

Assessing the value of Microwave Sounding Unit–radiosonde comparisons in ascertaining errors in climate data records of tropospheric temperatures

Carl A. Mears,¹ Frank J. Wentz,¹ and Peter W. Thorne²

Received 28 February 2012; revised 21 August 2012; accepted 23 August 2012; published 3 October 2012.

[1] Multidecadal-scale changes in atmospheric temperature have been measured by both radiosondes and the satellite-borne microwave sounding unit (MSU). Both measurement systems exhibit substantial time varying biases that need to be removed to the extent possible from the raw data before they can be used to assess climate trends. A number of methods have been developed for each measurement system, leading to the creation of several homogenized data sets. In this work, we evaluate the agreement between MSU and homogenized radiosonde data sets on multiyear (predominantly 5-year) time scales and find that MSU data sets are often more similar to each other than to radiosonde data sets and vice versa. Furthermore, on these time scales the differences between MSU data sets are often not larger than published internal uncertainty estimates for the RSS product alone and therefore may not be statistically significant when the internal uncertainty in each data set is taken into account. Given the data limitations it is concluded that using radiosondes to validate multidecadal-scale trends in MSU data, or vice versa, or trying to use such metrics alone to pick a ‘winner’ is an ill-conditioned approach and has limited utility without one or more of additional independent measurements, or methodological, or physical analysis.

Citation: Mears, C. A., F. J. Wentz, and P. W. Thorne (2012), Assessing the value of Microwave Sounding Unit–radiosonde comparisons in ascertaining errors in climate data records of tropospheric temperatures, *J. Geophys. Res.*, *117*, D19103, doi:10.1029/2012JD017710.

1. Introduction

[2] Multidecadal changes in global atmospheric temperature have primarily been estimated using measurements from two disparate measurement systems, balloon-borne radiosondes (beginning in the late 1950s) and satellite-borne microwave sounding instruments. The microwave measurements are constructed by merging together measurements from the Microwave Sounding Units (MSUs, late 1978–2005) and the Advanced Microwave Sounding Units (AMSUs, mid 1998 to the present). Hereafter we refer to the merged MSU/AMSU data sets as MSU data sets for brevity. Unfortunately, neither MSU nor radiosonde records have been designed with absolute calibration and traceability. Numerous changes in instrumentation, observing practice, time of observation and various other undesirable measurement aspects pervade these records [Thorne *et al.*, 2011a]. A number of techniques to characterize and remove these

problems have been developed and refined, resulting in a number of “homogenized” data sets for each type of data. Ideally, the resulting methodologically distinct data sets would report similar changes in atmospheric temperature during the post-1978 period (the satellite era) when both observing systems were in operation. This would yield confidence that data issues had been adequately understood and removed, leading to a good estimate of the true climate system evolution. Unfortunately, this has not been the case, which has led to intense debate about the details of recent changes in the Earth’s atmospheric temperatures. As the homogenization strategies have evolved over time, trends from radiosonde and MSU data have come into somewhat better agreement for global-scale averages (although still a substantial error as a percentage of the relatively small trend signal), while more substantial discrepancies remain in the tropics [Lanzante *et al.*, 2006]. However, overall trend agreement can hide interesting differences at shorter time-scales and may be a result of largely fortuitous cancellation of substantial differences on these shorter timescales.

[3] In the evaluation of satellite data, particularly in the early part of any mission, it is customary to “validate” the satellite data by comparing retrieved geophysical parameters with in situ measurements related to the variable in question. This is a useful exercise, particularly when the principles underlying the remote sensing techniques are still being

¹Remote Sensing Systems, Santa Rosa, California, USA.

²Cooperative Institute for Climate and Satellites, North Carolina State University, National Climatic Data Center, NOAA, Asheville, North Carolina, USA.

Corresponding author: C. A. Mears, Remote Sensing Systems, Santa Rosa, CA 95401, USA. (mears@remss.com)

©2012. American Geophysical Union. All Rights Reserved.
0148-0227/12/2012JD017710

tested, and when the accuracy of the in situ measurements is expected to exceed the accuracy of the satellite-based measurements. It is tempting to extend this approach to the evaluation of long-term trends in geophysical variables with well-established measurement techniques, such as atmospheric sounding. For example, one could use radiosonde/satellite inter-comparison studies to try to make determinations of satellite data set quality. In fact a number of studies have been used to suggest that one of the MSU data sets is more accurate than the others, based on a closer agreement with various radiosonde measurements. *Randall and Herman* [2008] compared MSU measurements with the results from a subset of a single radiosonde data set and concluded that the University of Alabama, Huntsville (UAH) satellite data set was more accurate than the RSS data. They focused on trends in the data set differences over 5-year and 10 year periods, and on a limited analysis period. *Christy et al.* [2010] reached a similar conclusion using a similar short-term trend analysis, but analyzed only within the deep tropics, and only one time period (1989–1995). *Christy et al.* [2007] used tropical radiosonde measurements (both raw soundings and a single homogenized data set) to argue that the RSS data set contains a spurious warming trend in the tropics during the early 1990s, with the bulk of the analysis of MSU-radiosonde differences focusing on this period. Conversely, *Po-Chedley and Fu* [2012] argued for a significant discontinuity in the early portion of the UAH record associated with the short life-time NOAA-9 satellite. All these papers used a limited number of radiosonde data sets, and focused their attention on a limited time period.

[4] Over the last decade, there have been numerous other studies that have included inter-comparisons between MSU and homogenized radiosonde data sets at both global [*Haimberger et al.*, 2008; *Lanzante et al.*, 2006; *Seidel et al.*, 2004] and regional scales [*Christy and Norris*, 2006; *Thorne et al.*, 2007, 2011b; *Titchner et al.*, 2009]. A recently completed study of 32-year trends in both MSU and homogenized radiosonde data [*Mears et al.*, 2011] serves in part to update such studies to the current time. When the internal uncertainty in the Remote Sensing Systems (RSS) data set is taken into account, the trends in this data set tend to be consistent with those from homogenized radiosonde data sets for the tropospheric channels considered in this paper. The exception to this finding is in the deep tropics, where data sets from RSS and the Center for Satellite Applications and Research (STAR) tend to show trends that are high compared to most adjusted radiosonde trends.

[5] The rest of this paper is structured as follows. Section 2 outlines the overarching methodological philosophy and approach. Section 3 provides details of the data used in our study. In Section 4, we compare time series of 5- and 10- year trends derived from each data set for various regions, and in Section 5 we investigate the impact of estimates of internal uncertainty on our findings. In Section 6 we revisit the analysis of interlayer differences performed by *Randall and Herman* [2008], and in Section 7 we conclude with a discussion of our findings.

2. Methodological Approach and Rationale

[6] The purpose of this paper is to comprehensively investigate whether it is appropriate to use MSU/radiosonde

intercomparisons as arbiters of MSU data set quality. Arguably this can only be done if two over-arching conditions are both adequately met. First, the radiosonde measurements being considered need to be a sufficiently unbiased representation of the true evolution of the climate system. Second, the results of the comparison need to be valid in the statistical sense, given the inevitable uncertainty in both the MSU data sets and homogenized radiosonde data sets.

[7] Several papers, often authored by the developers of the radiosonde data sets themselves, have addressed the first question and have mostly concluded that substantial decadal-scale errors may remain even in the homogenized data [*Lanzante et al.*, 2003; *Randel and Wu*, 2006; *Titchner et al.*, 2009]. A number of investigators have specifically caveated that it is probable that significant residual errors remain in the tropics where the network is sparse and most observations are daytime only when radiation effects are more important [*Randel and Wu*, 2006; *Sherwood et al.*, 2005, 2008; *Titchner et al.*, 2009].

[8] For the second question, for both MSU and radiosonde data sets, there are two types of important error [*Thorne et al.*, 2005a]. First is the structural uncertainty, which is the uncertainty caused by the choice of a single processing method from the set of all possible reasonable methods. This uncertainty can best be characterized by the spread of results from different data sets constructed by different research groups. Such an estimate is predicated upon the assumptions that all methods are reasonable (peer review being considered a necessary but not necessarily adequate condition) and that the very finite number of published estimates provides an unbiased estimate of the much larger spread of possible estimates. Second, there is internal uncertainty, unique to each product, that arises from uncertainty in parameters used to perform the various adjustments after a processing method is chosen. In the case of microwave sounders, this includes the uncertainty in calibration offsets, and the adjustments for instrument nonlinearity and drifts in local measurement time. This type of uncertainty can be determined by an analysis of the individual uncertainty in each of the adjustment steps used. In *Mears et al.* [2011], we performed such an analysis for the RSS data sets using a Monte-Carlo approach.

[9] This paper assesses the issue of robustness of MSU/radiosonde intercomparisons to inevitable uncertainty in radiosonde and MSU records. Specifically, it aims to comprehensively address the suitability of such comparisons given the known data limitations. It aims to address the following questions and in so doing update the results of the various intercomparisons [e.g., *Randall and Herman*, 2008; *Christy et al.*, 2007, 2010] that have specifically looked to assess the quality of different MSU products by reference to radiosonde data and data sets:

[10] 1. What is the impact of use of only 1 (as has been common), a subset, or all available radiosonde data sets?

[11] 2. What is the impact of different approaches to accounting for sampling mismatches between radiosonde and satellite data?

[12] 3. What are the implications of undertaking analyses for limited time periods rather than the whole period of record?

[13] 4. What impact recently published comprehensive parametric uncertainty estimates for both one MSU and one radiosonde product may have on the results.

[14] 5. Is the physical interpretation of the difference between MT and LT used in some previous studies correct?

[15] 6. Do updated versions of many of the data sets previously considered impact the results and implications of such analyses?

3. Data Sets Used in This Study

[16] We focus our attention on MSU deep-layer temperature measurements (and equivalent estimates from radiosondes) of the temperature of the atmosphere with the bulk of the signal arising from the troposphere. The first MSU product, TMT (Temperature Middle Troposphere) is an average of the atmospheric temperature with a weighting function that extends from the surface to the lower stratosphere, and peaks about 5 km above the surface. The second MSU product, TLT (Temperature Lower Troposphere) involves a mathematical recombination of several of the off-nadir view scenes of the TMT channel [cf. *Mears et al.*, 2011, Figure 4]. The TLT product has a weighting function that peaks much lower, about 2 km above the surface [*Mears and Wentz*, 2009a; *Spencer and Christy*, 1992].

[17] MSU/AMSU equivalent temperatures have been calculated from the radiosonde data (gridded temperatures on standard pressure levels) using a method based on radiative transfer calculations [*Rosenkranz*, 1975, 1993] and simple models of the surface emissivity [*Meissner and Wentz*, 2012; *Prigent et al.*, 2000].

3.1. Satellite-Based Products

[18] All satellite records considered here are derived from the MSU/AMSU series of microwave sounders, which make measurements near a complex of Oxygen absorption lines centered at 60 GHz [*Smith et al.*, 1979]. For TMT, we consider the most recent versions of the data set constructed by three different groups, the University of Alabama, Huntsville (UAH V5.4) [*Christy et al.*, 2003], the Satellite Technology and Research Division at NOAA (STAR 2.0) [*Zou et al.*, 2006, 2009; *Zou and Wang*, 2011], and Remote Sensing Systems (RSS V3.3) [*Mears and Wentz*, 2009a, 2009b]. For TLT, only two versions are available, UAH V5.4 and RSS V3.3. TLT is more influenced by radiation emitted by Earth's surface than TMT. Also, the diurnal cycle in surface skin temperatures is much larger than that in the free troposphere or even that of the near-surface air temperatures. Taken together with the uncertainty introduced through the weighted combination of view angles, this yields larger uncertainties in TLT than TMT [*Mears and Wentz*, 2005, 2009a; *Mears et al.*, 2011]. In all cases the independent groups have applied different methods to characterize and remove errors in the raw measurements associated with calibration errors and drifting local measurement times.

3.2. Radiosonde Based Products

[19] For radiosondes, we take the approach of considering all published homogenized radiosonde data sets: HadAT2 [*Thorne et al.*, 2005b], RAOBCORE (Version 1.4) [*Haimberger*, 2007], RICH (Version 1) [*Haimberger et al.*, 2008], IUK [*Sherwood et al.*, 2008], and RATPAC [*Free*

et al., 2005; *Lanzante et al.*, 2003]. The first four data sets are fully or partially (in the case of HadAT operational version) automated methods to find and estimate the size of “breakpoints” in the time series for a radiosonde station and create adjusted versions of the radiosonde data. The IUK data set ends in 2006. The version of the RATPAC data used here, RATPAC-B, uses a manual breakpoint detection and adjustment method for data from 1958 to 1997. After 1997, no further adjustments were made. All other data sets made adjustments over the entire period of record. We use RATPAC-B because it is a station-based data set, and thus has sufficient information to construct a gridded data set, which is required to perform the spatial sampling steps in our analysis. The companion data set RATPAC-A, contains automated adjustments up to the present time, but is not available in gridded form. We also show results for a subset of the RATPAC data set (RATPAC_RW) which was developed by *Randel and Wu* [2006] to remove radiosonde stations with the largest errors in the stratosphere. It is not known the extent to which these errors are also present in these stations' tropospheric measurements that are studied here. The RATPAC_RW subset is included here solely because it was the focus of the earlier satellite/radiosonde comparison study performed by *Randall and Herman* [2008].

3.3. Data Set Internal Uncertainty Estimates

[20] For the RSS (satellite) and HadAT (radiosonde) products there exist ensembles that attempt to quantify the data set uncertainty [*Mears et al.*, 2011; *Titchner et al.*, 2009; *Thorne et al.*, 2011b]. Such estimates can help to inform on whether differences between pairs of data sets arise through chance choices given the chosen methodological frameworks employed or reflect very real and substantial impacts of differences in methodological choices. The MSU error ensembles were calculated by combining estimated errors in the adjustments for measurement time drift and realizations of the spatial/temporal sampling noise introduced by incomplete sampling [*Mears et al.*, 2011]. The resulting error data sets are available at the same spatial and temporal resolution as the base data set. This makes it possible to construct 400 realizations of this data set that are consistent with the estimated uncertainty by adding them to the baseline, satellite derived temperature data set. The HadAT ensemble used herein are derived from perturbed versions of the Hadley Centre's automated neighbor based homogenization procedure. Here we utilize the 20-member seasonal ensemble described in *Thorne et al.* [2011b] that most closely recreated the real climate system behavior across a range of analogs to the real world created by *Titchner et al.* [2009]. Further details on the derivation of these uncertainty products are given in the auxiliary material and in the referenced papers.¹

[21] The radiosonde and MSU data ensembles being considered are fundamentally distinct. The Remote Sensing Systems MSU error model [*Mears et al.*, 2011] is a perturbed ensemble around an assumption of essentially zero mean bias in the operational product version. The radiosonde ensemble makes no assumption about correctness of

¹Auxiliary materials are available in the HTML. doi:10.1029/2012JD017710.

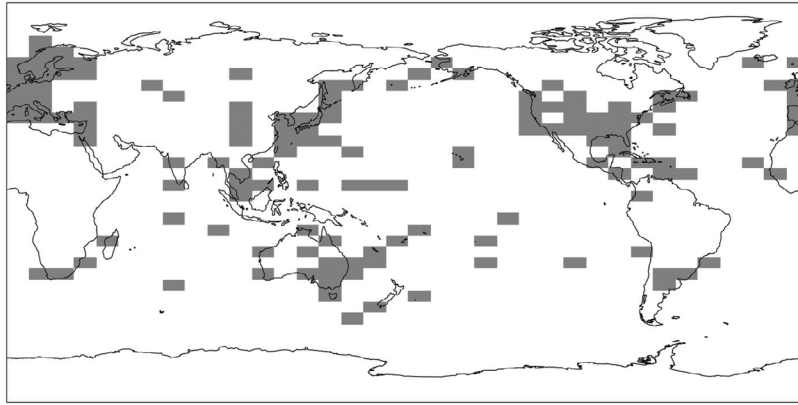


Figure 1. Sampling pattern for the HadAT data set for January, 2003. Grey boxes represent locations of valid data (5 degree latitude by 10 degree longitude gridboxes with at least one reporting radiosonde station for the month), and white regions are areas where data is not available.

the operational solution but rather undertakes fundamental end-to-end recalculations of the solution with a focus solely on breakpoint identification and adjustment issues to create the ensemble. Further, the sources of error are distinct for the MSU and radiosonde issues as would be expected given that they are fundamentally different observing technologies, each with unique issues. Therefore although the error estimates may be presented in a similar manner and may ostensibly appear very similar to the reader we would caution against over-interpretation as they are not strictly intercomparable.

[22] Despite these caveats, it is useful to combine the two error analyses to produce an estimate of the expected error in the RSS-HadAT difference time series that we evaluate herein, so that the statistical significance in the HadAT-RSS difference time series can be evaluated. To do this we difference pairs selected from the above ensembles to create an ensemble of difference time series. Because there are far fewer HadAT ensemble members available, each HadAT ensemble member is paired with 20 different RSS ensemble members, to yield an ensemble of 400 possible difference time series.

4. Time Series Comparisons

[23] We begin by comparing the MSU and homogenized radiosonde temperature anomalies time series of large spatial scale averages which are useful because of the significant uncertainties in the measurements from isolated radiosondes and in single MSU grid points. Over larger spatial scales, many components of these uncertainties are reduced by the averaging procedure. We choose to focus on monthly means, averaged over nearly the entire globe (75S to 75N, “global”), or deep tropics (20S to 20N, “tropical”). In both regions, the radiosonde spatial coverage is far from complete. Figure 1 shows a typical radiosonde sampling pattern. We also performed our analysis for the northern and southern extratropics separately. Summary plots from these analyses are presented in the auxiliary material.

[24] A simple (and common) way to construct time series from gridded data is to use area-weighted means of all available data. Comparing time series of simple area-

weighted global averages of radiosonde data with the area-weighted means of the more spatially complete MSU data can lead to substantial discrepancies, due to the large areas that are unsampled by radiosondes, and the changes in radiosonde sampling over time. Additional discrepancies occur because the radiosonde sampling patterns in the tropics exclude the eastern tropical Pacific Ocean, where the ENSO signal is often the strongest. A good approach to resolving these issues is to sample the MSU data at the actual radiosonde sampling for each month [Free and Seidel, 2005; Mears and Wentz, 2009b], and then compute an area-weighted average from the sub-sampled data for each month to produce a “global” average. This modifies the MSU means such that they more closely match the area-weighted radiosonde means, and automatically takes into account the presence or absence of a radiosonde measurement for a given location and month and thus changes in spatial sampling over time. We refer to the sampled satellite means as “sampled at radiosonde locations” (SRL).

[25] Because simple area-weighted averages have often been used to perform radiosonde/satellite intercomparisons, we show the results for both methods to assess sensitivity to this choice. In Figures 2a and 2b, we show area-weighted global time series for the operational HadAT product and each MSU data set for both TLT and TMT. The large month-to-month variations in all data sets make it difficult to draw conclusions from this plot. In Figures 2c and 2d, we show the area-weighted difference time series (HadAT – MSU) for each MSU data set, and in Figures 2e and 2f, we show the same difference series, except using SRL averaging to calculate the satellite time series. In all cases shown, the SRL time difference time series exhibit much less variance than the area-weighted time series. We made similar plots for the other 5 radiosonde data sets we consider (see auxiliary material). In all cases, the standard deviations of the difference time series were reduced, particularly on short time scales. Even when the difference time series were filtered to remove variability on time scales shorter than one year using a digital filter [Lynch and Huang, 1992], 11 of 12 TLT cases and 13 of 18 TMT cases showed reduced standard deviation, suggesting that the SRL procedure also tends to improve the agreement on interannual and longer

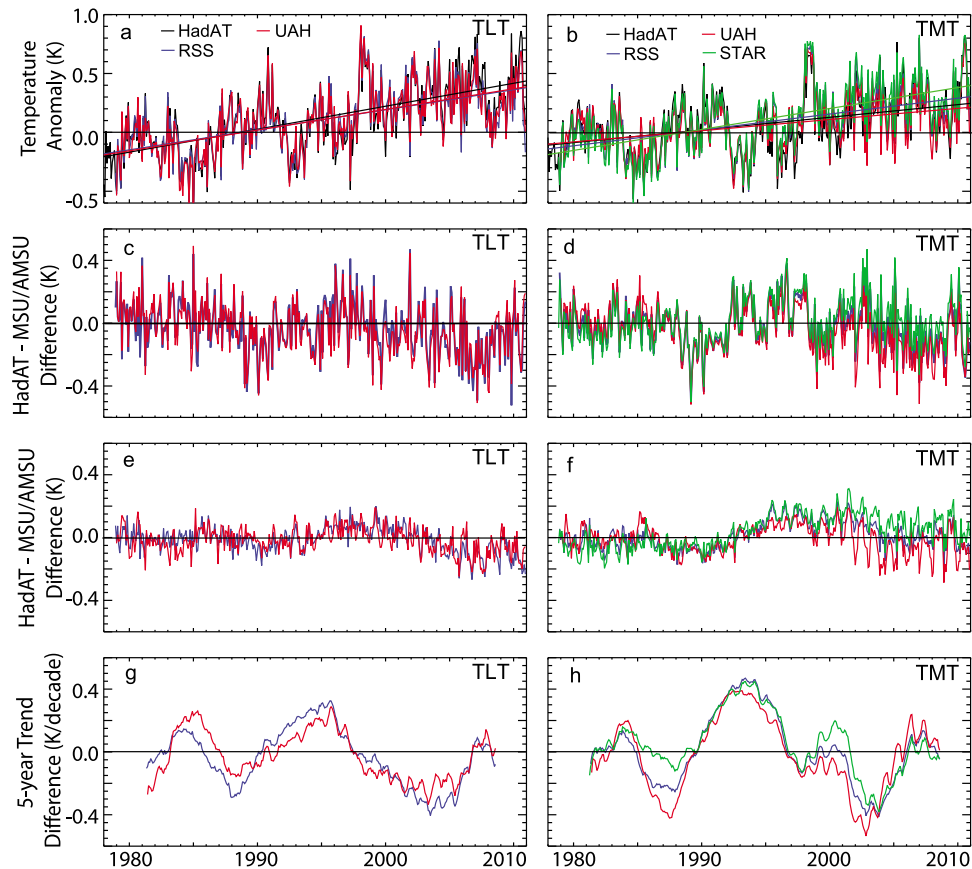


Figure 2. Graphical representation of various stages in processing the radiosonde and satellite data. (left) TLT data and (right) TMT. (a and b) Unmodified globally average time series for the HadAT radiosonde data set, and each of the satellite data sets. (c and d) The difference between HadAT and each of the satellite data sets (satellite - HadAT) using area-weighted averaging. (e and f) The difference between HadAT and each of the satellites data sets using SRL averaging for the satellite data. Note the large reduction of differences on most time scales. (g and h) 5-year trends calculated from the SRL differences in Figures 2e and 2f.

time scales. These results reinforce our previous conclusion [Mears and Wentz, 2009b] that MSU/radiosonde comparisons are best performed using SRL averaging.

[26] Even with SRL sampling, the difference time series show significant variability on intraannual time scales that makes it difficult to draw conclusions. One approach that has been used to help reduce the contribution of short-time scale variability is the analysis of trends in intermediate-length sub-samples of a longer time series. Randall and Herman [2008] used the Randel and Wu [2006] subset of the RATPAC radiosondes to analyze 5 and 10 year trends in tropospheric UAH and RSS data. In Figures 2g and 2h, we show plots of trends of rolling 5-year sub-samples of the difference time series, with each slope plotted at the location of the center of the sub-sample. When the 5-year trend is greater than zero, the MSU data showed warming relative to the radiosonde data over the 5-year period, and conversely, when the 5-year trend is less than zero, the MSU data cooled relative to the radiosonde data. It is immediately obvious that the 5 or 10-year trends can accentuate both the intermediate and long time-scale differences, as concluded by Randall and Herman. For example, in the TMT data (Figure 2f) the MSU data (for all 3 MSU data sets) warms relative to

HadAT over the 1990–1998 period. This is easier to see in Figure 2h as large maxima in the short term trends which reach their greatest magnitude at the center of this period.

[27] In Figures 3 (and 4), we plot the 5-year trends in global (tropical) difference time series for both TLT and TMT and for all combinations of radiosonde and MSU data sets (in the auxiliary material, we show similar sets of plots for the northern and southern extratropics, see Figures S6 and S7). There are several common features that stand out. For TLT, perhaps the most obvious is a peak in the RSS and UAH minus radiosonde differences centered near 1995 which is consistently somewhat larger for RSS. This feature, due to the warming of MSU data relative to radiosonde data, has been previously discussed in the literature as it occurs during the period where the RSS warms relative to the UAH data set [Christy and Norris, 2009; Christy et al., 2007; Randall and Herman, 2008]. The better agreement between UAH and the radiosondes (as shown by the lower peak for UAH) during this period, combined with the sign of temperature changes in the tropics during the period surrounding the eruption of Pinatubo, has been used by these authors to argue that the RSS data set contains a warming bias during this period. The analysis of this period is complicated by the

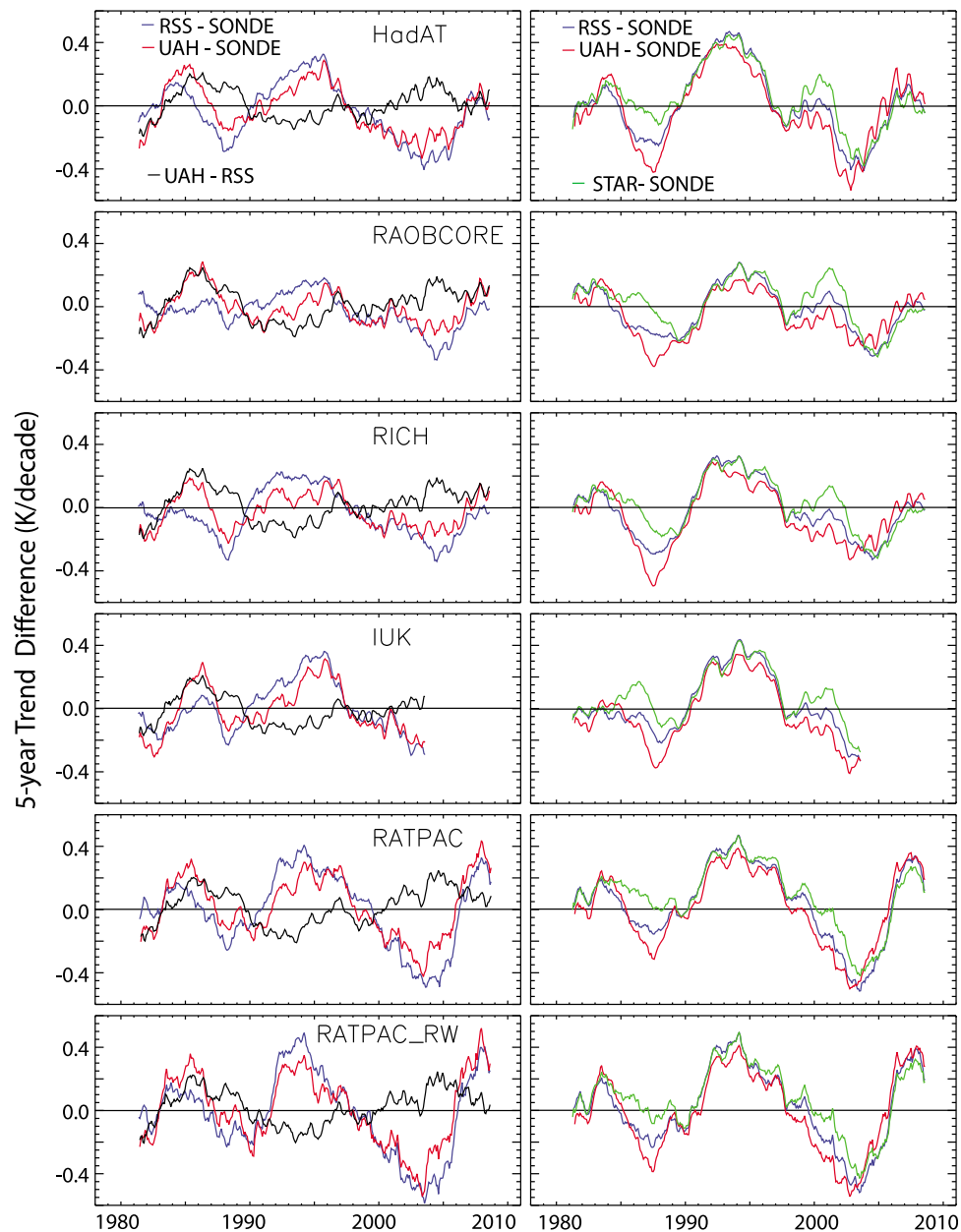


Figure 3. Difference in global SRL-averaged 5-year trends (Satellite – Radiosonde) for each radiosonde and satellite data set. (left) The TLT channel and (right) TMT. For TLT, we also plot the SRL difference between the two satellite data sets (UAH – RSS).

competing effects of volcano-induced cooling and ENSO-induced warming. This is also the period over which general improvements to radiosonde solar radiation shielding yielded an apparently artificial cooling across much of the radiosonde network [Sherwood *et al.*, 2005], an effect which may not entirely have been removed with available radiosonde data sets [e.g., Sherwood *et al.*, 2008]. Hence interpretation of these differences is fraught with physical and instrumental considerations that significantly inhibit a clean inference regarding which MSU product may be closer to the unknown true temperature evolution.

[28] We note that there are two other prominent features in the 5-year trend plots that have not received as much attention. First, there is a second common feature, a low point

centered near 2003. This indicates a period when the MSU data sets are cooling relative to the radiosonde data sets. This feature roughly coincides with the end of the data record for NOAA-14, the last MSU satellite, and may be related. We note that for TMT, there is an unexplained trend difference between MSU and AMSU measurements during 1999–2005 [Mears *et al.*, 2011]. Second, UAH tends to show a peak centered near 1986, which is either absent or much smaller for RSS and may be related to calibration problems with the NOAA-9 satellite, a finding confirmed by Po-Chedley and Fu [2012]. Over the entire time series, the effects of the relative warming in the 1990s and the relative cooling in the 2000s tend to cancel, leaving the 32-year trends for MSU

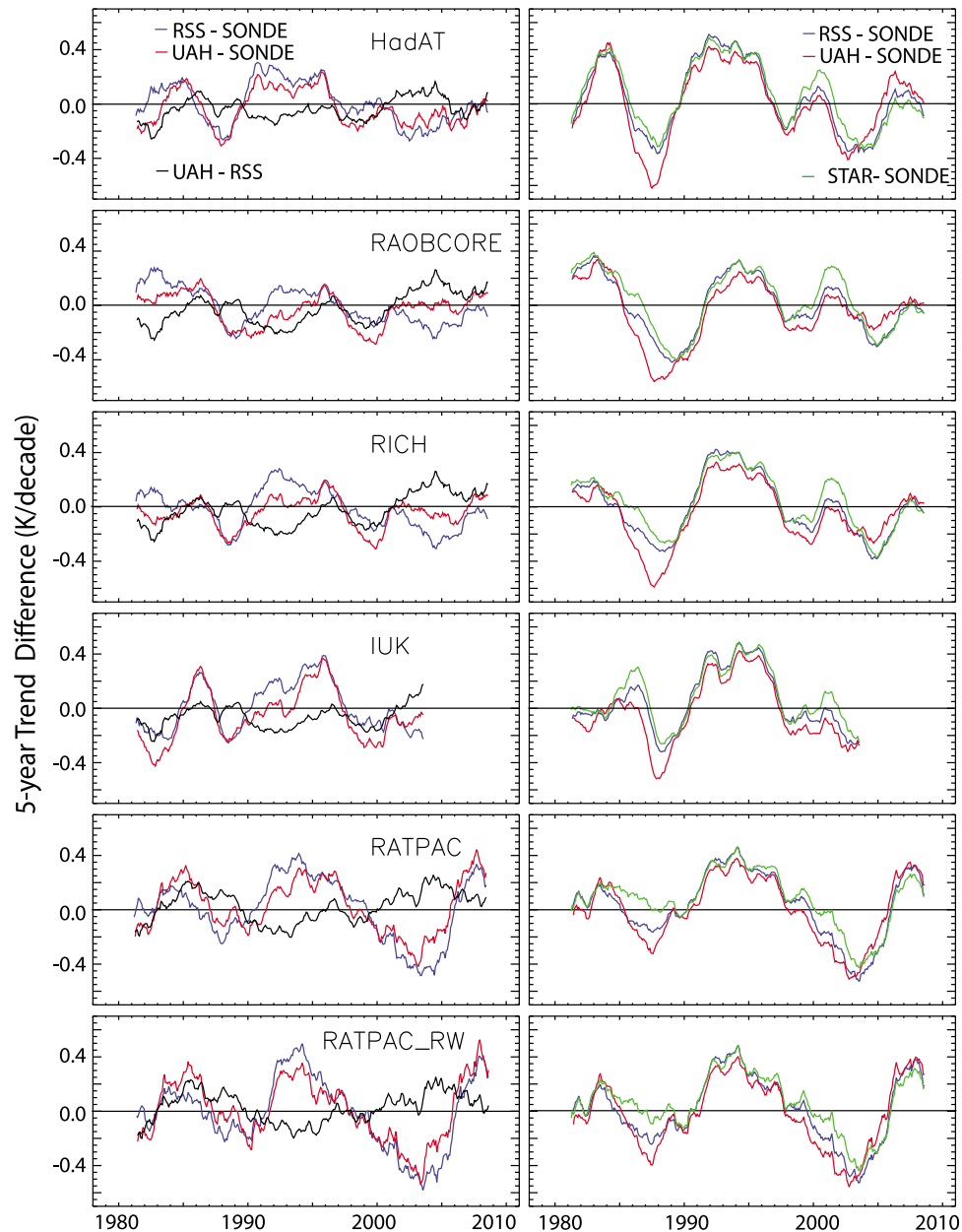


Figure 4. Same as Figure 3, except that the data are averaged over radiosonde locations in the deep tropics (20S-20N). Similar sets of plots for the southern and northern extratropics are available in Figures S6 and S7 in the auxiliary material.

and radiosonde data in relatively good agreement [Mears *et al.*, 2011].

[29] For the TLT plots in Figure 3 (left), we also plot the 5 year trend differences for RSS-UAH. Using the mean absolute value of each of the trend difference curves over the entire time period in Figure 5a in all cases except for RAOBCORE and RICH, the MSU data sets are in closer agreement with each other than they are with the radiosonde data sets. (We also evaluated the differences using the root-mean square difference as a difference metric, which yielded nearly identical results. See Figure S8 in the auxiliary material.) Note that the RAOBCORE data set, which has been asserted to be corrupted by anomalous warming due to underlying errors in the ERA-40 reanalysis used in its

construction [Christy *et al.*, 2010], is the radiosonde data set that agrees best with both satellite data sets when evaluated using our method. Given that the reanalysis field used to derive adjustments in RAOBCORE is strongly influenced by the MSU/AMSU data (among others) this is perhaps not surprising. The RICH data set generally agrees second best with the MSU data. This data set is more independent from the background reanalysis, and thus is less subject to the criticisms put forth in Christy *et al.* [2010].

[30] The ordering of the level of agreement between the radiosonde data sets and any of the MSU data sets (RAOBCORE, RICH, IUK, HadAT, RATPAC) is the same as the ordering of the radiosonde trends over the entire satellite era. If the 5-year trends were strongly influenced by the

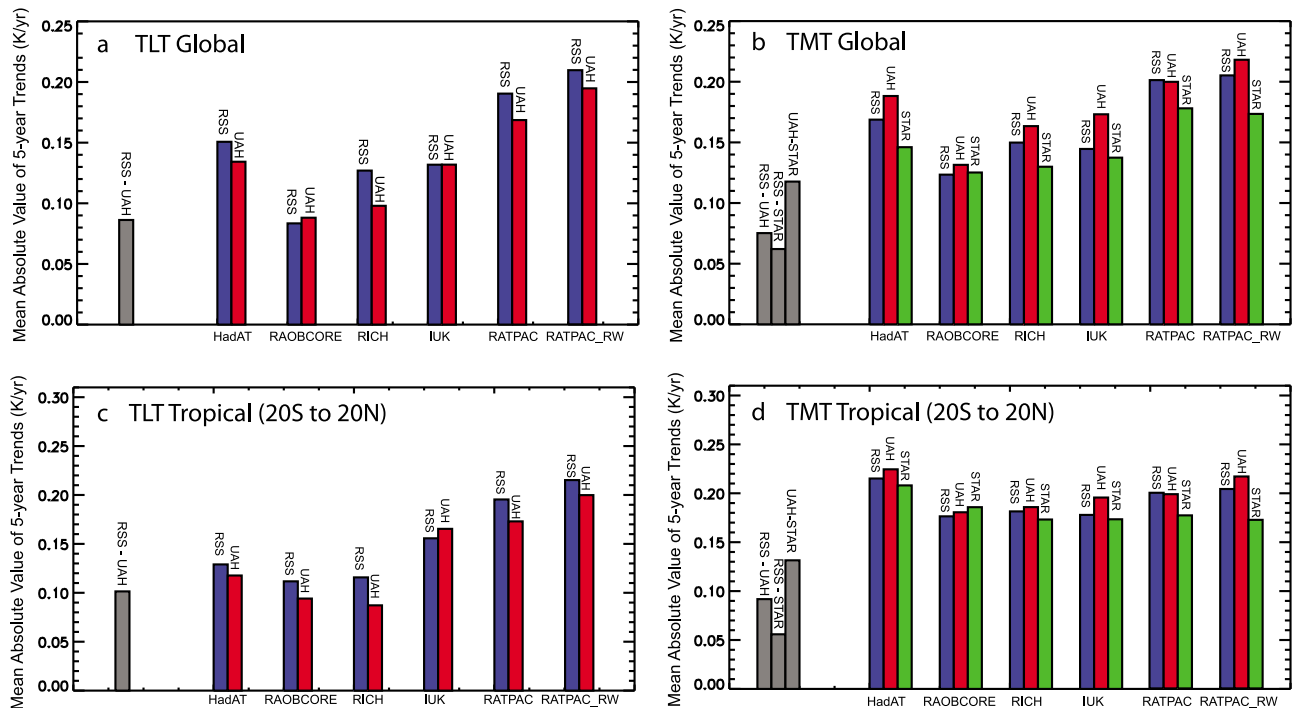


Figure 5. Summary of the mean absolute value of the overlapping 5-year trend difference time series from Figure 4. (a) TLT, global. (b) TMT, global. (c) TLT, deep tropics (20S–20N). (d) TMT, deep tropics. In most cases, the difference between different satellite data sets is substantially less than the difference between the satellite data sets and the radiosonde data sets. The exception to this is for the RAOBCORE and RICH data sets for TLT.

overall trend, this would be expected on mathematical grounds, as the overall trend is related to the accumulated 5-year trends. However, the difference between 5-year trends is dominated by differences on short time scales. We checked this by performing a second set of calculations with the overall trend removed from each time series before the 5-year trends were calculated. The resulting version of Figure 5 (see Figure S9 in the auxiliary material) is nearly identical.

[31] For TMT, we also include data from the STAR MSU data set. Again, there are several common features across all MSU and radiosonde data sets, including the peak in the mid 1990s, and a minimum in the mid 2000s. There is a second sharp minimum centered near 1987 in the UAH curve, a feature that is significantly reduced in the RSS data, and almost nonexistent in the STAR data. Also note the appearance of a strong seasonal cycle in the UAH data after 1998. (This feature is even more prominent in Figure 2, and may be related to the method used to combine MSU and AMSU measurements. AMSU measurements began in the middle of 1998). Again, the difference between the MSU data sets and the radiosonde data sets is larger than the spread between the MSU data sets themselves, as shown by the mean absolute values of the difference curves that are plotted in Figure 5b.

[32] Figure 4 is analogous to Figure 3, except the data are averaged over the deep tropics (20S to 20N) instead of the entire globe, with summary results presented in Figures 5c and 5d. The set of tropical plots shares many of the features of the global plots, such as the relative warming in the MSU data sets in the mid 1990s, and the relative cooling in

the late 1980s and early in the 21st century. One important difference is that there appear to be more differences between the radiosonde data sets in the 1990s, with the MSU data sets showing strong warming relative to radiosondes in IUK and RATPAC, with considerably less warming relative to RAOBCORE and (to a lesser extent) RICH. For HadAT, most of the relative MSU warming is shifted to a short time period in the early 1990s. It is not surprising that the differences between data sets are larger in the tropical case because the number of radiosonde stations in the sample has decreased, leading to more variability. Also, in the tropics, the radiosonde coverage is more sparse than in the northern extratropics, making it more difficult to perform the necessary adjustments for those methods that rely to some extent upon comparisons with near neighbors (HadAT, IUK, RICH, and to a lesser degree, RATPAC). Finally, many tropical sites have had daytime-only ascents which are most impacted by solar heating effects and these were significantly mitigated through the 1990s [Sherwood *et al.*, 2005; Randel and Wu, 2006] leading, on average, to a false cooling signal in the raw record, which may not have been entirely removed. Differences may relate to the efficacy of the various radiosonde products in removing this artifact and it is important to note that the sign and timing implies that possibly none of the radiosonde products have adequately removed this artifact rather than a consistent bias in MSU products. In Figure 5c (TLT), the best agreement between tropical satellite and MSU results is for the RAOBCORE and RICH data sets, with the UAH data set in better agreement than RSS except for the IUK data set, and in Figure 5d,

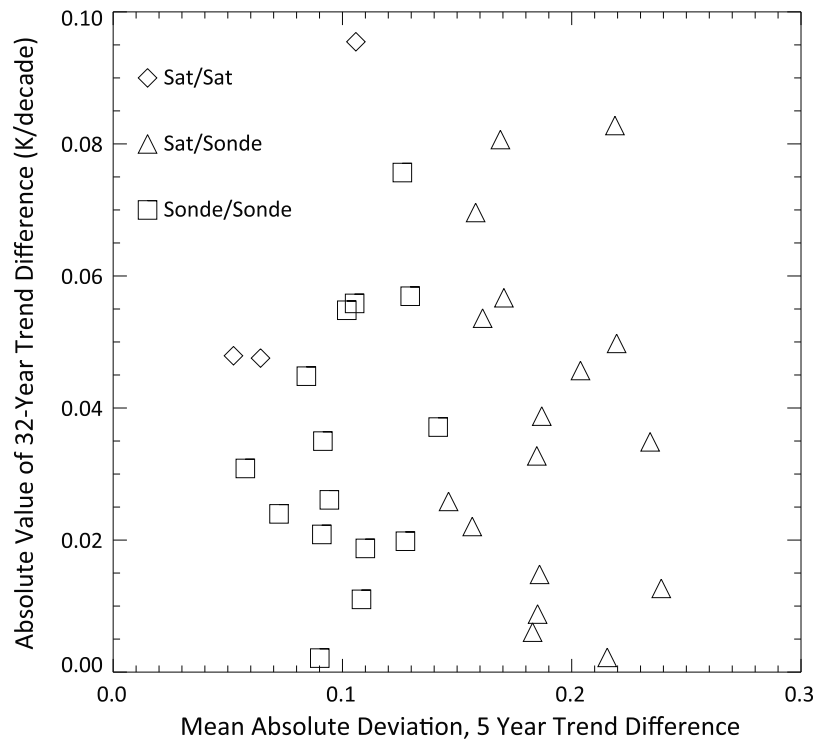


Figure 6. Scatterplot of the absolute value of the 32-year trend difference (1979–2010) in global means of TMT as a function of mean absolute deviation of the 5-year trend difference time series computed from the same difference time series. Each point represents the comparison of one data set (satellite or radiosonde) with another (satellite or radiosonde).

the best agreement again is with the ROABCORE and RICH data sets, but with the UAH data set typically showing more disagreement than the RSS and STAR data sets. For TMT, the RSS and STAR data sets are extremely close to each other on the 5-year scale, despite substantial differences in 32-year trends.

[33] The results for RATPAC-RW shown here differ from those presented in *Randall and Herman* [2008] for two important reasons. First, newer versions of the both the RSS and UAH MSU data sets were used in this analysis. Second, and more importantly, Randall and Herman did not subsample the MSU data at the radiosonde locations, but instead compared global radiosonde averages to area-weighted, land-only MSU averages. A majority (29 of 47) of the RATPAC-RW stations are located on islands (9 stations) or in coastal regions (20 stations) and thus are not representative of a land-only average (see auxiliary material for a map showing these results, and a precise description of our definition of land, coastal, or ocean).

[34] Despite extensive efforts we were unable to exactly replicate the results of Randall and Herman due to insufficient methodological clarity in their paper and thus no direct comparison is possible here. To illustrate the relative importance of variations in sampling method and data set updates, we present several alternative versions of Figure 5b in the auxiliary material (see Figure S10). We find that the changes in data set version are less important than changes in sampling method, and that while the use of land-only MSU data reduces the mean absolute difference relative to the use of global land-and-ocean averages, both are substantially worse than sampling at the radiosonde locations.

[35] Another question to investigate is the degree to which agreement on the 5-year time scale is useful for predicting agreement on a longer time scale. In Figure 6, we plot the absolute value of the difference between multidecadal trends in globally average TMT (1979–2010, except for IUK, which ends in 2006) as a function of the mean absolute value of each of the trend difference curves. This measure of 5-year trend agreement is the same as is plotted in Figure 5. The plot shows very little correlation (correlation coefficient = 0.025) between the level of agreement on 5 year time scales, and the agreement between multidecadal trends. This suggests that the agreement between 5 year trend time series is essentially useless for predicting agreement on longer time scales. This therefore calls into question many of the assertions made explicitly within or publicly (and often in a high profile manner) as a result of such inter-comparisons [e.g., *Randall and Herman*, 2008; *Christy et al.*, 2010; *Po-Chedley and Fu*, 2012] with regards to the fundamental quality or likely long-term trend errors in candidate products found to be ‘anomalous’ in a given sub-period.

5. Impact of Uncertainty Estimates

[36] The preceding section serves as an estimate of the structural (or between method) uncertainty in MSU/radiosonde comparisons. We now investigate the impact of internal uncertainty on both the MSU and radiosonde data sets. Errors in both types of data set are often correlated in both time and location. Only the RSS and HadAT data sets have associated internal uncertainty estimates that are

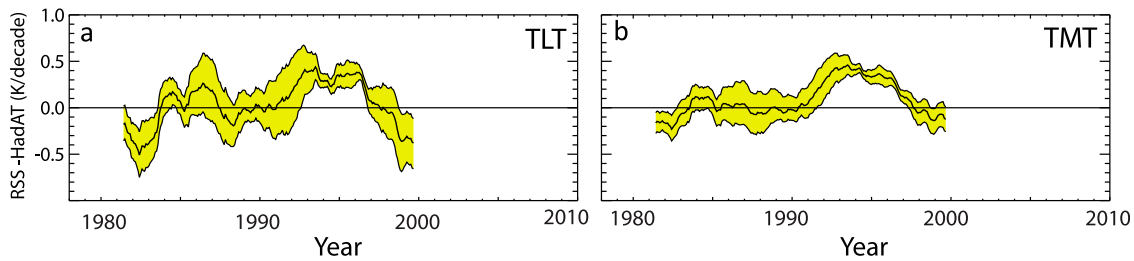


Figure 7. Differences between 5 year trends with uncertainty estimates for channels TLT and TMT. All time series and uncertainty estimates are global averages sampled at the location of HadAT radiosonde measurements (which end in 2003 in this ensemble and hence the series truncation). The thick black line is the median difference time series, and the yellow region surrounding it contains 95% of the instances in the uncertainty ensemble. The uncertainty ensemble includes the estimated uncertainty in both the RSS and HadAT data sets.

sufficiently detailed to accurately estimate the error in trends at various time scales (Section 2.3).

[37] For the RSS/HadAT case, in Figures 7a and 7b, we show the median difference between 5-year RSS - HadAT difference trend error ensemble (we emphasize that the median of the HadAT error ensemble is quite different from the operational version of HadAT considered in Section 3). We also plot the 95% confidence interval (CI) around the median difference. The 95% confidence interval was calculated from an ensemble of 400 realizations of the RSS-HadAT difference. The n th member of this ensemble was constructed by subtracting the $(n \bmod 20)$ th member of the HadAT ensemble from the n th member of the RSS ensemble. For both TLT and TMT, the 95% CI range for the differences between RSS and HadAT encompasses the zero line 59% of the time, implying that the differences between the 5-year trends are larger than can be easily accounted for by the combination of internal errors. We note that the HadAT adjustment procedure was not designed to remove errors on short time scales, so that the short-term error represented by the ensemble may underestimate the true error in the radiosonde data.

[38] Figures 8a–8c summarize a similar analysis of the differences between the RSS and UAH satellite data sets

(Figures 8a and 8b), and between the RSS and STAR satellite data sets (Figure 8c). Because detailed uncertainty ensembles are not available for the UAH and STAR data sets, this part of the analysis only includes uncertainty estimates for the RSS data sets. For TLT, the RSS-UAH 95% uncertainty range encompasses the zero line 63% of the time, and for TMT the uncertainty ranges encompass the zero line for 30% (UAH) and 62% (STAR) of the time period. We speculate that if the uncertainty in the UAH and STAR data sets were included, these percentages would increase, but the exact amount is difficult to reliably estimate without comparably comprehensive uncertainty analyses from the UAH and STAR groups.

6. Inter-channel Differences

[39] *Randall and Herman* [2008] (hereafter RH2008) also studied the differential trends between the TLT and TMT layers for the UAH, RSS, and RATPAC-RW data sets. They motivated this work as a method for diagnosing the impact of the diurnal cycle on the merged data set, since the adjustments made for changes in local measurement time are much larger for TLT than TMT. We recommend using such differences with caution. Because of the overlap between the

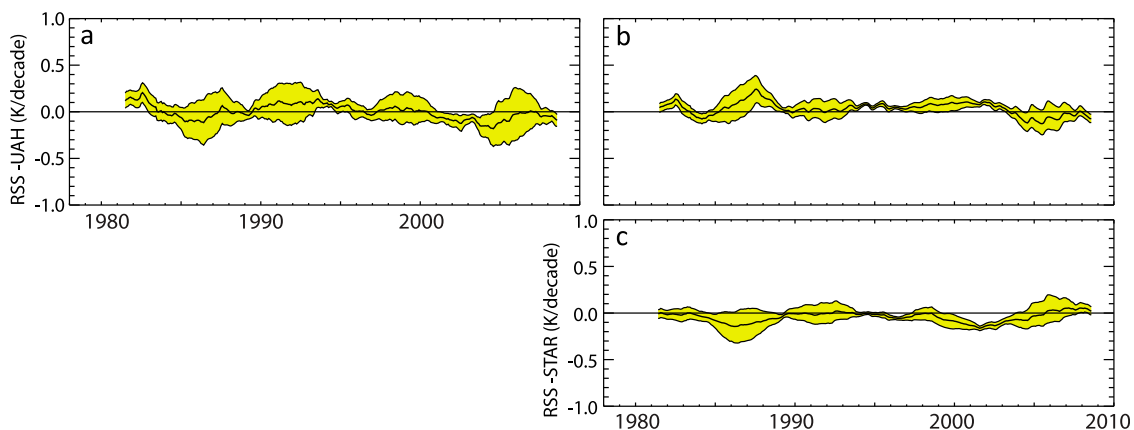


Figure 8. Differences between 5 year MSU trends with RSS-only uncertainty estimates for channels TLT and TMT. (a) RSS-UAH for TLT. (b) RSS-UAH for TMT. (c) RSS-STAR for TMT. In all panels, the thick black line is the median difference time series, and the yellow region surrounding it contains 95% of the instances in the RSS uncertainty ensemble.

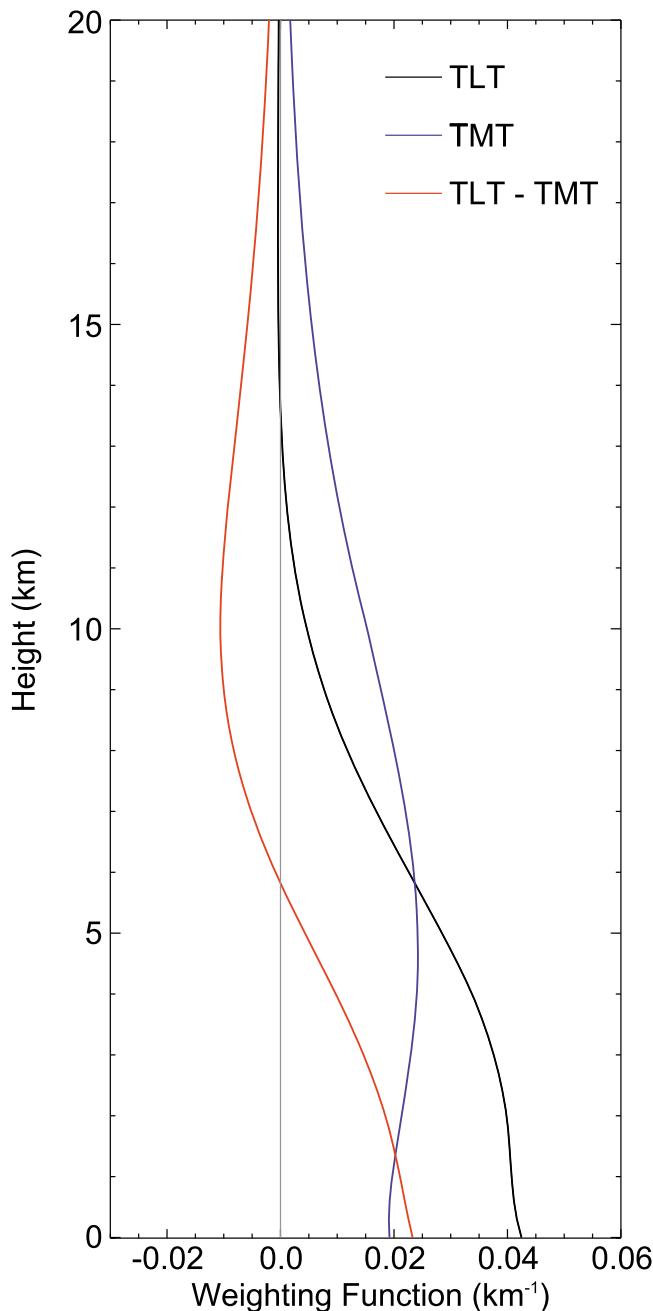


Figure 9. Weighting functions as a function of height for TLT, TMT, and the TLT minus TMT difference for ocean scenes. Because of the large absolute weight in the stratosphere and the surface, the TLT minus TMT difference should not be interpreted as the difference between the lower and middle tropospheric temperatures. There is also a non-negligible contribution from surface emission in each product: TLT 10.6%, TMT 4.8%, and TLT minus TMT 5.8%.

TLT and TMT weighting functions and the subsequent cancellation caused by differencing, the TLT – TMT difference contains a large amount of information from the surface and lower stratosphere. Furthermore, a large portion of the resulting weighting kernel (including most if not all of the atmosphere above 6 km) has a negative weighting which is hard to interpret in a physical manner. Figure 9 shows the

temperature weighting functions for TLT, TMT, and the TLT – TMT difference. About 12% of the total absolute value of the weight comes from surface emissions, and about 25% comes from above 12 km, or 200 hPa. These differences should not be thought of as the difference in temperature between the lower and middle troposphere. This is particularly a concern for radiosonde data because of the increase in the relative weight of radiosonde measurements at pressures ≤ 200 hPa where the adjusted radiosonde data sets may be less reliable [Randel and Wu, 2006].

[40] With these caveats in mind, here we update (to the extent we are able to replicate) and extend the RH2008 analysis. There are 4 important differences between the present analysis and RH2008. First, we use the most recent versions of the available data sets. Second, we consider all available homogenized radiosonde data sets (Section 2.2) rather than a single estimate. Third, we use SRL averaged MSU data instead of land-only area-weighted data (Section 3). And fourth, we consider results from outside the limited temporal ranges plotted in RH2008. Figure 10 shows the 5-year and 10-year trend differences between TLT and TMT for each radiosonde data set. In each case, the SRL MSU averages are plotted, which accounts for the difference in the RSS and UAH curves between plots. These plots correspond to Figure 4 in RH2008, which they used to argue that the UAH satellite data sets were more accurate than the RSS versions. The bottom row of plots shows results from the RATPAC-RW data set analyzed by RH2008. In Table 1, we present a summary of the mean absolute differences (MAD) and the number of months that each satellite data set is closer to the radiosonde data set. For the 5-year trends, the 3 data sets are in reasonably good agreement over the bulk of the time period, in agreement with the findings of RH2008, though in general, we find that the RSS data are in better agreement with the radiosonde data than the UAH data for these metrics (Table 1). For the 10 year time period, we find that the largest differences are outside the region plotted by RH2008, and that within the 1993–2002 period plotted, the RSS data set is in better agreement with the radiosonde data than UAH, in direct contradiction to the findings of RH2008. The probable reasons for this different result are both the use of MSU data sampled at the radiosonde locations (instead of land-averaged satellite data), and (to a lesser degree) the use of updated versions of the MSU data sets. The largest differences between the MSU data sets and radiosonde data sets tend to occur in the early part of the time series for both the 5-year and 10-year trends, outside the region plotted in RH2008. Again, we find that the two MSU data sets tend to be closer to each other than to any of the radiosonde data sets.

[41] For the other 5-year period plots, the conclusions are similar to that reached above for the RATPAC-RW data. The MSU data sets are in fairly good agreement with the radiosonde data sets during the period plotted in RH2008. During the pre-1990 period the analyses typically show TLT warming relative to TMT more in the radiosonde data than in the MSU data sets, with the MSU data sets being relatively similar. For the 10 year time period plots, this difference becomes more important, with TLT warming much more than TMT in the radiosonde data sets. During the 1993–2002 period plotted in RH2008, the results differ substantially from radiosonde data set to radiosonde data set,

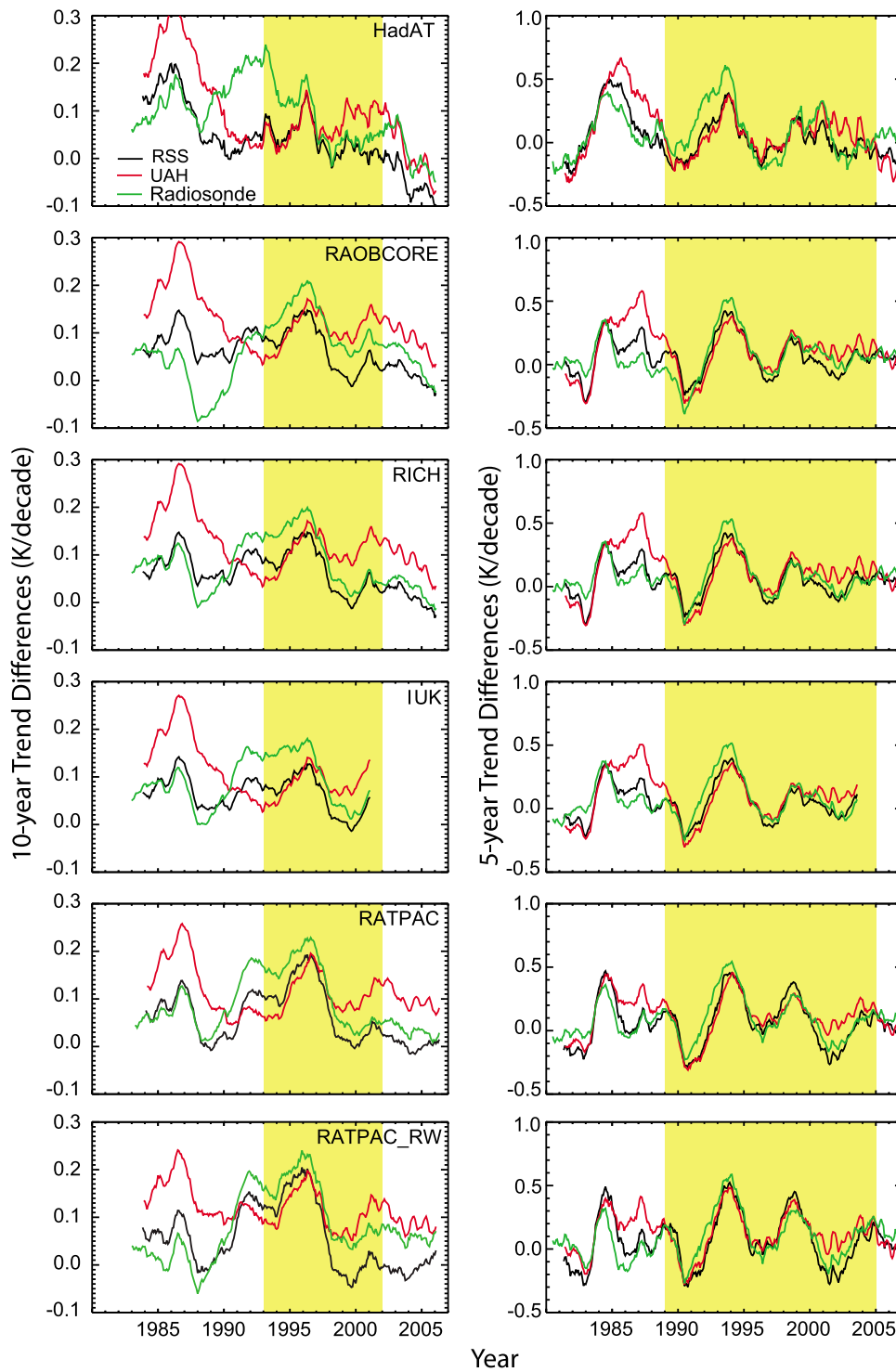


Figure 10. Differences (TLT minus TMT) in short-term trends in atmospheric temperature between adjacent layers for each radiosonde data set. In each case, the satellite data is sampled at the radiosonde locations, and the differences in spatial sampling lead to the differences between the satellite results for the different panels. The temporal range plotted in RH2008 is shaded in yellow.

even for the MSU data set SRL estimates, which makes it difficult to draw conclusions. Note that the only difference between the different versions of the MSU data is the spatial/temporal sampling used to construct the averages. After about 2000, TLT tends to warm relative to TMT more in the

radiosonde data sets, except for HadAT, where TLT cools slightly relative to TMT after about 2004.

[42] An improvement that can be made to the RH2008 method is to consider the difference between TLT and the “total troposphere” (TT) MSU product proposed by *Fu and*

Table 1. Interlayer Difference Statistics for 5 and 10 Year Trends

	5-Year Trend Differences				10-Year Trend Differences			
	Number of Months RSS Closer	Number of Months UAH Closer	MAD RSS-Sonde (k/decade)	MAD UAH-Sonde (K/decade)	Number of Months RSS Closer	Number of Months UAH Closer	MAD RSS-Sonde (k/decade)	MAD UAH-Sonde (k/decade)
HadAT	218	109	0.1006	0.1511	210	57	0.0628	0.0791
RAOBCORE	205	122	0.0745	0.1266	172	95	0.0505	0.0847
RICH	222	105	0.0597	0.1219	202	65	0.0291	0.0793
IUK	178	89	0.0647	0.1289	150	57	0.0373	0.0794
RATPAC	217	99	0.0741	0.0984	205	51	0.0339	0.0670
RATPAC_RW	215	101	0.0930	0.1009	174	82	0.0456	0.0690

Johanson [2005]. This product has reduced weight in the stratosphere, and thus is less affected by overall stratospheric cooling, and stratospheric warming events caused by volcanic eruptions. In the auxiliary material, we show an alternative version of Figure 10 calculated using TT instead of TMT. This replacement generally reduces the variability of the time series, probably due to the reduction of the influence of volcano-induced stratospheric warming, but does not alter the conclusions.

7. Discussion

[43] We have used methods similar to those presented in RH2008 to analyze 5- and 10-year trends in adjusted radiosonde and Microwave Sounding Unit (MSU) measurements of tropospheric temperature utilizing an inclusive range of MSU and radiosonde products. In all cases we find that there are several time periods during which there is substantial disagreement between 5-year trends in radiosonde data sets and 5-year trends in the MSU data sets. Sometimes these differences cancel over longer time periods, perhaps leading to false or overly confident conclusions about the agreement between satellite and radiosonde data sets on multidecadal time scales. When data from different MSU – radiosonde pairs are examined, the results indicate that all MSU-sonde differences share many common features, and that in most cases, the differences between radiosondes and MSU is much larger than between different MSU data sets, or between different radiosonde data sets. Given the current state of knowledge, we are unable to determine whether this commonality is due to shared problems in the MSU data sets, or to shared problems with the radiosonde data sets, or a combination of both. It is possible that both types of data sets retain substantial common biases within their respective types. For MSU data the three different versions are derived from identical raw source data. If there is a time-dependent bias in the raw data that none of the merging procedures is able to detect and remove, then the common bias would obviously remain in all three data sets. A similar argument holds for the radiosonde data sets, though in this case, the underlying, unadjusted data sets differ in the number and locations of radiosonde stations used.

[44] In addition, an analysis of the internal error in the MSU data sets suggests that the differences between RSS and UAH 5-year trends are possibly not statistically significant for TLT, while the differences between the RSS, STAR and UAH data sets may be significant for TMT. This type of

analysis is hampered by the lack of a detailed error analysis in the UAH and STAR products.

[45] Although radiosonde MSU comparisons have some information content, on their own they are ill-posed to assess satellite data set quality issues because both types of data almost certainly retain unknown and poorly quantified biases. Additional entirely independent measurements such as from the Hyperspectral Infrared Sounder, GPS Radio Occultation or reanalyses may help. But these are additionally fraught by a variety of issues relating to sampling (clear sky only for HIRS, a different atmospheric volume and temporal samples of opportunity for GPS-RO), interpretation (both satellite measures are responsive to more than just temperatures), period of record, and independence of record (particularly so for reanalyses). Despite this, bringing in such additional independent estimates may offer a future avenue of investigation. Additional insights may accrue from physical rather than wholly statistical interpretation of the records. Finally, real insights on biases and their causes will only accrue through additional in-depth analyses of the observations and accompanying metadata themselves to better understand the causes of biases and differences in the respective records.

[46] In conclusion, when the similarity of the MSU data sets relative to radiosonde data sets is combined with the lack of statistical significance in many of the difference findings, we conclude that trying to determine which MSU data set is “better” based on short-time period comparisons with radiosonde data sets alone cannot lead to robust conclusions. This is trivially true for any case where two poorly constrained and understood measurements of the same measurand exist. When they disagree the problem is under-constrained such that it is solely possible to conclude that one or both of the measurements is (are) biased relative to the true state of the measurand. Sadly, this is all too common in climate and is why SI traceable measurement programs such as the GCOS Reference Upper Air Network [Seidel *et al.*, 2009] are vital to our future ability to monitor the changing climate.

References

- Christy, J. R., and W. B. Norris (2006), Satellite and VIZ-radiosonde inter-comparisons for diagnosis of nonclimatic influences, *J. Atmos. Oceanic Technol.*, 23(9), 1181–1194, doi:10.1175/JTECH1937.1.
- Christy, J. R., and W. B. Norris (2009), Discontinuity issues with radiosonde and satellite temperatures in the Australian region 1979–2006, *J. Atmos. Oceanic Technol.*, 26(3), 508–522, doi:10.1175/2008JTECHA1126.1.
- 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(5), 613–629, doi:10.1175/1520-0426(2003)20<613:EEOVOM>2.0.CO;2.
- 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.
- Christy, J. R., B. M. Herman, R. Pielke Sr., P. Klotzbach, R. T. McNider, J. J. Hnilo, R. W. Spencer, T. Chase, and D. H. Douglass (2010), What do observational datasets say about modeled tropospheric temperature trends since 1979?, *Remote Sens.*, 2, 2148–2169, doi:10.3390/rs2092148.
- 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. R. 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.
- Fu, Q., and C. M. Johanson (2005), Satellite-derived vertical dependence of tropospheric temperature trends, *Geophys. Res. Lett.*, 32, L10703, doi:10.1029/2004GL022266.
- Haimberger, L. (2007), Homogenization of radiosonde temperature time series using innovation statistics, *J. Clim.*, 20(7), 1377–1403, doi:10.1175/JCLI4050.1.
- Haimberger, L., C. Tavalato, and S. Sperka (2008), Towards the elimination of warm bias in historic radiosonde records—Some new results from a comprehensive intercomparison of upper air data, *J. Clim.*, 21, 4587–4606, doi:10.1175/2008JCLI1929.1.
- Lanzante, J. R., S. Klein, and D. J. Seidel (2003), Temporal homogenization of monthly radiosonde temperature data. Part II: Trends, sensitivities, and MSU comparison, *J. Clim.*, 16, 241–262, doi:10.1175/1520-0442(2003)016<0241:THOMRT>2.0.CO;2.
- Lanzante, J. R., T. C. Peterson, F. J. Wentz, and K. Y. Vinnikov (2006), What do observations indicate about the change in temperatures in the atmosphere and at the surface since the advent of measuring temperatures vertically, in *Temperature Trends in the Lower Atmosphere: Steps for Understanding and Reconciling Differences*, pp. 47–70, Clim. Change Sci. Program, Washington, D. C.
- Lynch, P., and X. Y. Huang (1992), Initialization of the HIRLAM model using a digital filter, *Mon. Weather Rev.*, 120, 1019–1034, doi:10.1175/1520-0493(1992)120<1019:IOTHMU>2.0.CO;2.
- Mears, C. A., and F. J. Wentz (2005), The effect of drifting measurement time on satellite-derived lower tropospheric temperature, *Science*, 309, 1548–1551, doi:10.1126/science.1114772.
- Mears, C. A., and F. J. Wentz (2009a), Construction of the RSS V3.2 lower tropospheric dataset from the MSU and AMSU microwave sounders, *J. Atmos. Oceanic Technol.*, 26, 1493–1509, doi:10.1175/2009JTECHA1237.1.
- Mears, C. A., and F. J. Wentz (2009b), Construction of the Remote Sensing Systems V3.2 atmospheric temperature records from the MSU and AMSU microwave sounders, *J. Atmos. Oceanic Technol.*, 26, 1040–1056, doi:10.1175/2008JTECHA1176.1.
- Mears, C. A., F. J. Wentz, P. W. Thorne, and D. J. Bernie (2011), Assessing uncertainty in estimates of atmospheric temperature changes from MSU and AMSU using a Monte-Carlo estimation technique, *J. Geophys. Res.*, 116, D08112, doi:10.1029/2010JD014954.
- Meissner, T., and F. J. Wentz (2012), The emissivity of the ocean surface between 6–90 GHz over a large range of wind speeds and earth incidence angles, *IEEE Trans. Geosci. Remote Sens.*, 50(8), 3004–3026, doi:10.1109/TGRS.2011.2179662.
- Po-Chedley, S., and Q. Fu (2012), A bias in the midtropospheric channel warm target factor on the NOAA-9 microwave sounding unit, *J. Atmos. Oceanic Technol.*, 29, 646–652, doi:10.1175/JTECH-D-11-00147.1.
- Prigent, C., J. P. Wigneron, W. B. Rossow, and J. R. Pardo-Carrion (2000), Frequency and angular variations of land surface microwave emissivities: Can we estimate SSM/T and AMSU emissivities from SSM/I emissivities?, *IEEE Trans. Geosci. Remote Sens.*, 38(5), 2373–2386, doi:10.1109/36.868893.
- Randall, R. M., and B. M. Herman (2008), Using limited time period trends as a means to determine attribution of discrepancies in microwave sounding unit-derived tropospheric temperature time series, *J. Geophys. Res.*, 113, D05105, doi:10.1029/2007JD008864.
- Randel, W. J., and F. Wu (2006), Biases in stratospheric and tropospheric temperature trends derived from historical radiosonde data, *J. Clim.*, 19(10), 2094–2104, doi:10.1175/JCLI3717.1.
- Rosenkranz, P. W. (1975), Shape of the 5 mm oxygen band in the atmosphere, *IEEE Trans. Antennas Propag.*, 23(4), 498–506, doi:10.1109/TAP.1975.1141119.
- Rosenkranz, P. W. (1993), Absorption of microwaves by atmospheric gases, in *Atmospheric Remote Sensing by Microwave Radiometry*, edited by M. A. Janssen, pp. 37–90, John Wiley, New York.
- Seidel, D. J., et al. (2004), Uncertainty in signals of large-scale climate variations in radiosonde and satellite upper-air datasets, *J. Clim.*, 17, 2225–2240, doi:10.1175/1520-0442(2004)017<2225:UISOLC>2.0.CO;2.
- Seidel, D. J., et al. (2009), Reference upper-air observations for climate: Rationale, progress, and plans, *Bull. Am. Meteorol. Soc.*, 90, 361–369.
- Sherwood, S. C., J. R. Lanzante, and C. L. Meyer (2005), Radiosonde daytime biases and late 20th century warming, *Science*, 309, 1556–1559, doi:10.1126/science.1115640.
- Sherwood, S. C., C. L. Meyer, R. J. Allen, and H. A. Titcher (2008), Robust tropospheric warming revealed by iteratively homogenized radiosonde data, *J. Clim.*, 21(20), 5336–5352, doi:10.1175/2008JCLI2320.1.
- Smith, W. L., H. M. Woolf, C. M. Hayden, D. Q. Wark, and L. M. McMillan (1979), The TIROS-N operational vertical sounder, *Bull. Am. Meteorol. Soc.*, 60, 1177–1187.
- Spencer, R. W., and J. R. Christy (1992), Precision and radiosonde validation of satellite gridpoint temperature anomalies. Part II: A tropospheric retrieval and trends during 1979–1990, *J. Clim.*, 5, 858–866, doi:10.1175/1520-0442(1992)005<0858:PARVOS>2.0.CO;2.
- Thorne, P. W., D. E. Parker, J. R. Christy, and C. A. Mears (2005a), Uncertainties in climate trends: Lessons from upper-air temperature records, *Bull. Am. Meteorol. Soc.*, 86(10), 1437–1442, doi:10.1175/BAMS-86-10-1437.
- Thorne, P. W., D. E. Parker, S. F. B. Tett, P. D. Jones, M. McCarthy, H. Coleman, and P. Brohan (2005b), Revisiting radiosonde upper air temperatures from 1958 to 2002, *J. Geophys. Res.*, 110, D18105, doi:10.1029/2004JD005753.
- 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.
- Thorne, P. W., J. R. Lanzante, T. C. Peterson, D. J. Siedel, and K. P. Shine (2011a), Tropospheric temperature trends: History of an ongoing controversy, *Wiley Interdiscip. Rev. Clim. Change*, 2(1), 66–88, doi:10.1002/wcc.80.
- Thorne, P. W., et al. (2011b), A quantification of uncertainties in historical tropical tropospheric temperature trends from radiosondes, *J. Geophys. Res.*, 116, D12116, doi:10.1029/2010JD015487.
- Titchner, H. A., P. W. Thorne, M. P. McCarthy, S. F. B. Tett, L. Haimberger, and D. E. Parker (2009), Critically reassessing tropospheric temperature trends from radiosondes using realistic validation experiments, *J. Clim.*, 22(3), 465–485, doi:10.1175/2008JCLI2419.1.
- Zou, C.-Z., and W. Wang (2011), Intersatellite calibration of AMSU-A observations for weather and climate applications, *J. Geophys. Res.*, 116, D23113, doi:10.1029/2011JD016205.
- Zou, C. Z., M. D. Goldberg, Z. Cheng, N. C. Grody, J. T. Sullivan, C. Cao, and J. D. Tarpley (2006), Recalibration of microwave sounding unit for climate studies using simultaneous nadir overpasses, *J. Geophys. Res.*, 111, D19114, doi:10.1029/2005JD006798.
- Zou, C. Z., M. Gao, and M. D. Goldberg (2009), Error structure and atmospheric temperature trends in observations from the Microwave Sounding Unit, *J. Clim.*, 22, 1661–1681, doi:10.1175/2008JCLI2233.1.