The Reproducibility of Observational Estimates of Surface and Atmospheric Temperature Change
B. D. Santer, *et al.*
Science **334**, 1232 (2011);
DOI: 10.1126/science.1216273

technologies attract the attention of many investigators, especially in large consortia, thereby driving data reproducibility in a field. Funding incentives, reproducibility rewards and/or nonreproducibility penalties, and targeted requirements for repeatability checks may enhance the public availability of useful data and valid analyses.

References and Notes

1. J. P. A. Ioannidis, *PLoS Med.* **2**, e124 (2005).
2. S. J. Chanock *et al*., *Nature* **447**, 655 (2007).
3. R. D. Riley *et al*., *Health Technol. Assess.* **7**, 1 (2003).
4. A. Dupuy, R. M. Simon, *J. Natl. Cancer Inst.* **17**, 99 (2007).
5. P. J. Castaldi, I. J. Dahabreh, J. P. Ioannidis, *Brief. Bioinform.* **12**, 189 (2011).
6. M. Jelizarow, V. Guillemot, A. Tenenhaus, K. Strimmer, A. L. Boulesteix, *Bioinformatics* **26**, 1990 (2010).
7. S. M. Teutsch *et al*., EGAPP Working Group, *Genet. Med.* **11**, 3 (2009).
8. J. A. Sparano, S. Paik, *J. Clin. Oncol.* **26**, 721 (2008).
9. F. Cardoso *et al*., *J. Clin. Oncol.* **26**, 729 (2008).
10. K. Baggerly, K. R. Coombes, *Ann. Appl. Stat.* **3**, 1309 (2009).
11. K. Baggerly, *Nature* **467**, 401 (2010).
12. F. Andre *et al*., *Nat. Rev. Clin. Oncol.* **8**, 171 (2011).
13. J. P. A. Ioannidis, O. A. Panagiotou, *JAMA* **305**, 2200 (2011).
14. A. A. Al-Sheikh-Ali, W. Qureshi, M. H. Al-Mallah, J. P. Ioannidis, *PLoS ONE* **6**, e24357 (2011).
15. J. P. A. Ioannidis *et al*., *Nat. Genet.* **41**, 149 (2009).
16. A. D. Johnson, R. Leslie, C. J. O'Donnell, *PLoS Genet.* **7**, e1002269 (2011).
17. D. F. Ransohoff, *Clin. Chem.* **56**, 172 (2010).
18. A. W. Bell *et al*., HUPO Test Sample Working Group, *Nat. Methods* **6**, 423 (2009).
19. H. Mischak *et al*., *Sci. Transl. Med.* **2**, 46ps42 (2010).
20. M. Mann, *Nat. Methods* **6**, 717 (2009).
21. P. Legrain *et al*., *Mol. Cell. Proteom.* **10,** M111.009993 (2011).

Acknowledgments: The opinions expressed in this paper are those of the authors and do not reflect an official position of the Department of Health and Human Services.

10.1126/science.1211811

PERSPECTIVE

The Reproducibility of Observational Estimates of Surface and Atmospheric Temperature Change

B. D. Santer,[1]* T. M. L. Wigley,[2] K. E. Taylor[1]

Although concerns have been expressed about the reliability of surface temperature data sets, findings of pronounced surface warming over the past 60 years have been independently reproduced by multiple groups. In contrast, an initial finding that the lower troposphere cooled since 1979 could not be reproduced. Attempts to confirm this apparent cooling trend led to the discovery of errors in the initial analyses of satellite-based tropospheric temperature measurements.

"We have produced, using objective techniques, a long-term series of average Northern Hemisphere temperatures" (*1*, *2*). This innocuous sentence was published in 1982 by Phil Jones and colleagues at the University of East Anglia's Climatic Research Unit (CRU). The sentence was a prologue to the modern era of scientific attempts to estimate large-scale changes in the Earth's average surface temperature. Building on earlier work by American, British, Russian, and Japanese teams [reviewed in (*1*)], CRU researchers took on the difficult challenge of transforming surface temperature measurements from many hundreds of meteorological stations into credible scientific estimates of temperature changes over land areas of the planet. Temperature changes over ocean areas (*3*, *4*) and over the whole globe (land plus ocean) were soon to follow (*5*). Few could have imagined the far-reaching scientific, societal, and political repercussions of this seemingly routine research.

The CRU team soon joined forces with scientists at the UK Meteorological Office Hadley Centre (MOHC), who were refining estimates of observed changes in sea-surface temperature (SST). This scientific partnership led to the development of the Hadley Centre/CRU observational record of combined changes in SST and land-surface temperature (HadCRUT). The HadCRUT data set has provided hard scientific evidence for the warming of Earth's surface over the past 150 years (*6*).

The Hadley Centre and CRU efforts to construct successive versions of the HadCRUT data set were open and transparent, and are documented in many peer-reviewed papers. From the very beginning of this research, CRU and MOHC scientists recognized the difficulties involved in estimating the true (but unknown) temperature change in the physical climate system. In order to do this, it was necessary to account for the effects of nonclimatic factors, such as temporal changes in the type of thermometer used to make temperature measurements, the thermometer location and its immediate physical surroundings, and the time of observation. Jones and colleagues found that even if they made different (but reasonable) choices in data set construction, their bottom-line conclusion—that the surface of our planet experienced rapid warming over the second half of the 20th century—was rock solid.

"An extraordinary claim requires extraordinary proof" (*7*). The claim that the planet had warmed markedly during the 20th century had extraordinary societal implications, and was therefore subjected to extraordinary scrutiny. Groups at the NASA/Goddard Institute for Space Studies in New York (GISS) and at the National Oceanic and Atmospheric Administration's National Climatic Data Center (NCDC) in North Carolina independently attempted to reproduce the HadCRUT results. Although all three teams used raw temperature measurements from similar (but nonidentical) sets of observing stations, they made different choices in the treatment of these raw measurements and the calculation of area averages (*8*). In spite of these differences, the GISS and NCDC analyses confirmed the "warming Earth" findings of the CRU and MOHC scientists (*9*, *10*).

Other lines of evidence substantiated the CRU/MOHC, GISS, and NCDC estimates of planetary temperature increase. The surface warming was consistent with the independently monitored retreat of snow and Northern Hemisphere sea-ice cover, the widespread melting and retreat of glaciers, the rise in global-mean sea level, and the increase in the amount of water vapor in the atmosphere. These and many other independent observations provided the scientific underpinning for the finding that "warming of the climate system is unequivocal" (*11*).

Yet doubts about the reality of 20th-century surface warming remained, especially in the aftermath of the 2009 event colloquially referred to as "Climategate."

Reproducibility—the independent verification of prior findings—is at the core of "the spirit of science" (*12*). An additional verification effort has helped to underscore the robustness of the finding of "unequivocal" surface warming. Richard Muller, representing a fourth research group, testified to a Congressional committee in March of 2011 that his team had independently reproduced the land-surface warming found previously by the other three groups (*13*). The data set developed during this verification study has now been publicly released, and scientific papers describing the findings of this fourth group are currently undergoing peer review.

[1]Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA. [2]National Center for Atmospheric Research, Boulder, CO 80307–3000, USA.

*To whom correspondence should be addressed. E-mail: santer1@llnl.gov

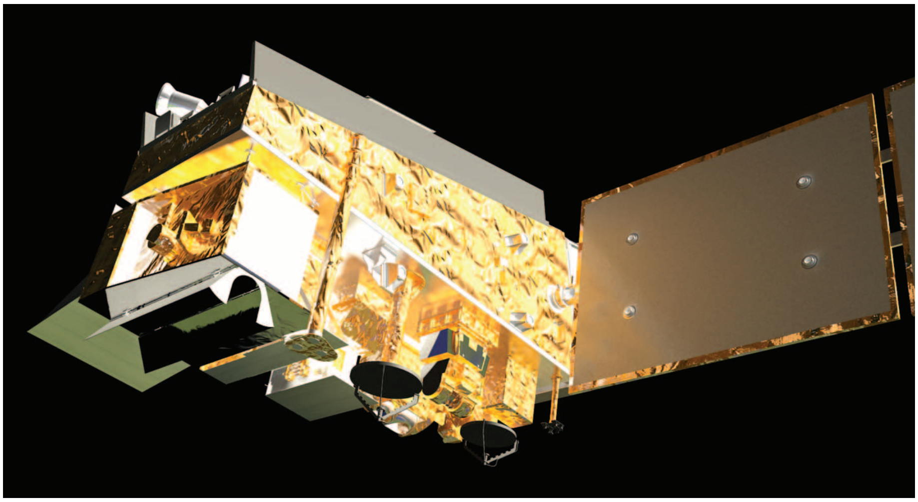

Fig. 1. The NPP weather and climate satellite. [Credit: National Oceanic and Atmospheric Administration]

Although "climate auditors" have closely scrutinized observational estimates of large-scale surface warming, comparatively little attention has been devoted to the "auditing" of satellite-based estimates of changes in the temperature of the lower troposphere (TLT) (Fig. 1). This is because such independent verification work requires substantial technical expertise in satellite microwave radiometry. Until 2005, Microwave Sounding Unit (MSU) measurements of TLT changes developed by scientists at the University of Alabama at Huntsville (UAH) suggested that the lower troposphere had cooled since 1979, which was in sharp contrast to all surface temperature results (*14*). This apparent cooling was used to cast doubt on the reality of surface warming, and on the credibility of climate model simulations of the tropospheric warming response to human-caused changes in atmospheric composition (*15*).

The UAH claims of a cooling troposphere did not withstand rigorous scientific scrutiny. Scientists at Remote Sensing Systems (RSS) in California identified two serious errors in the UAH TLT record. One UAH error arose from neglecting the effects of progressive orbital decay and altitude loss over the lifetimes of individual satellites (*16*). A second UAH error was caused by use of the wrong sign in adjusting for the effects of satellite orbital drift on the sampling of the diurnal temperature cycle (*17*). The effect of both errors was to introduce a spurious cooling trend into the UAH TLT record. The UAH team corrected these errors, and their latest results show a global-scale lower tropospheric warming of roughly 0.5°C over the full satellite era (1979 to present) (*18*). This is in good accord with the RSS TLT results and provides further independent corroboration of the reality of surface warming.

The scientists involved in attempts to reproduce initial claims of surface warming—and initial claims of lower tropospheric cooling—deserve considerable credit. Such work is difficult and time-consuming but of great scientific importance. These verification efforts have reinforced our confidence in the reality of surface warming, and have helped to identify the errors that led to initial "cooling troposphere" findings. Correction of these errors has yielded greater internal and physical consistency in observational estimates of surface and tropospheric warming trends (*19*).

Despite this scientific progress, there are still unresolved differences between UAH and RSS estimates of tropospheric temperature change in the tropics. These residual observational uncertainties have important implications for our ability to assess the true fidelity of climate model simulations of historical climate change. The UAH tropical TLT trend is roughly 0.1°C per decade smaller than that of RSS estimates—a difference large enough to have an impact on the statistical significance of model-versus-observed trend differences (*20*). Reducing this uncertainty will require better understanding of the underlying causes of differences between the UAH and RSS TLT retrievals. To achieve this, full public release of the codes used for generating tropospheric temperature retrievals would be highly desirable.

The development of homogeneous, climate-quality data records from satellite microwave radiometry is a challenging technical task. Analysts must account for complex, nonclimatic influences arising from drifts in satellite altitude, orbit, and instrument calibration; from biases between MSU instruments flown on different satellites; and from the merging of measurements obtained with older and more advanced MSUs. There is no single, demonstrably optimal way through the labyrinthine maze of plausible adjustment pathways for these and other factors. Unfortunately, independent measurements of atmospheric temperature change made by temperature sensors on weather balloons are also affected by a variety of nonclimatic factors, and are not an unambiguous "gold standard" for constraining differences in satellite-based estimates of tropospheric temperature change (*21*, *22*).

Improvements in our ability to reliably track changes in Earth's atmospheric temperature from space will require transitioning from an observing system designed for weather forecasting to a system geared toward long-term climate monitoring. At a minimum, this transition will involve ensuring adequate overlap between individual satellites and reliable means of establishing ground truth. Such a climate-observing system—which we do not yet have—represents an investment in our ability to measure and understand the evolving human imprint on climate.

References and Notes

1. P. D. Jones, T. M. L. Wigley, P. M. Kelly, *Mon. Weather Rev.* **110**, 59 (1982).
2. In view of the subsequent (and unjustified) portrayal of the first author of (*1*) as a "climate alarmist," the final sentence in the abstract is noteworthy: "There is, as yet, no statistical reason to associate the recent warming with atmospheric CO_2 increases."
3. C. K. Folland, D. E. Parker, F. E. Kates, *Nature* **310**, 670 (1984).
4. S. D. Woodruff, R. J. Slutz, R. L. Jenne, P. M. Steurer, *Bull. Am. Meteorol. Soc.* **68**, 1239 (1987).
5. P. D. Jones, T. M. L. Wigley, P. B. Wright, *Nature* **322**, 430 (1986).
6. P. Brohan, J. Kennedy, I. Harris, S. F. B. Tett, P. D. Jones, *J. Geophys. Res.* **111**, (D12), D12106 (2006).
7. M. Truzzi, *Zetetic Scholar* **1**, 11 (1978).
8. P. D. Jones, T. M. L. Wigley, *Clim. Change* **100**, 59 (2010).
9. J. Hansen, R. Ruedy, J. Glascoe, M. Sato, *J. Geophys. Res.* **104**, (D24), 30997 (1999).
10. T. M. Smith, R. W. Reynolds, T. C. Peterson, J. Lawrimore, *J. Clim.* **21**, 2283 (2008).
11. Intergovernmental Panel on Climate Change, in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, H. L. Miller, Eds. (Cambridge Univ. Press, Cambridge, UK and New York, 2007), p. 5.
12. www.guardian.co.uk/science/2011/feb/27/can-these-scientists-end-climate-change-war.
13. http://science.house.gov/sites/republicans.science.house.gov/files/documents/hearings/Muller%20Testimony%20rev2.pdf.
14. J. R. Christy, R. W. Spencer, W. B. Norris, W. D. Braswell, D. E. Parker, *J. Atmos. Ocean. Technol.* **20**, 613 (2003).
15. See, for example, (*23*). Schlesinger noted that "...the theory that increasing concentrations of greenhouse gases like carbon dioxide will lead to further warming is at least an oversimplification. It is inconsistent with the fact that satellite measurements over 35 years [sic] show no significant warming in the lower atmosphere, which is an essential part of the global-warming theory."
16. F. J. Wentz, M. Schabel, *Nature* **394**, 661 (1998).
17. C. A. Mears, F. J. Wentz, *Science* **309**, 1548 (2005).
18. B. D. Santer *et al.*, *J. Geophys. Res.*, **116**, D22105 (2011).
19. B. D. Santer *et al.*, *Science* **309**, 1551 (2005).
20. B. D. Santer *et al.*, *Int. J. Climatol.* **28**, 1703 (2008).
21. S. C. Sherwood, J. R. Lanzante, C. L. Meyer, *Science* **309**, 1556 (2005).
22. P. W. Thorne *et al.*, *J. Geophys. Res.* **116**, D12116 (2011).
23. J. Schlesinger, "Cold facts on global warming," *Los Angeles Times*, 22 January 2004.

10.1126/science.1216273