UNCERTAINTIES IN CLIMATE TRENDS
Lessons from Upper-Air Temperature Records

BY PETER W. THORNE, DAVID E. PARKER, JOHN R. CHRISTY, AND CARL A. MEARS

The range of trends in upper-air temperature datasets shows that dataset construction methodologies can add significant bias; we illustrate how this arises and offer potential means of reconciliation.

There has been much interest in understanding and reconciling rapid surface warming with upper-air temperatures, which were reported to have exhibited little if any warming over the satellite era (Folland et al. 2001; NRC 2000). This has led to a number of new satellite- (Christy et al. 2003; Mears et al. 2003; Grody et al. 2004), radiosonde- (Lanzante et al. 2003; Thorne et al. 2005), and reanalyses- (Uppala et al. 2005) based upper-air temperature datasets. Many of these agree in characterizing specific short-time-scale atmospheric features [e.g., El Niño–Southern Oscillation (ENSO), volcanic response], but diverge substantially in their long-term large-scale mean trends (Seidel et al. 2004). Their construction has actually served to increase the spread in reported long-term trends, nominally increasing our uncertainty.

We can no longer absolutely conclude whether globally the troposphere is cooling or warming relative to the surface. Clearly, however, the climate system has evolved in one unique way. Hence the challenge to the climate science community is to understand the reasons for the coherent differences between available datasets, and to discern the true climate evolution. The key first step is to understand the likely sources and causes of errors and biases. Only with this knowledge can we hope to truly reconcile the differences and gain a more complete and accurate picture of the true climate system evolution.

SOURCES OF UNCERTAINTY IN CLIMATE RECONSTRUCTIONS. An instrument is used to measure an atmospheric state variable, such as temperature. Once an observation has been made, it requires recording, sending, and collating into global digitized records. At all of these stages either errors or deliberate (but potentially incorrect and unrecorded) modifications to the raw data can be incorporated. Furthermore, observations have generally been made to satisfy the immediate demands of analysis for short-term weather forecasting with little emphasis on the long-term stability of the observing

system. Hence there have been numerous changes in instrumentation and observing practice to improve forecasts. Unfortunately these changes have often introduced spurious nonclimatic signals into long-term records, masking the true climate signal in the raw data. To make matters worse, many such changes have not been recorded and/or collated. The question of long-term stability of the network for climate monitoring purposes, and the specific requirements thereof, has only begun to be seriously addressed over the last decade or so.

Ideally, we would have an independent ground-truth measurement that did not change in characteristics over time, against which we could compare our historical databases so as to derive absolute adjustments for nonclimatic influences: a transfer standard. Unfortunately we have no effective transfer standard that we can use to unambiguously retrieve the true underlying time series from the available historical data for any climate variable. Therefore, we can only ever gain an estimate of the true climate evolution in creating climate datasets.

In any typical treatment to form a climate dataset, hundreds or thousands of individual adjustments can be applied. Because we have no absolute transfer standard, each of these individual adjustments will a priori retain a nonclimatic signal of unknown sign and magnitude regardless of how reasonable and physically plausible the chosen homogenization approach. There also remain serious issues of repeatability when seemingly unavoidable expert judgments have to be made to identify and adjust for spurious biases in the raw data. The probability of all the dataset treatment errors/biases summing exactly to zero in any final dataset, even on large space and time scales is infinitesimally small (and even then would not guarantee a zero-trend bias). Hence a number of seemingly physically acceptable methodologies for constructing a dataset from the same raw data will yield a range of solutions rather than converge to a single point solution.

In all observational datasets there exist two sources of uncertainty (error): structural uncertainty arises through the choice of approach; parametric uncertainty (also known as value or internal uncertainty) is the uncertainty given the chosen approach in the presence of a finite sample of data. Traditionally investigators have only considered parametric uncertainty when calculating error estimates. If they have considered structural uncertainty then it has been through tweaking their methodologies, but this yields a limited assessment of structural uncertainty. Different investigators starting from scratch may choose radically different approaches that still appear to be physically rigorous and yet come to significantly different solutions.

In each individual dataset, structural uncertainty adds systematic bias, but, aggregated over many independent, plausibly constructed datasets, it should be a random effect. So, increasing the number of datasets decreases the uncertainty; the "unmasking" of uncertainty by new analyses must not be used as an excuse not to produce independently derived datasets. When only a single dataset exists, and in the absence of any other information, our structural uncertainty must, by definition, be infinite. More datasets enable constraints to begin to be placed on the structural uncertainty and permit investigation of the causes of differences. So multiple independent efforts must be undertaken to create climate datasets. It is important that these efforts be truly independent to minimize the chances of results clustering around a single initial estimate or small set of estimates of the value of, for example, the long-term trend, which may eventually be proven wrong.

MICROWAVE SOUNDING UNIT CHANNEL-2 TIME SERIES: AN ILLUSTRATION OF STRUCTURAL UNCERTAINTY.

The (Advanced) Microwave Sounding Unit [(A)MSU] instrument is a passive microwave sensor that has been flown on consecutive National Oceanic and Atmospheric Administration (NOAA) polar orbiting satellites since mid-November 1978. The instrument measures upwelling microwave emissions from oxygen, and to a lesser extent, water vapor in deep atmospheric layers; these emissions are subsequently converted to brightness temperatures. Three groups (Christy et al. 2003, hereafter UAH; Mears et al. 2003, hereafter RSS; Grody et al. 2004, hereafter UM) have recently produced new time series based upon brightness temperatures from channel 2 (channel 5 on AMSU), which incorporates information from the surface to the lower stratosphere.

MSU structural trend uncertainty can be simplified to a flow diagram (Fig. 1). All three groups take identical input data from the satellite data archives at NOAA National Environmental Satellite, Data, and Information Service (NESDIS). These are then adjusted for the nonclimatic effects of known or strongly suspected causes of inter- and intrasatellite biases. In the absence of agreed to rigorous transfer standards between radiances measured by different satellites there is no objective way to specify the optimal correction procedure, so a degree of subjectivity is inevitably introduced.

It may prove dangerous to assume that our current range of estimates accurately characterizes structural uncertainty. Previous versions of UM and UAH have differed from their current versions by up to a similar magnitude to the currently reported trends as new effects have been recognized and accounted for, or adjustment procedures refined with improved physical understanding. For example, the initial version of the UM dataset reported a warming of 0.24 K decade^{-1} (Vinnikov and Grody 2003), but ignored the instrument body temperature effect, which both remaining groups found to be important. When this was taken into account the trend was reduced to 0.17 K decade^{-1} in closer agreement with other available estimates. Furthermore, even within a given approach choices can have a large effect. Using the "backbone satellite merging" method of UAH and repeating the sensitivity study of Christy et al. (1998) about choices of satellites to merge, but using channel 2 rather than 2LT, yields a range of 0.08 K decade^{-1} in the resulting trends. There are reasoned arguments that could be made for any of these merging pathway choices.

As a community we must assume that the latest dataset versions are the best estimates based upon investigators knowledge and experience using the data. However, clearly unknown, ignored, or poorly

characterized biases could explain at least part of the residual differences. Importantly, although the groups generally agree on the important nonclimatic influences there are some remaining disagreements. The true structural uncertainty may therefore be larger or smaller than the current range of estimates suggests.

Indeed, we could conceive of a very large number of physically plausible mechanisms for adjusting MSU time series for nonclimatic effects and hence create a distribution of plausible datasets. However, we have no information as to the trends that would result. A large spread in trends would imply that the satellite data are insufficiently constrained to improve our understanding of long-term climate change despite providing useful information on climate variability on shorter time scales. Conversely, a small spread of trends would imply that at least one of the current datasets is not physically plausible. Although creating such a range of datasets is highly desirable, there are serious resource implications that make such an explicit treatment of structural uncertainty impractical.

WHEN IS STRUCTURAL UNCERTAINTY LIKELY TO BE IMPORTANT? MSU is used in this essay solely as an example. Because there has been debate over the veracity of early UAH versions, a number of independent efforts have been made to construct and refine climate datasets from the raw MSU data. Unlike efforts to create climate datasets at the surface and from radiosonde records, the MSU groups have all used the same input data, meaning that differences will arise solely due to processing choices. Therefore MSU is the best current example that we can use to illustrate the problem of structural uncertainty. We stress that structural uncertainty will also be applicable to other climate variables and data from nonsatellite platforms.

We can generalize that from a hypothetical viewpoint structural uncertainty will be largest given the following "problems" with the raw data (in no particular order):

- if the variable is being measured by a single or very small number of instruments with little overlap;
- if the spatial sampling density is low, especially for variables with small spatial scales such as precipitation;
- if changes are pervasive across the observing network, especially so if they are undertaken coincidentally across large regions;
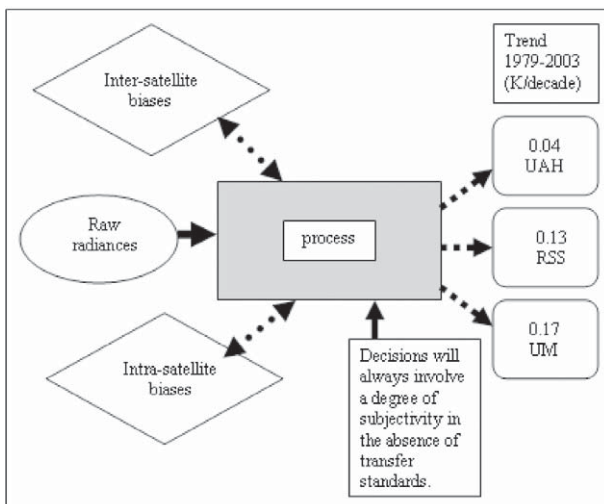- if the temporal sampling changes over time; or

Fɪɢ. I. Flow diagram to illustrate the processing of MSU channel-2 datasets from UAH, RSS, and UM. The processing of the raw data to account for the inter- and intrasatellite nonclimatic influences is uncertain and hence represented as a gray box. Arrows from the biases are bidirectional, since the processing typically both estimates and utilizes the various biases. Resulting decadal trend estimates are shown on the rhs and exhibit a spread of similar magnitude to the reported trend.

- if changes to observing procedures are poorly documented.

Plausibly the structural uncertainty will be larger for datasets constructed from relatively sparse "fire and forget" radiosondes or single-instrument satellite monitoring than from the relatively dense and stable surface observing network.

CAN WE CONSTRAIN STRUCTURAL UNCERTAINTY? Climate change detection studies are sensitive to the choice of a tropospheric temperature dataset (Thorne et al. 2003; Santer et al. 2003). As we do not know which observed upper-air temperature dataset is closest to the "truth" we are prevented from understanding the causal mechanisms behind observed climate change. Similar sensitivities are likely to be evident for many other climate variables. Hence, extensive efforts to understand climate model uncertainties (e.g., experiments such as www.climateprediction.net) will be greatly hindered without accompanying efforts to understand and rigorously constrain observational uncertainties in the quantities against which models are being compared.

We strongly caution that one cannot use model simulations to determine which observational datasets are "correct" as we may merely screen out observations that do not agree with our models, thereby making model errors invisible. An historical example of this was the automated screening out of low ozone levels over Antarctica that lead to a delay of several years in identifying and recognizing the importance of the ozone hole (e.g., Farman et al. 1985).

Independent data derived from other observing platforms or from reanalyses could be used to attempt to place useful constraints on our uncertainty in climate datasets. Such comparisons have already begun (Christy et al. 2003; Seidel et al. 2004), but their utility is severely limited if all datasets have serious flaws. Furthermore, any a priori weighting toward preferred datasets against which to compare could artificially bias intercomparison analyses and yield a false sense of agreement.

Returning to our MSU channel-2 example, analysis of differences between UAH and RSS MSU series suggests that much of the global-mean trend discrepancy arises over a series of very short satellite transitions (*NOAA-6* to *NOAA-9*) in 1987 (Mears et al. 2003). For very short overlaps the results are highly sensitive to the chosen intersatellite adjustment technique. Over these transitions we may be able to assume that nonclimatic changes in radiosonde and reanalyses products are small in comparison to, and independent of, the intersatellite bias effect. Therefore it might be possible to use these independent data to constrain this particular aspect of MSU long-term trend discrepancies (Christy and Norris 2004).

The global climate system is complex and an understanding of the evolution of a single variable alone is not a sufficient constraint to comprehensively understand the climate system evolution. In addition to temperature, it is also necessary to consider changes in a number of other variables, including vapor pressure/humidity, cloud cover and height, outgoing longwave radiation (OLR), atmospheric circulation, and ocean heat uptake. Much of the data required to create climate quality datasets in nontemperature variables already exist. Initial efforts to create climate datasets for some of these variables have also been made. However, many further efforts are required to bring these datasets up to the current state of temperature-based analyses, which have historically had the most attention paid to them. Geographical patterns of evolution in these complimentary variables will provide useful additional information and physical understanding that can help to constrain the structural uncertainty in temperature time series (and vice versa) and, perhaps more importantly, climate models. Continued construction and cross-validation of multivariate climate datasets must therefore be seen as a high priority.

SUPPORT FOR A FUTURE REFERENCE NETWORK. We can probably do little to provide retrospectively the absolute transfer standards that would allow us to remove much if not all of the structural uncertainty in historical climate records. However, we can learn from the past and instigate such standards for the future.

A reference network providing a multisite, multi-instrument calibration system consisting of surface, radiosonde, and satellite-based instruments should be instigated as a matter of urgency. This network will require a dedicated end-to-end management system with climate as the primary customer and result in a freely available database for use by the climate research community. Through having multiple independent measurements of the same variable colocated in space and time, a thorough understanding of instrumental biases and errors would be gained. If we had such a network then we would, at the very least, significantly reduce and likely eradicate uncertainty in our future climate monitoring activities.

What a reference network does not necessarily require is saturation (extensive) global coverage—the network acts as anchor points for the more global networks, which necessarily have operational weather requirements as their primary customers. A well-designed coverage of perhaps 50 sites globally would likely be sufficient. Although such sites will prove expensive, they are relatively inexpensive in comparison to our operational (particularly satellite based) programs. There would be additional benefits for instrument development, radiative transfer code development, and model development, among other applications.

Efforts underway at NOAA and through the World Meteorological Organization (WMO) Global Climate Observing System (GCOS) Implementation Plan (GCOS 2004) to plan and instigate a global reference network should be strongly supported by the climate community. Without such a reference network being established, it is likely that in 20 years time the climate community will still be struggling with uncertainty and unable to ascertain true climatic variations.

CONCLUSIONS. Following reports of a surface warming and a concurrent relative lack of warming aloft globally over the satellite period (Folland et al. 2001; NRC 2000) several groups have attempted to produce independent estimates of the true temperature variations aloft. There exists a pronounced spread in long-term temperature trends (e.g., Seidel et al. 2004)—the variable of most interest to policy makers. We argue that such a spread is inevitable, resulting from unintentional bias arising from the chosen methodological approaches: structural uncertainty. The challenge is to quantify the true spread of physically plausible solutions, given the limited number of datasets. If it is small then at least some of the current datasets must be physically unrealistic, whereas if it is sufficiently large then we cannot conclude anything meaningful about long-term trends aloft. Although we use upper-air temperatures as an example, we contend that structural uncertainty will be generic to all climate data records.

Production of a sufficiently large number of independent datasets to resolve structural uncertainty explicitly is unrealistic. However, an important lesson is that single datasets for a given variable will give a conservative estimate of the true uncertainty and hence a number of independently produced versions are both useful and required. The more independent versions the better, but we contend that three independently derived datasets is probably the minimum in order to get a handle on the magnitude of likely

structural uncertainty. Efforts to strictly quantify observational uncertainty are critical if current endeavors to understand and better quantify climate model uncertainties are to prove optimal.

We cannot use climate models to differentiate between observational datasets as the models may be wrong. We could use independent datasets for specified subperiods to constrain uncertainty. For example, over the short *NOAA-9* satellite transitions, when much of the long-term MSU trend uncertainty arises (Mears et al. 2003), radiosonde and reanalysis data could be used to constrain the intersatellite biases. More promising is the consideration of multivariate changes to constrain our uncertainty in individual variables. This requires much greater efforts than have been made to date to construct climate quality datasets of, for example, OLR, humidity, and ocean heat content. Although a necessary prerequisite for model validation, consideration of temperatures alone is not adequate. Showing that climate models capture multivariate changes across a wide range of variables would provide increased confidence in projections compared to traditional approaches that have, almost exclusively, considered temperature (or derived product) changes alone.

Looking forward, most of the uncertainty could be eradicated by setting up and maintaining a reference network from which absolute transfer standards can be derived. Efforts under way to implement such a network should be given the highest priority.

REFERENCES

Christy, J. R., and W. B. Norris, 2004: What may we conclude about global tropospheric temperature trends? *Geophys. Res. Lett.,* **31,** L06211, doi:10.1029/2003GL019361.

——, R. W. Spencer, and E. S. Lobl, 1998: Analysis of the merging procedure for the MSU daily temperature time series. *J. Climate,* **11,** 2016–2041.

——, ——, W. B. Norris, W. D. Braswell, and D. E. Parker, 2003: Error estimates of version 5.0 of MSU–AMSU

bulk atmospheric temepratures. *J. Atmos. Oceanic Technol.,* **20,** 613–629.

Farman, J. C., B. G. Gardiner, and J. D. Shanklin, 1985: Large losses of total ozone in Antarctica reveal seasonal CLOX/NOX interaction. *Nature,* **315,** 207–210.

Folland, C. K., and coauthors, 2001: Observed climate variability and change. *Climate Change 2001, The Scientific Basis: Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change,* J. T. Houghton et al., Eds., Cambridge University Press, 99–181.

GCOS, 2004: Implementation Plan for the Global Observing System for Climate in Support of the INFCCC. Tech. Rep. GCOS-92, 153 pp. [Available online at www.wmo.int/web/gcos/gcoshome.html.]

Grody, N. C., K. Y. Vinnikov, M. D. Goldberg, J. T. Sullivan, and J. D. Tarpley, 2004: Calibration of multisatellite observations for climaticstudies: Microwave Sounding Unit (MSU). *J. Geophys. Res.,* **109,** D24104, doi:10.1029/2004JD005079.

Lanzante, J. R., S. A. Klein, and D. J. Seidel, 2003: Temporal homogenization of monthly radiosonde temperature data. Part I: Methodology. *J. Climate,* **16,** 224–240.

Mears, C. A., M. C. Schabel, and F. J. Wentz, 2003: A Reanalysis of the MSU Channel 2 tropospheric temperature record. *J. Climate,* **16,** 3650–3664.

NRC, 2000: *Reconciling Observations of Global Temperature Change.* National Academy Press, 85 pp.

Santer, B. D., and Coauthors, 2003: Influence of satellite data uncertainties on the detection of externally forced climate change. *Science,* **300,** 1280–1284.

Seidel, D. J., and Coauthors, 2004: Uncertainty in signals of large-scale climate variations in radiosonde and satellite upper-air temperature datasets. *J. Climate,* **17,** 2225–2240.

Thorne, P. W., and Coauthors, 2003: Probable causes of late twentieth century tropospheric temperature trends. *Climate Dyn.,* **21,** 573–591.

Thorne, P. W., D. E. Parker, S. F. B. Tett, P. D. Jones, M. McCarthy, H. Coleman, and P. Brohan, 2005: Revisiting radiosonde upper-air temperatures from 1958 to 2002, *J. Geophys. Res.,* in press.

Uppala, S. M., and Coauthors, 2005: The ERA-40 reanalysis. *Quart. J. Roy. Meteor. Soc.,* in press.

Vinnikov, K. Y., and N. C. Grody, 2003: Global warming trend of mean tropospheric temperature observed by satellites. *Science,* **302,** 269–272.