

Tropical vertical temperature trends: A real discrepancy?

P. W. Thorne,¹ D. E. Parker,¹ B. D. Santer,² M. P. McCarthy,¹ D. M. H. Sexton,¹ M. J. Webb,¹ J. M. Murphy,¹ M. Collins,¹ H. A. Titchner,¹ and G. S. Jones¹

Received 2 March 2007; revised 2 July 2007; accepted 19 July 2007; published 16 August 2007.

[1] We examine the sensitivity of modeled and observed tropical tropospheric temperature trend amplification (the ratio of T_{2LT} “lower troposphere” to surface changes) to several sources of uncertainty. Model behaviour is robust across a large perturbed physics ensemble of HadCM3, yielding a smaller amplification range (1.44 ± 0.06) than a previous multi-model ensemble (1.41 ± 0.24). The uncertainty of inter-satellite calibration implied by available MSU T_2 (mid-troposphere) estimates ($\sigma = 0.035K$) is much greater than that required to adequately resolve the trend ($\sigma < 0.01K$), or the amplification behaviour (implied amplification range ± 0.95). Trend amplification uncertainty in both models and observations decreases as the timescale increases. Depending upon choice of dataset and time period, uncertainty in trend amplification estimates over 21 years lies between ± 1.5 and ± 0.2 . **Citation:** Thorne, P. W., D. E. Parker, B. D. Santer, M. P. McCarthy, D. M. H. Sexton, M. J. Webb, J. M. Murphy, M. Collins, H. A. Titchner, and G. S. Jones (2007), Tropical vertical temperature trends: A real discrepancy?, *Geophys. Res. Lett.*, 34, L16702, doi:10.1029/2007GL029875.

1. Background

[2] Since the original production of a tropospheric temperature dataset from satellite-based Microwave Sounding Units (MSU) [Spencer and Christy, 1992], there has been intense debate over whether climate models adequately capture observed changes in atmospheric vertical temperature structure [National Research Council, 2000; Folland et al., 2001; Karl et al., 2006]. Most early versions of upper air datasets (and even some more recent observational analyses) exhibited little if any tropospheric warming, while models amplified surface warming aloft, particularly in the tropics. This was the primary motivation for several national and international assessments; the most recent concluding that, “While these [observed] data are consistent with the results from climate models at the global scale, discrepancies within the tropics remain to be resolved” [Karl et al., 2006, p. 1]. If these discrepancies are real, they would cast significant doubt on the reliability of climate models and on the usefulness of model projections of future climate change.

[3] Within the tropical troposphere, vertical mixing is dominated by convective processes. Hence temperature anomalies within the tropics ($20^\circ N - 20^\circ S$) should amplify from the surface through the troposphere [Santer et al.,

2005, and references therein]. Santer et al. considered a suite of coupled climate model simulations and observational records from the surface and the troposphere. The modelled amplification (defined by a ratio $\Delta T_{\text{troposphere}} / \Delta T_{\text{surface}}$ whereby values > 1 indicate amplification) was found to be consistent with theoretical expectations on both high-frequency (monthly, inter-annual) and low-frequency (multi-decadal trend) timescales. Observations yielded similar high-frequency amplification, but a large spread in low-frequency behaviour, with only one satellite dataset exhibiting tropospheric amplification of surface trends.

[4] This analysis was key to the Karl et al. [2006] report. Although there was some overlap between tropical amplification behaviour in models and the available climate datasets, the panel were unable to rule out a fundamental discrepancy between the two. However, this analysis was not comprehensive in its consideration of uncertainties. Here, we re-assess these findings in the context of three additional specific sources of uncertainty: in selected physical parameters of one specific climate model, in the construction of satellite datasets, and in temporal sampling.

2. Assessing Climate Model Uncertainties

[5] Climate model amplification estimates for the tropics appear to be robust [Santer et al., 2005; Karl et al., 2006]. However, these analyses were restricted to an “ensemble of opportunity” for 1979–1999 consisting of 49 runs made using 19 different climate models. This ensemble combined structural differences in model physics with differences in the number and type of external forcings, forcing histories, methods of applying forcings, etc. It encapsulates our best current understanding of climate processes and of 20th Century climate change, but may represent an underestimate of the true uncertainty in model-based estimates. Each of the models represents each contributing group’s optimal climate model configuration, tuned to obtain a reasonably realistic mean climate state (but importantly, not to recreate the transient response to external forcings [Karl et al., 2006]).

[6] Other plausible model configurations can be generated by simultaneously varying empirical parameters within an individual model’s parameterisation schemes (which account for unresolved sub-gridbox scale physical processes). Parameter variations are typically made within reasonable ranges, while keeping all other aspects of model structure and physics constant [Murphy et al., 2004; Stainforth et al., 2005; Webb et al., 2006]. Such “perturbed physics” ensembles sample a wide range of climate feedbacks, and hence lead to a wide range of climate sensitivities [Webb et al., 2006].

[7] It is this source of model structural uncertainty that we consider here using a slab-ocean version of the HadCM3 climate model [Murphy et al., 2004]. Each of the

¹Hadley Centre, Met Office, Exeter, UK.

²Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Laboratory, California, USA.

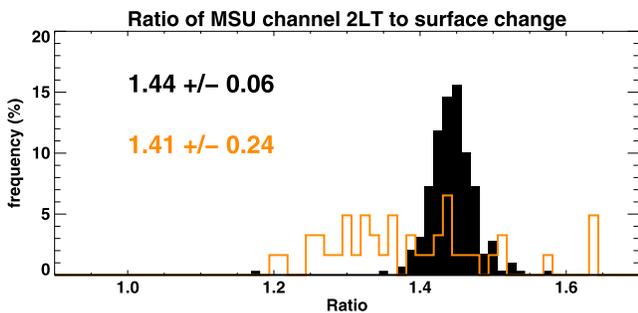


Figure 1. Frequency histogram for the modelled ratio of tropical T_{2LT} to surface temperature change for doubling of CO_2 in a perturbed physics HadCM3 ensemble [Murphy et al., 2004] (black) and from transient runs for 1979–1999 from 19 models [Santer et al., 2005] (orange). Those transient runs which exhibited absolute trends $<0.05K/decade$ in surface temperatures (4 of 49) are excluded. The legend shows means and 2σ ranges. For the transient model runs these were calculated by averaging all ensemble members (ensemble sizes range from 1 to 5) for each model, then taking the mean and standard deviations from these averaged populations, giving each model equal weighting [Santer et al., 2005].

230 members of the ensemble has a unique set of parameter choices. To estimate amplification factors from this ensemble, we calculate changes in tropically-averaged surface and lower tropospheric temperatures. The latter are derived by using static weighting functions [Spencer and Christy, 1992] to compute synthetic MSU T_{2LT} temperatures from model vertical profiles of temperature. For each ensemble member we consider the difference between two 20-year averages obtained from a pre-industrial integration and an equilibrium doubled CO_2 integration [Murphy et al., 2004; Webb et al., 2006].

[8] Figure 1 shows tight clustering of amplification results from the “perturbed physics” ensemble. The perturbed physics results suggest that HadCM3 model uncertainties are highly unlikely to encompass damping aloft, as is implied by some observational estimates [Santer et al., 2005; Karl et al., 2006]. The perturbed physics ensemble range is much smaller than in the “ensemble of opportunity” considered previously [Santer et al., 2005; Karl et al., 2006]. The larger spread in the transient runs is likely because these involve smaller anthropogenic forcing than the doubled CO_2 experiments, so that estimates of the true amplification behaviour are noisier and more sensitive to temporal sampling effects (see penultimate section). It may also relate, in part, to use of a slab ocean model here instead of a fully coupled ocean model.

3. A Simple Model for Assessing Satellite Climate Dataset Uncertainties

[9] A small number of groups have constructed climate records from raw MSU radiances [Prabhakara et al., 1998; Mears and Wentz, 2005; Mears et al., 2003; Christy et al., 2003; Grody et al., 2004; Vinnikov and Grody, 2003; Zou et al., 2006]. The construction methods differ in important ways. Each group has tried to assess uncertainties arising

from their own specific processing choices. These uncertainty estimates have been derived in very different ways, and none is exhaustive in the sources of uncertainty considered. Structural uncertainty dominates [Thorne et al., 2005b], and is poorly bounded [Karl et al., 2006].

[10] Without an independent reference set of high-quality collocated data, there will always be ambiguity in climate dataset construction [Thorne et al., 2005b]. For satellites, the problem is not easy to resolve: periods of overlap between satellites are often of limited duration, and only a very small fraction of measurements are collocated in space and time [Zou et al., 2006]. There are also poorly quantified non-climatic influences, such as orbital and instrument drift, which affect each satellite differently [Christy et al., 2003; Mears et al., 2003]. Typically, MSU dataset production has involved attempting to remove all suspected non-climatic influences from each individual satellite record, and then merging these individual records in a way that minimizes remaining absolute offsets. Errors in all other aspects of dataset construction are thus projected strongly onto the calculation of satellite offset factors. These can be as large as 1K [Mears et al., 2003]. Offset adjustment uncertainties calculated internally by each group [e.g., Mears et al., 2003] are much smaller than dataset inter-comparisons imply, at least for some transitions [Mears et al., 2003; Karl et al., 2006; Christy et al., 2007]. The uncertainty cannot be resolved solely from the statistics of the inter-satellite differences, as these are dependent upon the chosen intra-satellite homogenisation procedure. The addition of each new satellite therefore introduces to the overall time series a constant non-climatic offset of unknown magnitude. Further, the precise selection of satellites (and sub-sets of data from particular satellites) to be inter-calibrated differs between research groups.

[11] We construct a statistical model of the uncertainties in MSU trends, assuming that dataset construction uncertainties project entirely onto the calculation of inter-satellite offset adjustments. We then calculate the effect of offset uncertainties of specified magnitudes on the resulting trends. This approach is purely statistical, and does not replicate the labor-intensive process of explicitly producing multiple datasets. Our approach simply allows us to place bounds on the likely magnitude of trend uncertainties for different residual errors in inter-satellite offsets. We assume that the error model applies to each MSU retrieval (T_4 , T_2 , and T_{2LT}).

[12] Between 1979 and 2004, there were a total of 13 satellite transitions. We assume that all 13 satellites were used in the generation of MSU datasets, and that the properties of the errors in the offsets do not vary systematically over the satellite era. We use a Monte Carlo approach to produce 10,000 error time series. This involves randomly sampling an offset error from a normal distribution with mean zero and a prescribed standard deviation σ_{sat} at each satellite transition, and then creating timeseries based on these errors. Each realisation therefore consists of a random walk with 13 steps, one at each satellite introduction. We then calculate a median of pairwise slopes linear fit for each synthetic time series [Lanzante, 1996]. We vary σ_{sat} within the range 0.005K to 0.05K. This is based on published satellite-satellite and satellite-radiosonde inter-comparisons, which have yielded maximum differences between RSS (Remote Sensing Systems) and UAH (Uni-

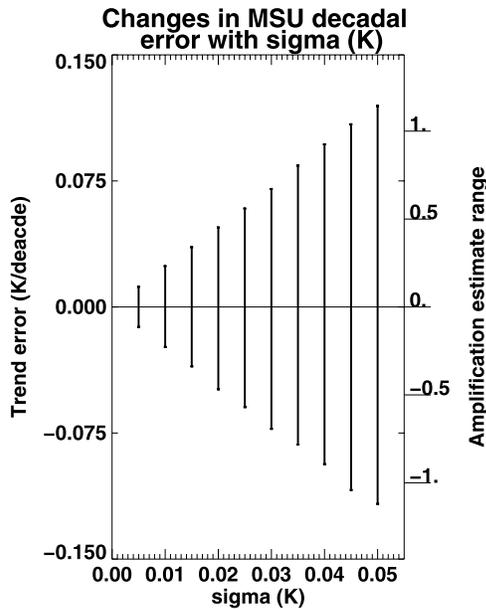


Figure 2. Impacts of different inter-satellite bias offsets σ_{sat} (sigma) on resulting 95% ranges of estimates of tropical trends (left hand y-axis) and amplification factors (right hand y-axis). Amplification is based on temperature trends at the surface [Brohan *et al.*, 2006].

versity of Alabama in Huntsville) of about 0.1K for NOAA-9 transitions in T_2 [Mears *et al.*, 2003] and NOAA-11 to NOAA-14 in $T_{2\text{LT}}$ [Christy *et al.*, 2007].

[13] In this model, the uncertainty in the long-term temperature trend grows linearly with increasing σ_{sat} . For MSU for 1979–2004 the 95% confidence interval is 4.3 σ_{sat} K/decade (Figure 2, this will differ for other instruments/satellites). Robustly resolving any tropospheric warming trend in MSU (uncertainty <20% of $\sim 0.15\text{K}/\text{decade}$ expected given the surface trend) requires minimal uncertainty ($\sigma_{\text{sat}} \leq 0.01\text{K}$) in inter-satellite offset biases. As the tropical amplification estimates involve division by the surface trend (0.11K/decade), uncertainties in the amplification factor grow much more quickly than those for the $T_{2\text{LT}}$ trends (Figure 2). So adequately resolving the amplification behaviour requires even stricter criteria to be met. More groups have tried to create trend estimates for T_2 than for $T_{2\text{LT}}$, so we use their range to estimate what the true value of σ_{sat} may be for $T_{2\text{LT}}$. The range of estimates for global T_2 of 0.16K/decade implies a σ_{sat} of 0.035K, and amplification uncertainty of ± 0.95 . Therefore, unless several of these datasets are fatally flawed, we are nowhere near attaining the required degree of accuracy from the historical data.

4. Temporal Sampling Uncertainty

[14] Amplification should be relatively independent of time scale [Santer *et al.*, 2005; Karl *et al.*, 2006]. Reliable estimation may be problematic for relatively short trends, however, since the slowly-evolving externally-forced signals are embedded in background noise arising from natural internal climate variability. This noise receives contributions from modes of variability operating on a range of time-scales. Hence trends are sensitive to the choice of trend

length L and trend period relative to the exact phasing of the noise, a problem that is compounded when one considers the ratio of noise-contaminated surface and $T_{2\text{LT}}$ trends. Furthermore, linear trends are probably a sub-optimal model of longer-term changes in surface and upper-air temperatures [Seidel and Lanzante, 2004; Thorne *et al.*, 2005a].

[15] We assume that the temperature response to slowly-evolving changes in anthropogenic forcings can be approximated by a linear signal plus noise with complex structure. As L increases, the anthropogenic forcing, signal amplitude, and signal-to-noise ratio also tend to increase [Santer *et al.*, 1996]. For each value of L , therefore, we expect to see a spread of amplification estimates, characterized by the mean $\bar{A}(L)$ and the standard deviation $\sigma_A(L)$, with $\sigma_A(L)$ decreasing as L increases.

[16] We have estimated tropical amplification using HadCRUT3 surface temperatures [Brohan *et al.*, 2006] and $T_{2\text{LT}}$ changes obtained from various radiosonde and satellite products and from an ensemble of the HadCM3 climate model run with anthropogenic and natural forcings over 1958–1999 [Stott *et al.*, 2000]. The radiosonde records begin in 1958, while satellite records commence in 1979, and all continue through 2006. For the HadAT radiosonde record and HadCM3 model we calculated all possible overlapping trends of length L , with L varying from 30 to 10 years in length, and with overlaps by all but one year (e.g., the first 30-year trend for the HadAT2 data is over 1958 to 1987, the second is over 1959 to 1988, etc.) For the remaining radiosonde records, the same calculations were made for 11, 16, 21, and 26 year periods. Trends for the satellite records were calculated over 11, 16, and 21 year periods only.

[17] Both the model and observational results in Figure 3 exhibit the expected decrease in $\sigma_A(L)$ with increasing L . The HadCM3 data exhibit relatively little variation with L in the central values of $\bar{A}(L)$, consistent with previous results showing timescale-invariant behaviour [Santer *et al.*, 2005] (Figure 3, top plot). In contrast, $\bar{A}(L)$ decreases with increasing L in the HadAT data.

[18] Over the radiosonde era as a whole there is better agreement between all radiosonde records and the model than for the satellite era alone (Figure 3, bottom plots). But the satellite era contains at most a few degrees of freedom in the observational results given the substantially overlapping segments, and the ranges over this era from all datasets are likely to be an under-estimate. The non-overlap between the UAH and RSS “envelopes” of amplification estimates on the longest timescales implies pervasive differences in low-frequency characteristics between them, consistent with our MSU error model (Figure 2). Depending upon choice of period and dataset, uncertainty on previously considered 21-year timescales ranges between ± 1.5 (radiosondes for the full radiosonde era) to ± 0.2 (RSS). HadCM3, for which the range will solely result from model internal and forced variability, exhibits a range of comparable magnitude to the observations.

[19] If the physical mechanisms that control “real world” amplification factors have changed markedly over time, then it is necessary to examine amplification behaviour over specific periods (rather than over all possible periods of a particular trend length). But we currently lack any evidence of changes in the basic physics governing tropical amplification behaviour. We do, however, have ample

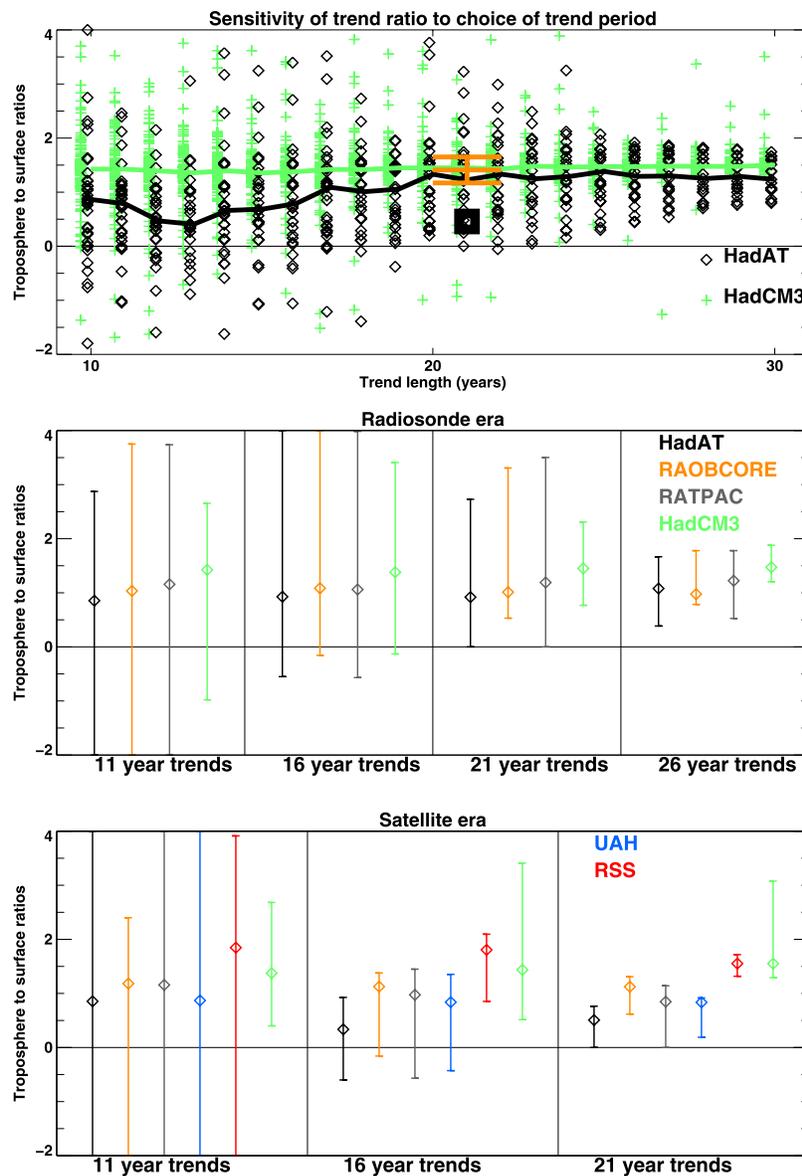


Figure 3. Sensitivity of estimates of tropical trend amplification ratio to period length and start date. Start dates are separated by a calendar year. In the top panel lines connect median values of the estimates for each trend length. The orange error bar at 21 year trend length denotes the mean and 2σ range of model trend ratios calculated from transient runs from a range of climate models (see Figure 1). The black cube denotes the 1979–1999 value for radiosondes. The bottom plots incorporate information from RATPAC [Free *et al.*, 2005], RAOBCORE [Haimberger, 2007] (radiosondes), RSS and UAH datasets, and consider (middle) the full radiosonde era (1958–2006) and (bottom) satellite era (1979–2006). The 90% range is indicated by the whiskers and the median value by diamonds.

evidence of the existence of observational measurement and sampling errors [Free and Seidel, 2005] and of residual non-climatic influences in satellite and radiosonde data [Karl *et al.*, 2006; Sherwood *et al.*, 2005; Randel and Wu, 2006; Christy *et al.*, 2007]. Clearly, these factors must contribute both to the “between dataset” differences in amplification behaviour shown in Figure 3, and amplification differences across timescales in individual datasets.

5. Discussion

[20] A perturbed physics ensemble, encompassing a large range of climate sensitivities, shows robust amplification

behaviour within the tropical troposphere. Conversely, we find much greater uncertainty in amplification behaviour when the same model is run in transient mode with 20th century forcings. We conclude from this that much of the spread in previous tropical amplification estimates from transient model runs [Santer *et al.*, 2005] arises from natural climate variability in the models rather than from structural differences between them. Hence our results imply that tropical amplification of surface temperature changes is very strongly constrained in all current climate models, and unlikely to arise through choices of sub-gridbox scale parameterisations.

[21] Uncertainties in observed MSU dataset trends are large relative to the emerging climate change signal. This is despite the satellite era being the best-observed period in Earth's climate history. There are important implications for the design of current and future satellite-based temperature monitoring systems. For current polar-orbiters, individual inter-satellite offsets after removing all intra-satellite non-climatic influences need to be constrained to well within 10% of the expected decadal trend magnitude for any 25-year emerging signal to be robustly detected. To ascertain amplification behaviour, which depends upon relative trends between the troposphere and surface, requires even smaller uncertainties to be attained.

[22] Consideration of amplification behaviour across a range of different periods yields large uncertainties in both models and observations. The choice of period and period length impacts conclusions regarding the existence of a discrepancy. Assessment in the context of all possible choices yields lower confidence in the significance of any discrepancy over the satellite era.

[23] Analyses of differences in trends between the surface and the troposphere for an emerging climate change signal remain highly uncertain. Although we cannot rule out "real world" amplification factors being different on different timescales, and hence problems common to all climate models, uncertainty arising from residual observational errors and choice of analysis period need to be carefully discounted if such a discrepancy is to be proven.

[24] **Acknowledgments.** This work benefited from access to MSU data and radiosonde data from the various dataset providers. This work was funded by the UK Department for Environment, Food and Rural Affairs under Contract PEC7/12/37 and is British Crown copyright. Simon Tett provided valuable contributions and insights during internal review.

References

- Brohan, P., J. J. Kennedy, I. Harris, S. F. B. Tett, and P. D. Jones (2006), Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850, *J. Geophys. Res.*, *111*, D12106, doi:10.1029/2005JD006548.
- Christy, J. R., R. W. Spencer, W. B. Norris, W. D. Braswell, and D. E. Parker (2003), Error estimates of version 5.0 of MSU-AMSU bulk atmospheric temperatures, *J. Atmos. Oceanic Technol.*, *20*, 613–629.
- Christy, J. R., W. B. Norris, R. W. Spencer, and J. J. Hnilo (2007), Tropospheric temperature change since 1979 from tropical radiosonde and satellite measurements, *J. Geophys. Res.*, *112*, D06102, doi:10.1029/2005JD006881.
- Folland, C. K., et al. (2001), Observed climate variability and change, in *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment of the Intergovernmental Panel on Climate Change*, edited by J. T. Houghton et al., pp. 99–181, Cambridge Univ. Press, New York.
- Free, M., and D. J. Seidel (2005), Causes of differing temperature trends in radiosonde upper air data sets, *J. Geophys. Res.*, *110*, D07101, doi:10.1029/2004JD005481.
- Free, M., D. J. Seidel, J. K. Angell, J. Lanzante, I. Durre, and T. C. Peterson (2005), Radiosonde Atmospheric Temperature Products for Assessing Climate (RATPAC): A new data set of large-area anomaly time series, *J. Geophys. Res.*, *110*, D22101, doi:10.1029/2005JD006169.
- Grody, N. C., K. Y. Vinnikov, M. D. Goldberg, J. T. Sullivan, and J. D. Tarpley (2004), Calibration of multisatellite observations for climatic studies: Microwave Sounding Unit (MSU), *J. Geophys. Res.*, *109*, D24104, doi:10.1029/2004JD005079.
- Haimberger, L. (2007), Homogenization of radiosonde temperature time series using innovation statistics, *J. Clim.*, *20*, 1377–1403.
- Karl, T. R., S. J. Hassol, C. D. Miller, and W. L. Murray (Eds.) (2006), *Temperature trends in the lower atmosphere: Steps for understanding and reconciling differences*, Climate Change Sci. Program, Washington, D. C.
- Lanzante, J. R. (1996), Resistant, robust and non-parametric techniques for the analysis of climate data: Theory and examples, including applications to historical radiosonde station data, *Int. J. Climatol.*, *16*, 1197–1226.
- Mears, C. A., and F. J. Wentz (2005), The effect of diurnal correction on satellite-derived lower-tropospheric temperature, *Science*, *309*, 1548–1551.
- Mears, C. A., M. C. Schabel, and F. J. Wentz (2003), A reanalysis of the MSU channel 2 tropospheric temperature record, *J. Clim.*, *22*, 3650–3664.
- Murphy, J. M., et al. (2004), Quantification of modelling uncertainties in a large ensemble of climate change simulations, *Nature*, *430*, 768–772.
- National Research Council (2000), *Reconciling Observations of Global Temperature Change*, 85 pp., Washington, D. C.
- Prabhakara, C., R. Iacovazzi Jr., J.-M. Yoo, and G. Dalu (1998), Global warming deduced from MSU, *Geophys. Res. Lett.*, *25*(11), 1927–1930.
- Randel, W. J., and F. Wu (2006), Biases in stratospheric and tropospheric temperature trends derived from historical radiosonde data, *J. Clim.*, *19*, 2094–2104.
- Santer, B. D., et al. (1996), A search for human influences on the thermal structure of the atmosphere, *Nature*, *382*, 39–46.
- Santer, B. D., et al. (2005), Amplification of surface temperature trends and variability in the tropical atmosphere, *Science*, *309*, 1551–1556.
- Seidel, D. J., and J. R. Lanzante (2004), An assessment of three alternatives to linear trends for characterizing global atmospheric temperature changes, *J. Geophys. Res.*, *109*, D14108, doi:10.1029/2003JD004414.
- Sherwood, S. C., J. R. Lanzante, and C. L. Meyer (2005), Radiosonde daytime biases and late-20th century warming, *Science*, *309*, 1556–1559.
- Spencer, R. W., and J. R. Christy (1992), Construction and radiosonde validation of satellite gridpoint temperature anomalies. 2. A tropospheric retrieval and trends during 1979–1990, *J. Clim.*, *8*, 858–866.
- Stainforth, D. A., et al. (2005), Uncertainty in predictions of the climate response to rising levels of greenhouse gases, *Nature*, *433*, 403–406.
- Stott, P. A., et al. (2000), External control of 20th century temperature by natural and anthropogenic forcings, *Science*, *290*, 2133–2137.
- Thorne, P. W., D. E. Parker, S. F. B. Tett, P. D. Jones, M. McCarthy, H. Coleman, and P. Brohan (2005a), Revisiting radiosonde upper air temperatures from 1958 to 2002, *J. Geophys. Res.*, *110*, D18105, doi:10.1029/2004JD005753.
- Thorne, P. W., D. E. Parker, J. R. Christy, and C. A. Mears (2005b), Uncertainties in climate trends—Lessons from upper-air temperature records, *Bull. Am. Meteorol. Soc.*, *86*, 1437–1443.
- Vinnikov, K. Y., and N. C. Grody (2003), Global warming trend of mean tropospheric temperature observed by satellites, *Science*, *302*, 269–272.
- Webb, M. J., et al. (2006), On the contribution of local feedback mechanisms to the range of climate sensitivities in two GCM ensembles, *Clim. Dyn.*, *27*, 17–38.
- Zou, C.-Z., M. D. Goldberg, Z. Cheng, N. C. Grody, J. T. Sullivan, C. Cao, and D. Tarpley (2006), Recalibration of microwave sounding unit for climate studies using simultaneous nadir overpasses, *J. Geophys. Res.*, *111*, D19114, doi:10.1029/2005JD006798.

M. Collins, G. S. Jones, D. E. Parker, M. P. McCarthy, J. M. Murphy, D. M. H. Sexton, P. W. Thorne, H. A. Titchner, and M. J. Webb, Hadley Centre, Met Office, Exeter EX1 3PB, UK. (peter.thorne@metoffice.gov.uk)
B. D. Santer, Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Laboratory, CA 94550, USA.