National Aeronautics and Space Administration
Goddard Institute for Space Studies

Goddard Space Flight Center
Sciences and Exploration Directorate
Earth Sciences Division

May 9 2018

# Reproducibility and Replication in Climate Science

Gavin Schmidt, NASA GISS

For the National Academies of Sciences, Engineering, and Medicine, Committee on Reproducibility and Replicability in Science

Goddard Institute for
Space Studies

# Topics covered in response to cmtee's Qs.

GCM results and robustness

Local vs. general replicability/reproducibility

CMIP process and archives

Raw vs. derived data archives

Code archives: Analyses & GCMs

Operational data products (e.g. GISTEMP)

Paleo-data time-series/collection challenges

Thoughts & Recommendations

# Examples of non-robustness in GCMs?

Goddard Institute for
Space Studies

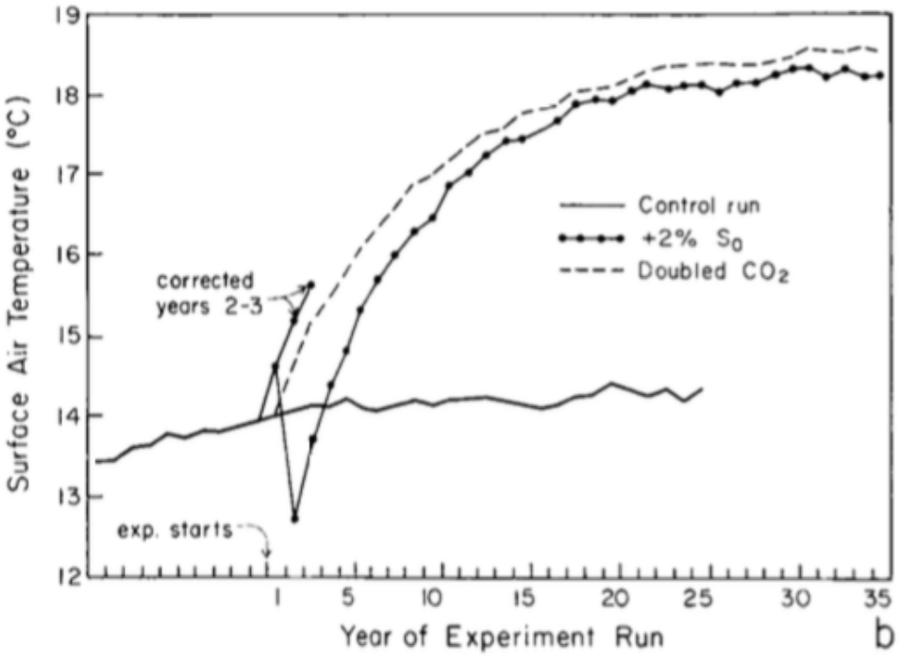
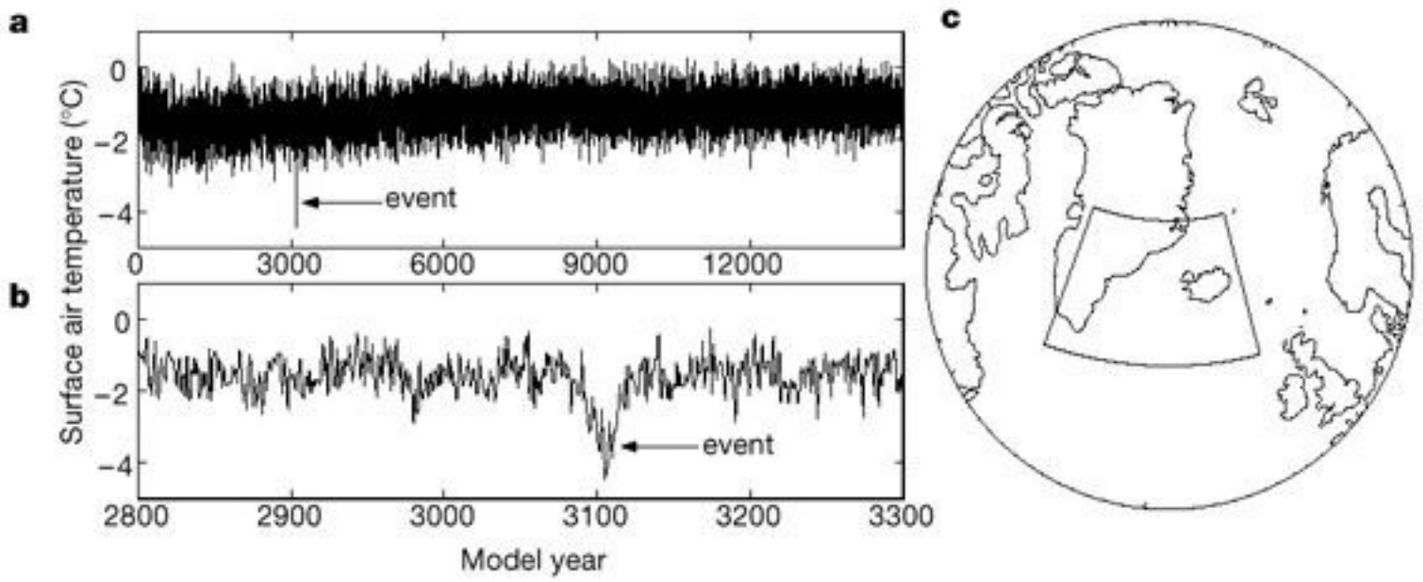
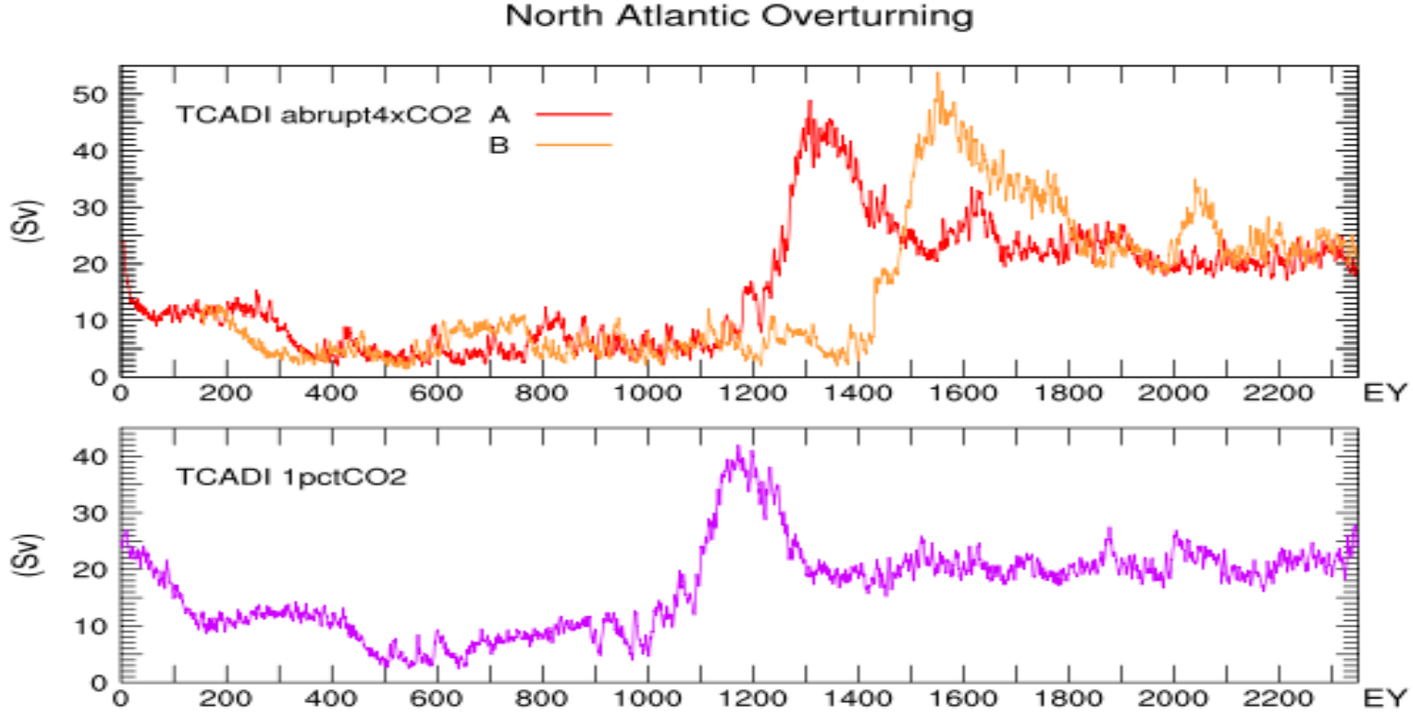



Fig. 3. Global net heat flux into planetary surface (a) and global surface air temperature (b). On April 1 of year 2 in the $S_0$ experiment the computer was hit by a cosmic ray or some other disturbance which caused improper numbers to be stored in the ground temperature array. This affected the temporal development of that run, but should not influence its equilibrium results. In order to determine the maximum heat flux into the ocean, the $S_0$ experiment was rerun for years 2 and 3 from March 31 year 2 thus eliminating the computer error for that period.

Hansen et al (1984)




Hall and Stouffer (2001)




Rind et al (2018, in press)

Goddard Institute for
Space Studies

# GCM R&R challenges

GCMs are locally replicable, but not generally

- Bit-replicability on local system is built-in & essential for debugging/testing.
- Environment is fragile (compilers/libraries/OS are not stable over time)
- e.g. even with code/inputs, old simulations are not replicable without environment ‘containers’.

Reproducibility of results more important (w/diff models/ parameter choices/background climate etc.), but rarely done outside of organized multi-model ensembles.

Goddard Institute for Space Studies

# GCM R&R challenges: The CMIP solution

Coupled Model Intercomparison Project (CMIP) (& predecessor/sub-projects) are community-driven standardized simulations w/common publicly available outputs

- Complete buy-in from global modelling groups (since CMIP3 ca. 2004)
- Funded mostly via uncosted mandate + agency support for ESGF
- Increasingly complex/time consuming/massive (CMIP5: 4 PB, CMIP6: ~100 PB)
- Very easy to test reproducibility/robustness of results across ensemble, but unclear how to interpret (i.e. it isn't a pdf).
- Data access open, but ability to do complex multivariate analysis limited by bandwidth
- No official support for archiving of analysis code/derived or intermediate data
- No server side analytics (yet)

# Raw and derived data archives

Goddard Institute for Space Studies

Climate science has 3 massive data streams:

- Remote sensing (NASA/NOAA/Japan/ESA/etc.): continuous global coverage/multi-variable
- Analyses/Reanalyses: Operational weather forecasts and Hindcast weather forecasts w/fixed model.
- Coupled GCMs (i.e. via CMIP etc.)

Almost all ‘raw’ data (Level 2+ for remote sensing) available quickly. No joint archives/derived data storage.

Biggest need? Process-based diagnostics across all three archives

Goddard Institute for
Space Studies

# Code archives

Many standard toolkits for looking at climate data:

- nco, cdat, python/R/matlab/IDL libraries

Many standard code repositories (github, Jupyter, etc.)

Archiving analysis code is haphazard

Most (not all) GCMs have public releases of frozen codes (i.e. CMIP versions). No GCM is open source(?)

Specific repository tags for experimental/development versions exist, but not generally publicly accessible.

No standardized archive of specific versions (w/ configuration files/input data).

Goddard Institute for Space Studies

# Operational data product: GISS Surface Temperatures (GISTEMP)

Product originally developed in 1981, updated methods, data inputs, homogenizations, urban corrections over time.

Only uses publicly available data. Currently:

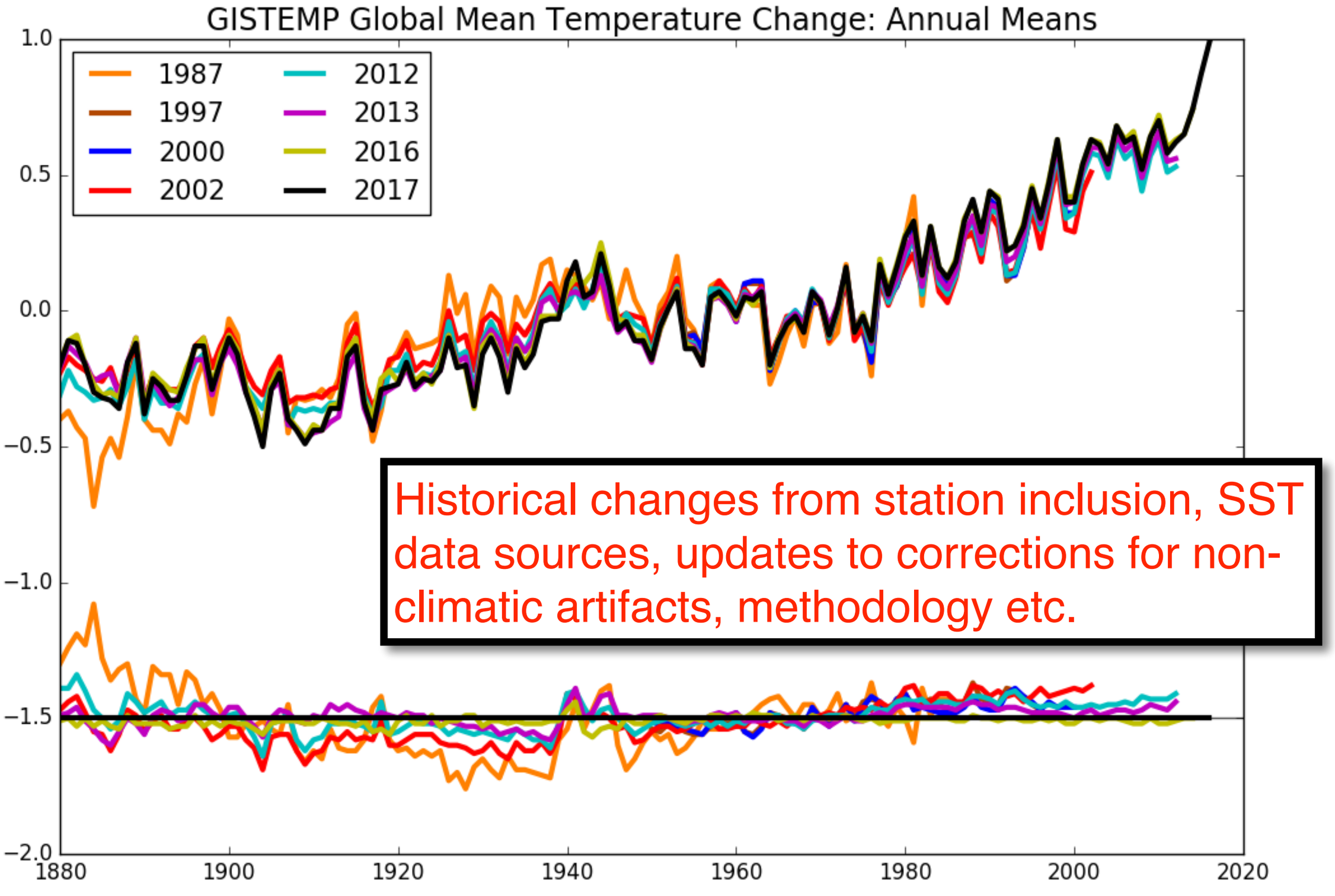
- NOAA NCEI GHCN, NOAA ERSSTv5, SCAR
- New open source data sets soon: ISTI/GHCN4

Analysis code available since 2007. Recoded in python by external ‘citizen scientists’ ~ 2011, now basis for operational analysis.

Homogenization process calculated de novo every month

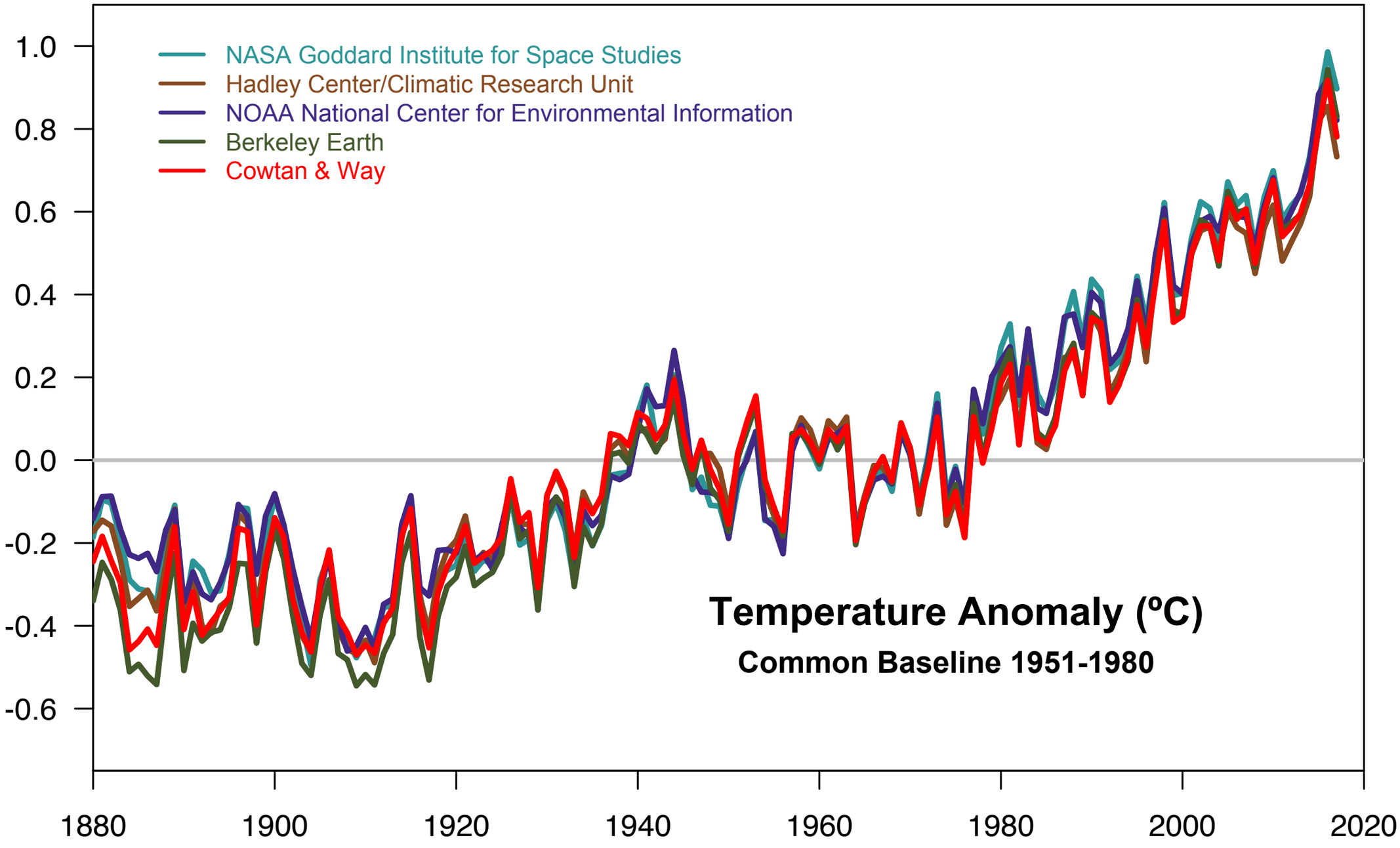
# History of GISTEMP product over time

Goddard Institute for
Space Studies



# GISTEMP: Replicable and Reproducible

Goddard Institute for
Space Studies



Goddard Institute for Space Studies

# The unique challenges of paleo data archiving

Proxy systems do not necessarily have stable relationships to ‘standard variables’ over time.

Age models are preliminary and often subject to change

Down-core archiving is much easier than time period archives

Multi-proxy records needed for spatial coverage/bias reduction

Current archives are too ‘dumb’ - freezing in place ‘as published’ (often obsolete) data, not automatically machine readable etc.

Goddard Institute for Space Studies

# The unique challenges of paleo data archiving: Solutions

Better standards for meta-data:

- Linked Paleo Data (LiPD.net)
- PAGES-2K (Kaufman et al, 2018; doi: 10.5194/cp-14-593-2018 - see peer review discussion too)

Intelligent archives that can update age models, take account of uncertainties, recalculate syntheses interactively.

- Smithsonian-led “Phantastic” project
- paleo-CO2 (RCN)
- Liesecki ‘stack’

Goddard Institute for
Space Studies

# Some general thoughts

Reproducibility has a cost to community (less de novo explorations of parameter/structure space), benefit is understanding robustness of results (modulo common assumptions).

Replicability cost is borne by provider, but many benefits to community (less time reinventing the wheel), easier sensitivity tests.

Organized homes for replications, variations, code, and derived data are needed.

Goddard Institute for Space Studies

# Specific recommendations…

1) Greater use of institutionally supported thematic/multi-project archives of code and derived/intermediate data

- Current practice is haphazard, piecemeal and incoherent

2) Easier routes to publication comments and replications

- See RealClimate post


**Someone C.A.R.E.S.**

Filed under: Climate Science Scientific practice — gavin @ 25 February 2017 Edit This

**Do we need a new venue for post-publication comments and replications?**

Social media is full of commentary (of varying degrees of seriousness) on the supposed replication crisis in science. Whether this is really a crisis, or just what is to be expected at the cutting edge is unclear (and may well depend on the topic and field). But one thing that is clear from all the discussion is that it's much too hard to publish replications, or even non-replications, in the literature. Often these efforts have to be part of a new paper that has to make its own independent claim to novelty before it can get in the door and that means that most attempted replications don't get published at all.

3) Standardized repositories for snapshots of GCM code/config/inputs

- Is entire runtime environment useful?

4) Intelligent archives for paleo-data

- Build in versioning/age model construction/uncertainty