The 20th Century Reanalysis (20CR): Capturing 200 years of weather using surface observations

Laura C. Slivinski[1],

Gilbert P. Compo[1,2], Jeffrey S. Whitaker[1], Prashant D. Sardeshmukh[1,2]

[1]NOAA Physical Sciences Laboratory, Boulder CO

[2]Cooperative Institute for Research in Environmental Sciences, CU Boulder

SYNOPTIC WEATHER MAP
NORTHERN HEMISPHERE
SEA LEVEL 1300 GMT
AUG 16 1915

CIRES

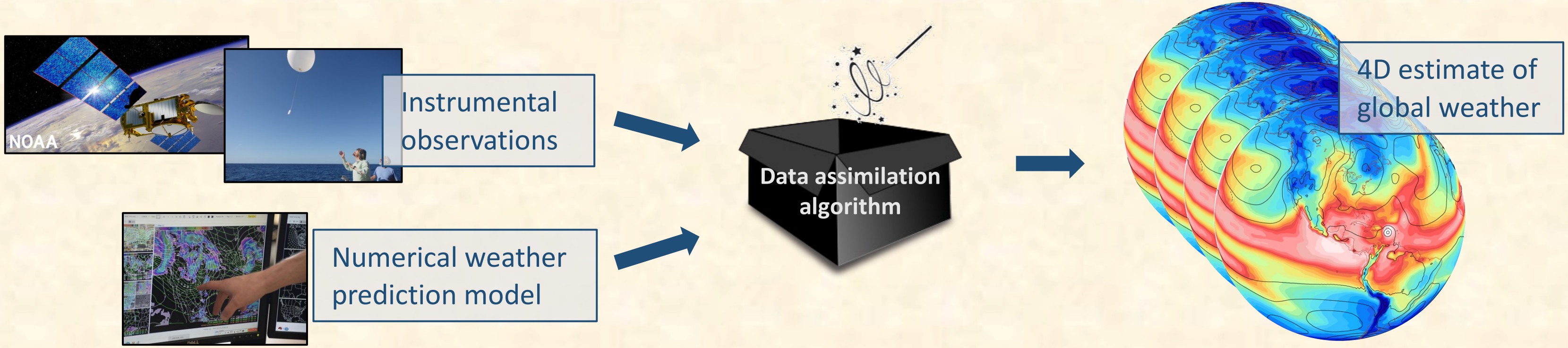
NOAA
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
U.S. DEPARTMENT OF COMMERCE

Outline

- Reanalysis fundamentals
- Full input vs sparse input reanalysis
- 20CRv3: methods
- 20CRv3: performance
- Outlook: possibilities for 20CRv4

Reanalysis fundamentals

- Reanalyses provide a consistent, gridded record of weather & climate by assimilating historical observations into a modern weather forecast model (“re-analyzing” the data.)
- To achieve consistency, a single forecast model and single data assimilation method are fixed.
  - To some degree, the observing network may also be fixed.



Full input vs sparse input reanalysis

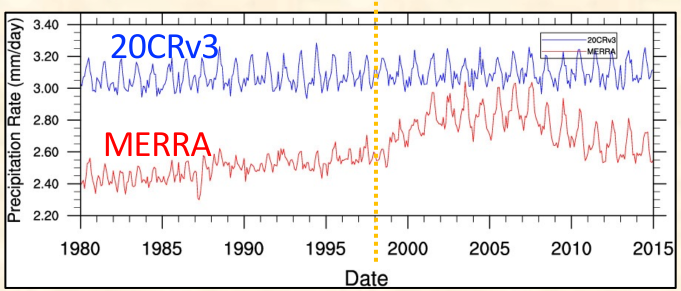
- **Full input**
  - ERA-interim, ERA5, MERRA, MERRA2
  - Assimilate most observations that are available *(in-situ, satellite, upper-air, aircraft)*
  - Cover latter half of 20th century to avoid spurious trends and signals arising from significant changes in the observing system
  - …Can still be impacted by instruments coming online
- **Sparse input**
  - 20th Century Reanalysis, CERA-20C
  - Assimilate only surface observations (surface pressure, optionally marine winds)
  - Extend 100+ years into the past
  - Less impact from changes in observing network

Image provided by the NOAA-ESRL Physical Sciences Laboratory from their website at https://psl.noaa.gov/data/writ

Full input vs sparse input reanalysis

- **Full input**
  - ERA-interim, ERA5, MERRA, MERRA2
  - Assimilate most observations that are available *(in-situ, satellite, upper-air, aircraft)*
  - Cover latter half of 20th century to avoid spurious trends and signals arising from significant changes in the observing system
  - …Can still be impacted by instruments coming online
- **Sparse input**
  - 20th Century Reanalysis, CERA-20C
  - Assimilate only surface observations (surface pressure, optionally marine winds)
  - Extend 100+ years into the past
  - Less impact from changes in observing network

Monthly mean global precipitation



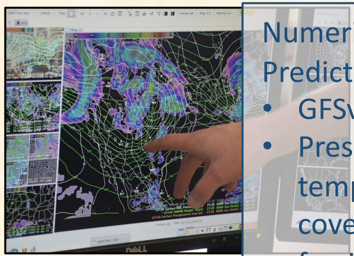
Example: spurious trend in global precipitation from observing network change

- MERRA: model may have dry bias that was corrected by water vapor sensitive radiances from AMSU, which came online in 1998
- 20CRv3 does not assimilate any radiances, therefore does not exhibit this discontinuity

Image provided by the NOAA-ESRL Physical Sciences Laboratory from their website at https://psl.noaa.gov/data/writ

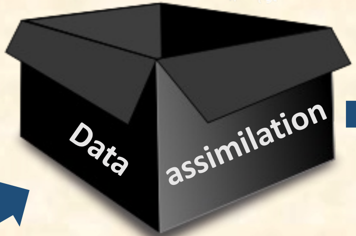
20th Century Reanalysis (20CR)

Only surface pressure observations are assimilated *(other obs necessary for SSTs, ice)*



Numerical Weather Prediction model:

- GFSv14, ~0.75deg
- Prescribed sea surface temperature (SST), ice coverage, and radiative forcings

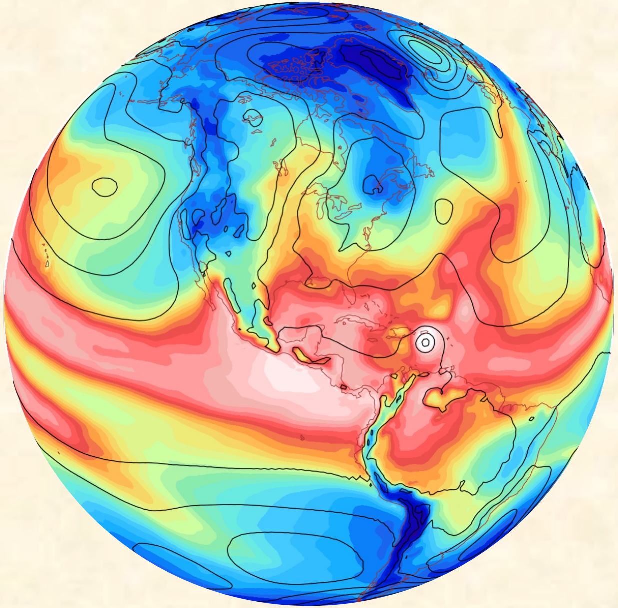


Ensemble Kalman Filter with 80 ensemble members.

➢ EnKF can dynamically move sparse information around via covariances.

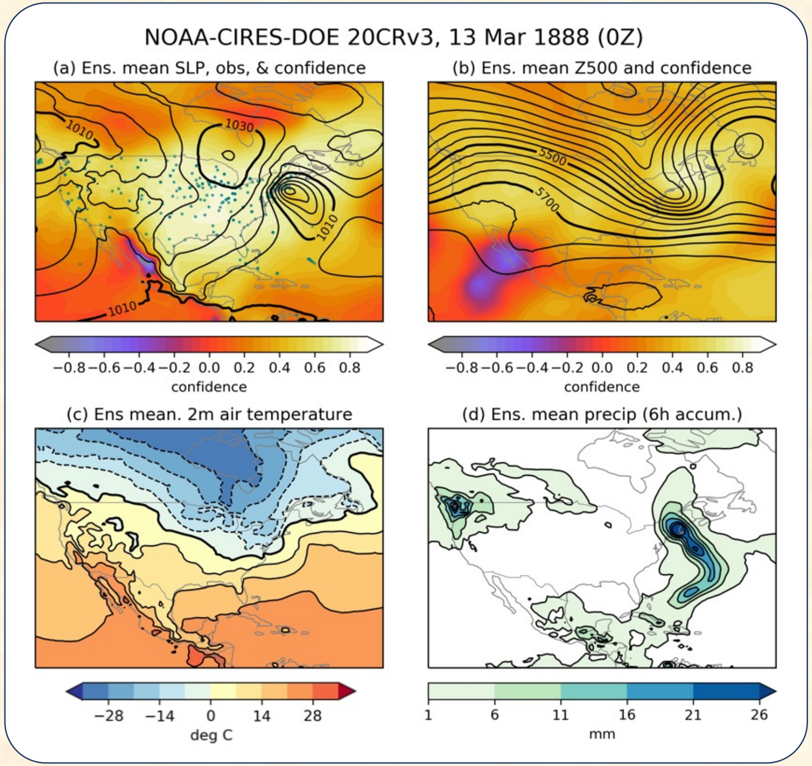
200 years of weather (1806-2015):

20th Century Reanalysis Version 3



The 20th Century Reanalysis (20CR) provides a global, 200-year history of sub-daily weather

by assimilating *only* surface pressure observations into a modern weather model



NOAA-CIRES-DOE 20th Century Reanalysis Version 3

- Estimates temperature, wind, precipitation, pressure, humidity, & other variables, from the ground to the top of the atmosphere
- Prescribed sea surface temperature *(quasi-weakly coupled)*, sea ice concentration, and radiative forcing
- Global 75km grid
- 3-hourly resolution
- Spans 1836-2015 [1806-1835 experimental]
- Data assimilation: Ensemble Kalman Filter with 80 ensemble members to quantify uncertainty

- Publicly available: https://go.usa.gov/XTd

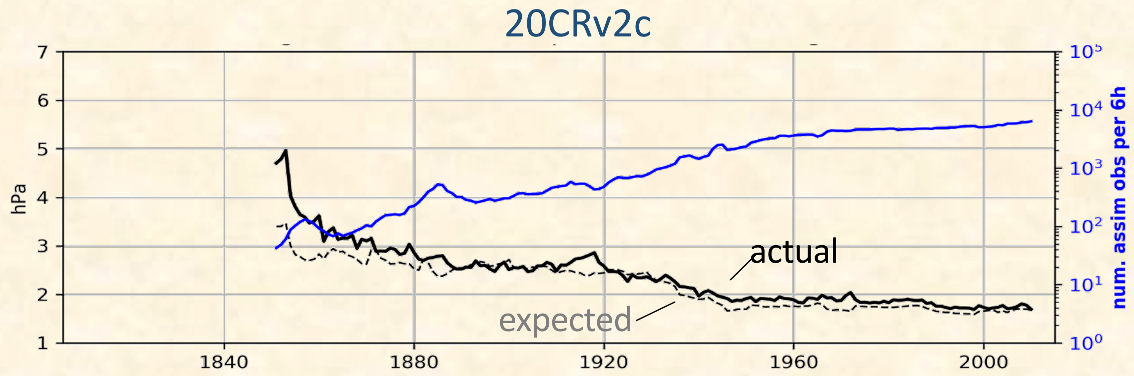
20th Century Reanalysis version 3 (20CRv3)
System improvements over 20CRv2c

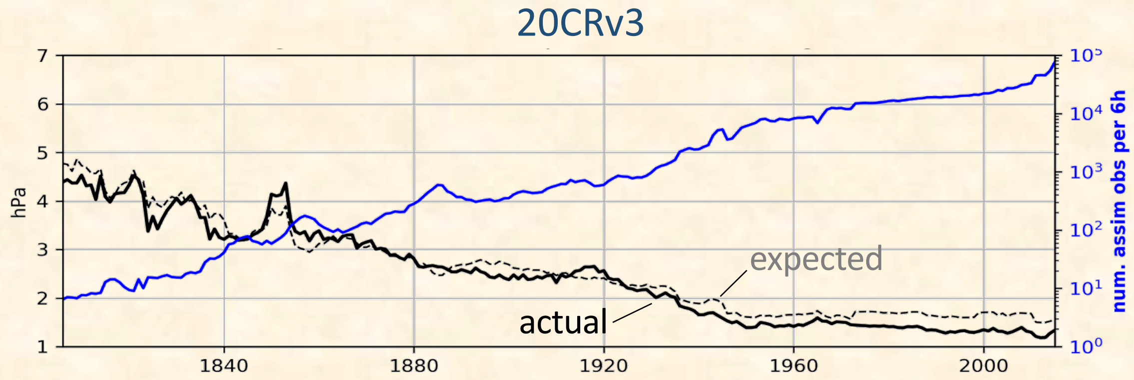
- Newer, higher-resolution forecast model
- Larger available set of observations to be assimilated
- Improved observation quality control methods
- Improved DA methods (4D incremental analysis update, adaptive localization, adaptive inflation)
  - ➢Improved confidence estimates

Slivinski, L.C., et. al. (2019) Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system. *Quarterly Journal of the Royal Meteorological Society,* **145**: 2876– 2908. https://doi.org/10.1002/qj.3598

20CRv3 outperforms 20CRv2c:
fit to surface pressure obs

Global annual first-guess root mean squared errors in surface pressure



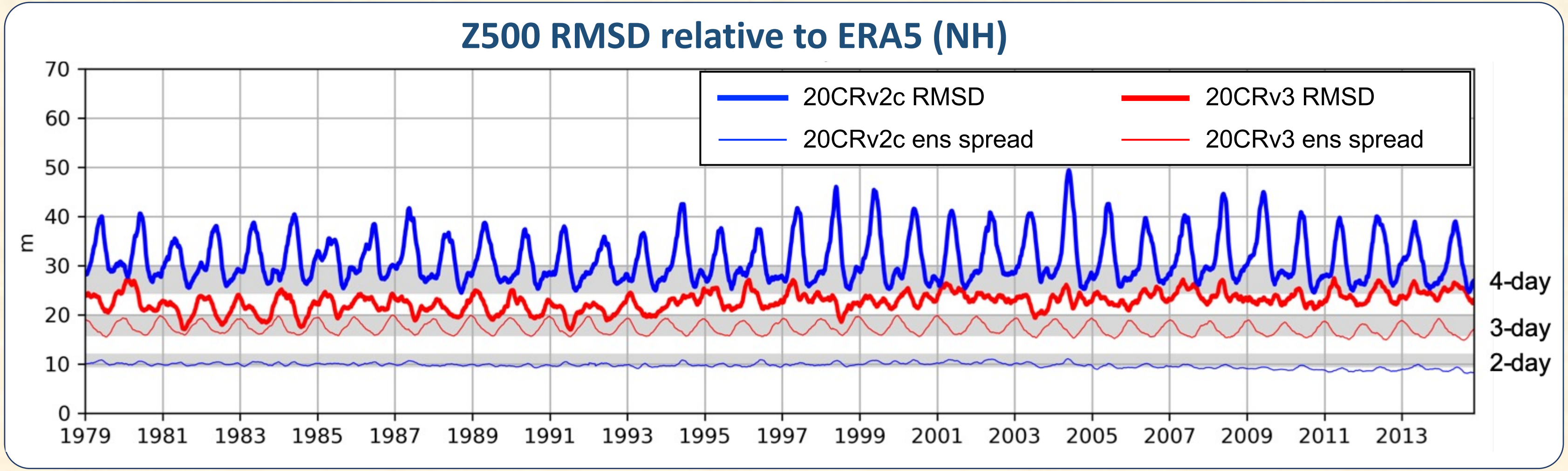


***Actual* error** is the difference between 6-hour forecasts and not-yet-assimilated observations: $\langle (ob - fg)^2 \rangle^{1/2}$

***Expected* error** is the root-mean of the sum of ob error variance and background ensemble covariance at ob time/location: $\left\langle (\sigma^2{}_{ob} + \sigma^2{}_{fg})^2 \right\rangle^{1/2}$

- **20CRv3 errors that are lower and more consistent with ensemble spread than 20CRv2c.**

20CRv3 outperforms 20CRv2c:
500hPa geopot. height agreement with ERA5

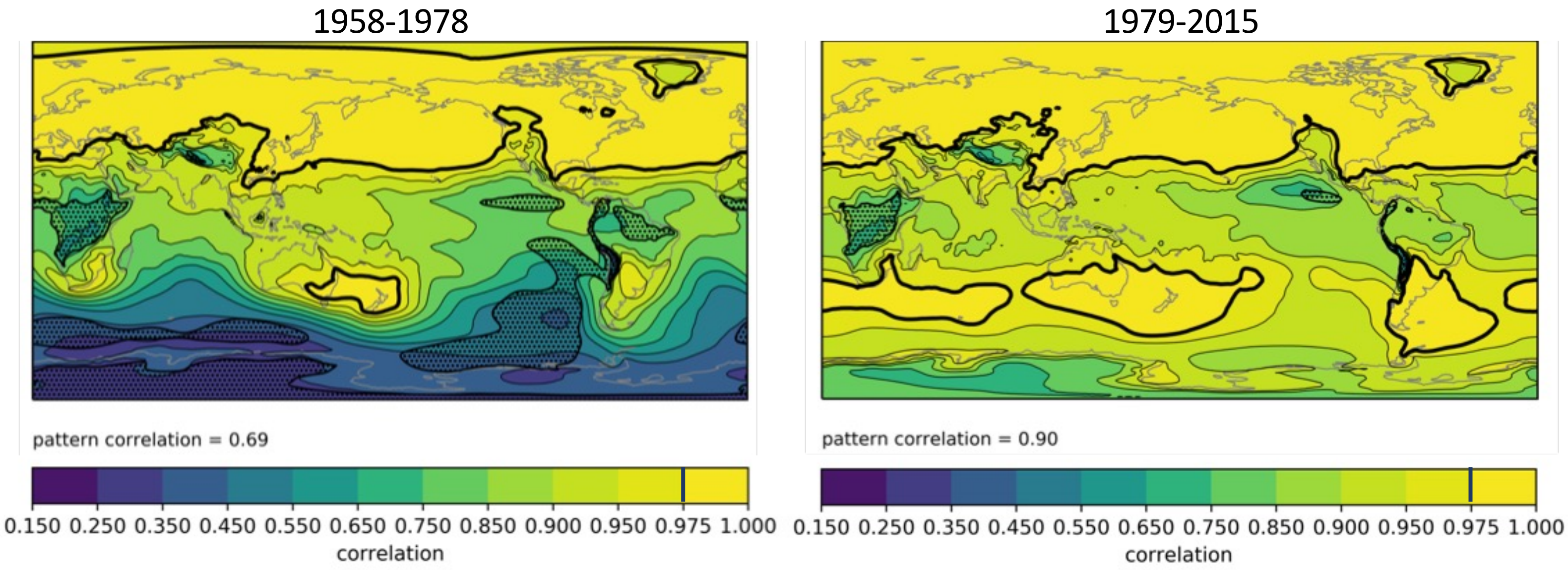


- **20CRv3 Z500 agrees with ERA5 better than 20CRv2c**, and has **more consistent ensemble spread**
- Gray shading shows range of ECMWF and NCEP oper. forecast errors in 2019 for 2-, 3-, 4-day leads
- **20CRv3 Z500 errors are comparable to modern 3-4 day operational forecast skill in the NH** (consistent with earlier predictions by Compo et al, 2006)

Slivinski, L.C., et. al. (2021) An Evaluation of the Performance of the Twentieth Century Reanalysis Version 3. *Journal of Climate,* 34(4): 1417-1438. https://doi.org/10.1175/JCLI-D-20-0505.1

20CRv3 correlates well with other reanalyses, and can "predict" that correlation

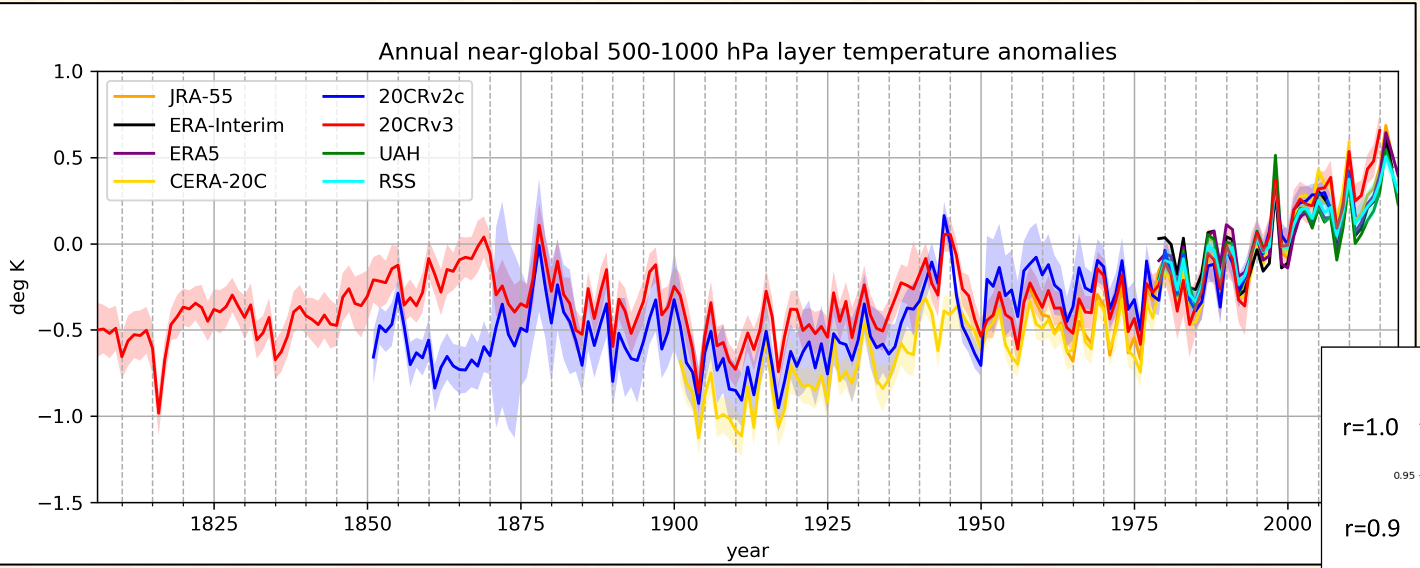
SLP anomaly correlation between JRA-55 and 20CRv3

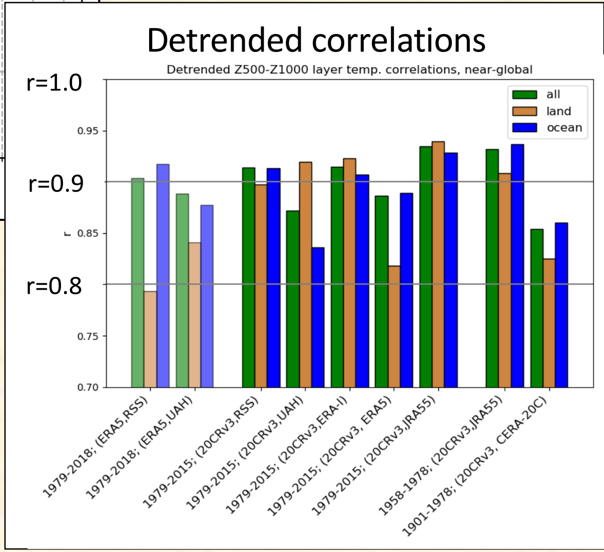


- Stippling indicates regions of low confidence (large ensemble spread) in 20CRv3
- Pattern correlation is given between confidence field and correlation field
- ➢ **20CRv3 uncertainty estimates are a good predictor of skill relative to JRA-55**

Slivinski et. al. (2021)

20CRv3 can capture trends & variability in temperature

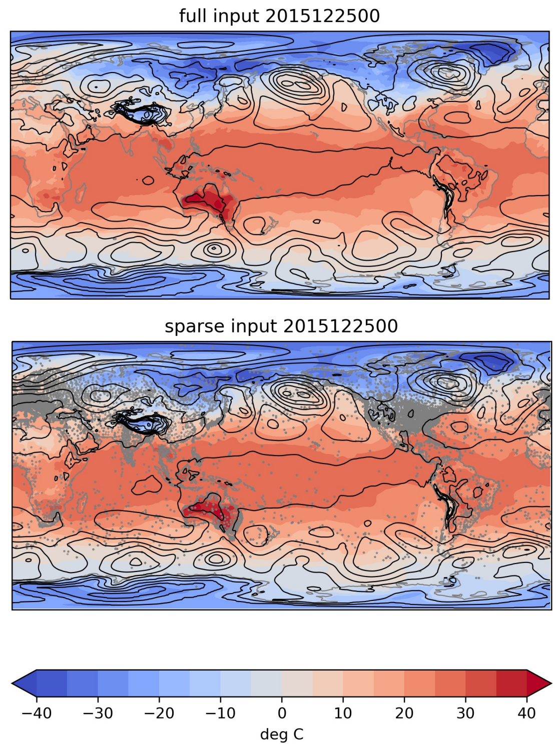


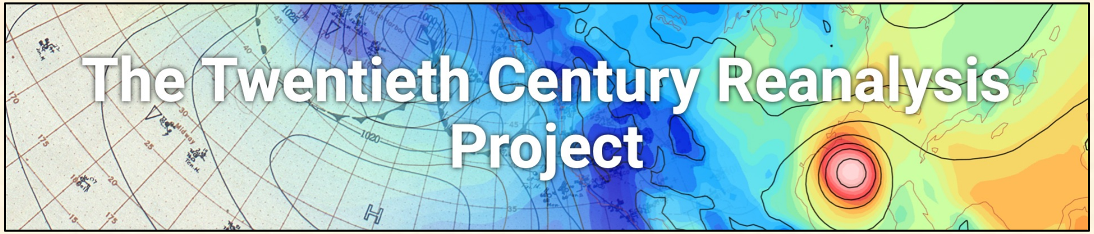


UAH & RSS: two satellite-based temperature reconstructions

Future of 20CR – Possibilities

- ✓ Larger set of available observations *(smaller errors, greater confidence, maybe extend further back in time)*
- ✓ Up-to-date forecast model
- Coupled ocean-atmosphere
- Additional observation types *(SST, wind direction)*
- Data-driven models *(incorporate Linear Inverse Model [LIM] for ocean)*
- No-DA counterpart simulation





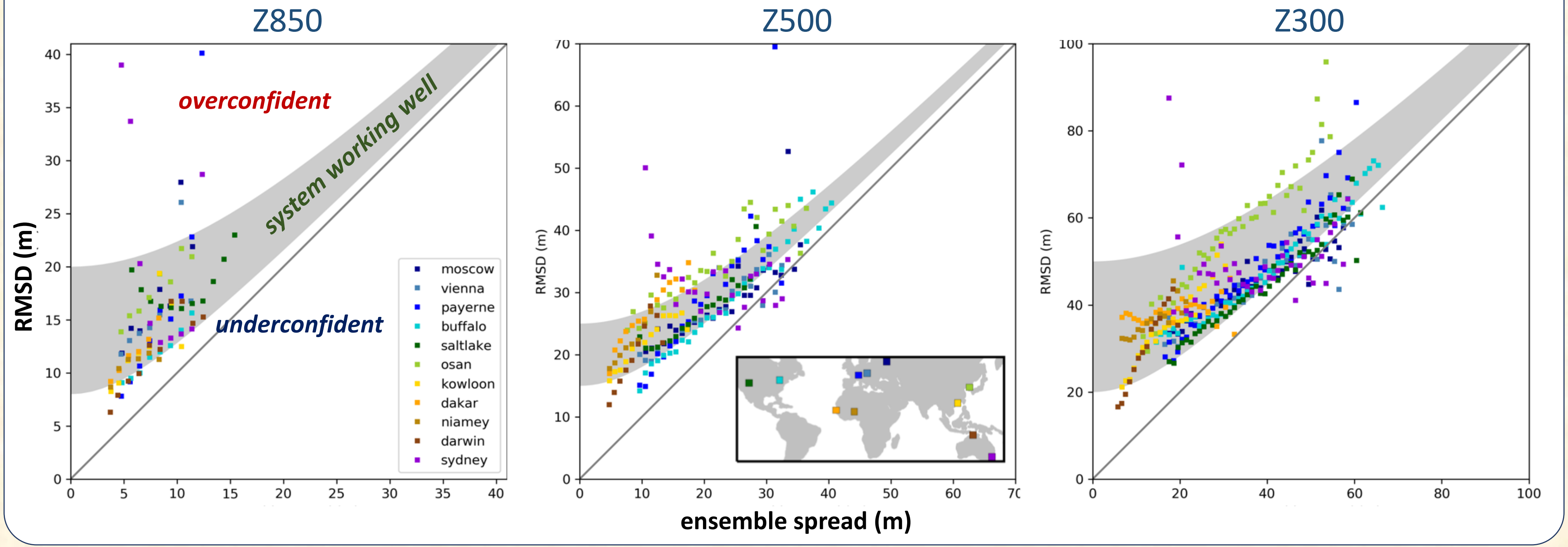
- The 20th Century Reanalysis version 3 (20CRv3) *only* assimilates surface pressure observations in order to consistently extend 200 years into the past
- It agrees well with other reanalyses, independent upper-air observations, and station/satellite-based datasets
- Confidence estimates are reliable and 20CRv3 can often predict its own skill
- More information:
  - Compo, G.P., et. al. (2011) The Twentieth Century Reanalysis Project. *Q.J.R. Meteorol. Soc.*, 137: 1-28. https://doi.org/10.1002/qj.776
  - Slivinski, L.C., et. al. (2019) Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system. *Quarterly Journal of the Royal Meteorological Society*, 145: 2876–2908. https://doi.org/10.1002/qj.3598
  - Slivinski, L.C., et. al. (2021) An Evaluation of the Performance of the Twentieth Century Reanalysis Version 3. *Journal of Climate*, 34(4): 1417-1438. https://doi.org/10.1175/JCLI-D-20-0505.1

For data access, visualization tools, and references, please visit https://go.usa.gov/XTd

laura.slivinski@noaa.gov

20CRv3 performs well relative to indep. upper air obs

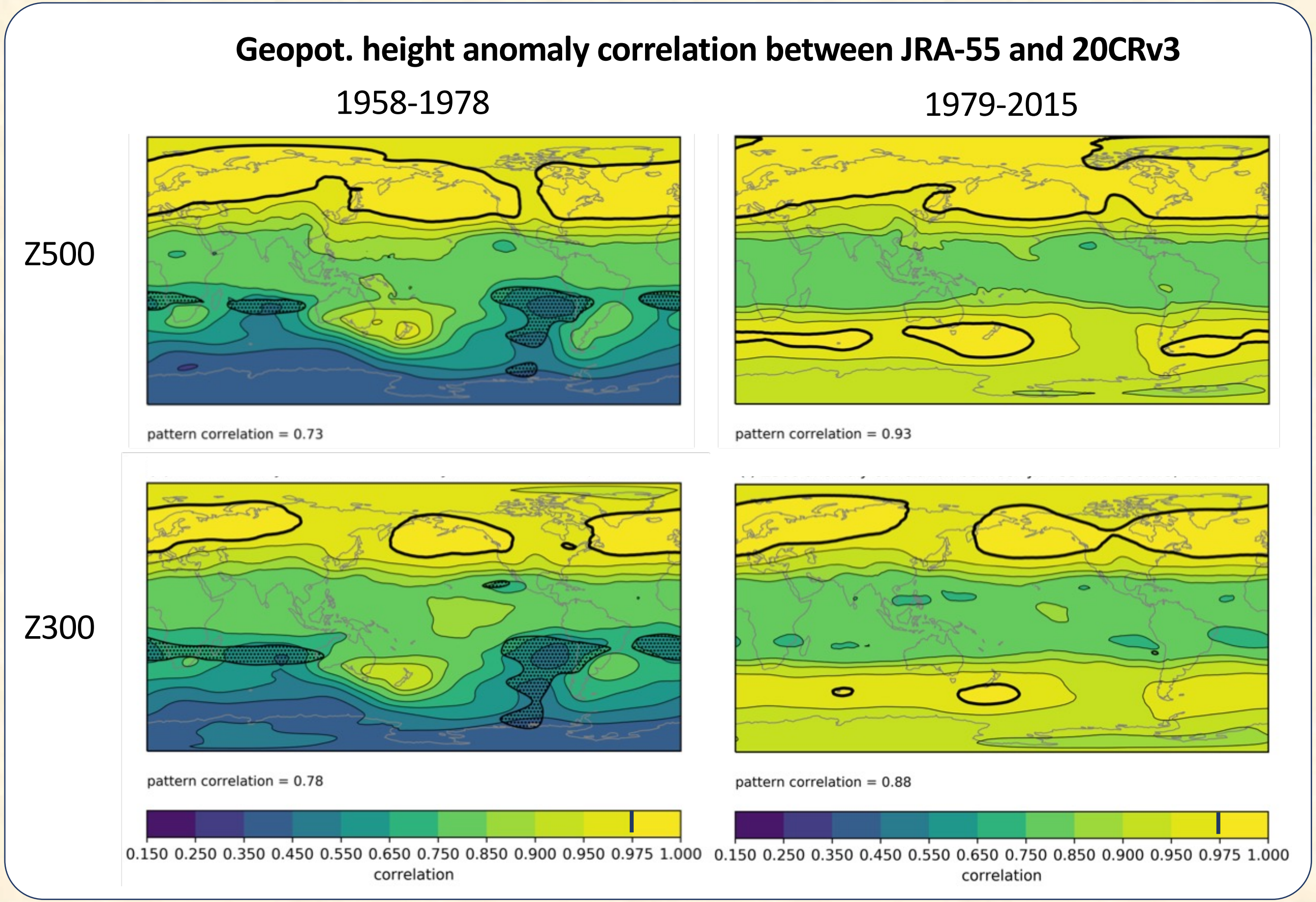
Difference between observed and analyzed values, as function of 20CRv3 ensemble spread; 1943-2015



- If obs were perfect (zero error), then RMSDs should fall on diagonal.
- If ob error range estimated accurately and system works well, RMSDs ideally fall in gray swath.
- Above swath: 20CRv3 is overconfident. Below swath: underconfident.
- **20CRv3 geopot. height analysis performs well globally at several vertical levels**

Slivinski et. al. (2021)

20CRv3 correlates well with other reanalyses, and can "predict" that correlation



(Similar results for 20CRv3/ERA5)

Slivinski et. al. (2021)

Like any EnKF, this system will suffer from filter divergence and/or spurious long-distance correlations without inflation and localization.

- **Inflation**: prevents "ensemble collapse" by artificially spreading out ensemble members.
  - Simple example: multiplying the ensemble covariance by a predefined (often tuned) factor larger than 1
- **Localization**: prevents an observation from incrementing the state at unreasonably long distances.
  - Simple example: Gaspari-Cohn localization, a function applied to the background covariance matrix which smoothly tapers long-distance correlations to zero

Like any EnKF, this system will suffer from filter divergence and/or spurious long-distance correlations without inflation and localization.

- **Inflation**: prevents "ensemble collapse" by artificially spreading out ensemble members.
  - Simple example: multiplying the ensemble covariance by a predefined (often tuned) factor larger than 1
- **Localization**: prevents an observation from incrementing the state at unreasonably long distances.
  - Simple example: Gaspari-Cohn localization, a function applied to the background covariance matrix which smoothly tapers long-distance correlations to zero

Inflation

- Previous version of reanalysis used **pre-defined multiplicative inflation** factors based on year and location
- Needed larger inflation factors for densely-observed times and places; smaller inflation for sparsely-observed times and places

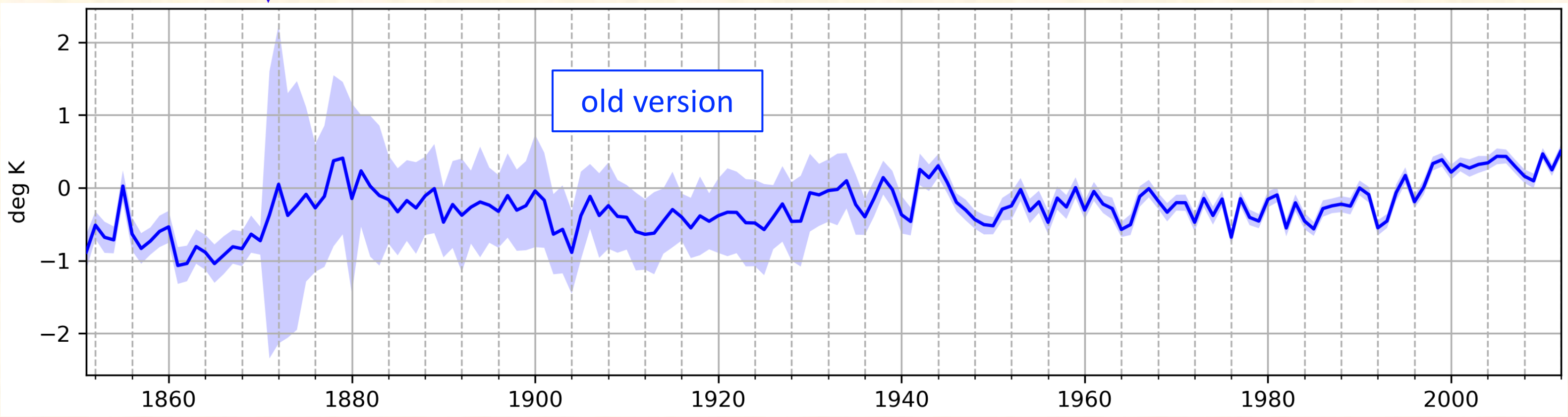
	Northern Hemisphere	Tropics	Southern Hemisphere
1851 – 1870	1.01	1.01	1.01
1871 – 1890	1.05	1.01	1.01
1891 – 1920	1.09	1.02	1.01
1921 – 1950	1.12	1.03	1.02
1951 – 2012	1.12	1.07	1.07

Simple adaptive inflation

	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



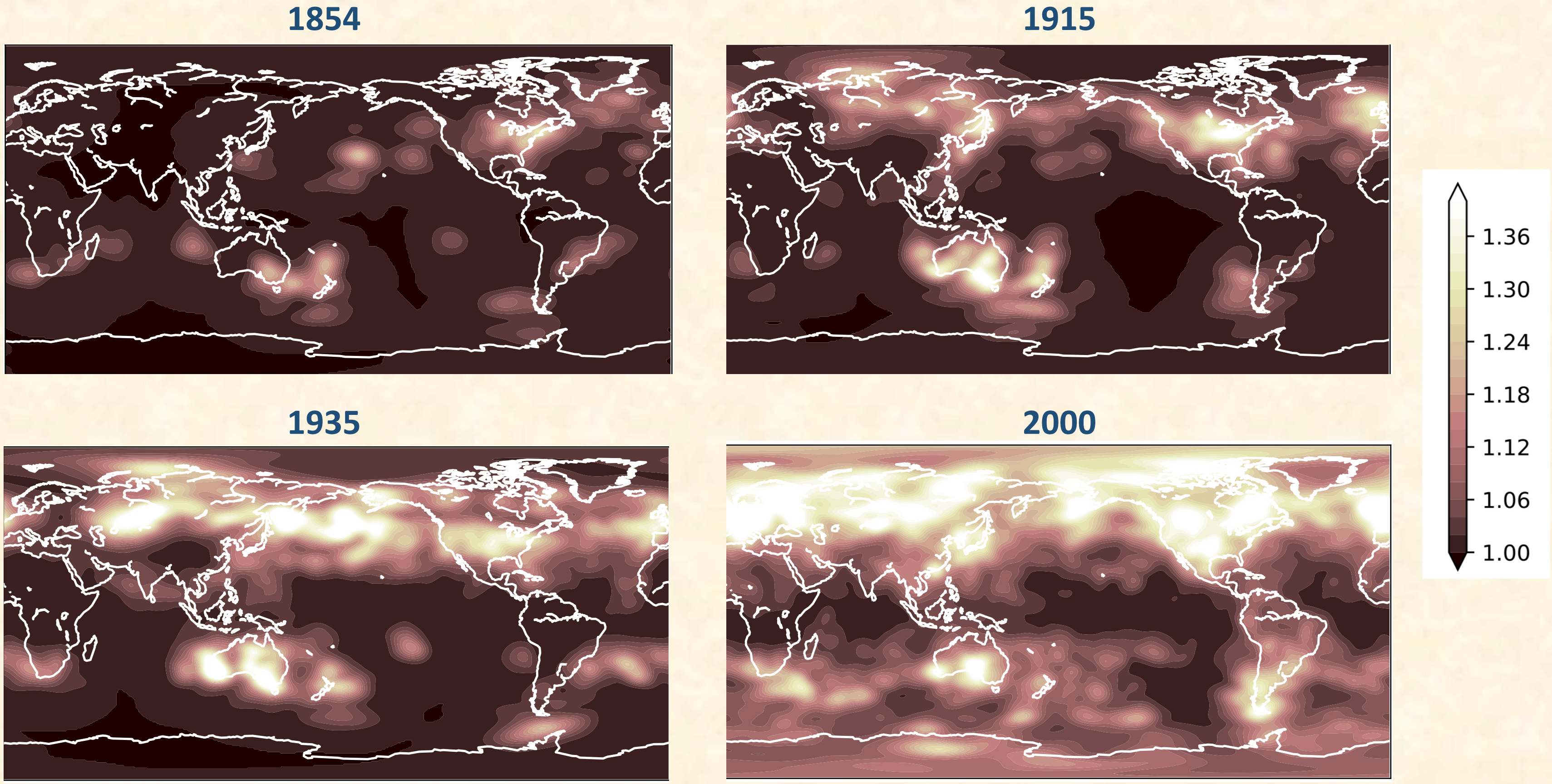
New inflation: relaxation-to-prior-spread

- Adaptive inflation: larger inflation when observations are dense, smaller inflation when observations are sparse
- Inflation parameter λ_{inf} is defined as function of individual gridpoints (x,y) and timesteps (t):

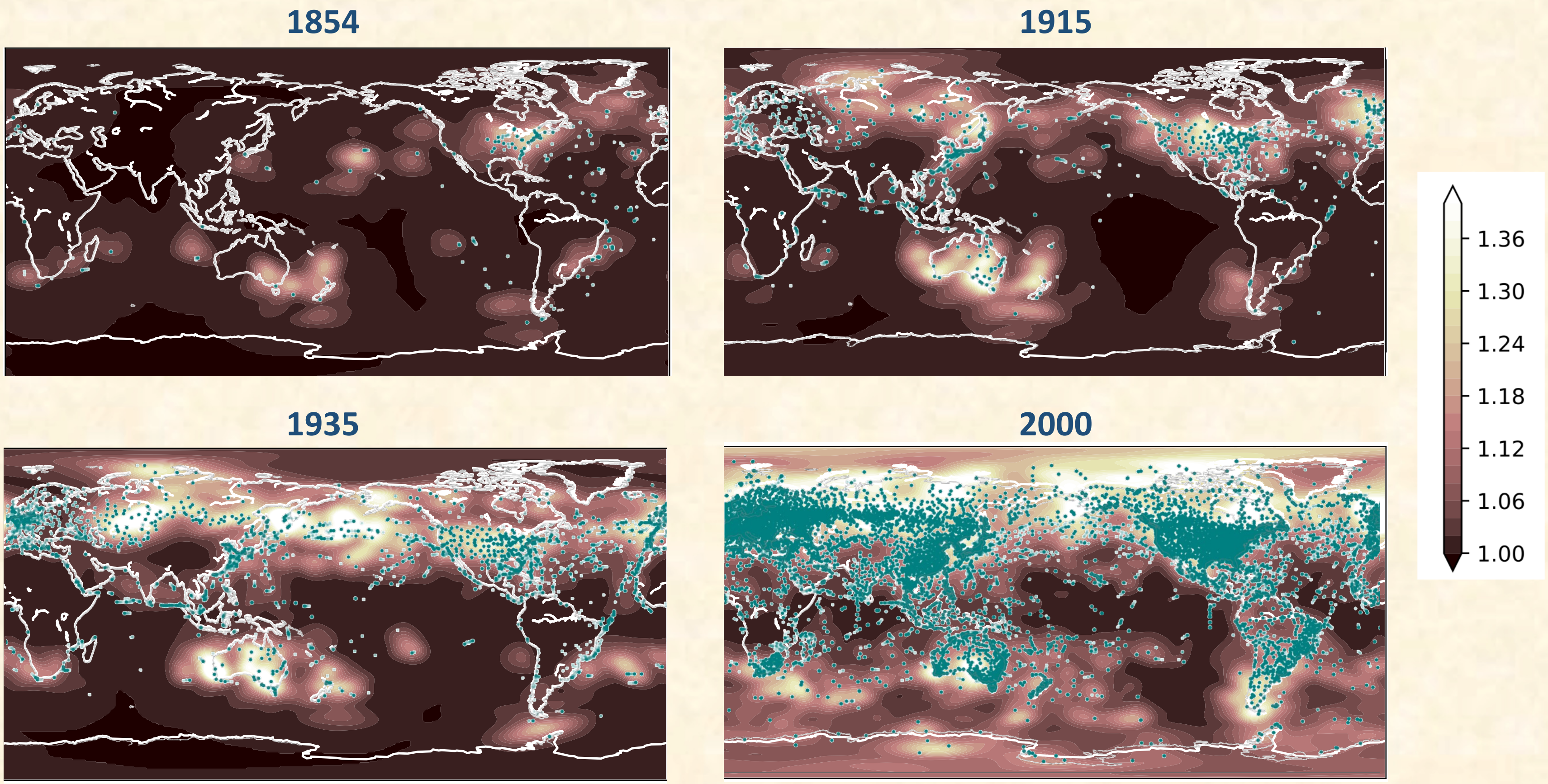
$$\lambda_{inf}(x, y, t) = p_{relax}\left(\frac{\sigma_b(x, y, t) - \sigma_a(x, y, t)}{\sigma_a(x, y, t)}\right) + 1,$$

σ_b is background ensemble standard deviation,
σ_a is analysis ensemble standard deviation, and
p_{relax} is a parameter varying from 0 (no inflation) to 1 (inflate fully to prior spread)

New inflation: relaxation-to-prior-spread



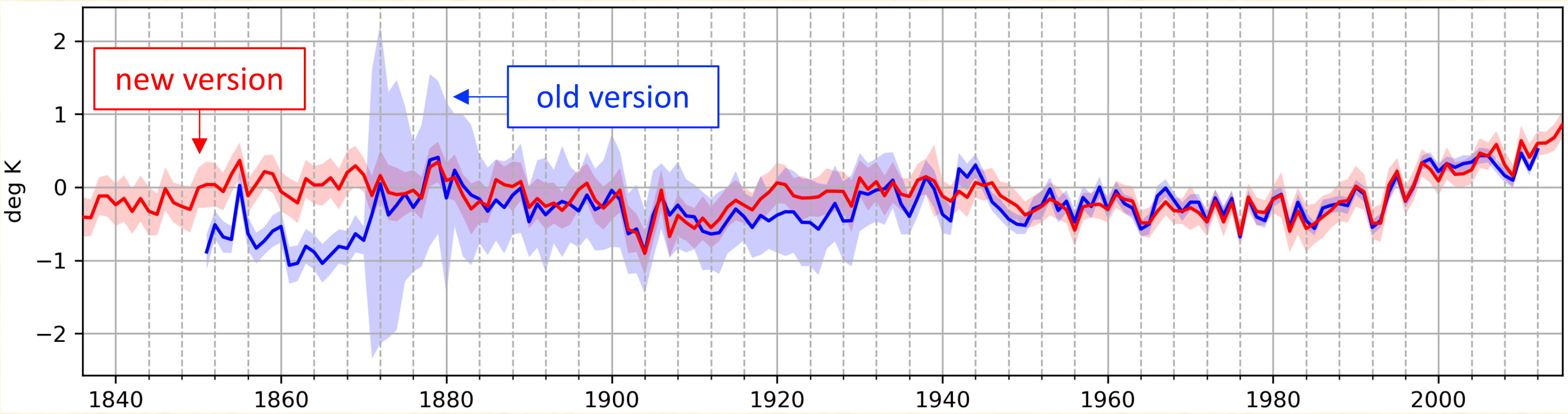
New inflation: relaxation-to-prior-spread



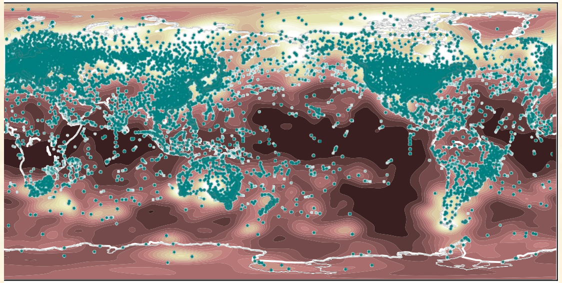
New inflation: relaxation-to-prior-spread

- More accurate, consistent estimates of uncertainty
- Can make stronger statements about trends

Atmospheric layer temperature anomalies, Northern Hemisphere



- **Inflation**: prevents "ensemble collapse" by artificially spreading out ensemble members.
    - Simple example: multiplying the ensemble covariance by a predefined (often tuned) factor larger than 1
- **Localization**: prevents an observation from incrementing the state at unrealistically long distances.
    - Simple example: Gaspari-Cohn localization, a function applied to the background covariance matrix which smoothly tapers long-distance correlations to zero
    - "Long-distance" defined in terms of a localization radius (can be fixed)



- **Inflation**: prevents “ensemble collapse” by artificially spreading out ensemble members.
  - Simple example: multiplying the ensemble covariance by a predefined (often tuned) factor larger than 1
- **Localization**: prevents an observation from incrementing the state at unrealistically long distances.
  - Simple example: Gaspari-Cohn localization, a function applied to the background covariance matrix which smoothly tapers long-distance correlations to zero
  - “Long-distance” defined in terms of a localization radius (can be fixed)
  - **Adaptive localization: allow the localization radius to change for each observation, based on that observation’s expected influence**

Adaptive localization

- **Hypothesis**: for a given single observation, the optimal localization length scale is proportional to the reduction of ensemble variance in observation space
  - In other words, the more a given observation would reduce the analysis covariance, the longer localization length it is given.
- Define ρ as the reduction of ensemble covariance in observation space:

$$\rho = \mathbf{H}\mathbf{P}^a\mathbf{H}^T/\mathbf{H}\mathbf{P}^b\mathbf{H}^T = \mathbf{R}/(\mathbf{H}\mathbf{P}^b\mathbf{H}^T+\mathbf{R})$$

where $\mathbf{H}$ is the linearized observation operator, $\mathbf{P}^a$ is the analysis ensemble covariance, $\mathbf{P}^b$ is the background ensemble covariance, and $\mathbf{R}$ is the observation error covariance

- Small $\rho \longleftrightarrow$ large reduction in variance $\longleftrightarrow$ useful observation $\Rightarrow$ larger optimal localization radius
- Conversely, ρ close to 1 implies a smaller optimal localization radius

Adaptive localization

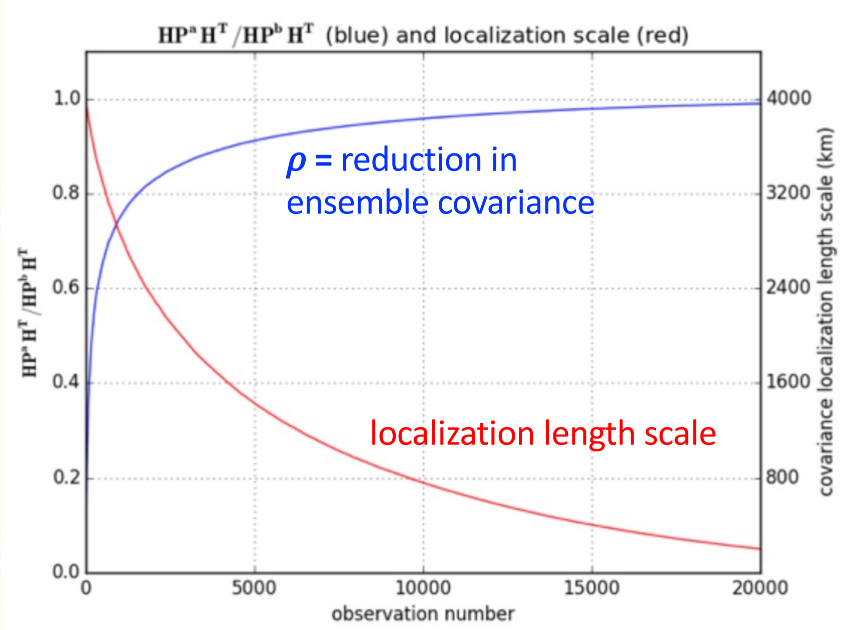
Empirically define

$$L = L_0(1 - e^{-(1-\rho)/r})$$

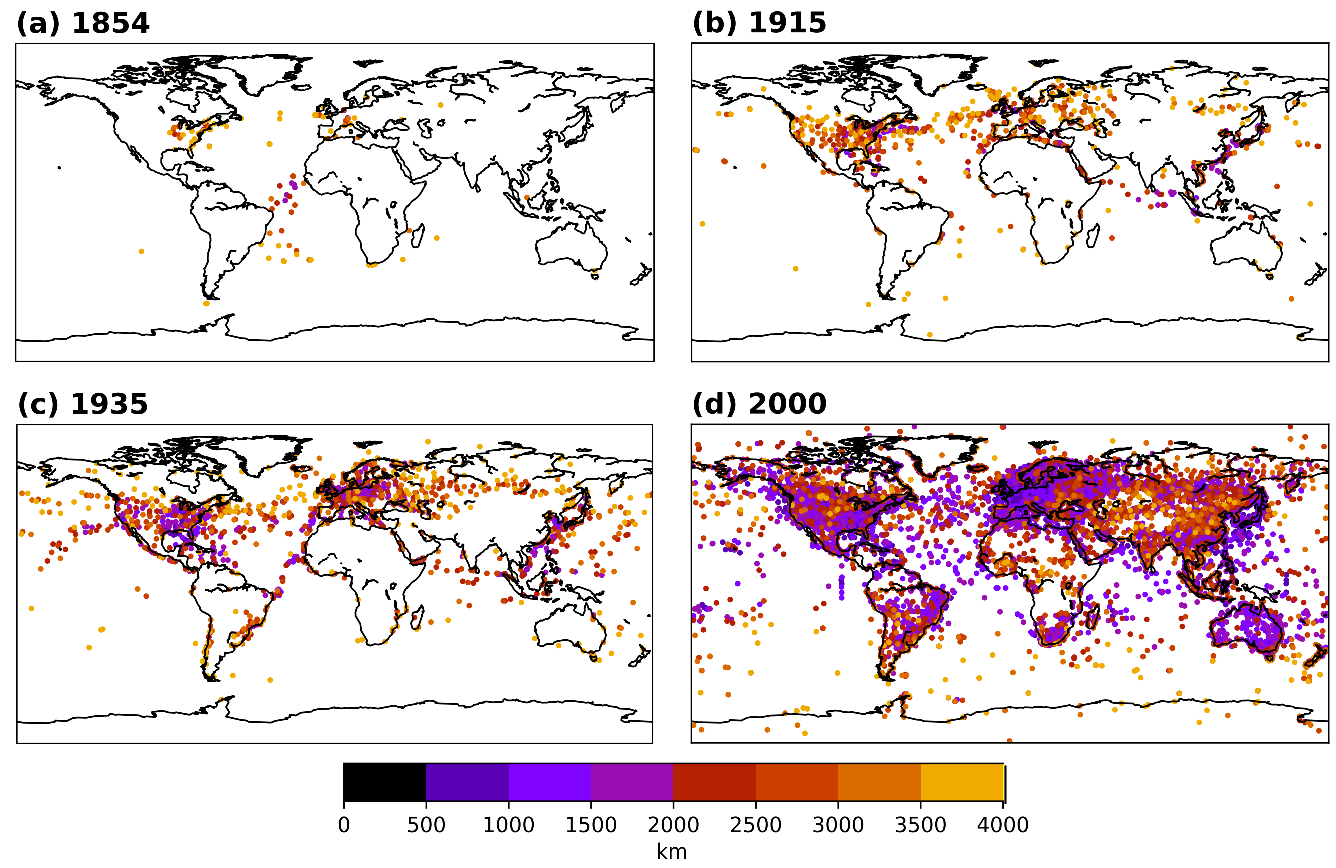
where

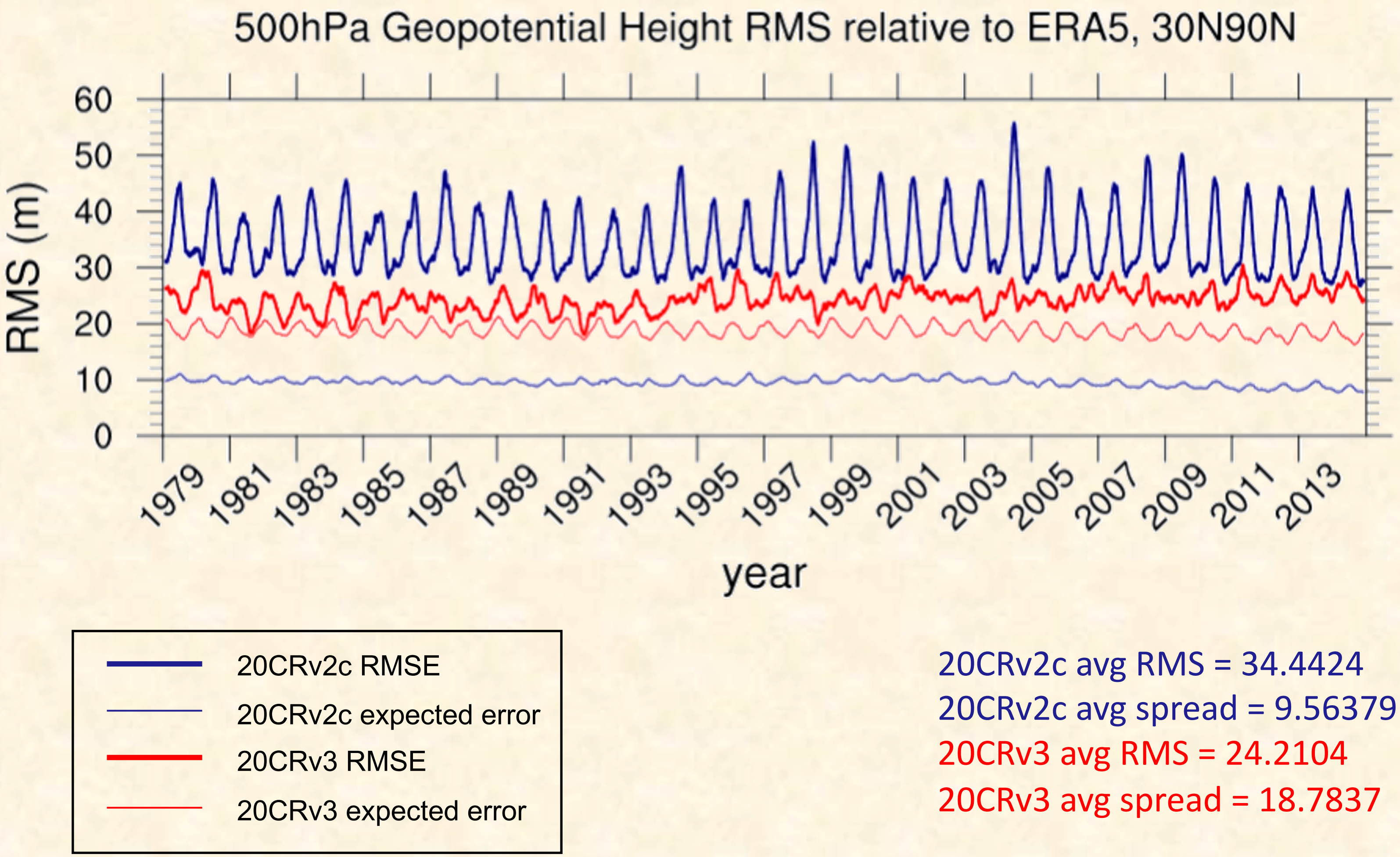
- L is the localization length scale
- L_0 is the maximum allowed localization length scale
- ρ is the reduction of ensemble variance in ob space (prev. slide); $\rho \in (0,1]$
- r is a parameter governing how tight the relationship between ρ and L is

Observations are assimilated serially in order of increasing ρ, and ρ is recomputed after each ob is assimilated.



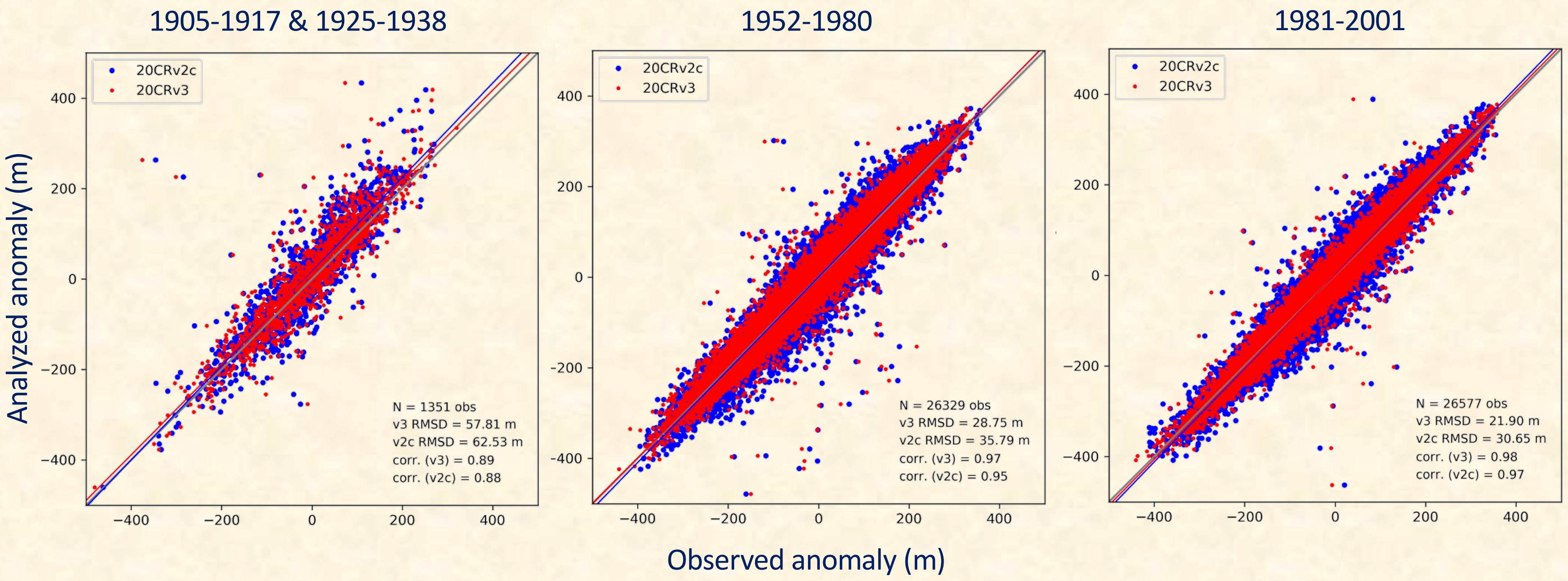
Adaptive localization





20CRv2c avg RMS = 34.4424
20CRv2c avg spread = 9.56379
20CRv3 avg RMS = 24.2104
20CRv3 avg spread = 18.7837

500hPa geopotential height analyzed anomalies versus observed anomalies from upper-air measurements at Lindenberg, Germany



Summary

- The 20th Century Reanalysis version 3 reconstructs nearly 200 years of sub-daily, global weather history by assimilating *only* surface pressure observations
- Reliable confidence and uncertainty estimates provided by 80 ensemble members in EnKF
- Updated inflation algorithm (relaxation-to-prior-spread) yields more consistent estimates of uncertainty
- Adaptively-varying localization makes observation-thinning unnecessary; more observations can be assimilated
- Overall, the 20CRv3 system includes many improvements over the previous 20CRv2c, leading to improved performance of the dataset (Slivinski et al 2019)

References

- 20th Century Reanalysis:
  - Homepage: https://go.usa.gov/XTd
  - Slivinski, L.C., G.P. Compo, J.S. Whitaker, P.D. Sardeshmukh, et. al. (2019) Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system. *Quarterly Journal of the Royal Meteorological Society,* **145**: 2876– 2908
  - Compo, G.P., Whitaker, J.S., Sardeshmukh, et. al. (2011) The Twentieth Century Reanalysis project. *Quarterly Journal of the Royal Meteorological Society*, **137**, 1–28
  - Compo, G.P., J.S. Whitaker, and P.D. Sardeshmukh, 2006: Feasibility of a 100-Year Reanalysis Using Only Surface Pressure Data. *Bulletin of the American Meteorological Society,* **87**, 175–190
  - Whitaker, J.S., G.P. Compo, X. Wei, and T.M. Hamill, 2004: Reanalysis without Radiosondes Using Ensemble Data Assimilation. *Monthly Weather Review,* **132**, 1190–1200
- Inflation:
  - Anderson, J.L. and Anderson, S.L. (1999) A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*, **127**, 2741–2758
  - RTPS: Whitaker, J.S. and Hamill, T.M. (2012) Evaluating methods to account for system errors in ensemble data assimilation. *Monthly Weather Review*, **140**, 3078–3089
- Localization:
  - Houtekamer, P.L. and Mitchell, H.L. (1998) Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review*, **126**, 796–811
  - Houtekamer, P.L. and Mitchell, H.L. (2001) A sequential ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, **129**, 123–137
  - Hamill, T.M., Whitaker, J.S. and Snyder, C.(2001) Distance-dependent filtering of background-error covariance estimates in an ensemble Kalman filter. *Monthly Weather Review*, **129**, 2776–2790
  - Anderson, J.L. (2007) Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. *Physica D*, **230**, 99–111
  - Adaptive inflation: Appendix D of Slivinski et. al. (2019)