RMSD (dots) binned by ensemble spread (x-axis).
- Diagonal line expected RMSD for perfect obs
- Shaded area expected RMSD for 15>ob error>25m assuming accurate ensemble spread
- Dots above overconfident, below underconfident.

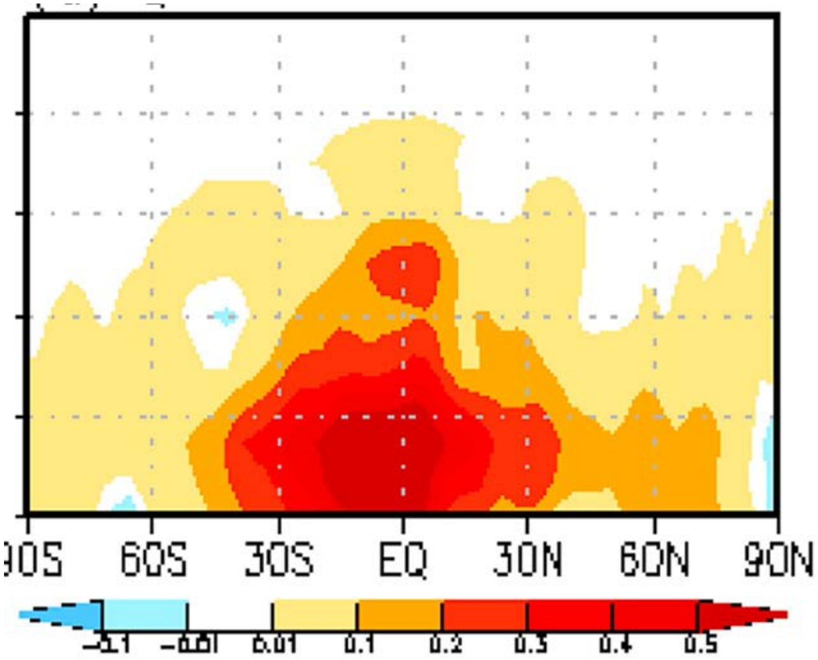


Interaction between systematic bias/obs network on time series

Spurious jumps apparent when obs systems come online that correct for model biases

CFS has dry bias in tropics - expressed in reanalysis prior to AMSU in 1999

PWAT (1997-2007 mean - 1987-1997 mean)



PWAT increment

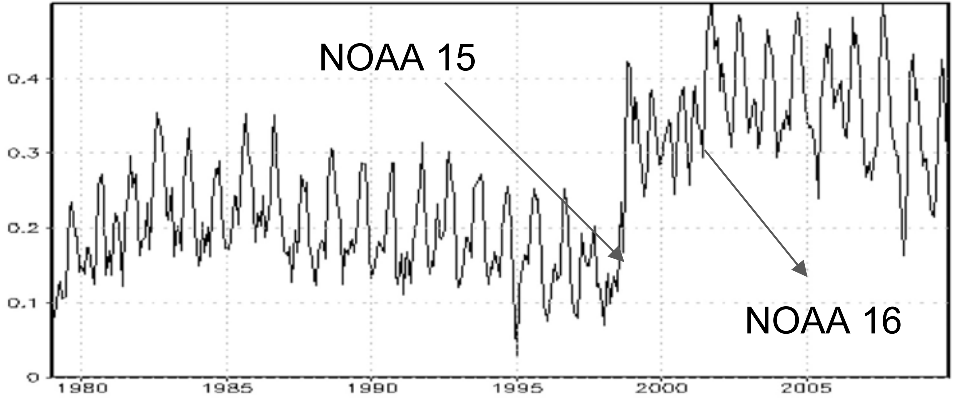


Figure 6. Monthly mean of global average of precipitable water increment for CFSR.

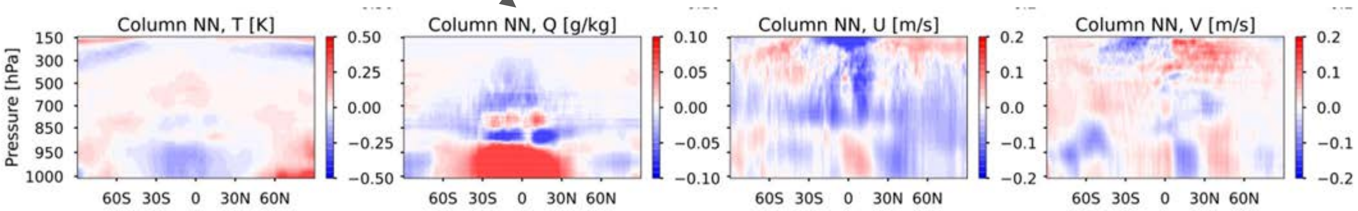
Influence of changes in observations on precipitation: A case study for the Climate Forecast System Reanalysis (CFSR)
Zhang et al 2012 (http://dx.doi.org/10.1029/2011JD017347)

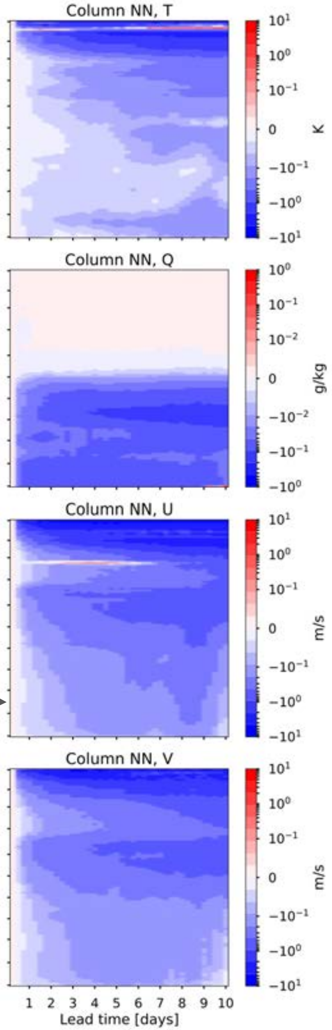
Using ML to learn model errors, build in-line model error correction (Tse-Chun Chen PSL)

- Following approach outlined in Bonavita and Laloyaux 2020: (https://doi.org/10.1029/2020MS002232) train column-based NN correction using analysis increments in data-rich period.
- Approach applied to UFS by Chen et al 2022 (manuscript in preparation).
- Currently building an inline correction interface for UFS model.
- For reanalysis, applying in-line correction should reduce spurious jumps in time series when new observations are introduced that correct for model biases.

Zonal mean/time mean cross-section of column NN corrections

Time mean forecast RMSE % change as a function of lead time (neg is RMSE reduction)





Critical requirements for a consistent (representation of uncertainty in) Earth system reanalysis

- For a consistent representation of (random) uncertainty:
  - An ensemble data assimilation system that adapts to the changing observing network to maintain consistent spread/error relationships.
    - Ensemble DA system spread is very sensitive to inflation (& localization) parameters. The optimal values of the parameters are sensitive to the observing network.
- To avoid spurious jumps/trends in time series due to the interplay between model bias and changes in the observing network:
  - An unbiased background forecast for quantities that are constrained by observations in the era of densest observations.
    - Systematic model bias is expressed in analysis increments if obs are not sufficient to correct for it.
    - ML algorithms trained on analysis increments in the dense observation period can be used to develop an in-line model error correction to reduce background forecast biases.