Chapter 4

What is our understanding of the contribution made by observational or methodological uncertainties to the previously reported vertical differences in temperature trends?

Convening Lead Author: Carl Mears

Lead Authors: Chris Forest, Roy Spencer, Russell Vose, and Dick Reynolds

Contributing Authors

Peter Thorne and John Christy

Chapter 4: Key Findings

Surface

It is likely that errors in the homogenized surface air temperature data do not contribute substantially to the large-scale differences between trends for different levels because these errors are very likely to be smaller than those for the upper air data.

- Systematic local biases in surface trends may exist due to changes in station exposure or instrumentation over land, and due to the small number of measurements over a number of regions of the earth, including parts of the oceans, sea ice areas, and some land areas. Such biases have been documented at the local and regional scale, but no such effect has been identified in the zonal and global averages presented in this report. On large spatial scales, sampling studies suggest that these local biases in trends are likely to mostly cancel through the use of many observations with differing instrumentation.
- Since all known bias adjustments have not yet been applied to sea surface temperature data, it is likely that errors remain in these data, though it is generally agreed that these errors are likely to be small compared to errors in radiosonde and satellite measurements of the upper air, especially for the satellite era.

Troposphere

While all datasets indicate that the troposphere has warmed over both the radiosonde era and the satellite era, uncertainties in the tropospheric data make it difficult to determine whether the troposphere has warmed more than or less than the surface. Some tropospheric datasets indicate that the troposphere has warmed more than the surface, while others indicate the opposite.

- It is very likely that errors remain in the homogenized radiosonde datasets in the troposphere since the methods used to produce them are only able to detect and remove the more obvious errors, and involve many subjective decisions. It is likely that a net spurious cooling corrupts the area-averaged homogenized radiosonde data in the tropical troposphere in at least one and probably both of the datasets, causing the data to indicate less warming than has actually occurred.
- For tropospheric satellite data (T_2 and T_{2LT}), the primary cause of trend discrepancies between different versions of the datasets is differences in how the data from the different satellites are merged together.
- A secondary contribution to the differences between these datasets is the difference between the diurnal adjustments that are used to account for drifting measurement times. These differences in the diurnal adjustment are more important for regional trends than for global trends, though regional trend differences are also partly influenced by differences in merging methods.

- Each tropospheric satellite dataset has strengths and weaknesses that are coming into better focus. Improvements have occurred in several datasets even during the drafting of this report, each moving it closer to the others, suggesting that further convergence in the not-too-distant future is a strong possibility.
- Comparisons between radiosonde data and satellite data for T_2 are very likely to be corrupted by the excessive cooling in the radiosonde data from the stratosphere which are used to help construct the radiosonde-derived T_2 data. Trend discrepancies between radiosonde and satellite datasets are reduced by considering a multi-channel retrieval that estimates and removes the stratospheric influence ($T*_G$).

Stratosphere

Despite their large discrepancies, all datasets indicate that the stratosphere has cooled considerably over both the radiosonde era and the satellite era.

- The largest discrepancies between datasets are in the stratosphere, particularly between the radiosonde and satellite-based datasets. It is very likely that the satellite-sonde discrepancy arises primarily from uncorrected errors in the radiosonde data.
- There are also substantial discrepancies between the satellite datasets in the stratosphere, indicating that there remain unresolved issues with these datasets as well.

Chapter 4 recommendations

All of the surface and atmospheric temperature datasets used in this report require ongoing assessment to further quantify uncertainty and to identify and remove any possible systematic biases that remain after the appropriate homogenization methods have been applied.

- *The diurnal cycles in both atmospheric and surface temperature need to be accurately determined and validated to reduce uncertainties in the satellite data due to the diurnal adjustment. Possible approaches include examining more model or reanalysis data to check the diurnal adjustments currently in use, concerted in situ measurement campaigns at a number of representative locations, or operating a satellite-borne sounder in a non sun-synchronous orbit. Information about the surface skin temperature diurnal cycle may be obtained by studying data from existing satellites, or the upcoming Global Precipitation Mission.*
- *The relative merits of different merging methods for satellite data for all relevant layers need to be diagnosed in detail. Possible approaches include comparison with other temperature data sources (radiosondes or IR satellites) over limited time periods where the discrepancies between the satellite results are the greatest, comparison with other ancillary data sources such as winds and integrated water vapor, and comparison of trends on regional spatial scales, particularly in regions where trends are large or well characterized by radiosonde data.*

- *The methods used to remove radiosonde inhomogeneities and their effects on trends need to be rigorously studied. The detailed intercomparisons of the methods used by different groups to construct satellite based climate records has been beneficial to our understanding of these products, and similar parallel efforts to create climate records from radiosonde data would be likely to provide similar benefits.*
- *Possible errors in trends in spatially averaged surface temperature need to be assessed further. On land these errors may arise from local errors due to changes in instrumentation or local environment that do not completely cancel when spatial averaging is performed. Over the ocean, these errors may arise from the small number of samples available in many regions, and long-term changes in measurement methods. For historical data, these assessments may benefit from the recovery of additional metadata to better characterize possible non-climatic signals.*
- *Tools and methods need to be developed to help reduce structural uncertainty by providing methods to objectively differentiate between different datasets and construction methods. To the extent possible, such tools should be based on generally accepted physical principles, such as consistency of the temperature changes at adjacent levels in the atmosphere, include physically-based comparisons with external ancillary data, and take account of the consistency of intermediate data generated while producing the datasets.*

1. Background

In the previous chapter, we have discussed a number of estimates of vertically resolved global temperature trends. Different sources of data (e.g., surface measurements, vertical profiles from radiosondes, and data from satellite borne sounding radiometers), as well as different analysis methods applied to the same data, can yield long term (multi-decadal) temperature trends that differ by as much as several tenths of a ºC per decade. This is of comparable magnitude to the actual climate change signal being searched for. In this chapter we discuss these discrepancies in light of the observing system capabilities and limitations described in Chapter 2. We note the degree to which estimates of uncertainty can account for the differences in reported values for the temperature trends in given layers, and differences in the trends of adjacent layers. Most of the time our focus will be on the period from 1979-2004, during which atmospheric temperatures were observed using multiple observing systems.

We begin our discussion in the stratosphere, and move to successively lower layers until we reach the Earth's surface. We proceed in this order because the largest discrepancies in trends between data sources occur in upper atmospheric layers, especially the stratosphere. As mentioned in Box 2.2, when satellite-equivalent measures are made from vertically resolved radiosonde data to facilitate comparisons between the two systems, large stratospheric errors can significantly influence measures centered much lower in the atmosphere.

2. Uncertainty in stratospheric temperature trends

Long-term observations of the stratosphere have been made by two observing systems: radiosondes and satellite-borne sounders. On both the global and the zonally averaged scale, there is considerably less variation between datasets derived from the same type of observing system for this layer than between those from different observing systems. This can be seen in the leftmost panel of Figure 3.5, which shows the zonally averaged trends over the satellite era (1979-2004) for two radiosonde-based datasets, and two satellite-based datasets. The radiosonde data ($T_{4\text{-HadAT2}}$ and $T_{4\text{-NOAA}}$) show more cooling than datasets based on satellite data ($T_{4\text{-UAH}}$ and $T_{4\text{-RSS}}$), and also do not show the reduced cooling in the tropics relative to the mid-latitudes that is seen in the satellite data.

2.1 Radiosonde Uncertainty

As discussed in Chapter 2, radiosonde data are plagued by numerous spurious discontinuities in measured temperature that must be detected and removed in order to construct a homogenized long-term record of atmospheric temperature, a task that is particularly difficult in the absence of reliable metadata describing changes in instrumentation or observing practice. A number of physical sources of such discontinuities have larger effects in the stratosphere. The lower atmospheric pressure in the stratosphere leads to reduced thermal contact between the air and the temperature sensor in the radiosonde package. This in turn leads to increased errors due to daytime solar heating and lags between the real atmospheric temperature and the sensor response as the instrument rises through atmospheric layers with rapidly varying temperatures.

Such systematic errors are not important for trend studies provided that they do not change over the time period being studied. In practice, as noted in Chapter 2, radiosonde design, observing practices, and procedures used to attempt to correct for radiation and lag errors have all changed over time.

Past attempts to make adjustments to radiosonde data using detailed physical models of the instruments (Luers and Eskridge, 1998) improved data homogeneity in the stratosphere, but not in the troposphere (Durre et al., 2002). Since it is important to use the same methods for all radiosonde levels for consistency, scientists have tended to instead use empirical methods to deduce the presence and magnitude of any suspected discontinuity. Both of the homogenized radiosonde datasets used in this report make these estimates using retrospective statistical analyses of the radiosonde data without input from other measurements. The investigators who constructed these datasets have attempted to identify and to adjust for the effects of suspected change points, either by examination of station time series in isolation (NOAA), or by comparison with nearby stations (UK). Both approaches can most successfully identify changes that are large and step-like. While based in statistics, both these methods also include significant subjective components. As a result, different investigators with nominally the same sets of radiosonde data can calculate different trend estimates because of differences in adjustment procedures (Free et al., 2002). The lack of sensitivity to small or gradual changes may bias the resulting homogenized products if such changes are numerous and predominantly of one sign or the other[1]. The relative frequency of large step-like changes and smaller changes that may be statistically indistinguishable from natural variability remains an open question.

Since the adjustments needed to remove the resulting discontinuities tend to be larger for the stratosphere than for lower levels (Parker et al., 1997; Christy et al., 2003; Lanzante et al., 2003), the uncertainty associated with the homogenization procedures is very likely to be larger in the stratosphere than at lower levels, as has been shown for the UK radiosonde dataset (Thorne et al., 2005). The best estimate of the size of this source of uncertainty is obtained by comparing the statistics (e.g., the trends) from the two adjusted radiosonde datasets that are currently available. However, the UK group analysis is partly based upon the NOAA dataset, so we may be under-estimating the uncertainty. Only through increasing the number of independently produced datasets under different working assumptions can we truly constrain the uncertainty (Thorne et al., 2005).

Differences in trends between daytime and nighttime observations in the uncorrected radiosonde data used in constructing the NOAA and UK radiosonde datasets, suggest that the biases caused by solar heating[2] have been reduced over time, leading to a spurious cooling trend in the raw daytime data (Sherwood et al., 2005). Many of the changes in

[1] It is speculated that gradual changes could result from the same changes in instrumentation or practices that cause the step like changes, provided that these changes are implemented gradually (Lanzante et al., 2003).

[2] For some types of radiosondes, radiation adjustments based on information provided by the manufacturer are made as part of routine processing of radiosonde data by the observing station. The findings cited here refer to data that has already had these corrections performed. The reduction in daytime biases is likely to be due to a combination of improvements in instrument design, and improvements in the radiation adjustment procedure.

observing practice will affect both day and night time observations; e.g., a change in practice may yield a spurious 0.5ºC daytime cooling and 0.4ºC night time cooling, so day-night differences cannot be used in isolation to correct the observations. Whether the NOAA and UK methods have successfully removed day-night and other effects, or if sufficiently targeted are capable of doing so, is a matter for ongoing research. Randel and Wu (2005) have shown for a subset of tropical stations in the NOAA dataset, there is strong evidence for step-like residual cooling biases following homogenization, which will cause a spurious cooling in the tropical area-averaged NOAA time series considered here. They find that the effect is not limited to daytime launches, as would be expected from discussions above, and that it is likely to affect at least the upper-troposphere as well as the stratosphere. Finally, the balloons that carry the instruments aloft have improved over time, so they are less likely to burst at high altitudes or in extreme cold. This could also lead to a warm sampling bias within the stratosphere in early radiosondes which has gradually ameliorated with time, introducing a spurious stratospheric cooling signal (Parker et al., 1997). Taken together these results imply that any residual systematic errors in the homogenized radiosonde products will likely lead to a spurious cooling bias.

Since the radiosonde stations selected for inclusion in the homogenized datasets do not cover the entire globe, there can be a bias introduced in to the global mean trend depending on the locations of the chosen stations. On a global scale, this bias has been estimated to be less than 0.02ºC/decade for T_4 by sub-sampling globally complete satellite or reanalysis datasets at the station locations[3], and thus it is not an important cause of the differences between the datasets on large spatial scales (Free and Seidel, 2004). Though they have not been explicitly calculated, sampling errors are likely to be more important for the zonal radiosonde trends plotted in Figure 3.5, and may account for some of the zone-to-zone variability seen in the radiosonde data in that figure that is not duplicated in the smoother satellite data. The sampling effects also permeate in the vertical – above 100hPa there is a significant reduction in the number of valid measurements whereas below this level the number of measurements is relatively stable. Because the trends vary with height, this can lead to errors, particularly when calculating satellite-equivalent measures.

2.2 Satellite Uncertainty

The two satellite-based stratospheric datasets ($T_{4\text{-UAH}}$ and $T_{4\text{-RSS}}$) have received considerably less attention than their tropospheric counterparts (see section 4.3 below), though they differ in estimated trend by roughly the same absolute amount (~0.1ºC/decade) as the corresponding tropospheric datasets produced by the same institutions. However the importance of the differences is perceived to be much less because the trend is much larger (a cooling over 1979-2004 of approximately 0.8ºC). A detailed comparison of the methods used to construct the two datasets has not yet been performed. Despite the lack of such a study, it is very likely that in the stratosphere, like the troposphere (discussed in section 4.3), structural uncertainty is the most important

[3] This estimate is valid for the NOAA dataset and a previous version of the UK dataset. The estimated bias increases to about 0.05K for a tropical average. In the cited work the tropics were defined to be 30S to 30N – we would expect the sampled error to be a few hundredths of a degree per decade larger for the 20S to 20N definition of the tropics used in this report.

source of uncertainty. Two important types of structural uncertainty are likely to dominate: those associated with the method of correcting for drifts in diurnal sampling time, and those associated with the method of correcting calibration drifts associated with the temperature of the hot calibration target. Section 3 discusses how these uncertainty sources are treated in the troposphere.

Despite unresolved problems in the satellite datasets, the similarity of the satellite measurement and homogenization methods suggest that the satellite measurements of the stratosphere are no more uncertain than those of the mid-troposphere, where satellites and radiosondes are in much closer agreement. This assessment, coupled with the evidence presented above that residual artificial cooling is likely to exist in the stratospheric radiosonde data, particularly in the tropics, implies that the discrepancy between radiosondes and satellite estimates of stratospheric trends (see Table 3.3) during the satellite era is very likely to be mostly due to uncorrected biases in the radiosonde measurements.

3. Uncertainty in tropospheric trends

In contrast to the stratosphere, differences in reported tropospheric trends from the same type of measurement are as large as or larger than differences in trends reported from different data sources. This can be seen in Figure 3.5 and Tables 3.3 and 3.4. Also note that the radiosonde data for the two tropospheric layers show the same general north-south pattern (i.e. more temperature increase in the mid-latitudes than at the poles or in the tropics) as the satellite data, in contrast to the stratospheric results.

3.1 Radiosonde uncertainty

The main sources of error in tropospheric radiosonde trends are similar to those encountered in the stratosphere. The challenge is to assess to what extent these types of errors, which in the stratosphere likely result in artificial cooling even in homogenized datasets, extend down into the troposphere. Another important issue is that when performing calculations to directly compare radiosonde data with satellite trends for the T_2 layer, the contribution of errors in the stratospheric trends to the results for this layer become important, since 10% to 15% of the weight for this layer comes from the stratosphere.

3.1.1 Removing non-climatic influences.

There are several pieces of evidence that suggest that any residual bias in tropospheric radiosonde data will be towards a cooling. First, the more obvious step-like inhomogeneities that have been found tend to predominantly introduce spurious cooling into the raw time series, especially in the tropics. This suggests that any undetected change points may also favor spurious cooling (Lanzante et al., 2003). Second, solar-heating-induced errors, while largest in the stratosphere have been found to bias daytime measurements to higher temperatures at all levels, particularly in the tropics. Periodic radiosonde intercomparisons (most recently at Mauritius in Feb. 2005) undertaken under the auspices of WMO imply that the magnitude of these errors has been reduced over time, and that radiosondes from independent manufacturers have become increasingly similar (and presumably more accurate) over time[4] (Silveira et al., 2003; Pathack et al., 2005). If these effects have on average been uncorrected by the statistical procedures used to construct the homogenized radiosonde datasets discussed in this report, they would introduce an artificial cooling signal into the radiosonde records. Of course on an individual station basis the picture is likely to be much more ambiguous and many stations records, even following homogenization efforts, are likely to retain large residual warm or cold biases. But on average, the evidence outlined above suggests that if there is a preferred sign it is likely to be towards a residual cooling. It is important to stress that to date the quantitative evidence to support such an argument, at least away from a small number of tropical stations (Randel and Wu, 2005), is at best ambiguous.

3.1.2 Sampling uncertainty

The fact that most radiosonde data are primarily collected over Northern Hemispheric land areas naturally leads to uncertainties about whether or not averages constructed from radiosonde data can faithfully represent global trends. However, (Wallis, 1998) and (Thorne et al., 2005) show that stations can be representative of much larger scale averages above the boundary layer, particularly within the deep tropics. Spatial and temporal sampling errors for the radiosonde datasets have been assessed by sub-sampling trends in reanalyses or satellite data at the locations of radiosonde stations used in the production of global datasets, and comparing the results to the full global average of the reanalysis or satellite data (Free and Seidel, 2004). Typically, errors of a few hundredths of a ºC per decade have been estimated for global averages, too small to fully account for

[4] These intercomparisons provide a source data about the differences between different type of sondes that have not yet been used to homogenize sonde data.

the differences between radiosonde and satellite trends, though it has been suggested that the existing sampling could lead to a warm bias in the radiosonde record (Agudelo and Curry, 2004). As is the case for the stratosphere, sampling errors may be part of the cause for the zone to zone variability seen in the radiosonde data. Residual differences between two radiosonde dataset global means are assessed to be approximately equally caused by sampling error, choice of raw data, and choice of adjustments made[5].

3.1.3 The influence of uncertainty in stratospheric measurements

To compare data that represent identical layers in the atmosphere, "satellite-equivalent" radiosonde data products are constructed using a weighted average of radiosonde temperatures at a range of levels (see Box 2.2). The T_2 radiosonde datasets are constructed to match the weighting function for Microwave Sounding Unit (MSU) channel 2. Since 10% to 15% of the weight for this channel comes from the stratosphere (see Figure 2.1), it is important to keep in mind the suspected relatively large errors in the stratospheric measurements made by radiosondes. It is possible that stratospheric errors could cause the trends in the radiosonde-derived T_2 to be as much as 0.05ºC/decade too cool, particularly in the tropics, where the suspected stratospheric errors are the largest (Randel and Wu, 2005) and therefore have a large impact on area-weighted averages. This error source may be partly eliminated by considering the multi-channel tropospheric retrievals discussed in section 5 below.

[5] This comparison was made using a previous version of the UK dataset (HadRT), which uses a different set of stations than the current version. This difference is very unlikely to substantially alter these conclusions.

3.2. Satellite uncertainty

Satellite-derived temperature trends in the middle and upper troposphere have received considerable attention. In particular, the causes of the differences between $T_{2\text{-UAH}}$ and $T_{2\text{-RSS}}$ have been examined in detail; less work has been done concerning $T_{2\text{-U.Md.}}$ because this dataset is newer. There are two potentially important contributions to the residual uncertainty in satellite estimates of global trends for the satellite-based datasets: (1) corrections for drifts in diurnal sampling, and (2) different methods of merging data from the series of different satellites.

3.2.1 Diurnal Sampling Corrections

During the lifetime of each satellite, the orbital parameters tend to drift slowly with time. This includes both a slow change of the local equator crossing time (LECT), and a decay of orbital height over time due to drag by the upper atmosphere. The LECT is the time at which the satellite passes over the equator in a northward direction. Changes in LECT indicate corresponding changes in local observation time for the entire orbit. Because the temperature changes with the time of day (e.g., the cycle of daytime heating and nighttime cooling), slow changes in observation time can cause a spurious long-term trend. These diurnal sampling effects must be estimated and removed in order to produce a climate-quality data record.

The three research groups that are actively analyzing data from microwave satellite sounders first average together the ascending and descending orbits, which has the effect of removing most of the first harmonic of the diurnal cycle. For the purposes of this report, "diurnal correction" means the removal of the second and higher harmonics. Each group uses a different method to perform the diurnal correction.

The UAH group calculates mean differences by subtracting the temperature measurements on one side of the satellite track from the other (Christy et al., 2000). This produces an estimate of how much, on average, the temperature changes due to the difference in local observation times from one side of the satellite swath to another, typically about 40 minutes. This method has the advantage of not relying on data from other sources to determine the diurnal cycle, but it has been shown to be sensitive to satellite attitude errors (Mears and Wentz, 2005), and is too noisy to produce a diurnal adjustment useable on small spatial scales.

The RSS group uses hourly output from a climate model in a microwave radiative transfer model to estimate the diurnal cycle in brightness temperature at each grid point in the satellite dataset (Mears et al., 2003). This method has the advantage that a diurnal adjustment can be made at the data resolution. However, it is likely that the climate model-based adjustment contains errors, both because models are often unable to accurately represent the diurnal cycle (Dai and Trenberth, 2004), and because the parameterization of the ocean surface temperature used as a lower boundary for the atmospheric model used does not include diurnal variability. The model has been shown to represent the first harmonic of the diurnal cycle for MSU channel 2 with less than 10% error, but less is known about the accuracy of the second and higher harmonics that are more important for adjusting for the diurnal sampling errors (Mears et al., 2003).

Both groups use their diurnal cycle techniques to adjust the satellite data before merging the data from the different satellites. In contrast, the Maryland group averaged the ascending and descending satellite data to remove only the first harmonic in the diurnal cycle before merging, and used a fitting procedure to account for both the first and second harmonic diurnal components when performing the trend analysis *after* merging the data from different satellites (Vinnikov and Grody, 2003; Vinnikov et al., 2005). Since they only accounted for the first harmonic diurnal component during the merging of satellite data, errors in the diurnal cycle can cause errors in the data analysis following the merging procedure. However, the removal of the diurnal cycle before merging may also introduce some error into UAH and RSS merging procedures if the assumed diurnal cycle is inaccurate, but physically, the removal of the diurnal harmonics before merging seems to be a more logical approach as the diurnal harmonics will tend to add noise unless removed.

On a global scale, the total impact of the diurnal correction applied by the RSS and UAH groups to the microwave sounding data for the RSS data is to increase the decadal trend by about 0.03ºC/decade for T_2 (Christy et al., 2003; Mears et al., 2003). The impact of the Maryland group's adjustment is almost negligible. For the RSS T_2 data, when a diurnal correction is applied that is 50% or 150% as large as the best estimate, these adjustments significantly worsen the magnitude of the intersatellite differences. Changes of this magnitude in the diurnal cycle lead to temperature trends that differ by 0.015ºC; so we estimate that the uncertainty in trends due to uncertainty in the diurnal correction is about

0.015ºC/decade for T_2. The UAH group estimates that the diurnal correction for T_2 is known to 0.01ºC/decade (Christy et al., 2000). These estimates of residual uncertainty are relatively small, and are considerably less than the structural uncertainties associated with the satellite merging methodology described in the next section. Despite the global agreement for the diurnal adjustment for the RSS and UAH results, significant differences in the adjustments exist as a function of location (Mears and Wentz, 2005), which may explain some of the difference on smaller spatial scales between these two datasets[6].

3.2.2 Satellite merging methodology

It is very likely that the most important source of uncertainty in microwave sounding temperature trends is due to inter-satellite calibration offsets, and calibration drifts that are correlated with the temperature of the calibration target (Christy et al., 2000; Mears et al., 2003). When results from supposedly identical co-orbiting satellites are compared, intersatellite offsets are immediately apparent. These offsets, typically a few tenths of a ºC, must be identified and removed or they will produce errors in long-term trends of several tenths of a ºC per decade. When constant offsets are used to remove the inter-satellite differences, the UAH group found that significant differences still remain that are strongly correlated with the temperature of the calibration target[7] (Christy *et al.,* 2000). This effect has since been confirmed by the RSS group (Mears *et al.,* 2003). Both

[6] See for example Figure 3.5 versus Figure 4.3.
[7] The calibration target can change temperature by tens of ºC over the course of the life of the satellite due to orbit- and season-dependent solar heating.

the UAH and RSS groups remove the calibration target temperature effect using a model that includes a constant offset for each satellite, and an additional empirical "target factor" multiplied by the calibration target temperature.

Despite the similarity in methods, the RSS and UAH groups obtain significantly different values for the global temperature trends (see Table 3.3). In particular, the difference between the trends for T_2 has received considerable attention. A close examination of the procedures suggests that about 50% of the discrepancy in trends is accounted for by a difference between the target factor for the NOAA-09 instrument deduced by the two groups. This difference mainly arises from the subsets of data used by the two groups when determining the satellite merging parameters (i.e., offsets and target factors). The UAH group emphasizes pairs of satellites that have long periods of overlap, and thus uses data from six pairs of satellites, while RSS uses all available (12) overlapping pairs of satellites. Most of the remainder of the difference is due to a smaller difference in the calibration target temperature proportionality constant for NOAA-11, and to small differences in the diurnal correction. Both these differences primarily affect the measurements made by NOAA-11 and NOAA-14, due to their large drifts in local measurement time.

In Figure 4.1a, we plot the difference ($T_{2\text{-RSS}} - T_{2\text{-UAH}}$) between the RSS and UAH time series. There is an obvious step that occurs in 1986, near the end of the NOAA-09 observation period, and a gradual slope that occurs during the observation periods of NOAA-11 and NOAA-14. Note that the trend difference between these two datasets is statistically significant at the 1% level, even though the error ranges quoted in Table 3.3 overlap, due to the presence of nearly identical short term fluctuations in the two datasets (see Appendix A for more details).

Figure 4.1 (a) Time series of the difference between global averages of satellite-derived T_2 datasets. Both the RSS and UMD datasets show a step-like feature relative to the UAH dataset during the lifetime of NOAA-09. The difference between the RSS and the UAH datasets shows a slow drift during the NOAA-11 and NOAA-14 lifetimes. Both these satellites drifted more than 4 hours in observations time. (b) Time series difference between global averages of satellite derived T_{2LT} datasets. A slow drift is apparent during the lifetime of NOAA-11, but the analysis during the NOAA-14 lifetime is complicated because the $T_{2LT-RSS}$ dataset does not include data from the AMSU instruments on NOAA-15 and NOAA-16, while the $T_{2LT-UAH}$ dataset does. All time series have been smoothed using a Gaussian filter with width = 7 months.

The Maryland group data set ($T_{2-U.Md.}$), in its most recent version (Grody et al., 2004; Vinnikov et al., 2005), implemented a more detailed, physically based error model to describe the errors that correlated with a nonlinear combination of the observed brightness temperature measurements and the warm target temperature used for calibration[8]. They use a substantially different merging procedure to deduce values of the parameters that describe the intersatellite differences. First, as noted above, only the first harmonic diurnal component is accounted for during the satellite merging, possibly causing errors in the retrieved parameters. Second, they only use the spatial variation seen by the different MSU instruments to derive the calibration adjustments and perform long-time-scale temporal averaging of the measured temperatures to reduce the noise in the overlapping satellite measurements. This averaging procedure may attenuate the time dependent signal that the UAH empirical error model was introduced to explain. The large step in the $T_{2\text{-U.Md.}} - T_{2\text{-UAH}}$ difference time series that occurs in 1986 (see Figure 4.1a) suggests that uncertainty in the parameters for the NOAA-09 satellite are also important for this dataset[9]. The cause of the large fluctuations in the difference during the 2000-2004 time period is not known, but may be related to the absence of AMSU data in the $T_{2\text{-U.Md.}}$ dataset. Due to its relatively recent appearance, considerably less is known about the reasons for the differences between the Maryland dataset and the RSS and UAH datasets, thus the comments about these differences should be viewed as more speculative than the statements about the RSS-UAH differences.

[8] The Maryland group accounted for uncertainties in the radiometers non-linearity parameter as well as errors in the warm target radiation temperature (due to uncertainties in its emissivity and physical temperature) and errors in the cold space radiation temperature (due to uncertain antenna side lobe contributions for example). However, while all of these error sources are accounted for, they are assumed to be constant during the lifetime of a given instrument and thus do not take into account the possibility of contributions to the side lobe response from the earth or warm parts of the satellites whose temperature varies with time. These error sources lead, when globally averaged and linearized, to an expression where the target temperature is the most important factor. Thus while the exact physical cause of the observed effect is not known precisely, it is possible to accurately model and remove it on a global scale from the data using either method

[9] The trend in this difference time series is statistically significant at the 1% level.

These differences are an excellent example of structural uncertainty, where identical input data and three seemingly reasonable methodologies lead to trends that differ significantly more than the amount expected given their reported internal uncertainties. Since methodological differences yield data products showing differences in trends in T_2 of about 0.1 ºC per decade, it is clear that the most important source of uncertainty for satellite data are structural uncertainties and that these need to be included in any overall uncertainties assessed for tropospheric temperature trends and lapse rates.

3.2.3 Differences in spatial pattern.

Only $T_{2\text{-}UAH}$ and $T_{2\text{-}RSS}$ have provided gridded results. Maps of gridded trends for these products are shown in Fig 4.2, along with a map of the difference between the trends. The overall pattern in the trends is very similar between the two datasets, aside from difference in the globally averaged trends. Differences in the latitude dependence are due to the use of zonally varying intersatellite offsets in the construction of $T_{2\text{-}UAH}$ (in contrast to the constant offsets in $T_{2\text{-}RSS}$) and to differences in the applied diurnal adjustment as a function of latitude. Other differences may be caused by the spatial smoothing applied to the $T_{2\text{-}UAH}$ during the construction of the data set, and to differences in spatial averaging performed on the diurnal adjustment before it was applied. This last difference will be discussed in more detail in section 4.4 below because the effects are more obvious for the T_{2LT} layer.

Figure 4.2 Global maps of trends from 1979-2004 for (a) $T_{2\text{-UAH}}$ and (b) $T_{2\text{-RSS}}$. Except for an overall difference between the two results, the spatial patterns are very similar. A map of the difference $T_{2\text{-UAH}} - T_{2\text{-RSS}}$ between trends for the two products shown in (c) reveals more subtle differences in the trend.

4 Uncertainty in Lower Tropospheric Trends

4.1 Radiosonde Uncertainty

Uncertainties in lower tropospheric trends measured by radiosondes are very similar to those discussed above for the middle-upper troposphere. The most important difference is that when comparing to the T_{2LT} satellite product, the contribution of the stratospheric radiosonde trends, which is suspected to be erroneous to some extent, is substantially less than for the T_2 data records. This decreases the likelihood that T_{2LT} data products constructed from radiosonde data are biased toward excess cooling. However, it is possible that undetected negative trend bias remains in all tropospheric levels (see section 3.1 above for more details), so radiosonde trends may still be biased cold.

4.2 Satellite Uncertainty

Currently, there are two lower tropospheric satellite data records, $T_{2LT\text{-}UAH}$ and $T_{2LT\text{-}RSS}$. As discussed in the Preface, both datasets are relatively recent, thus little is known about the specific reasons for their differences. Because of the noise amplification effects of the differencing procedure[10] used to construct the data record, the merging parameters tend to be more sensitive to the methods used to deduce them. A number of different methods were explored in the creation of $T_{2LT\text{-}RSS}$, leading to an estimate of the structural uncertainty of 0.08ºC/decade for global trends. When combined with internal uncertainty, the estimated total global trend uncertainty for this dataset is 0.09ºC/decade (Mears and Wentz, 2005). Note that the difference between the global trends for $T_{2LT\text{-}RSS}$

[10] The T_{2LT} datasets are constructed by subtracting 3 times the average temperature measured by the outermost 4 (near-limb) views from 4 times the average temperature measured by the 4 adjacent views, which are closer to nadir. This has the effect of removing most of the stratospheric signal, and moving the effective weighting function lower in the troposphere (Spencer and Christy 1992). Assuming that the errors is each measurement are uncorrelated, this have the effect of amplifying these errors by a factor of about 5 relative to T_2 (Mears and Wentz 2005). Even if some of the error is correlated between view, this argument still applies to the uncorrelated portion of the error.

(0.189ºC/decade) and $T_{2LT\text{-}UAH}$ (0.115ºC/decade) shown in Table 3.3 is less than this estimated uncertainty. The estimated global trends in the radiosonde datasets are also within the $T_{2LT\text{-}RSS}$ error range. In Figure 4.1b we plot the difference ($T_{2LT\text{-}RSS} - T_{2LT\text{-}UAH}$) between the RSS and UAH time series. This time series shows more variability than the corresponding T_2 difference time series, making it more difficult to speculate about the underlying causes of the differences between them. The step-like feature during the 1985-1987 period is less obvious, and while there appears to be a slow drift during the NOAA-11 lifetime, a corresponding drift during the NOAA-14 lifetime is less obvious, perhaps because the RSS data do not yet include data from the more recent AMSU satellites. We speculate that the drift during NOAA-11 is in part due to differences in the diurnal correction applied. The UAH diurnal correction is based on a parameterization of the diurnal cycle which is constrained by measurements made during a time period with 3 co-orbiting satellites, , while RSS uses a model-based diurnal correction analogous to that used for TMT.

In Figure 4.3, we show global maps of the gridded trends for $T_{2LT\text{-}UAH}$ and $T_{2LT\text{-}RSS}$, along with a map of the trend differences. The spatial variability in the trend differences between the two datasets is much larger than the variability for T_2, though both datasets show similar patterns in general, with the greatest temperature increase occurring in the Northern Hemisphere, particularly over Eastern Asia, Europe, and Northern Canada. The two datasets are in relatively good agreement north of 45°N latitude. In the tropics and subtropics, the largest differences occur over land, particularly over arid regions.

Figure 4.3 Global maps of trends from 1979-2004 for (a) $T_{2LT-UAH}$ and (b) $T_{2LT-RSS}$. Except for an overall difference between the two results, the spatial patterns are similar. A map of the difference $T_{2LT-UAH} - T_{2LT-RSS}$ between trends for the two products shown in (c) shows that the largest differences are over tropical and subtropical land areas. Data from land areas with elevation higher than 2000m are excluded from the $T_{2LT-RSS}$ dataset and shown in white.

We speculate that this may be in part due to differences in how the diurnal adjustment is done by the two groups. The UAH group applies an averaged diurnal adjustment for each zonal band, based on different adjustments used for land and ocean. The RSS group uses a grid-point resolution diurnal correction. The UAH method may lead to errors for latitudes where the diurnal cycle varies strongly with longitude. More arid regions (e.g., subtropical Africa), which typically have much larger surface diurnal cycles, may be under-adjusted when the zonally averaged correction is applied, leading to long-term trends that are too low. Correspondingly more humid regions and oceans may be over-adjusted, in some cases making up for the overall difference between the two datasets, perhaps accounting for the good agreement in regions such as Southeast Asia, Southern India, and Northern South America. Further analysis is required using a range of alternative diurnal correction estimation techniques for definitive conclusions to be reached. Other differences, such the north-south streaking seen in the RSS data, may be caused by differences in spatial smoothing, and by the inclusion of AMSU data in $T_{2LT\text{-}UAH}$, but not in $T_{2LT\text{-}RSS}$.

The decay of orbital height over each satellite's lifetime can cause substantial errors in satellite-derived T_{TLT} because changes in height lead to changes in the earth incidence angles for the near-limb observations used to construct the data record Wentz and Schabel, 1998). Both the RSS and UAH groups correct for this error by calculating the expected change in observed temperature as a function of incidence angle, and then using this estimate to remove the effect of orbital decay. The straight-forward method used to make these corrections, combined with its insensitivity to assumptions about the vertical structure of the atmosphere, leads to the conclusion that errors due to orbital decay have been accurately removed from both datasets and are not an important cause of any differences between them.

4.3 Comparison between satellite and well characterized radiosonde stations

Point-by-point comparisons between radiosonde and satellite data eliminate many sources of sampling error normally present in radiosonde data. Also, since uniform global coverage is less important when using radiosondes to validate satellite data locally, stations can be chosen to minimize the contribution due to undocumented changes in radiosonde instrumentation or observing practice. For instance, if one restricts comparisons of the satellite and radiosonde data to 29 Northern Hemisphere radiosonde stations that have consistently used a single type of instrumentation (the Viz sonde) since 1979, the average difference between these radiosonde trends and $T_{2LT-UAH}$ trends since 1979-2004 is only 0.03ºC/decade (Christy et al., 2003; Christy et al., 2005). Similarly, when this set of radiosondes is extended to include a set of Southern Hemisphere stations where instrument changes were well documented, agreement between $T_{2LT-UAH}$ and radiosonde trends is almost as good (Christy and Norris, 2004; Christy et al., 2005). This suggests that, for the T_{2LT} layer, where the stratospheric problems with radiosonde data are minimized, some level of corroboration can be attained from these two diverse measurement systems.

<u>5 Multi-channel retrievals of tropospheric temperature.</u>

As mentioned above, the single channel statellite measurements commonly identified as tropospheric temperature (T_2) are impossible to interpret as solely tropospheric temperatures because 10% to 15% (seasonally and latitudinally varying) of the signal measured by MSU channel 2 arises from the stratosphere. In principle, it is possible to reduce the stratospheric contribution to Channel 2 by subtracting out a portion of the stratospheric Channel 4, though the exact values of the weights used in this procedure are controversial (see Chapter 2 for more details). Despite this controversy, there is little doubt that the resulting trends are more representative of the troposphere than the T_2 datasets. The reduction in stratospheric signal also reduces the difference between trends in the satellite data and the radiosonde data (see Table 3.3), because the error-prone stratospheric levels in the stratosphere have reduced (but still non-zero) weight.

The existence of a stratosphere-corrected tropospheric retrieval allows tests for consistency of temperature trends among the different datasets constructed by a research group for different atmospheric layers. One test, when applied to an earlier version (v5.1) of the UAH global average trends, did not prove inconsistency on the global scale, because the difference between the $T_{2LT\text{-}UAH}$ trend and the retrieval-calculated T_{2LT} trend was well within the published margin of error. However, a clearer inconsistency was found for the tropics (Fu and Johanson, 2005; Johanson and Fu, 2005). In this case, the difference between the retrieval-calculated trend and $T_{2LT\text{-}UAH}$ trend was larger than its estimated error range, an indication of uncharacterized error in at least one of the UAH products, or more generally that $T_{2LT\text{-}UAH}$, $T_{2\text{-}UAH}$ and $T_{4\text{-}UAH}$ were not strictly self-consistent as a set. This inconsistency is now resolved (within error estimates) with the introduction of a new version of the $T_{2LT\text{-}UAH}$ dataset. The RSS versions of the T_2, T_4 and T* datasets were found to be consistent for both global and tropical averages (Fu and

Johanson, 2005). The trends in the RSS version of the TLT dataset (produced after Fu and Johanson was submitted) is also consistent with the other RSS based datasets.

6 Uncertainty in Surface Trends.

6.1 Sea surface temperature uncertainty

Temperature analyses over the ocean are produced from sea surface temperatures (SST) instead of marine air temperatures. This is because marine air temperatures are biased from daytime ship deck heating (Folland and Parker, 1995; Rayner et al., 2003) and because satellite observations are available for SST beginning in November 1981 to augment *in situ* data (Reynolds and Smith, 1994). Spatially complete analyses of SSTs can be produced by combining satellite and *in situ* data (from ships and buoys) (Reynolds et al., 2002; Rayner et al., 2003), from *in situ* data alone (Smith and Reynolds, 2004), or from satellite data alone (Kilpatrick et al., 2001).

6.1.1 Satellite SST uncertainties

Climate comparison analyses based on infra-red satellite data alone are not useful because of possible large time-dependent biases. These biases have typically occurred near the end of a satellite's life time when the instrument no longer works properly, or during periods when assumptions made about the atmospheric profile in the satellite algorithm are no longer valid, e.g., during periods immediately following volcanic eruptions, when a large amount of dust from the eruption is present in the stratosphere (Reynolds, 1993; Reynolds et al., 2004). These problems may be partially mitigated in the future by use of the microwave SST sensors that became available starting with the launch of the Tropical Rainfall Measuring Mission (TRMM) in 1987 (Wentz et al., 2000), but these microwave SST data have not been available long enough to derive meaningful trends, and are difficult to calibrate absolutely due to various instrument related problems (Wentz et al., 2001; Gentemann et al., 2004). Thus, analyses now use multiple satellite instruments blended with or anchored to *in situ* data that reduce the overall analysis errors (e.g., Reynolds et al., 2002, Rayner et al., 2003).

6.1.2 In Situ SST uncertainties

As discussed in Chapter 2, the primary sources of uncertainty in *in situ* SST measurements are non-climatic signals caused by changes in the mix of instrumentation over time and sampling errors. Over time the measurements have typically evolved from insulated bucket measurements to engine intake, through hull, and buoy mounted sensors – these changes are not necessarily accurately recorded in the metadata Both non-climatic signals and sampling error are thought to be largest in sparsely sampled regions, such as the southern oceans, where a single erroneous or unrepresentative measurement could bias the average for an entire measurement cell for the month in question. Both types of errors have been calculated for the ERSST dataset and included in the quoted error range (see Figure 4.4).

Figure 4.4. SST, Land Surface Air Temperature, and the Combined Temperature Data Record anomaly averaged annually and between 60ºS and 60ºN (purple), with its estimated 95% confidence intervals (dashed). Data are from the NOAA GHCN-ERSST dataset (Smith and Reynolds 2005).

6.2 Land surface air temperature uncertainty

The three surface temperature analyses exhibit similar warming rates since 1958. As the surface data sets have many stations in common, they are not totally independent.

However, the MSU series take identical input, and radiosonde datasets have common data also, so this problem is not unique to the surface records. The fact that the range in trends is much smaller for the surface datasets than for these other datasets implies that the structural uncertainty arising from dataset construction choices is much smaller at the surface, in agreement with the arguments made in Thorne et al. (2005a). Also, a number of studies e.g., (Peterson et al., 1999; Vose et al., 2004) suggest that long-term, large-scale trends are not particularly sensitive to variations in choice of station networks. But because most land networks were not designed for climate monitoring, the data contain biases that dataset creators address with different detailed methods of analysis. The primary sources of uncertainty from a land-surface perspective are (a) the construction methods used in the analyses and (b) local environmental changes around individual observing stations that may not have been addressed by the homogeneity assessments.

Because the stations are not fully representative of varying-within-area land surface, coastal, and topographical effects, global data sets are produced by analyzing deviations of temperature from station averages (anomalies) as these deviations vary more slowly with a change in location than the temperatures themselves (Jones et al., 1997). Random errors in inhomogeneity detection and adjustments may result in biased trend analyses on a grid box level. However, on the relatively large space scales of greatest importance to this Report, such problems are unlikely to be significant in current data sets in the period since 1958 except where data gaps are still serious, e.g., in parts of central Africa, central South America, and over parts of Antarctica. Note that for the contiguous United States, the period 1958-2004 uses the greatest number of stations per grid box anywhere on the

Earth's land surface, generally upwards of 20 stations per grid box. For regions with either poor coverage or data gaps, trends in surface air temperature should be regarded with considerable caution, but do not have serious effects on the largest of scales as most of the spatial variability is well sampled.

Local micro-climatological environmental changes around observing stations may be problematic, particularly if a similar change occurred near many observing stations (e.g., Davey and Pielke, 2005). For instance, urbanization may have increased temperatures in many locations. Numerous investigators have used a variety of approaches to study these effects and most have shown that any bias is likely to be small in comparison to the warming signal for large-scale means (e.g., Peterson, 2003; Parker, 2004). To insure that potential urbanization effects do not impact analyses, NASA adjusts the data from all urban stations so that their long-term trends are consistent with those from neighbouring rural stations (Hansen et al., 2001). It is generally accepted that local biases in trends mostly cancel through the use of many stations or ocean observations. Because such a cancellation has not been rigorously proved, partly due to the lack of adequate metadata, it is conceivable that systematic changes in many station exposures of a similar kind may exist over the land during the last few decades, which may give biases in trends of one sign over large land regions.

6.3 Combined land-ocean analyses uncertainty.

Global combined surface temperature products are computed by combining ocean and land gridded datasets. The latest version of the UK surface dataset, HadCRUT2v, (Jones and Moberg, 2003) has been optimally averaged with uncertainties for the globe and hemispheres. The NOAA surface temperature dataset produced by (Smith and Reynolds, 2005), uses Global Historical Climatology Network (GHCN), merged with the *in situ* Extended Reconstruction SST (ERSST) analysis of (Smith and Reynolds, 2004). The analyses are done separately over the ocean and the land following the ERSST methods. Error estimates include the bias, random and sampling errors.

As an example of uncertainties in a combined land-ocean analysis, near-global time series (60^{o}S to 60^{o}N) are shown in Figure 4.4 for SST, land-surface air temperature, and the combined SST and land-surface air temperature (Smith and Reynolds, 2005). (The combined product is the GHCN-ERSST product used in Chapter 3). The SST has the tightest (95%) uncertainty limits (upper panel). The land-surface air temperature (middle panel) has a larger trend over the period since 1958, but its uncertainty limits are also larger than for SST. Land surface air temperature uncertainty is larger than the uncertainty for SST because of higher variability of surface air temperature over land (see Chapter 1), persistently un-sampled regions, including central Africa and interior South America, and because the calculations include an increasing urbanization bias-error estimate. Merged temperature anomalies and their uncertainty (lower panel) closely resemble the SST result, since oceans cover most of the surface area. Similar uncertainty was found by (Folland et al., 2001) using different methods.

7. Interlayer comparisons.

7.1 Troposphere/Stratosphere

All data sources agree that on a global scale, the stratosphere has cooled substantially while the troposphere has warmed over both the 1958-2004 and the 1979-2004 time periods (note that this is not true for all 25-year time periods within the longer 1958-2004 time period). We suspect that the stratospheric cooling trends estimated from radiosondes are larger in magnitude than the actual trend. Despite the uncertainty in the exact magnitude of stratospheric cooling, we have very high confidence that the lower stratosphere has cooled relative to the troposphere by several tenths of a ºC per decade over the past 5 decades.

7.2 Lower Troposphere/Mid-Upper Troposphere

The difference in trend between the lower troposphere and mid-upper troposphere is not well characterized by the existing data. On a global scale, all data sets suggest that T_{2LT} is warming relative to T_2, but it is important to note that the T_2 data records have significant stratospheric contributions that reduce their warming trends. Radiosonde measurements suggest that the $T_{(850\text{-}300)}$ layer (which does not include the stratosphere) is warming at about the same rate as T_{2LT}, while satellite data suggest that T^*_G is warming more rapidly than T_{2LT}. The magnitude of these inter-dataset differences are typically less than their individual estimates of uncertainty, substantially reducing confidence in our ability to deduce even the sign of the lower troposphere-mid-upper troposphere trend difference.

7.3 Surface/Lower Troposphere

On a global scale, both radiosonde datasets and one of the satellite datasets ($T_{2LT\text{-}UAH}$) indicate that the surface warmed more than the lower troposphere between 1979 and 2004, while one satellite dataset ($T_{2LT\text{-}RSS}$) suggests the opposite. The magnitude of these differences is less than the uncertainty estimates for any one data record. The situation is similar in the tropics. However, in some regions, such as North America and Europe (regions where the most reliable radiosonde stations are located), the warming in the surface and lower troposphere appears to be very similar in all datasets.

7.4 Surface/Mid Troposphere

It is also interesting to consider the trend differences between the surface and mid troposphere since more satellite datasets are available for T_2. Here, mostly due to the large structural uncertainty in the trends in T_2, the various datasets are unable to agree on the sign of the trend difference over the 1979-2004 period. On a global scale, the two radiosonde datasets and two of the satellite datasets suggest that T_2 has warmed less than the surface, but the other satellite dataset suggests that the opposite is true. The situation is similar in the tropics. For the longer 1958-2004 period, all available datasets agree that T_2 warms more than the surface. When T^*_G is considered, the difference between the surface and tropospheric trends is reduced, with two satellite datasets indicating more warming than at the surface.

8. Resolution of Uncertainty

In almost all of the tropospheric and stratospheric data records considered, our uncertainty is dominated by structural uncertainty arising through dataset construction choices (Thorne et al., 2005). Differences arising as a result of different, seemingly plausible correction models applied by different groups to create a climate-quality data record are significantly larger than the uncertainties internal to each method, in the raw data measurements, or in the sampling uncertainties. These structural uncertainties are difficult to assess in an absolute sense. The best estimates we can currently make come from examining the spread of results obtained by different groups analyzing the same type of data. This "all datasets are equal" approach has been employed in our present analysis. As outlined in Chapter 2, this estimate of uncertainty can either be too small or too large, depending on the situation. Given this caveat, it is always better to have multiple (preferably at least three) data records that purport to measure the same aspects of climate with the same data, so we can get some idea of the structural uncertainty.

In reality, all datasets are not equally plausible realizations of the true climate system evolution. The climate system has evolved in a single way, and some datasets will be closer to this truth than others. Given that the importance of structural uncertainty, particularly for trends aloft, has only recently been recognized, it is perhaps unsurprising that we are unable to quantify this at present. We could make value-based judgments to imply increased confidence in certain datasets, but these would not be unambiguous, may eventually be proven wrong, and are not a tenable approach in the longer term from a scientific perspective. Therefore tools need to be developed to objectively discriminate between datasets. These may include (1) measures of the internal consistency of the construction methods, (2) assessment of the physical plausibility of the merged products, including consistency of vertically resolved trends, and (3) comparisons with vicarious data – for example, changes in temperature need to be compared with changes in water vapor, winds, clouds, and various measures of radiation to assess consistency with the expected physical relationships between these variables. Taken together such a suite of indicators can be used to provide an objectively based way of highlighting residual problems in the datasets and gaining a closer estimate of the truth. Such an audit of current datasets should be seen as very high priority and preferably undertaken independently of the dataset builders in a similar manner to the model intercomparisons performed at Lawrence Livermore National Laboratory. In addition to an agreed set of objective analysis tools, such an effort requires full and open access to all of the datasets including a full audit trail.

Some specific suggestions for resolving some of the issues brought forward in this chapter are mentioned here, but these are not exhaustive and further investigation is required.

8.1 Radiosondes.

A significant contribution to the long-term inhomogeneity of the radiosonde record appears to be related to changes in radiative heating of the temperature sensor for various radiosonde models, and changes in the adjustments made to attempt to correct for these changes. Recent work suggests that such problems may account for much of the tropical cooling shown in unadjusted data. Other recent work suggests that step-like changes in bias may still remain, even in adjusted datasets. Suitable tests on radiosonde products may therefore include: stability of day-night differences, spatial consistency, internal consistency (perhaps including wind data that to date have not been incorporated), and consistency with MSU-derived and other independent estimates.

8.2 Satellites.

The most important contributions to satellite uncertainty are merging methodology and the diurnal adjustment. The satellite data are simple enough that considerable understanding can result from examination of intermediate results in the merging process, including intersatellite differences that remain after the merging adjustments are complete. Consistent reporting of such results can help differentiate between methods. It appears that the differences in merging methodology often result in sharp step-like features in difference time series between datasets. Other datasets, such as spatially averaged adjusted radiosonde data, might be expected to show more slowly changing errors, since their errors are due to the overlap of many different, potentially step-like errors that occur at different times. So comparisons of satellite data with radiosonde data over short time periods may help differentiate between satellite datasets. The diurnal adjustment can be improved by a more rigorous validation of model-derived diurnal cycles, or by further characterization of the diurnal cycle using the TRMM satellite or concerted radiosonde observing programs designed to characterize the diurnal cycle at a number of representative locations.

8.3 Surface.

The uncertainty in the historical near-surface temperature data is dominated by sampling uncertainty, systematic changes in the local environment of surface observing stations, and by difficult-to-characterize biases due to changes in SST measurement methods. The relative maturity of the surface datasets suggest that to a large degree, these problems have been addressed to the extent possible for the historical data, due to the absence of the required metadata (for the bias-induced uncertainties) or the existence of any observations at all. However, it is likely that much of the relatively recent SST data can be adjusted for measurement type as some of the needed metadata is available or can be estimated.

References

Agudelo, P. A. and J. A. Curry, 2004: Analysis of spatial distribution in tropospheric temperature trends. *Geophysical Research Letters,* **31**: L22207.

Christy, J. R. and W. B. Norris, 2004: What may we conclude about global tropospheric temperature trends? *Geophysical Research Letters*, **31**: L06211.

Christy, J. R., R. W. Spencer, et al., 2000: MSU Tropospheric Temperatures: Dataset Construction and Radiosonde Comparisons. *Journal of Atmospheric and Oceanic Technology*, **17**(9): 1153-1170.

Christy, J. R., R. W. Spencer, W. Norris, W. Braswell, D. Parker, 2003: Error Estimates of Version 5.0 of MSU/AMSU Bulk Atmospheric Temperatures. *Journal of Atmospheric and Oceanic Technology*, **20**: 613-629.

Dai, A. and K. E. Trenberth, 2004: The Diurnal Cycle and Its Depiction in the Community Climate System Model. *Journal of Climate* **17**: 930-951.

Davey, C. A. and R. A. Pielke, 2005: Microclimate exposures of surface-based weather stations - implications for the assessment of long-term temperature trends. *Bull. Amer. Meteor. Soc.* **86**(4): 497-504.

Durre, I., T. C. Peterson, R. Vose, 2002: Evaluation of the Effect of the Luers-Eskridge Radiation Adjustments on Radiosonde Temperature Homogeneity. *Journal of Climate* **15**(6): 1335-1347.

Folland, C. K. and D. E. Parker, 1995: Correction of instrumental biases in historical sea surface temperature data. *Q. J. Roy. Meteor. Soc.* **121**: 319-367.

Folland, C. K., N. A. Rayner, S. J. Brown,T. M. Smith,S. S. P. Shen, D. E. Parker, I. Macadam, P. D. Jones, R. N. Jones,N. Nicholls,D. M. H. Sexton, 2001: Global temperature change and its uncertainties since 1861. *Geophysical Research Letters* **28**(13): 2621-2624.

Free, M., I. Durre,E. Aguilar, D. Seidel,T.C. Peterson, R.E. Eskridge, J.K. Luers, D. Parker, M. Gordon, J. Lanzante, S. Klein, J. Christy, S. Schroeder, B. Soden, L.M. McMillin, 2002: Creating climate reference datasets: CARDS workshop on adjusting radiosonde temperature data for climate monitoring. Bulletin of the. American Meteorological. Society **83**: 891-899.

Free, M. and D. Seidel, 2004: Accounting for Differences between Radiosonde Upper-air Temperature Datasets. Under review.

Fu, Q. and C. M. Johanson, 2004: Stratospheric influences on MSU-derived tropospheric temperature trends: A direct error analysis. *Journal of Climate* **17**: 4636-4640.

Fu, Q. and C. M. Johanson, 2005: Satellite-Derived Vertical Dependence of Tropospheric Temperature Trends. *Geophysical Research Letters* **32**: L10703.

Fu, Q., C. M. Johanson, F. Warren, D. Seidel, 2004: Contribution of stratospheric cooling to satellite-inferred tropospheric temperature trends. *Nature* **429**: 55-58.

Gentemann, C. L., F. J. Wentz, C. Mears, D. Smith, 2004: In situ validation of Tropical Rainfall Measuring Mission microwave sea surface temperatures. *Journal of Geophysical Research* **109**: C04021.

Gillett, N. P., B. D. Santer, A. Weaver, 2004: Quanitfying the influence of stratospheric cooling on satellite-derived tropospheric temperature trends. *Nature* **432**: Published online December 2, 2004 (doi: 10.1038/nature03209).

Grody, N. C., K. Y. Vinnikov, M. Goldberg, J.. Sullivan, J. Tarpley, 2004: Calibration of Multi-Satellite Observations for Climatic Studies: Microwave Sounding Unit (MSU). *Journal of Geophysical Research* **109**.

Hansen, J. E., R. Ruedy,M. Sato, M. Imhoff, W. Lawrence, D. Easterling,T. Peterson,T. Karl, 2001: A Closer look at United States and global surface temperature change. *Journal of Geophysical Research* **106**: 23947-23963.

Jones, P. D. and A. Moberg, 2003: Hemispheric and Large-Scale Surface Air Temperature Variations: An Extensive Revision and an Update to 2001. *Journal of Climate* **16**: 206–223.

Jones, P. D., T. J. Osborn, K. Briffa, 1997: Estimating Sampling Errors in Large-Scale Temperature Averages. *Journal of Climate* **10**(10): 2548–2568.

Kilpatrick, K. A., G. P. Podesta, R. Evans, 2001: Overview of the NOAA/NASA advanced very high resolution radiometer pathfinder algorithm for sea surface temperature and associated matchup database. .*Journal of Geophysical Research* **106**(C5): 9179-9198.

Lanzante, J., S. Klein, D. Seidel, 2003: Temporal Homogenization of Monthly Radiosonde Temperature Data. Part II: Trends, Sensitivities, and MSU comparison. *Journal of Climate* **16**: 241-262.

Luers, J. K. and R. E. Eskridge, 1998: Use of Radiosonde Temperature Data in Climate Studies. *Journal of Climate* **11**: 1002-1019.

Mears, C. and F. Wentz, 2005: The Effect of Drifting Measurement Time on Satellite-Derived Lower Tropospheric Temperature. *Science* **309**: 1548-1551.

Mears, C. A., M. C. Schabel, F. W. Wentz, 2003: A reanalysis of the MSU channel 2 tropospheric temperature record. *Journal of Climate* **16**: 3650-3664.

Parker, D., 2004: Large scale warming is not urban. *Nature* **432**: 290.

Parker, D. E., M. Gordon, D. P. N Cullum, D. M. H. Sexto, C. K. Folland, N. Rayner, (1997: A new global gridded radiosonde temperature data base and recent temperature trends. *Geophysical Research Letters* **24**: 1499-1502.

Pathack, B., J. Nash, R. Smout, and S. Kurnosenko, 2005: Preliminary Results of WMO Intercomparison of high quality radiosonde systems, Mauritius, February 2005. www.wmo.ch/web/www/IMOP/meetings/Upper-Air/RSO-ComparisonMauritius/3(16)_UK_Nash_Vacoas.pdf

Peterson, T. C., 2003: Assessment of Urban Versus Rural In Situ Surface Temperatures in the Contiguous United States: No Difference Found. *Journal of Climate* **16**(18): 2941-2959.

Peterson, T. C., K. P. Gallo, J. Lavrimore, T. W. Owen, A. Huang, D.A. McKittrick, (1999: Global rural temperature trends. *Geophysical Research Letters* **26**(3): 329-332.

Randel, W. J. and F. Wu, 2005: Biases in stratospheric temperature trends derived from historical radiosonde data. *Journal of Climate* (in press).

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century.*Journal of Geophysical Research* **108**(d14).

Reynolds, R. W., 1993: Impact of Mt Pinatubo aerosols on satellite-derived sea surface temperatures. *Journal of Climate* **6**: 768-774.

Reynolds, R. W., C. L. Gentemann, F. W. Wentz, 2004: Impact of TRMM SSTs on a Climate-Scale SST Analysis. *Journal of Climate* **17**: 2938-2952.

Reynolds, R. W., N. A. Rayner, et al., 2002: An Improved In Situ and Satellite SST Analysis for Climate. *Journal of Climate* **15**: 1609-1625.

Reynolds, R. W. and T. M. Smith, 1994: Improved global sea surface temperature analyses using optimum interpolation. *Journal of Climate* **7**: 929-948.

Sherwood, S., J. Lanzante, Cathryn Meyer, 2005: Radiosonde daytime biases and late 20th century warming. Science **309**: 1556-1559.

da Silveira, R. B., G. Fisch, L Machato, A.Dall'Antonia, L Sapuchi, D. Fernandes, J. Nash, 2003: *Executive Summary of the WMO Intercomparison of GPS Radiosondes*. Geneva, World Meteorological Organization**:** 17.

Smith, T. M. and R. W. Reynolds, 2004: Improved Extended Reconstruction of SST (1854-1997). *Journal of Climate* **17**: 2466-2477.

Smith, T. M. and R. W. Reynolds, 2005: A global merged land and sea surface temperature reconstruction based on historical observations (1880-1997). *Journal of Climate* **18**(12): 2021-2036.

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. *Journal of Climate* **5:** 858-866.

Spencer, R.J., W.D. Braswell and J.R. Christy, 2005: Document describing v5.2 lower tropospheric retrieval.

Tett, S. and P. Thorne, 2004: Atmospheric Science: Tropospheric temperature series from satellites. *Nature* **432** .

Thorne, P. W., D. E. Parker, et al., 2005: Causes of differences in observed climate trends. *Bulletin of the American Meteorological Society.* (in press).

Thorne, P. W., D. E. Parker, S. F. B. Tett, P. D. Jones, M. McCarthy, H. Coleman, P. Brohan, 2005: Revisiting radiosonde upper-air temperatures from 1958 to 2002. *Journal of Geophysical Research.* (in press).

Vinnikov, K. Y. and N. C. Grody, 2003: Global Warming Trend of Mean Tropospheric Temperature Observed by Satellites. *Science* **302**: 269-272.

Vinnikov, K. Y., N. C. Grody, et al., 2005: Temperature Trends at the Surface and in the Troposphere. *Journal of Geophysical Research* (submitted).

Vose, R. S., D. Wuertz, T. C. Peterson, P. D. Jones, 2004: An intercomparison of land surface air temperature analyses at the global, hemispheric, and grid-box scale. *Geophysical Research Letters* **32**: L18718.

Wallis, T. W. R., 1998: A subset of core station from the Comprehensive Aerological Reference Dataset (CARDS). *Journal of Climate* **11**: 272-282.

Wentz, F., M. Schabel, 1998: Effects of satellite orbital decay on MSU lower tropospheric temperature trends. *Nature* **394**: 661-664.

Wentz, F., C. Gentemann, D. Smith, D. Chelton, 2000: Satellite measurements of sea surface temperature through clouds. *Science* **288**: 847-850.

Wentz, F. J., P. D. Ashcroft, C. Gentemann, 2001: Post-launch calibration of the TMI microwave radiometer. *IEEE Transactions on Geoscience and Remote Sensing* **39**: 415-422.