Assessing uncertainty in estimates of atmospheric temperature changes from MSU and AMSU using a Monte-Carlo estimation technique

Carl A. Mears,[1] Frank J. Wentz,[1] Peter Thorne,[2,3] and Dan Bernie[2]

Received 25 August 2010; revised 7 February 2011; accepted 9 February 2011; published 20 April 2011.

[1] Measurements made by the Microwave Sounding Unit (MSU) and the Advanced Microwave Sounding Unit (AMSU) provide a multidecadal record of global atmospheric temperature change, which have been used by several groups to produce long-term temperature records of thick layers of the atmosphere from the lower troposphere to the lower stratosphere. Here we present an internal uncertainty estimate for the Remote Sensing Systems data sets made using a Monte Carlo approach that includes contributions to the total uncertainty from sampling error, premerge adjustments to each individual satellite, and the merging procedure. The results can be used to estimate uncertainties in this product at all space and time scales of interest to any specific application. On small space and time scales sampling effects dominate. On the longer time scales intersatellite merging is important at all levels and the diurnal adjustment is a critical uncertainty for the two layers that have a significant surface component, particularly over land. A comparison of trends for the globe, tropics, and extratropics between the best estimate data set along with these error estimates and homogenized radiosonde estimates and available MSU/AMSU estimates from other groups is undertaken. This shows consistency between our product and those produced by others within the stated uncertainty for many regions and layers. In almost as many cases, however, the interdata set differences of the estimated trends are too large be accounted for by the internal uncertainty estimates derived herein.

Citation: Mears, C. A., F. J. Wentz, P. Thorne, and D. 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.

1. Introduction

[2] Temperature sounding microwave radiometers flown on polar-orbiting weather satellites provide an important record of atmospheric temperatures throughout the troposphere and into the lower stratosphere beginning with the Microwave Sounding Unit (MSU) on *TIROS-N* in late 1978. Subsequently a series of eight additional MSU instruments provided a continuous record up to the present, with the MSU on *NOAA-14* still in operation. Beginning in 1998, the first of a similar but, crucially from a climate perspective, nonidentical, follow-on series of instruments, the Advanced Microwave Sounding Units (AMSUs) was launched. In order to provide continuous records of atmospheric temperatures, data from the AMSU instruments has been merged with data for the previous MSU series of instruments by several groups to create candidate [*Christy et al.*, 2003; *Grody et al.*, 2004a; *Mears and Wentz*, 2009a, 2009b; *Prabhakara et al.*, 2000; *Zou et al.*, 2006] climate data records. These have been used in many studies and major scientific assessments over the past 2 decades [e.g., *Karl et al.*, 2006; *Solomon et al.*, 2007]. The state of our knowledge has evolved substantially over this time from a ill-founded position of certainty when there was only one data set available to one that is now far more nuanced [*Thorne et al.*, 2010].

[3] There has been substantial previous discussion of uncertainty in merged microwave sounder data sets [*Christy et al.*, 2000, 2003; *Fu and Johanson*, 2005; *Grody et al.*, 2004b; *Mears et al.*, 2003; *Mears and Wentz*, 2005; *Zou et al.*, 2006]. These studies have primarily focused on uncertainty in trends in global scale averages, though some work has been done on gridbox scale data [*Christy and Norris*, 2006, 2009]. In this work, we extend these earlier estimates by describing uncertainty on a number of spatial and temporal scales and more fully accounting for the uncertainty introduced by sampling error and the diurnal adjustment procedures. Since much of the focus of current research is on changes in temperature associated with global climate change, we focus our attention on uncertainty in temperature anomalies, rather than on bias in the absolute temperature measurements.

[1]Remote Sensing Systems, Santa Rosa, California, USA.
[2]U.K. Met Office, Exeter, U. K.
[3]Now at Cooperative Institute for Climate and Satellites-North Carolina, NOAA NCDC, Asheville, North Carolina, USA.

0148-0227/11/2010JD014954

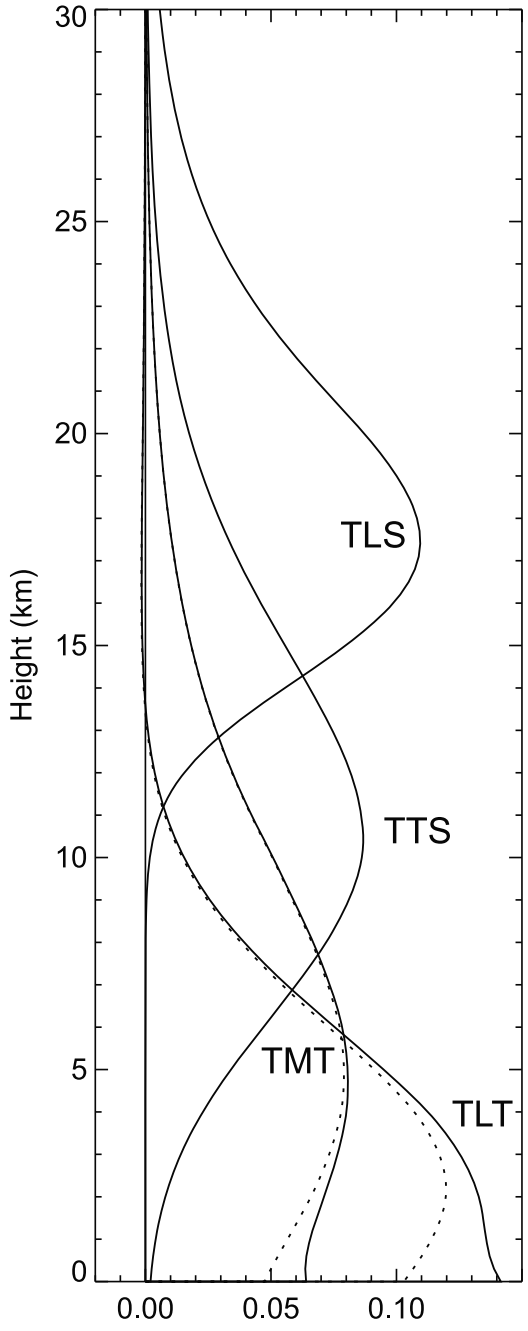

Figure 1. Temperature weighting functions as a function of height for each channel. For TMT and TLT, the weighting functions for land scenes are shown by dotted lines and for ocean scene is shown by solid lines. For TTS and TLS, the land and ocean weighting functions are nearly identical. For TLT and TMT, there is also a significant contribution from emission by the material surface. The surface weight for TLT land, ocean measurements is 0.193, 0.106, and for TMT land, ocean measurements the surface weight is 0.083, 0.043.

[4] In two previous papers [*Mears and Wentz*, 2009a, 2009b] we described the revised methods we have used to merge data from the MSU and AMSU satellites together to form the latest version of our TLT (temperature lower troposphere), TMT (temperature middle troposphere), TTS (temperature troposphere stratosphere), and TLS (temperature lower stratosphere) data sets. The temperature weighting function for each channel is plotted in Figure 1, and a flowchart showing the most important parts of the merging procedure is shown in Figure 2. In this paper, we estimate the uncertainty in these data sets by combining the contributions from various sources of error that result from a finite sample and very many unknowns. This uncertainty is the internal uncertainty, that arising given the explicit methodological choices we have made to create the data set (Figure 2). These estimates apply to the RSS Version 3.2 MSU/AMSU data sets [*Mears and Wentz*, 2009a, 2009b], which are available online at www.remss.com. These data sets are monthly averages of temperature, gridded on a 2.5° by 2.5° grid. These uncertainty estimates are important both when performing comparisons of atmospheric temperature estimates from different measurement sources and when undertaking scientific studies using these data.

[5] Another source of uncertainty is the structural uncertainty [*Thorne et al.*, 2005] imparted unintentionally by our methodological choices. This effect can only be robustly ascertained by comparison to estimates derived from a sufficiently large number of other independent groups who have made reasonable choices. We undertake such a comparison along with recourse to a similar suite of data sets derived from radiosonde records to inform on the relative importance of these two distinct sources of uncertainty to the MSU/AMSU data sets. However, it must be recognized that this sample size is small and that it is not guaranteed that all data set creators (including by definition ourselves) have made reasonable methodological choices at each step or accounted for all nonclimatic factors. So, such a comparison comes with substantial caveats attached as it cannot be guaranteed a priori that all of the sample in such an intercomparison is "plausible."

[6] It is important to note that the trend errors reported for MSU/AMSU data in the major assessment reports [*Lanzante et al.*, 2006; *Solomon et al.*, 2007] are typically generated using information about how well the reported time series fits a linear trend. This type of error provides information about how much the trend might differ if the Earth had undergone a different set of short time scale variations, such as those caused by ENSO. They do not provide information about the uncertainty internal to the data set due to measurement and construction error or the structural uncertainty that would be imparted by applying a different set of reasonable processing choices. While all these types of error can be important, depending upon the type of analysis being undertaken, our analysis here is restricted to determining the internal uncertainty in our data sets

[7] In section 2 we describe and estimate the magnitude of the various sources of internal uncertainty, and in section 3 we compare the results from our data set to results from other approaches used to estimate changes in atmospheric temperature.

2. Estimation of Internal Uncertainty

2.1. Approach

[8] There are a number of sources of uncertainty in our merged MSU/ASMU temperature data sets. The most

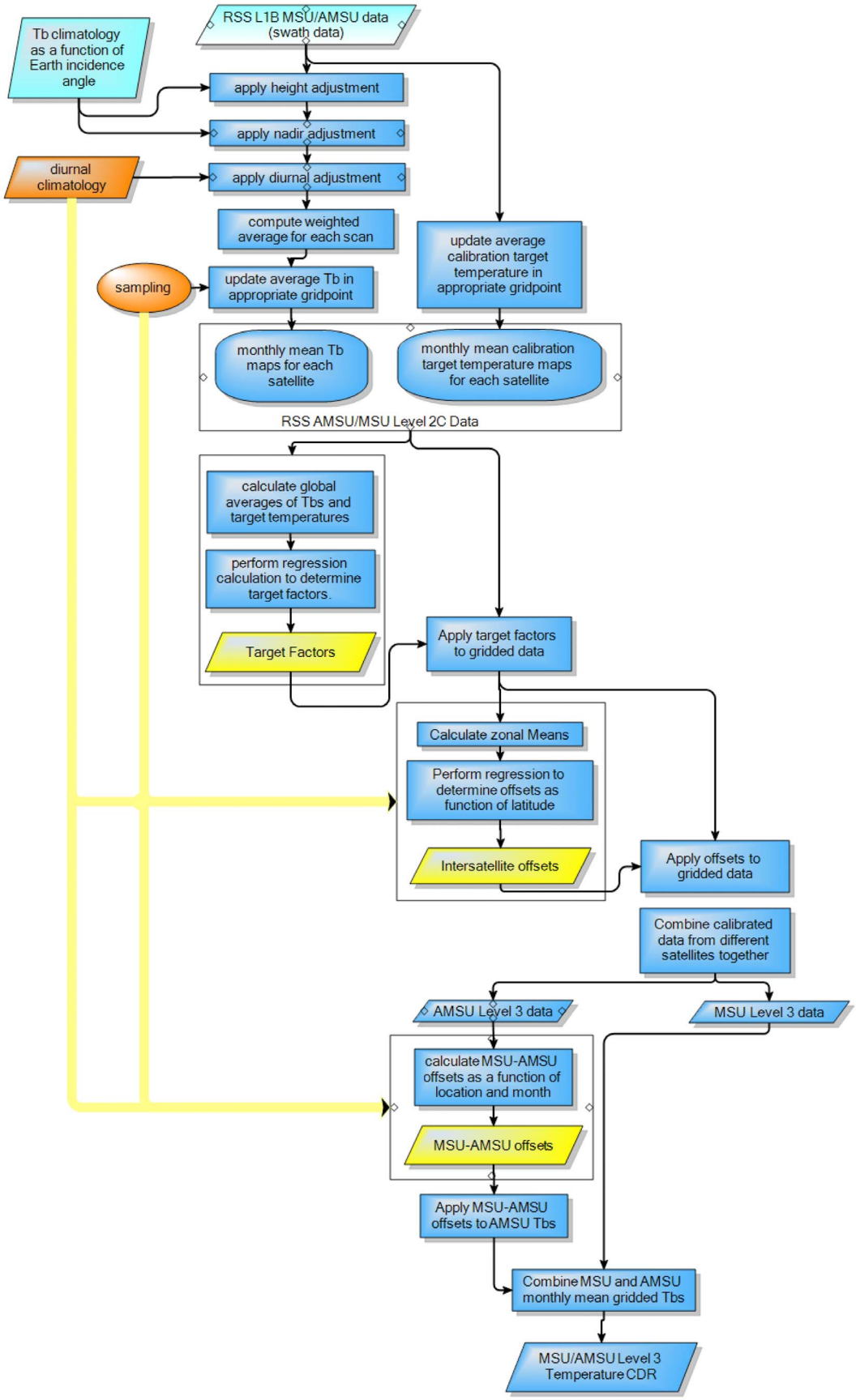
RSS L1B MSU/AMSU data (swath data)
Tb climatology as a function of Earth incidence angle
apply height adjustment
apply nadir adjustment
diurnal climatology
apply diurnal adjustment
compute weighted average for each scan
update average calibration target temperature in appropriate gridpoint
sampling
update average Tb in appropriate gridpoint
monthly mean Tb maps for each satellite
monthly mean calibration target temperature maps for each satellite
RSS AMSU/MSU Level 2C Data
calculate global averages of Tbs and target temperatures
perform regression calculation to determine target factors.
Target Factors
Apply target factors to gridded data
Calculate zonal Means
Perform regression to determine offsets as function of latitude
Intersatellite offsets
Apply offsets to gridded data
Combine calibrated data from different satellites together
AMSU Level 3 data
MSU Level 3 data
calculate MSU-AMSU offsets as a function of location and month
MSU-AMSU offsets
Apply MSU-AMSU offsets to AMSU Tbs
Combine MSU and AMSU monthly mean gridded Tbs
MSU/AMSU Level 3 Temperature CDR

Figure 2

important fundamental uncertainty sources are spatial-temporal sampling errors, and the uncertainty in the diurnal adjustment used to account for the effects of drifting local measurement time. These uncertainties in turn lead to uncertainty in the merging parameters because these merging parameters are deduced from the actual measurements (which include sampling uncertainty) after the diurnal adjustments are performed. The remaining methodological steps were found in an initial analysis to exhibit very minimal error and so are discounted as important components of the overall error budget.

[9] It is difficult to determine the exact structure of the uncertainty in the absence of any reference data set that is known to be free from error. Sadly, such a data set has not existed, although efforts to create precisely calibrated ground-based [*Seidel et al.*, 2009] and space-based [*Space Studies Board*, 2007] observing systems could help resolve this in the future. We therefore provide an informed estimate of the uncertainty internal to our data set, often based on an analysis of intersatellite differences and our knowledge of the merging procedures we used to produce the data set. In the following sections, we outline the overall approach and the contribution of each source of error to the total error.

[10] Different uncertainty sources are important for different spatial and temporal scales. For example, radiometer noise is important only for short time scales and small spatial scales because its effects are rapidly diminished by averaging multiple observations together and thus it is not considered further in this analysis. Sampling uncertainty tends to dominate other uncertainty sources for a single monthly average over a 2.5° times 2.5° cell (the smallest temporal and spatial resolution we consider) but decreases rapidly when larger temporal or spatial averages are considered. Diurnal and merging parameter uncertainties are spatially and temporally correlated and thus do not decrease rapidly with averaging and thus become dominant for largest spatial and longest temporal scales.

[11] Because of the complex interplay between various parts of the adjustment system (Figure 2) we cannot a priori determine the degree to which errors in various parts of the system are interdependent. Therefore a stochastic model is required to be employed that implicitly allows any such interdependencies to be expressed rather than making assumptions which would be hard to justify and could lead to either too liberal or too conservative a set of estimates. We therefore use Monte-Carlo methods to produce a large number of instances of estimated error in the final data set.

[12] These instances are constructed so that they have spatial and temporal correlations that are consistent with those we expect to be present in the final data set and thus can be interrogated to produce estimates of the estimated error on a variety of spatial or temporal scales. To generate each instance, we start with a gridded monthly data set for each satellite that is set to zero for each month if this given satellite has valid data and is set to missing otherwise. We then add to this data set estimated realizations of the sampling and diurnal adjustment uncertainty. Both these estimates are constructed so that their ensemble averages have zero mean. (Our methods for the construction of these uncertainties are described in the following sections.)

[13] The data sets with uncertainty added, which we refer to as "noise datasets," are then analyzed using merging procedures that are identical to those used for the real data. Each noise data set results in a set of noisy merging parameters (intersatellite offsets and target factors). Since the noise-free data set is constructed with all zeros, we know that these merging parameters should be zero, and thus any differences from zero are due to the influence the underlying uncertainties, i.e., sampling and diurnal adjustment. Thus our approach describes the total effect of the underlying uncertainties upon the final merged product, including both their direct effect and their indirect effect via uncertainty in the merging parameters. In Figure 2, we show both the direct and indirect paths of the uncertainties through the merging procedure.

[14] In undertaking this analysis approach we are making an implicit assumption that the expression of the errors will be independent of the underlying spatiotemporal evolution of the climate system state (although this is partially captured arguably in the sampling error derivation, see section 2.2). That is to say that the overall magnitude of the internal uncertainty errors is independent of the timing and magnitude of natural modes of variability such as ENSO and the presence or otherwise of an underlying long-term climate change signal. This would include any long-term evolution of the magnitude of the diurnal cycle. Intuitively, any such impact would be very much a second-order effect and this assumption is likely to be valid.

[15] A further, more tenuous assumption made when adding this Monte Carlo simulation to the operational data set products is that these products themselves are essentially unbiased. Several studies have been published that suggest that our TLT and TMT data sets may contain residual biases [e.g., *Christy et al.*, 2007; *Christy and Norris*, 2009; *Randall and Herman*, 2008]. These biases could either be due to the type of errors discussed here or to other unknown error sources that are not addressed in this analysis. The internal error analysis produced here cannot, by construction, provide any information on this aspect of any error in our products. Instead, we view this error analysis as a prerequisite to performing detailed comparisons with complementary data sets, since it is critical to be able to assess the statistical significance of any discrepancies found. All of the radiosonde comparisons to date neglect such uncertainty in both the radiosonde measurements and in the satellite data and thus the true quantitative significance of their results is impossible to determine.

Figure 2. A flowchart that shows the most important aspects of the merging algorithm we use to generate long-term records from MSU and AMSU observations. For more details on the algorithm, see *Mears et al.* [2009a, 2009b]. Inputs to the algorithm that are held constant in our uncertainty analysis are shown in light blue, while inputs that are explicitly allowed to vary are shown in orange. The uncertainty caused by these variation can be thought of as flowing in two paths, directly along the black arrows, and indirectly via the influence on the merging parameters deduced from the data (shown as yellow parallelograms) along the yellow arrows.

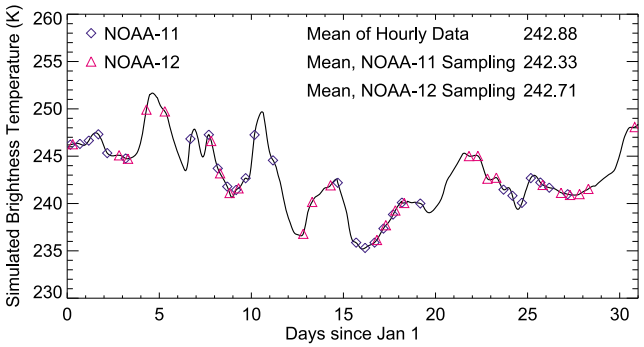

Figure 3. The effect of sampling on the mean of a simulated brightness temperature time series. The black line is the simulated TMT brightness temperature for January 1980 from the CCM3 climate model, for a point in the North Pacific Ocean (40°N, 170°E). The symbols represent the times the hourly time series would be sampled by the NOAA-11 and NOAA-12 satellites during a typical month. By chance, the means found using the sampled data are both less than the true mean of the hourly data; other sampling patterns could result in the opposite bias.

2.2. Sampling Errors

[16] Satellites make measurements at discrete times. Because the satellites that MSU/AMSU are carried upon are polar orbiters that orbit pole-to-pole every 90 min there is a distinct geographical sampling imparted upon the data. At the very highest latitudes sampled (around 85° latitude) each location is sampled multiple times per day. In the deep tropics the sampling at a given location can be as infrequently as every third day. The sampling grades between these two extremes at other latitudes. When estimating sampling error this effect must be combined with the synoptic variability so it need not follow that the magnitude of this effect would be directly proportional to the sampling rate or that it would be seasonally invariant.

[17] When forming a monthly average, all measurements made during a given month that fall within a given 2.5° by 2.5° cell were averaged together. Sampling errors occur when this average of discrete measurements does not accurately represent the true monthly average. Figure 3 shows the simulated hourly TMT brightness temperature (which uses the central five fields of view) from the CCM3 atmospheric model [*Kiehl et al.*, 1996] for a point in the North Pacific (50°N, 170°W). The symbols on the line represent the times at which the temperature would be sampled by overflights of two coorbiting satellites. The monthly means obtained by averaging the temperature at these sampling times can be quite different than the true monthly mean obtained by averaging all the hourly temperatures. Note that the mean diurnal cycle is very small for this location (as it is over the vast majority of the world's oceans [*Kennedy et al.*, 2007]) and thus is not an important source of error at this location.

[18] The lower tropospheric data set (TLT) is formed by calculating a weighted difference of different fields of view (FOVs) [*Spencer and Christy*, 1992]. For the left side of the MSU swath, this difference is given by

$$T_{2LT-MSU} = 2.0(T_3 + T_4) - 1.5(T_1 + T_2), \quad (1)$$

where T_n is the temperature measured by the nth FOV, with T_1 denoting the left-hand, near-limb view. Figure 4 shows the spatial pattern of measurement footprints and weights for example scans on a map of Western Europe and the North Atlantic. A more complex, but similar differencing procedure is used for AMSU [*Mears and Wentz*, 2009b]. On average, this differencing procedure has the effect of extrapolating the measurement lower in the troposphere. However, since each individual view makes a measurement at a different place (see Figure 4), an undesirable spatial

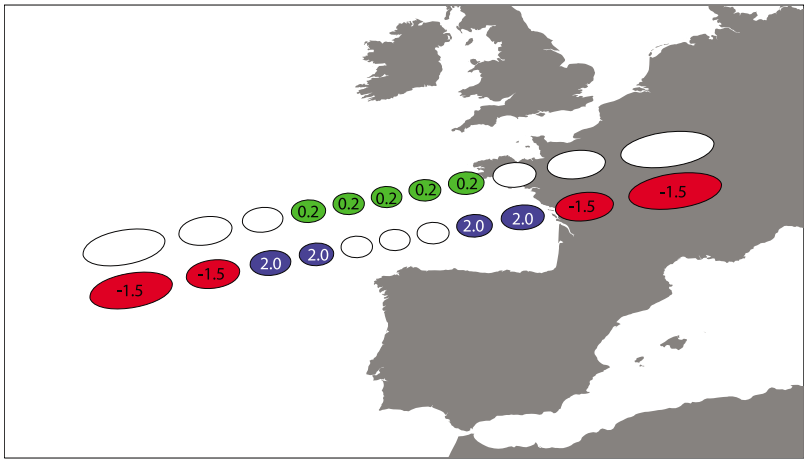

Figure 4. The spatial pattern of measurement footprints for two example scans. The top scan shows the weighting for the nonextrapolated products (TMT, TTS, and TLS). For the products, the average of the five green FOVs is computed and assigned to each of the footprint locations that contribute to the average. The bottom scan shows the weights for the TLT lower tropospheric extrapolation, where we compute the weighted average separately for each set of four FOVs on each side of the scan separately.

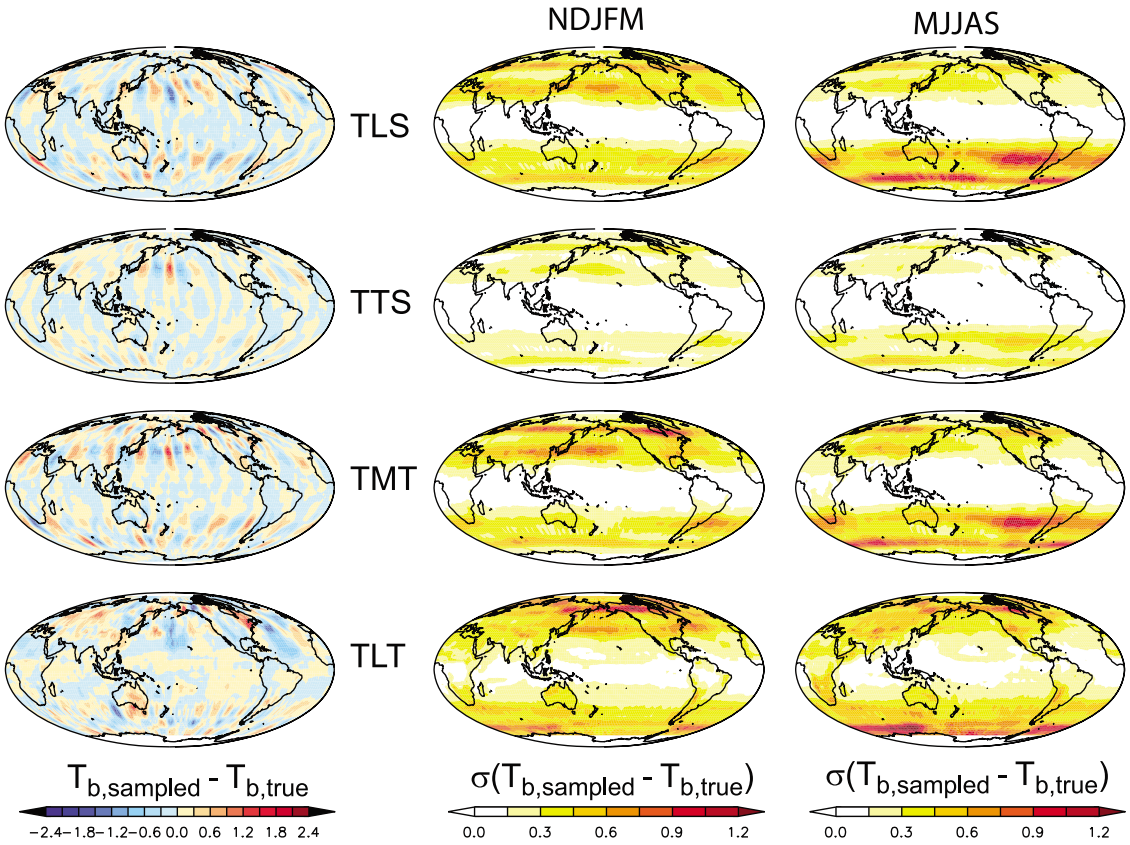

Figure 5. Maps of the estimated temperature sampling error for each channel. The left column is an example of the sampling error for the month of January. The middle column is the standard deviation of the sampling error averaged over all Northern Hemisphere winter months (November, December, January, February, and March). The right column is the standard deviation averaged over the Northern Hemisphere summer months (May, June, July, August, and September).

derivative is also included in the extrapolated value. Here we consider the effects of this derivative to be part of the sampling error.

[19] To estimate this error, we used the hourly CCM3 results to calculate the simulated MSU channel two brightness temperature at each FOV for a simulated satellite orbit and then used the weighted difference of values to deduce a satellite-equivalent TLT that includes the spatial derivative effect. This is then compared to the true average (without spatial derivative or temporal sampling effects) of the modeled TLT-equivalent brightness temperature at the location in question. We find that the spatial derivative effect increases the sampling noise for the TLT data set. The amount of the increase ranges from about a factor of 1.5 in the tropics, to as high as a factor of 4 in midlatitude storm tracks (where the extrapolation is likely to straddle a midlatitude cyclone) and near coastlines (where the extrapolation is likely to cross the land-ocean boundary). As noted by *Mears and Wentz* [2009b], the extrapolation procedure also results in a location-dependent bias. This bias can be quite large near the poles due to a net north-south spatial derivative in each monthly average and near coastlines. Since the focus of this work is errors in temperature trends and anomalies, we do not evaluate this bias in detail here.

[20] We formulated a model-based estimate of sampling noise for each channel and for each satellite type (MSU and AMSU) by calculating the differences between the monthly sampled mean and the true monthly mean for three different sampling patterns (i.e., the locations and times of each measurement during the month. These are derived from sampling patterns in the observed data.), for 5 years of hourly model output from the CCM3 model which was also used to derive the operational diurnal cycle adjustments (see section 2.3). Thus for any given month of the year, there are 15 different possible realizations of the sampling error. In Figure 5, we show an example of the simulated sampling error for each MSU channel for the month of January and the standard deviation of the MSU sampling error averaged over the winter and summer seasons separately. Sampling errors are largest in the midlatitudes, where the effects of large day-to-day variability and gaps in temporal sampling are combined together. The gaps in temporal sampling are even larger in the tropics, but there is usually much less day-to-day variability so that the magnitude of the sampling error is less. Sampling errors in the tropospheric channels tend to be largest in the winter hemisphere, due to the presence of more intense midlatitude cyclones that yield much greater synoptic scale variability

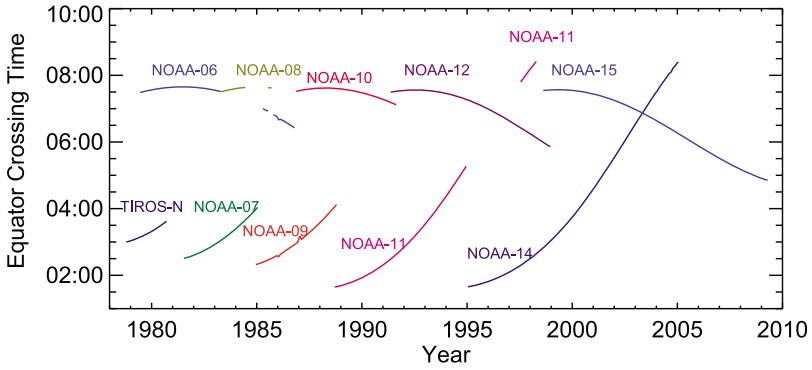

Figure 6. Descending node local equator crossing time for each satellite carrying a MSU or AMSU instrument used in our study.

than is evident in the summer season. Figure 5 also shows that the errors at a given grid point tend to be strongly correlated with those in other nearby grid points due to similar temporal sampling patterns combined with the spatial correlation in the synoptic-scale variability.

2.3. Diurnal Adjustment Errors

[21] During the lifetime of most of the MSU and AMSU instruments, the orbit of each instrument's satellite platform slowly changed as a function of time, leading to drifts in local equator crossing time (see Figure 6). These drifts cause the diurnal cycle to be aliased into the long-term records unless their effects are characterized and removed. As discussed by *Mears and Wentz* [2009a, 2009b], we used model-based diurnal cycle to make adjustments for drifting local measurement times. We estimate the uncertainty in this diurnal adjustment by evaluating the diurnal adjustments derived from different climate models. These are used to generate plausible realizations of the diurnal adjustment error D, given by

$$D = \sum_{i=1}^{N} a_i (D_i - \overline{D}). \tag{2}$$

Here, D_i is the adjustment calculated using the ith model, $\overline{D}$ is the mean adjustment, and the a_i' s are normally distributed random numbers with zero mean and unit variance.

[22] For the data products derived from MSU channel 2 and AMSU channel 5, two models are available for calculating the diurnal adjustment, the NCAR Community Climate Model-3 (CCM3) [*Kiehl et al.*, 1996] and the Hadley Centre Global Environmental Model (HadGEM1) [*Martin et al.*, 2006]. In our previous work, the CCM3 model was used to formulate our diurnal correction. For the TTS and TLS data products, which have temperature weighting functions centered higher in the atmosphere, we also considered a diurnal adjustment derived from the Canadian Middle Atmosphere Model (CMAM) [*Beagley et al.*, 2000]. Our version of the data from this climate model lacks a surface temperature, making it inappropriate for TLT and TMT, which have a substantial contribution due to surface emission. In Table 1, we show the global and tropical trends that results when each off these diurnal adjustments are applied to the actual satellite data.

[23] The limited number of climate models available is far from ideal. We note that the very small number of models makes tenuous our implicit assumption that the differences between the specific models we have available span the range of possible error. It would be highly desirable to gain further estimates from other modeling centers and from reanalyses of their diurnal cycle estimates so that this plausible solution space can be more adequately sampled. Despite this limitation, we proceed with our analysis, since it nevertheless provides a significant step forward from our earlier attempts to characterize the uncertainty in our diurnal adjustments, where we simply assumed a +/− 50% error independent of location [*Mears and Wentz*, 2005]. In Figure 7, we show a typical realization of the diurnal adjustment and error in the diurnal adjustment for each channel. In Figure 8 we show global averages of the diurnal adjustments applied to the NOAA-14 instrument. On a global scale, the HadGEM1 model tends to results in larger diurnal adjustments and a larger seasonal cycle in these adjustments, while the CMAM model tends to result in smaller adjustments than our operational product.

[24] Note that there are large spatial correlations in the diurnal adjustment errors (see the right side of Figure 7) particularly for the TLT and TMT channels. For these channels, the HadGEM1 diurnal cycle is significantly larger in arid land regions than the CCM3 diurnal cycle. We do not definitively know the reason for this difference, though we suspect it is influenced by differences between the land surface parameterizations used by the two models.

Table 1. Long-Term Trends Using Different Diurnal Models[a]

Channel	Diurnal Model	Global (80°S to 80°N)	Tropics (20°S to20°N)
TLT (70°S to 80°N)	CCM3	0.150	0.147
	HADGEM	0.184	0.169
TMT	CCM3	0.088	0.110
	HADGEM	0.108	0.121
TTS (1987–2008)	CCM3	−0.022	−0.013
	HADGEM	−0.011	−0.002
	CMAM	−0.018	−0.008
TLS	CCM3	−0.324	−0.310
	HADGEM	−0.313	−0.299
	CMAM	−0.335	−0.323

[a]Long-term trends are from 1979 to 2009.

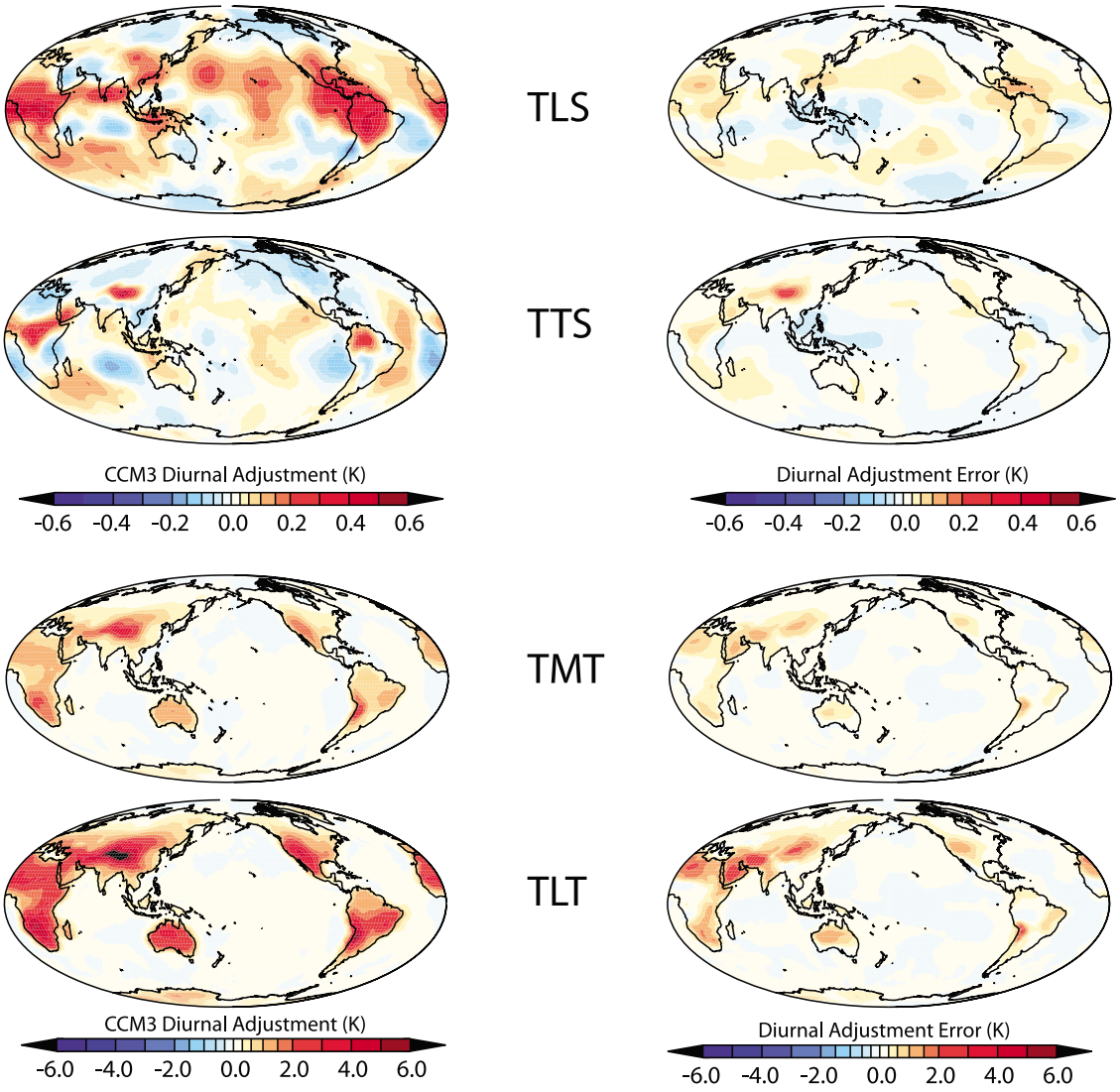

Figure 7. Maps of the diurnal adjustment applied to each channel. The left column is an example of the diurnal adjustment applied to a single monthly average for each data set, calculated using the CCM3-derived diurnal cycles. (The adjustments shown here are for the NOAA-14 satellite, September 2003). The right column is a deviation from this adjustment chosen from one of the 400 realizations of possible diurnal error. In general, the largest deviations occur in regions with the largest adjustments. Note the large change in scale between the tropospheric channels and TTS and TLS.

[25] These realizations of the diurnal uncertainty are added to the realizations of the sampling uncertainty we calculated in the previous sections, yielding a set of 400 "noise realizations" for each satellite in the series that are consistent in the estimated combined uncertainty from both sources. Note that for a given Monte Carlo noise realization, all satellites are adjusted using adjustments calculated from the same diurnal cycle error realization from equation (2) so as to not artificially inflate this estimate as the real world cannot logically exhibit multiple coincident diurnal cycles.

2.4. Uncertainty in Merging Parameters

[26] Earlier work found that global averages of simultaneous measurements made by co-orbiting MSU instruments differ by both a time-invariant intersatellite offset and an additional term that is strongly correlated with the variations in temperature of the hot calibration target (which is an integral component of the instrument and measurement technology) for each satellite [*Christy et al.*, 2000]. To describe these differences, we use an empirical error model for brightness temperature incorporating the target temperature and scene temperature correlation [*Mears and Wentz*, 2009a],

$$T_{MEAS,i} = T_0 + A_i + \alpha_i T_{TARGET,i} + \beta_i T_{SCENE} + \varepsilon_i \qquad (3)$$

where $T_{\mathrm{MEAS,i}}$ is the brightness temperature measured by the ith instrument, T_0 is the true brightness temperature, A_i is the temperature offset for the ith instrument, α_i is a small multiplicative "target factor" describing the correlation of the measured antenna temperature with the temperature

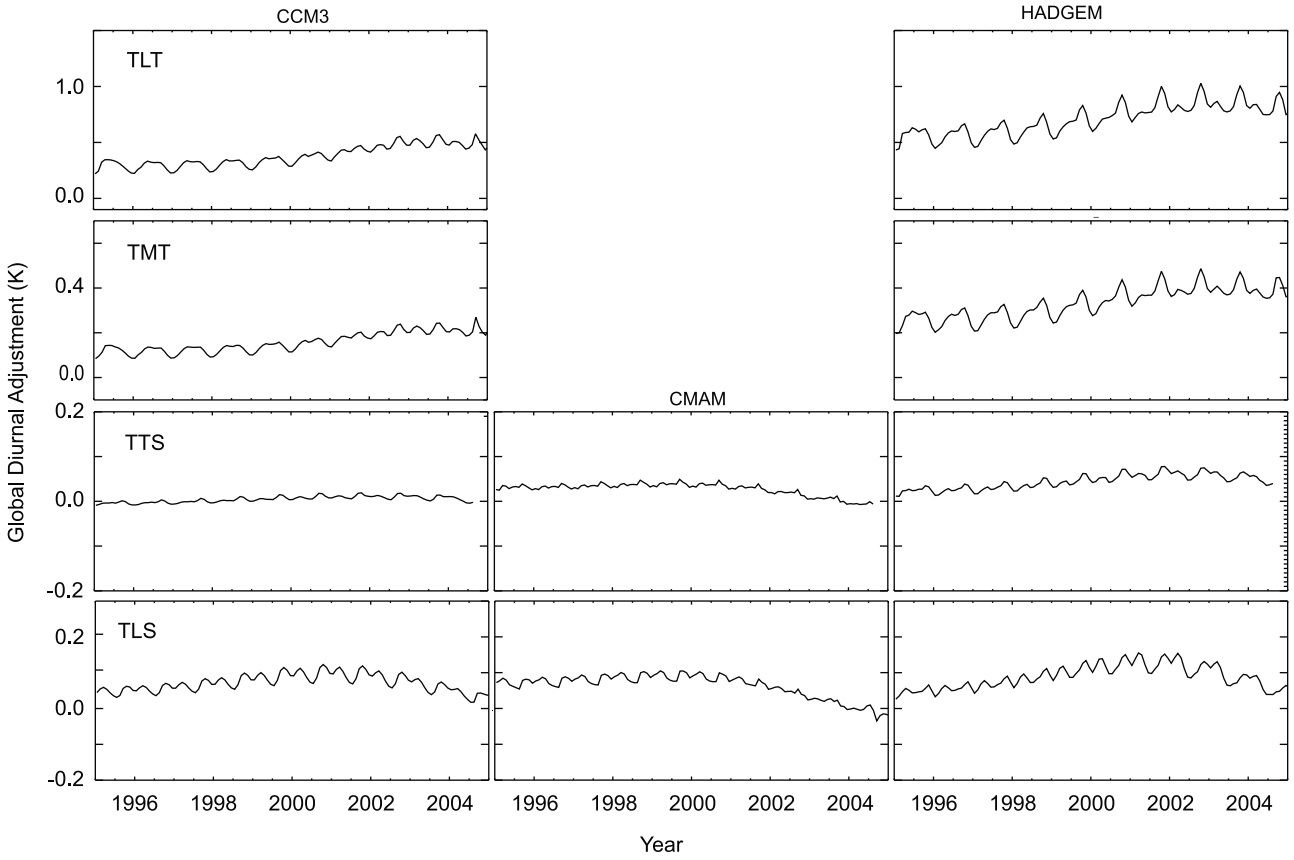

Figure 8. Global diurnal adjustments applied to the MSU instrument on NOAA-14. The rows correspond to the different channels, and the columns to the different models used. Over this period, the LECT for NOAA-14 drifted over 6 h (see Figure 3).

anomalies of the hot calibration target, $T_{TARGET,i}$. The parameter β_i describes the correlation of the calibration error with the scene temperature anomaly T_{SCENE}, and ε_i is an error term that contains additional uncorrelated, zero-mean errors due to instrumental noise and sampling effects. The merging parameters used for the published data sets were found using a regression procedure that minimizes intersatellite differences between monthly averages. Since these monthly averages contain errors from various sources (dominated by the sampling and diurnal sources mentioned above) there is uncertainty in the calculated merging parameters. This uncertainty leads to additional uncertainty in the final merged data set. Because the intersatellite offset is a single invariant delta estimate for each satellite pair any error in this step is equivalent to adding units of red noise to the series. This error will therefore increase rather than decrease with length of record. In our analysis we found that the effect of the uncertainty in the β_i's was negligible compared to other error sources and is ignored from this point forward to simplify the analysis.

[27] The α_i's (target factors) are the most important parameters for long-term behavior of the merged data set since they are difficult to determine and they multiply the target temperatures, which often show large long-term changes. Figure 9 shows the standard deviation of the fitted α_i's for each satellite and channel calculated using the 400 noisy realizations from the intrasatellite error determination above. For TLT and TMT, we find that the largest uncertainty is in the value of $\alpha_{NOAA\text{-}09}$, consistent with our earlier findings [*Mears et al.*, 2003]. For TLS, the uncertainty in $\alpha_{NOAA\text{-}09}$ is less because the period during which MSU channel 4 was functioning for both NOAA-09 and NOAA-10 was longer. For TTS, satellites before NOAA-10 are not used because both NOAA-09 and NOAA-06 have large, unexplained drifts in the data for MSU channel 3.

[28] Once the α_i's are determined, we then find the latitude dependent offsets using a regression procedure for each 2.5°-wide latitude band. These offsets also vary due to the sampling noise, diurnal adjustment error, and target factors error determined in previous steps. As the diurnal adjustment, and thus any error in the diurnal adjustment, is typically much larger over land, the errors in these offsets tend to be dominated by errors in the land diurnal cycle. We note that since we use a single offset for each latitude band for both land and ocean regions, any land-caused error will affect the merged data set in ocean regions at the same latitude.

2.5. Merging MSU and AMSU Together

[29] The MSU and AMSU measurement bands for the corresponding channels for the two instruments differ, leading to small differences in vertical weighting, which in turn

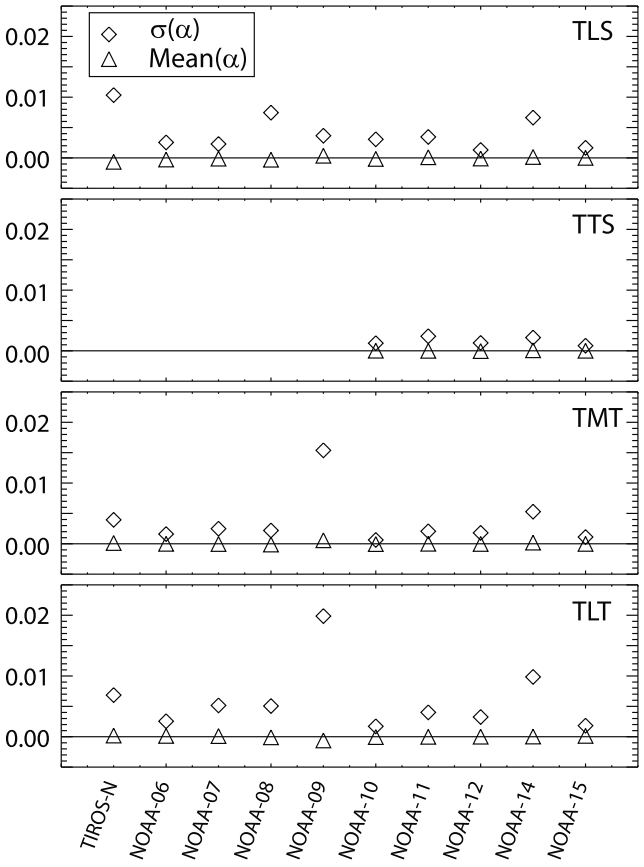

Figure 9. Mean and standard deviations of the target factors calculated for each error realization. For the synthetic data we analyze here, we expect the target factors to be zero, so nonzero value we calculate is due to the effect of sampling and diurnal adjustment error in the target factors. The difference between the mean values and zero is due to the finite number of noise data sets we evaluate. For an infinite sample, the target factors would average to zero.

lead to small differences in brightness temperature that depend on location and time of year. We remove these differences empirically by using location and time-of-year dependent offsets to adjust the AMSU measurements so that they match the MSU measurements during the 1999–2004 period. These offsets also contain error, which is modeled by our Monte-Carlo process.

[30] This enables an instantaneous estimate of the effect given the instantaneous mean atmospheric state at the time of changeover. Given that there is ample evidence that the mean climate state is nonstationary and that changes in the mean climate state are not the same for different atmospheric layers there is a further insidious aspect that will come into play over time for our data set and for all other MSU/AMSU data sets which we have not explicitly modeled here but warrants further attention. If, as is widely predicted, the climate system continues to warm then there will be a mismatch that grows over time between the change the AMSU instruments see and that which would have been seen had MSU frequencies been maintained. Taking HadAT data [*Thorne et al.*, 2005] and calculating MSU TMT and the AMSU equivalent shows a trend mismatch over 1958 to date of approximately 0.003 K/decade, roughly 4% of the long-term signal. Particularly for assessing differences between adjacent layers or the surface and the troposphere effects of this magnitude may be important. Looking forward, the transition to the Advanced Technology Microwave Sounder (ATMS) sensors will introduce a further, but smaller, such impact onto the long-term record because the measurement bands are well-matched to the AMSU measurement bands.

2.6. MSU/AMSU Drift for TMT

[31] Examination of the differences between TMT from MSU channel 2 on NOAA-14 and AMSU channel 5 on NOAA-15 shows a long-term trend difference, with NOAA-15 cooling at a rate of 0.2 K per decade relative to NOAA-14 over the July 1998 to December 2004 period of overlap. This trend difference is not present for the other channel pairs, including, to our surprise, TLT, and is more than 2.5 times larger than the trend difference for any pair of MSU satellites with more than 18 months of overlapping observations. This trend difference is too large to explain using the difference between the MSU and AMSU TMT weighting functions. It is about 100 times larger than the trend difference simulated using HadAT data over the overlap period. The cause is not known and could be a drift in calibration in one or both of the satellites that is not explained by our calibration error model (equation (3)). Since we do not know which satellite is closer to being correct, we treat this drift as an additional source of uncertainty. It is unlikely that the drift is caused by errors in the diurnal adjustment because the magnitude of the drift is similar for land-only and ocean-only averages, which is unlikely to be the case for errors in the diurnal cycle.

[32] To estimate the effect of this source of uncertainty, we add in artificial ascending (descending) location independent trends of 0.2 K per decade to the MSU (AMSU) noise realizations during the overlap period. Then, we pick 2-year portions of this overlap from a set of four possible periods (1999–2000, 2000–2001, 2001–2002, and 2002–2003) at random. MSU data from after the end of the period, and AMSU data from before the beginning of the period, are ignored. This process results in a significant contribution to the overall uncertainty in TMT and is the reason that the total uncertainty in TMT is comparable to the uncertainty in TLT, which has larger diurnal and sampling uncertainty.

[33] In the future, we will resolve this issue by detailed comparisons with subsequent AMSU instruments, the most important of these being AMSU on AQUA. Data from the AMSU instrument on NOAA-16, which would be useful if valid, appear to suffer from a large unexplained drift in several channels, including channel 5 [*Mears and Wentz*, 2009a].

2.7. Results

[34] At the end of the Monte Carlo merging process, we have 400 realizations of the expected error in our merged MSU/AMSU data sets. In Figure 10, we show maps of a realization of the gridded error for a typical month, along with maps of the standard deviation of the gridded errors calculated across the 400 realizations. For TLT and to a lesser extent, TMT, the errors are dominated by errors in the

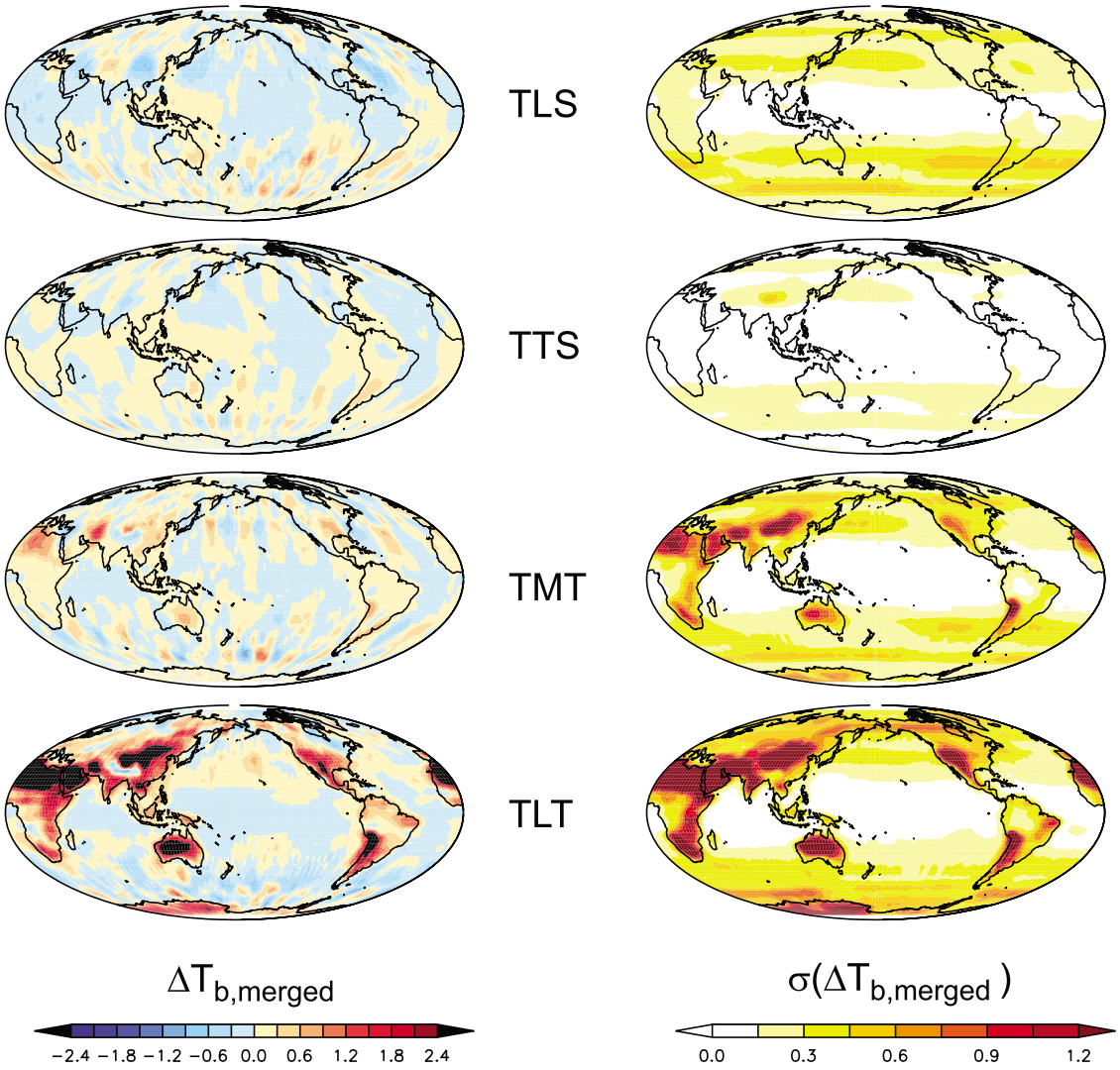

Figure 10. Gridded estimated error for each channel. The left column shows maps of a typical realization of estimated error from a single month (June). The right column shows the standard deviation of the gridded errors got the month of June calculated over the 400 realizations of error.

diurnal adjustment over land. Errors in TTS are significantly less than for the other channels, due to both the smaller diurnal adjustment and the relatively small sampling error given the generally low levels of synoptic variability in this atmospheric region. Errors tend to be spatially correlated indicating the dominance of sampling effects and diurnal uncertainties at these timescales. Both these error sources have strong spatial coherence.

[35] Figure 11 shows the range of global time series for each channel. Errors grow away from the tie-point NOAA-10 satellite in 1989. Errors are largest for the two lowest channels that have a significant surface signal. Table 2 shows the 2-σ estimated error in the global and tropical (20S to 20N) trends, with and without including the uncertainty due to the diurnal adjustment. For TMT, the MSU-AMSU drift is also excluded at the same time as the uncertainty in the diurnal adjustments. It is clear from the reduction in uncertainty when omitted that the largest factor in the error in our product for these lowermost channels arises from uncertainty in how we account for orbital diurnal drift effects and the associated knock-on impacts on determining intersatellite offsets. In addition to these first-order findings there are interesting high- and low-frequency aspects to the error estimates that may warrant future investigation but are outside the scope of this present analysis.

3. Comparison With Complementary Long-Term Atmospheric Temperature Data Sets

[36] Our set of gridded estimated error realizations will allow for the first time a rigorous analysis of the uncertainty inherent in a number of data set intercomparisons and climate model validation activities. These include comparisons of microwave satellite data with in situ data sets, comparisons with other satellite-derived temperature data sets such as those from COSMIC and comparison with reanalysis and climate model output, including fingerprinting studies. Here we provide an example of this type of analysis by briefly

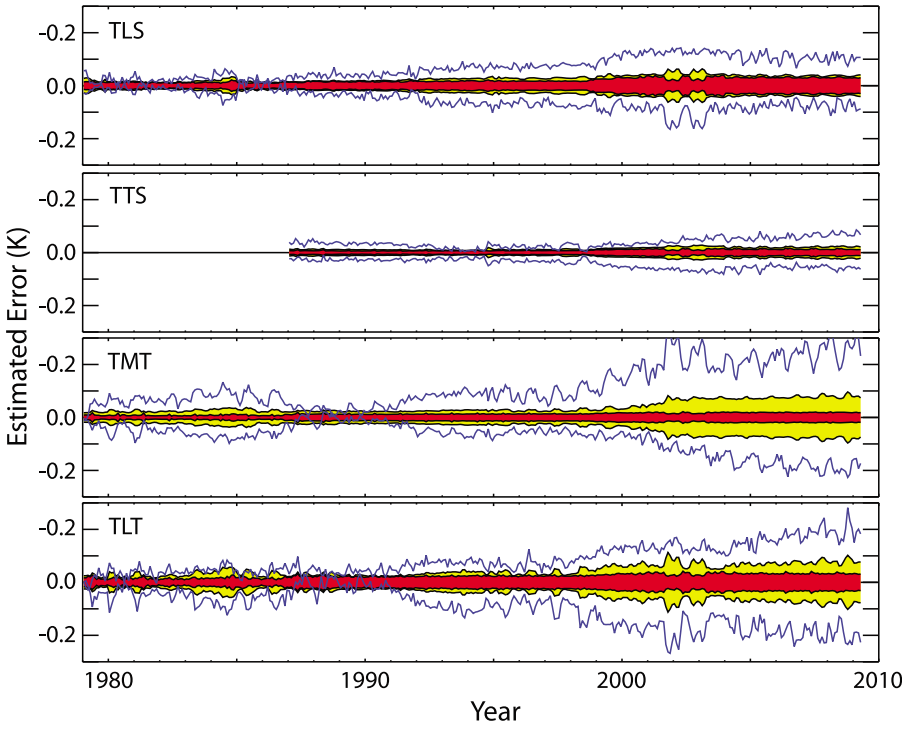

Figure 11. Range of globally averaged error time series for each channel. The yellow region is ±1 standard deviation for the globally averaged (75°S to 75°N) estimated error for each month. The red region is the same, except that the effects of the diurnal adjustment and, for TMT,the MSU/AMSU drift have been ignored. The blue lines are the global time series with the largest and smallest trend in the error realization set.

describing the intercomparison of our data with four homogenized radiosonde data sets and the two other current real-time MSU products.

[37] Here we take an approach of considering all published homogenized radiosonde data sets HadAT [*Thorne et al.*, 2005], RAOBCORE [*Haimberger*, 2007], RICH [*Haimberger et al.*, 2008], IUK [*Sherwood et al.*, 2008], and RATPAC [*Lanzante et al.* 2003; *Free et al.*, 2005]. The first four data sets were constructed using automated methods to find and estimate the size of "breakpoints" in the time series for radiosonde station which are then used to create adjusted versions of the radiosonde data with the effects of the detected breakpoints removed. The version of the RATPAC data used here, RATPAC-B, uses a similar but nonautomated method for data from 1958 to 1997. After 1997, no adjustments were made. We also show results for a subset of the RATPAC data set (RATPAC_RW) defined by *Randel and Wu* [2006], using short-term comparisons with satellite data with the intention of removing radiosonde stations from RATPAC with previously undetected inhomogeneities. These data sets are either available as gridded measurements vertically weighted to correspond to each channel or contain enough information so that it is possible for us to construct such a gridded data set by weighting individual levels.

[38] Several similar analyses and intercomparisons have previously been undertaken [*Christy et al.*, 2010] which question the veracity at given times of several of these products, including but by no means limited to our own [e.g., *Christy et al.*, 2007; *Christy et al.*, 2007; *Randall and Herman*, 2008]. We would caution that all such comparisons are two or more point comparisons between instruments or products that are not absolutely calibrated. Any such comparisons between uncertain measures cannot de facto provide absolute conclusions as to the veracity of any of the individual products. Indeed, on a scientific basis such comparisons have limited quantitative value without making recourse to defensible quantified uncertainty estimates in each comparator series when we know a priori that each is in fact uncertain. Rather than making at best semiobjective decisions based upon such imperfectly scientifically posed prior comparisons to preclude data products we prefer to include all products. We caution that readers should not consider our analysis in isolation of others which provide potentially valuable insights.

[39] We have already completed an intercomparison of TLT with these radiosonde data sets [*Mears and Wentz*,

Table 2. Uncertainty Estimates for Trends for Each Channel[a]

		TLT	TMT	TTS	TLS
Global (75°S–75°N)	all errors	0.044	0.042	0.014	0.028
	no diurnal	0.022	0.012[b]	0.008	0.020
Tropical (20°S–20°N)	all errors	0.034	0.038	0.020	0.030
	no diurnal	0.026	0.012[b]	0.008	0.020

[a]Uncertainty estimates are 2σ and trends for each channel are K/decade.
[b]The "no diurnal" results for TMT also does not include the effects of MSU/AMSU drift in this channel.

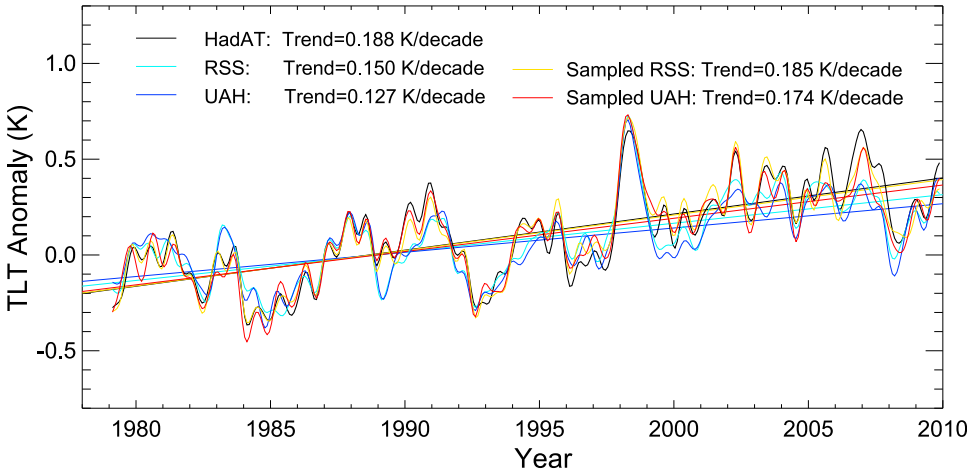

Figure 12. An example of the intercomparison of satellite data from RSS, UAH, and data from the HadAT radiosonde-based data set. For the satellite data, we show both the true globally averaged time series, and time series found by averaging together only those locations that have radiosonde data. These data have been smoothed to remove variability on time scales shorter than 6 months.

2009b], and thus this work only serves to add error estimates and results from RATPAC to this previous work. For TMT, TTS, and TLS we follow the earlier analysis method exactly. For each month, the satellite data is sampled at the location of the available radiosonde stations in each data set. This allows us to make a direct comparison with the radiosonde products, without needing to worry about whether or not the radiosonde sampling is dense enough to faithfully represent a global average. This means that the appearance and disappearance of radiosonde stations over time is also automatically taken into account. Note that this procedure results in a separate sampled satellite data set for each radiosonde data set. We then construct area-weighted global and regional time series from each data set for intercomparison. An example set of time series, for globally averaged data is shown in Figure 12. In Figure 12, we show for the satellite data both the true global time series and the time series calculated using the sampling procedure. The sampling procedure improves the agreement between the radiosonde data and the satellite data on short time scales in almost all cases. For long-term trends, agreement is also often improved by such subsampling. Subsampling is therefore important in making a fair comparison between the spatially complete satellite data and spatially incomplete radiosonde products [*Free and Seidel*, 2005].

[40] Similarly, trend error estimates for the radiosonde-sampled RSS satellite data were obtained by constructing error time series sampling the gridded error realizations at the radiosonde locations and then calculating the standard deviation of the trends in these time series. In general, the estimated trend error is larger than it is for more spatially complete averages, reflecting the influence of spatial sampling error. This is particularly true for the southern extratropics, where the radiosonde sampling is poor.

[41] Figure 13 shows a summary of trends for each channel, radiosonde data set, and averaging region. We also include sampled trends from the UAH [*Christy et al.*, 2003] and STAR [*Zou et al.*, 2006] satellite data sets, calculated using identical methods. We use the most recent versions of each data set available in gridded form, versions 5.3 (TMT, TLS) and 5.3 (TLT) for UAH and version 2.0 for STAR.

[42] Agreement between radiosonde and satellite data sets is best for TLT. For TLT, both the radiosonde and UAH trends lie within our error bars, except for the tropics and HadAT in the southern extratropics. All data sets agree that the largest tropospheric warming is in the northern extratropics, with least warming in the southern extratropics. The exception to this is the UAH data set that exhibits approximately the same warming rate in the deep tropics and the southern extratropics. Overall UAH exhibits more extratropical warming and less tropical warming than RSS which leads to a near-cancellation of differences at the global average.

[43] For TLT, changes in radiosonde trends as a function of latitude (i.e., the gradient between these zones and not the absolute values) are in better agreement with RSS than UAH. This finding contrasts with many recent publications which suggest UAH better matches with subsets of the radiosonde network than does RSS. But as the radiosondes may retain common biases we would strongly argue that both this result and the suite of recently published radiosonde-satellite comparisons cannot be used to make concrete inferences about the relative quality of the two satellite TLT products. We include this observation largely to highlight that there are multiple potential diagnostics which one could use to evaluate MSU series against radiosondes and that the choice of diagnostic can significantly impact conclusions about apparent MSU data set quality and a naïve choice as to a "winner" on such a basis.

[44] In the tropics, the RSS trends tend to be higher than the radiosonde or UAH data sets. In this region three of the radiosonde data sets and the sampled UAH data set all fall outside the range of RSS +/− our internal uncertainty estimates with an exception for the RATPAC_RW sampling. TLT trends in the tropics have been the subject of much recent controversy [*Douglass et al.*, 2008; *Santer et al.*,

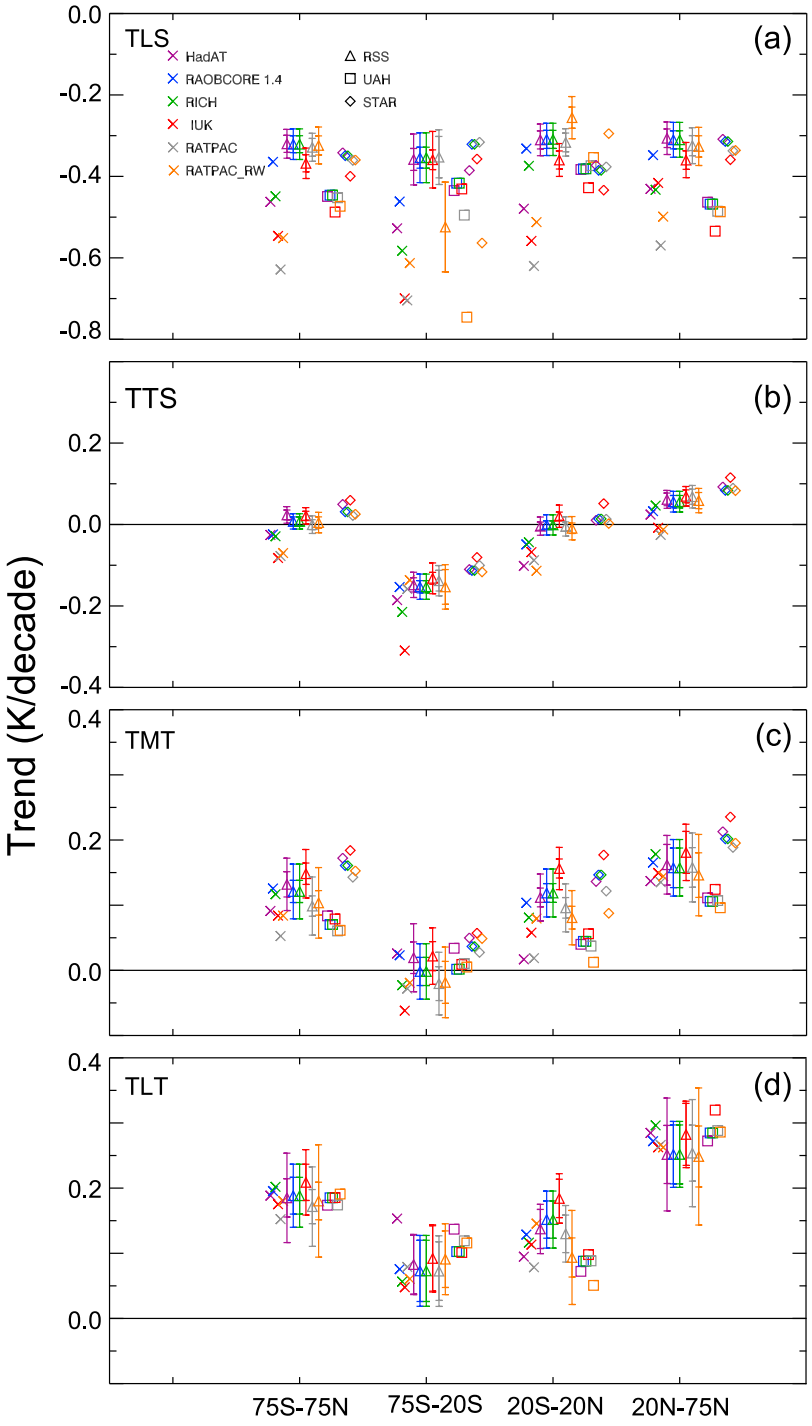

Figure 13. Summary of radiosonde, satellite, and spatially sampled satellite trends for each channel. All trends are calculated for 1979–2009 with two exceptions. TTS trends start in 1987, and all trends involving IUK end in 2005. Error bars on the satellite data were calculated using the methods described in this work. The outer error bars are the total 2-σ errors including the contributions from the diurnal adjustment and the MSU/AMSU drift, and the inner error bars include only the contributions from sampling error.

2008]. Our error bars increase the probability of concurrence between our data set and model expectations, reducing the chances of proving a discrepancy if our method of correcting the data is adequate, an aspect which we stress again our current analysis does not address.

[45] As we move higher in the atmosphere, the radiosonde-satellite trend differences tend to increase. For TMT, only about 50% of the radiosonde trends lie within our error bars, except in the northern extratropics, where the radiosonde network is the most spatially complete, and thus the

adjusted data sets, which in almost all cases rely upon a neighbor constraint, are most likely to be reliable. Here the agreement remains good. For TMT, the STAR trends are consistently larger than RSS but generally just within our uncertainty bounds when diurnal adjustment uncertainties are included, while the UAH trends are less and tend to be outside our calculated error margins with the exception of the southern extratropics where the two data sets are in good agreement.

[46] It is important to note that for both TLT and TMT, the different satellite data sets agree most well in the southern extratropics, where the effects of diurnal drift are small because of the small amount of land. This further confirms our finding for our data set that unambiguously resolving the diurnal drift effect correction and its impacts is likely to be a key determinant in reducing the uncertainty in long-term tropospheric temperature changes from MSU/AMSU records.

[47] For TTS, the RSS data show trends that are negatively biased relative to the STAR data and positively biased relative to the radiosonde data. Almost all other data sets lie outside the range of our error bars in all regions, suggesting that our uncertainty analysis may be missing some important error sources for this channel. However, as they lie either side of our estimate it does not imply that our best-guess product is necessarily biased.

[48] For TLS, the results are similar to TTS, except RSS and STAR are in better agreement with most STAR trends lying inside our error bars. The radiosonde data show much more cooling than the RSS satellite data, particularly in the southern extratropics. UAH data in this region are consistently and significantly cooling faster than our estimates and in the Northern Hemisphere extratropics also at a greater rate than all available radiosonde estimates. The increasing discrepancy between satellite and radiosonde based estimates as we move higher in the atmosphere is likely to be caused at least in part by residual errors in the homogenized radiosonde data sets [*Mears et al.*, 2006; *Randel and Wu*, 2006].

[49] It is notable that agreement between all radiosonde and satellite data sets at all levels is best in the northern extratropics, where the radiosonde spatial sampling is much more complete. Spatially, more complete sampling is likely to improve the accuracy of the radiosonde homogenization methods, all of which are based either in their breakpoint identification and/or in their adjustment step in some sense upon neighbor comparisons.

4. Conclusions

[50] We have performed a comprehensive internal error analysis of our MSU/AMSU based data sets of atmospheric temperature. This work improves upon earlier work in that it provides uncertainty information on a variety of spatial and temporal scales, which is critical for meaningful application of the data sets to the study of climate change on both global and regional scales. The fundamental results of our calculations, a set of 400 realizations of estimated error for each MSU/AMSU channel, is available for download on our Web site, http://www.remss.com.

[51] We find that on a month to month basis or for very local scale studies sampling effects are likely to be important. At larger time and space scales the dominant sources of uncertainty are intersatellite merging uncertainty arising from a finite nonoverlapping sample and how to adjust for diurnal cycle aliasing introduced through satellite drift for the two tropospheric layers with a surface contribution. This latter effect is the dominant source of uncertainty for our tropospheric estimates. Its effect is twofold in that it is a substantial uncertainty in itself but also because the processing system is sequential and the satellite merge is the last step it also substantially impairs our ability to cleanly undertake this step. Getting a better handle on this effect should be a priority and requires many more diurnal cycle climatology estimates at 3-hourly or preferably hourly resolution from a larger suite of climate models and reanalysis products from the skin temperature all the way to the top-of-atmosphere. Reanalyses, which are constrained by observations, would be a particularly useful resource.

[52] A comparison to alternative estimates from radiosondes and independent producers of MSU/AMSU products was made in the context of our new uncertainty analysis. It is clear from this comparison that many hitherto unexplained differences between the data sets, many of which have been previously documented, remain. Although the internal uncertainty estimates derived herein lead to consistency between a number of estimates there are nearly as many cases where differences between the RSS product and competing estimates cannot be reconciled as being caused solely by RSS data set internal uncertainties. An inescapable conclusion from this is that the methodological choices that we and others have made have lead to a substantial and significant impact upon the resulting estimates. This reinforces the importance of creating multiple independent estimates from the raw data which is known both to contain nonclimatic influences and lack metrological traceability if we are to avoid the possibility of reaching false conclusions.

[53] **Acknowledgments.** Work at Remote Sensing Systems was supported by the NOAA Climate Program Office, NOAA award NA08OAR4310674. The U.K. Met Office authors were supported by the Joint DECC and Defra Hadley Centre Climate Programme – DECC/Defra (GA01101).

References

Beagley, S. R., C. McLandress, V. I. Fomichevand, and W. E. Ward (2000), The Extended Canadian Middle Atmosphere Model, *Geophys. Res. Lett.*, *27*(16), 2529–2532, doi:10.1029/1999GL011233.

Christy, J. R., and W. B. Norris (2006), Satellite and VIZ-radiosonde intercomparisons 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 radiosondes and satellite temperatures in the Australian region, *J. Atmos. Oceanic Technol.*, *26*(3), 508–522, doi:10.1175/2008JTECHA1126.1.

Christy, J. R., R. W. Spencer, and W. D. Braswell (2000), MSU tropospheric temperatures: Dataset construction and radiosonde comparisons, *J. Atmos. Oceanic Technol.*, *17*(9), 1153–1170, doi:10.1175/1520-0426(2000)017<1153:MTTDCA>2.0.CO;2.

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. Herman, R. Peilke Sr., P. Klotzbach, R. T. McNider, J. J. Hnilo, R.W. Spencer, T. Chase, and D. Douglass (2010), What do obser-

vational datasets say about modeled tropospheric temperature trends since 1979?, *Remote Sens.*, *2*, 2148–2169, doi:10.3390/rs2092148.

Douglass, D. H., J. R. Christy, B. D. Pearson, and F. S. Singer (2008), A comparison of tropical temperature trends with model predictions, *Int. J. Climatol.*, *28*(13), 1693–1701, doi:10.1002/joc.1651.

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 dataset 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.

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

Grody, N. C., K. Y. Vinnikov, M. D. Goldberg, J. T. Sullivan, and J. D. Tarpley (2004b), 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*(7), 1377–1403, doi:10.1175/JCLI4050.1.

Haimberger, L., C. Tavolato, 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.

Karl, T. R., S. J. Hassol, C. D. Miller, and W. L. Murray (2006), Temperature trends in the lower atmosphere: Steps for understanding and reconciling differences, report, U. S. Clim. Change Sci. Progr., Washington D. C. (Available at http://www.climatescience.gov/Library/sap/sap1-1/finalreport/default.htm)

Kennedy, J. J., P. Brohan, and S. F. B. Tett (2007), A global climatology of the diurnal variations in sea-surface temperature and implications for MSU temperature trends, *Geophys. Res. Lett.*, *34*, L05712, doi:10.1029/2006GL028920.

Kiehl, J. T., J. J. Hack, G. B. Bonan, B. A. Boville, B. P. Briegleb, D. L. Williamson, and P. J. Rasch (1996), Description of the NCAR Community Climate Model (CCM3), *Rep. TN-420*, Natl. Cent. for Atmos. Res., Boulder, Colo.

Lanzante, J. R., S. Klein, and D. J. Seidel (2003), Temporal homogenization of monthly radiosonde temperature data. Part I: Methodology, *J. Clim.*, *16*, 224–240, doi:10.1175/1520-0442(2003)016<0224: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*, edited by T. R. Karl et al., pp. 47–70, U.S. Clim. Change Sci. Progr., Washington, D. C. (Available at http://www.climatescience.gov/Library/sap/sap1-1/finalreport/default.htm)

Martin, G. M., M. A. Ringer, V. D. Pope, A. Jones, C. Dearden, and T. J. Hinton (2006), The physical properties of the atmosphere in the new Hadley Centre Global Environmental Model, HadGEM1. Part 1: Model description and global climatology, *J. Clim.*, *19*(7), 1274–1301, doi:10.1175/JCLI3636.1.

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 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., and F. J. Wentz (2009b), 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., M. C. Schabel, and F. J. Wentz (2003), A reanalysis of the MSU channel 2 tropospheric temperature record, *J. Clim.*, *16*(22), 3650–3664, doi:10.1175/1520-0442(2003)016<3650:AROTMC>2.0.CO;2.

Mears, C. A., C. E. Forest, R. W. Spencer, R. S. Vose, and R. W. Reynolds (2006), What is our understanding of the contribution made by observational or methodological uncertainties to the previously reported vertical differences in temperature trends?, in *Temperature Trends in the Lower Atmosphere: Steps for Understanding and Reconciling Differences*, edited by T. R. Karl et al., pp. 71–88, U.S. Clim. Change Sci. Progr., Washington, D. C. (Available at http://www.climatescience.gov/Library/sap/sap1-1/finalreport/default.htm)

Prabhakara, C., R. Iaacovazzi, J. M. Yoo, and G. Dalu (2000), Global warming: Evidence from satellite observations, *Geophys. Res. Lett.*, *27*(21), 3517–3520, doi:10.1029/2000GL011719.

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.

Santer, B. D., et al. (2008), Consistency of modelled and observed temperature trends in the tropical troposphere, *Int. J. Climatol.*, *28*(13), 1703–1722, doi:10.1002/joc.1756.

Seidel, D. J., et al. (2009), Reference upper-air observations for climate: Rationale, progress, and plans, *Bull. Am. Meteorol. Soc.*, *90*, 361–369, doi:10.1175/2008BAMS2540.1.

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.

Solomon, S. D., D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, and H. L. Miller (2007), *Climate Change 2007: The Physical Scientific Basis: Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, 996 pp., Cambridge Univ. Press, Cambridge, U.K.

Space Studies Board (2007), *National Research Council of the National Academies, National Imperatives for the Next Decade and Beyond, Committee on Earth Science and Applications from Space: A Community Assessment and Strategy for the Future*, National Academies Press, Washington, D.C.

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, 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.*, *110*, D18105, doi:10.1029/2004JD005753.

Thorne, P. W., J. R. Lanzante, T. C. Peterson, D. J. Seidel and K. P. Shine, (2010), Tropospheric temperature trends: History of an ongoing controversy, *Wiley Interdisciplinary Rev. Clim. Change*, *2*(1), 66–88, doi:10.1002/wcc.80

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.

D. Bernie, U. K. Met Office, FitzRoy Road, Exeter EX1 3PB, UK.

C. A. Mears and F. J. Wentz, Remote Sensing Systems, 438 First Street, Suite 200, Santa Rosa, CA 95401, USA. (mears@remss.com)

P. Thorne, Cooperative Institute for Climate and Satellites-North Carolina, NOAA NCDC, 151 Patton Avenue, Asheville, NC 28801, USA.