Chapter 2

What kinds of atmospheric temperature variations can the current observing systems detect and what are their strengths and limitations, both spatially and temporally?

Convening Lead Author: John Christy

Lead Authors: Dian Seidel, Steve Sherwood

Contributing Authors: Adrian Simmons, Ming Cai, Eugenia Kalnay, Chris Folland, Carl Mears, Peter Thorne, John Lanzante

Findings and Recommendations

- The observing systems available for this report are able to detect small surface and upper air temperature variations from year to year, for example, those caused by El Niño or volcanic eruptions.

- The data from these systems also have the potential to provide accurate trends in climate over the last few decades (and over the last century for surface observations), once the raw data are successfully adjusted for changes over time in observing systems, practices, and micro-climate exposure to produce usable climate records. Measurements from all systems require such adjustments and this report relies on adjusted datasets.

- Adjustments to the land surface temperature record have been sufficiently successful that trends are reasonably similar on large (e.g., continental) scales, despite the fact that spatial sampling is uneven and some errors undoubtedly remain. This conclusion holds to a lesser extent for the ocean surface record, which suffers from more serious sampling problems and changes in observing practice.

- Adjustments for changing instrumentation are most challenging for upper-air datasets. While these show promise for trend analysis, it is likely that current upper-air climate records give reliable indications of directions of change (e.g. warming of troposphere, cooling of stratosphere) but some questions remain regarding the precision of the measurements.
    - Upper-air datasets have been subjected to less scrutiny than surface datasets.

- Adjustments are complicated, sometimes as large as the trend itself, involve expert judgments, and cannot be stringently evaluated because of lack of traceable standards.
- Unlike surface trends, reported upper-air trends vary considerably between research teams beginning with the same raw data owing to their different decisions on how to remove non-climatic factors.
- The diurnal cycle, which must be factored into some adjustments for satellite data, is well observed only by surface observing systems.
- No available observing system has reference stations or multi-sensor instrumentation that would provide stable calibration over time.
- Most observing systems have not retained complete metadata describing changes in observing practices which could be used to identify and characterize non-climatic influences.

- Relevant satellite datasets measure broad vertical layers and cannot reveal the detailed vertical structure of temperature changes, nor can they completely isolate the troposphere from the stratosphere. However, retrieval techniques can be used both to approximately isolate these layers and to check for vertical consistency of trend patterns. Consistency between satellite and radiosonde data can be tested by proportionately averaging radiosonde profiles.

- Reanalyses and other multi-system products have the potential for addressing issues of surface and atmospheric temperature trends by making better use of available information and allowing analysis of a more comprehensive, internally consistent, and spatially and temporally complete set of climate variables. At present, however, they contain biases, especially in the stratosphere, that affect trends and that cannot be readily removed because of the complexity of the data products.

- There are as yet under-exploited data archives with potential to contribute to our understanding of past changes, and new observing systems that may improve estimates of future changes if designed for long-term measurement stability and operated for sufficient periods.

Recommendation: Current and future observing systems should adhere to the principles for climate observations adopted internationally under the Framework Convention on Climate Change and documented in "NRC 2000b" and the "Strategic Plan for the U.S. Climate Change Science Program (2003)" to significantly mitigate the limitations listed above.

Recommendation: The ability to fully and accurately observe the diurnal cycle should be an important consideration in the design and implementation of new observing systems.

Recommendation: When undertaking efforts to retrieve data it is important to also to collect detailed metadata which could be used to reduce ambiguity in the timing, sign and magnitude of non-climatic influences in the data.

Recommendation: New climate-quality reanalysis efforts should be strongly encouraged and specifically designed to minimize small, time-dependent biases arising from imperfections in both data and forecast models.

Recommendation: Some largely overlooked satellite datasets should be reexamined to try to extend, fortify or corroborate existing microwave-based temperature records for climate research, e.g. microwave data from NEMS (1972) and SCAMS (1975), infrared from the HIRS suite and radio occultation from GPS.

1. MAIN OBSERVING SYSTEMS AND SYNTHESIS DATA PRODUCTS

Temperature is measured in three main ways; (1) *in situ,* where the sensor is immersed in the substance of interest; (2) by *radiative emission*, where a remote sensor detects the intensity or brightness of the radiation emanating from the substance; and (3) *radiative transmission*, where radiation is modified as it passes through the substance in a manner determined by the substance's temperature. All observations contain some level of random measurement error, which is reduced by averaging; bias, which is not reduced by averaging; and sampling errors (see Appendix).

a) Surface and near-surface air temperatures

Over land, “near-surface” air temperatures are those commonly measured about 1.5 to 2.0 meters above the ground level at official weather stations, at sites run for a variety of scientific purposes, and by volunteer (cooperative) observers (e.g., Jones and Moberg, 2003). These stations often experience relocations, changes in instrumentation and/or exposure and changing observing practices all of which can introduce biases into their long-term records. These changes are often undocumented.

“Near-surface” air temperatures over the ocean (“Marine Air Temperatures” or MATs) are measured by ships and buoys at various heights from 2 to more than 25m, with poorer coverage than over land (e.g., Rayner et al., 2003). To avoid the contamination of daytime solar heating of the ships’ surfaces that may affect the MAT, it is generally preferred to limit these to night MAT (NMAT) readings only. Observations of the water temperature near the ocean surface or “Sea Surface Temperatures” (SSTs) are widely used and are closely tied to MATs; ships and buoys measure SSTs within a few meters below the surface.

Incomplete geographic sampling, changing measurement methods, and land-use changes all introduce errors into surface temperature compilations. The spatial coverage, indicated in Figure 2.1, is far from uniform over either land or ocean areas. The southern oceans, polar regions and interiors of Brazil and Africa are not well sampled by in-situ networks. However, creating global surface temperature analyses involves not only merging land and ocean data but also considering how best to represent areas where there are few or no observations. The most conservative approach is to use only those grid boxes with data, thus avoiding any error associated with interpolation. Unfortunately, the areas without data are not evenly or randomly distributed around the world, leading to considerable uncertainties in the analysis, though it is possible to make an estimate of these uncertainties. Using the conservative approach, the tropical land surface areas would be under-represented, as would the southern ocean. Therefore, techniques have been developed to interpolate data to some extent into surrounding data-void regions. A single group may produce several different such datasets for different purposes. The choice may depend on whether the interest is a particular local region, the entire globe, or use of the dataset with climate models (Chapter 5). Estimates of global and hemispheric scale averages of near-surface temperatures generally begin around 1860 over both land and ocean.

Figure 2.1 Top: Location of radiosonde stations used in the HadAT upper air dataset with those also in the LKS as crosses. Bottom: Distribution of land stations (green) and SST observations (blue) reporting temperatures used in the surface temperature datasets over the period 1979-2004. Darker colors represent locations for which data were reported with greater frequency.
See chapter 3 for definitions of datasets.

Datasets of near-surface land and ocean temperatures have traditionally been derived from *in-situ* thermometers. With the advent of satellites, some datasets now combine both *in-situ* and remotely sensed data (Reynolds et al., 2002; Smith and Reynolds, 2005), or use exclusively remotely sensed data (Kilpatrick et al., 2001) to produce more geographically complete distributions of surface temperature. Because the satellite sensors measure infrared or microwave emission from the earth's surface (a "skin" typically tens of microns thick that may have a temperature different from either the air above or material at greater depths), calculations are required to convert the skin temperature into the more traditional near-surface air or SST observation (in this context SSTs are called "bulk sea surface temperatures", Chelton, 2005.) Typically, in-situ observations are taken as "truth" and satellite estimates (which may be affected by water vapor, clouds, volcanic aerosols, etc.) are adjusted to agree with them (Reynolds, 1993.) With continued research, datasets with surface temperatures over land, ice, and ocean from infrared and microwave sensors should provide expanded coverage of surface temperature variations (e.g., Aires et al., 2004).

Sampling errors in ship and buoy SST data typically contribute more to large-scale averages than random measurement errors as shown in Smith and Reynolds (2004), especially as the temperature record extends backward in time. Biases depend on observing method. Most ship observations since the 1950s were made from insulated buckets, hull contact sensors, and engine intake temperatures at depths of one to several meters. Historic correction of ship data prior to 1942 is discussed in (Folland and Parker, 1995) and bias and random errors from ships are summarized by (Kent and Taylor, 2004) and (Kent and Challenor, 2004). They report that engine intake temperatures are typically biased 0.1-0.2ºC warmer than insulated buckets. This is primarily due to engine room heating of the water temperatures although there is also evaporative cooling of the water in the insulated buckets. Hull contact sensors are the most accurate though much less common. The bias correction of the ship SST data (Kent and Kaplan, 2004) requires information on the type of measurement (e.g. insulated bucket, etc.) which becomes more difficult to determine prior to 1990s due to incomplete documentation. Kent and Kaplan (2005) also found that insulated bucket temperatures may be to cold by 0.12 to 0.16ºC. When the bucket bias is used, engine intake temperatures in the mid-to-late 1970s and the 1980s were found to be smaller than that suggested by previous studies, ranging from 0.09 to 0.18ºC. In addition, their study indicates that engine intake SSTs may have a cold bias of -0.13ºC in the early 1990s. The reliability of these biases are subject to revision due to small sample sizes that sample sizes for these comparisons tend to be small with large random errors. Buoy observations became more plentiful following the start of the Tropical Ocean Global Atmosphere (TOGA) Program (McPhaden, 1995) in 1985. These observations are typically made by an immersed temperature sensor or a hull contact sensor, and are more accurate because they do not have the bias errors of ship injection or insulated bucket temperatures.

The global surface air temperature data sets used in this report are to a large extent based on data readily exchanged internationally, e.g., through CLIMAT reports and the WMO publication *Monthly Climatic Data for the World*. Commercial and other considerations prevent a fuller exchange, though the United States may be better represented than many other areas. In this report we present three global surface climate records, created from available data by NASA Goddard Institute for Space Studies (GISS), NOAA National Climatic Data Center

(NOAA/NCDC) and the cooperative project of the U.K. Hadley Centre and the Climate Research Unit of the University of East Anglia (HadCRUT2v). These will be identified as $T_{Sfc\text{-}G}$, $T_{Sfc\text{-}N}$ and $T_{Sfc\text{-}U}$ respectively.

b) Atmospheric "upper air" temperatures

1. Radiosondes

Radiosonde or balloon-based observations of atmospheric temperature are *in-situ* measurements as the thermometer (often a thermistor or a capacitance-based sensor), suspended from a balloon, is physically carried through the atmospheric column. Readings are radio-transmitted back to a data recorder. Balloons are released once or twice a day (00 and/or 12 Coordinated Universal Time or UTC) at about 1,000 stations around the globe, many of which began operations in the late 1950s or 1960s. These sites are unevenly distributed, with only the extratropical northern hemisphere land areas and the Western Pacific Ocean/Indonesia/Australia region being well-sampled in space and time. Useful temperature data can be collected from near the surface through the lower and middle stratosphere (though not all balloons survive to these heights). Radiosonde data in the first hundred meters or so above the surface are sometimes erroneous if the sensors have not been allowed to reach equilibrium with the atmosphere before launch, and may not be representative of regional conditions, due to microclimatic and terrain effects.

Although most radiosonde data are transmitted to meteorological centers around the world and archived, in practice many soundings do not reach this system and are collected later. No definitive archive of radiosonde data exists, but several archives in the U.S. and abroad contain nearly complete collections, though several different schemes have been employed for quality control. To monitor climate, it is desirable to have a long, continuous record of measurements from many well-distributed fixed sites. There are about 700 radiosonde stations that have operated in the same location for at least three decades; many of these are clustered in a few areas, further reducing the effective coverage (Figure 2.1). Thus, a dilemma exists for estimating long-term changes: whether to use a smaller number of stations having long segments of continuous records, or a larger number of stations with shorter records that do not always overlap well. Various analysis groups have approached this differently (see Chapters 3 and 4).

Typically, radiosonde-based datasets are developed for specific atmospheric pressure surfaces known as “mandatory reporting levels” (Figure 2.2). Such data at discrete vertical levels provide unique information for assessing changes in the structure of the atmosphere. Two such datasets are featured in this report, The Hadley Centre Atmospheric Temperatures from the U.K. (HadAT) and Radiosonde Atmospheric Temperatures Products for Assessing Climate (RATPAC) from NOAA. A product such as $T_{850\text{-}300}$, for example, will be identified as $T_{850\text{-}300\text{-}U}$ and $T_{850\text{-}300\text{-}N}$ for HadAT and RATPAC respectively. [1]

---

[1] A third radiosonde dataset was generated by comparing radiosonde observations against the first-guess field of the ERA-40 simulation forecast model (Haimberger, 2004). Adjustments were applied when the relative difference between the radiosonde temperatures and the forecast temperatures changed by a significant amount. The data were not yet in final form for consideration in this report, although the tropospheric values appear to have general agreement with HadAT and RATPAC

Figure 2.2 Terminology and vertical profiles for the temperature products referred to in this report. Radiosonde-based layer temperatures ($T_{850\text{-}300}$, $T_{100\text{-}50}$) are height-weighted averages of the temperature in those layers. Satellite-based temperatures (T_{2LT}, T_2, and T_4) are mass-weighted averages with varying influence in the vertical as depicted by the curved profiles, i.e., the larger the value at a specific level, the more that level contributes to the overall satellite temperature average. The subscript simply indicates the layer where 90% of the information for the satellite average originates.

Notes: (1) because radiosondes measure the temperature at discrete (mandatory) levels, their information may be used to create a temperature value that mimics a satellite temperature (Text Box 2.1), (2) layer temperatures vary from equator to pole so the pressure and altitude relationship here is based on the atmospheric structure over the conterminous U.S., (3) about 10% (5%) of the value of T_{2LT} (T_2) is determined by the surface character and temperature, (4) T^*_T and T^*_G are simple retrievals, being linear combinations of 2 channels, T_2 and T_4.

Throughout the radiosonde era there have been numerous changes in stations, types of instrumentation, and data processing methods that can create data discontinuities. Because radiosondes are expendable instruments, instruments are more easily changed than for the more permanent surface sites. The largest discontinuities appear to be related to solar heating of the temperature sensor and changes in design and/or data adjustments intended to deal with this problem. These discontinuities have greatest impact at stratospheric levels (the stratosphere's lower boundary is ~16 km in the tropics, dropping to < 10 km in the high latitudes, Figure 2.2), where direct sunlight can cause radiosonde-measured temperatures to rise several °C above ambient temperatures. For example, when Australia and U.S. stations changed instrumentation to Vaisala RS-80, processed stratospheric temperatures shifted downward by 1 to 3°C (Parker et al., 1997, Christy et al., 2003). Many other sources of system-dependent bias exist (which often affect the day and night releases differently), including icing of the sensors in regions of super-cooled water, software errors in some radiosonde systems, poor calibration formulae, and operator errors. Documentation of these many changes is limited, especially in the earlier decades.

2. Passive Satellite Instrumentation

Unlike radiosondes, passive satellite observations of microwave and infrared brightness temperatures sample thick atmospheric layers (and may include surface emissions), depicted as weighting functions in Figure 2.2. These measurements may be thought of as bulk atmospheric temperatures, as a single value describes the entire layer. Although this bulk measurement is less informative than the detailed information from a radiosonde, horizontal coverage is far superior, and consistency can be checked by comparing the appropriate vertical average from a radiosonde station against nearby satellite observations (see Box 2.2). Furthermore, because there are far fewer instrument systems than in radiosonde datasets, it is potentially easier to isolate and adjust problems in the data.

The space and time sampling of the satellites varies according to the orbit of the spacecraft, though the longer satellite datasets are based on polar orbiters. These spacecraft circle the globe from pole to pole while maintaining a nominally constant orientation relative to the sun (sun-synchronous). In this configuration, the spacecraft completes about 14 roughly north-south orbits per day as the earth spins eastward beneath it, crosses the equator at a constant local time, and provides essentially global coverage. Microwave measurements utilized in this report begin in late 1978 with the TIROS-N spacecraft using a 4-channel radiometer (Microwave Sounding Unit or "MSU") which was upgraded in 1998 to a 16-channel system (advanced MSU or "AMSU") with better calibration, more stable station-keeping (i.e., the timing and positioning of the satellite in its orbit – see discussion of "Diurnal Sampling" below), and higher spatial and temporal sampling resolution.

Laboratory estimates of precision (random error) for a single MSU measurement are 0.25 °C. Thus with 30,000 observations per day, this error is inconsequential for global averages. Of far more importance are the time varying biases which arise once the spacecraft is in orbit; diurnal drifting, orbital decay, intersatellite biases and calibration changes due to heating of the instrument in space (see section 3 below.)

While bulk-layer measurements offer the robustness of a large-volume sample, variations within the observed layer are masked. This is especially true for the layer centered on the mid-troposphere (T_2) for which the temperatures of both lower stratospheric and tropospheric levels, which generally show opposite variations, are merged (Figure 2.2). Three MSU/AMSU-based climate records are presented in this report, prepared by Remote Sensing Systems (RSS) of Santa Rosa, California, The University of Alabama in Huntsville (UAH), and The University of

Maryland (UMd). Subscripts identify the team, for example, T_2 will be listed as $T_{2\text{-}R}$, $T_{2\text{-}A}$ and $T_{2\text{-}M}$ for RSS, UAH and UMd respectively.

Some polar orbiters also carry the Stratospheric Sounding Unit (SSU), an infrared sensor for monitoring deep layer temperatures above about 15 km. SSU data have been important in documenting temperature variations at higher elevations than observed by MSU instruments on the same spacecraft (Ramaswamy et al., 2001). Generally, the issues that complicate the creation of long-term MSU time series also affect the SSU, with the added difficulty that infrared channels are more sensitive to variations in atmospheric composition (e.g., volcanic aerosols, water vapor, etc.).

Future observing systems using passive-satellite methods include those planned for the National Polar-orbiting Operational Environmental Satellite System (NPOESS) series: the microwave sensors Conical scanning Microwave Imager/Sounder (CMIS) (which will succeed the Special Sensor Microwave/Imager [SSM/I]), Special Sensor Microwave Imager/Sounder (SSMI/S) and Advanced Technology Microwave Sounder (ATMS) (which will succeed the AMSU), and the infrared sensor Cross-track Infrared Sounder (CrIS) (following the High-resolution Infrared Radiation Sounder [HIRS]). Each of these will follow measuring strategies that are both similar (polar orbit) and dissimilar (e.g., CMIS's conical scanner vs. AMSU's cross-track scanner) but add new spectral and more detailed resolution.

3. "Active" satellite instrumentation

A relatively recent addition to temperature monitoring is the use of Global Positioning System (GPS) radio signals, whose time of transmission through the atmosphere is altered by an amount proportional to air density and thus temperature at levels where humidity can be ignored (Kursinski et al., 1997). A key advantage of this technique for climate study is that it is self-calibrating. Current systems are accurate in the upper troposphere and lower to middle stratosphere where moisture is insignificant, but at lower levels, humidity becomes a confounding influence on density. Future versions of this system may overcome this limitation by using shorter wavelengths to measure humidity and temperature independently. Because of the relatively short GPS record and limited spatial coverage to date, its value for long-term climate monitoring cannot yet be definitively demonstrated.

c) Operational Reanalyses

Operational reanalyses (hereafter simply "reanalyses") will be discussed here in chapter 2, but their trends presented only sparingly in the following chapters because of evidence that they are not always reliable, even during the recent period. All authors expressed concern regarding reanalyses trends, a concern that ranged from unanimous agreement that stratospheric trends were likely spurious to mixed levels of confidence regarding tropospheric trends (see chapter 3). Surface temperature trends are a separate issue as reanalyses values are indirectly *estimated* rather than *observed* (see below). However, reanalyses products hold significant potential for addressing many aspects of climate variability and change.

Reanalyses are not separate observing systems, but are mathematically blended products based upon as many observing systems as practical. Observations are assimilated into a global weather forecasting model to produce analyses that are most consistent with both the available data (given their imperfections) and the assimilation model. The model, which is constrained by known but parameterized atmospheric physics, generates a result that could be more accurate and physically self-consistent than can be obtained from any one observing system. Some data are rejected or adjusted based on detected inconsistencies. Importantly, the operational procedure optimizes only the accuracy of each near-instantaneous ("synoptic") analysis. Time-varying biases of a few hundredths or tenths of a degree, which contribute little to short time scale weather error, present a major problem for climate trends, and these are not minimized (e.g., Sherwood, 2000). The two main reanalyses available at this time are the National Centers for Environmental Prediction (NCEP)/National Center for Atmospheric Research (NCAR) reanalysis of data since 1948 (Kalnay et al., 1996) and the European Center for Medium-Range Weather Forecasts Re-Analysis-40 (ECMWF ERA-40) beginning in 1957 (Simmons, 2004).

Because many observational systems are employed, a change in any one will affect the time series of the final product. Reanalyses would be more accurate than lower-level data products for climate variations only if the above shortcomings were outweighed by the benefits of using a state-of-the-art model to treat unsampled variability. Factors that would make this scenario likely include a relatively skillful forecast model and assimilation system, large sampling errors (which are reduced by reanalysis), and small systematic discrepancies between different instruments. However, current models tend to have significant intrinsic biases that can particularly affect reanalyses when sampling is sparse.

Reanalysis problems that influence temperature trend calculations arise from changes over time in (a) radiosonde and satellite data coverage, (b) radiosonde biases (or in the corrections applied to compensate for these biases), (c) the effectiveness of the bias corrections applied to satellite data and (d) the propagation of errors due to an imprecise formulation of physical processes in the models. For example, since few data exist for the Southern Hemisphere before 1979, temperatures were determined mainly by model forecasts; a cold model bias (in ERA-40, for example) then produces a spurious warming trend when real data become available. Indirect effects may also arise from changes in the biases of other fields, such as humidity and clouds, which affect the model temperature (Andrae et al., 2004; Simmons et al., 2004.).

Different reanalyses do not employ the same data. NCEP/NCAR does not include surface temperature observations over land but the analysis still produces estimated near-surface temperatures based on the other data (Kalnay and Cai, 2003). On the other hand, ERA-40 does incorporate these but only indirectly through their modeled impacts on soil temperature and surface humidity (Simmons et al., 2004). Thus, the 2-meter air temperatures of both reanalyses may not track closely with surface observations over time (Kalnay and Cai, 2003). SSTs in both reanalyses are simply those of the climate records used as input.

For upper air reanalyses temperatures, simultaneous assimilation of radiosonde and satellite data is particularly challenging because the considerably different instrument characteristics and products make it difficult to achieve the consistency possible in theory. Despite data adjustments, artifacts still remain in both radiosonde and satellite analyses; these produce the largest differences in the lower stratosphere in current reanalysis datasets (e.g., Pawson and Fiorino, 1999; Santer et al., 1999; Randel, 2004). Some of these differences can now be explained, so that future reanalyses will very likely improve on those currently available. However any calculation of deep-layer temperatures from reanalyses which require stratospheric information are considered in this report to be suspect (see Figure 2.2, T_T, T_2, T_4, and $T_{100\text{-}50}$).

d.) Simple retrieval techniques

A problem in interpreting MSU (i.e., broad-layer) temperature trends is that many channels receive contributions from both the troposphere and stratosphere, yet temperatures tend to change oppositely in these two layers with respect to both natural variability and predicted climate change. In particular, MSU Channel 2 (T_2) receives 10-15% of its emissions from the stratosphere (Spencer and Christy, 1992), which is a significant percentage because stratospheric cooling in recent decades far exceeds tropospheric warming. It is impossible to eliminate all physical stratospheric influences on MSU 2 by simply subtracting out MSU 4 (T_4) influences because any linear combination of these two channels still retains stratospheric influence (Spencer et al., 2005), which will lead to errors. However, it is possible to rely upon radiosonde-measured correlations between tropospheric and stratospheric temperature fluctuations in order to find what linear combination of these two channels leads to a near-cancellation of these errors, i.e., where *y* is determined by regression:

Tropospheric Retrieval = $(1+y)\cdot(T_2) - (y)\cdot(T_4)$. The challenge here is that the resulting relationship depends on the training dataset (radiosondes) being globally or tropically representative (i.e., the troposphere/stratosphere boundary varies spatially and thus the relationship between T_2 and T_4 does as well) and free from significant biases.

Fu et al. (2004) used a radiosonde dataset to estimate values for y (for the globe, tropical region, and Northern and Southern Hemispheres) that most closely reproduced the monthly variability of mean temperature from 850 to 300 hPa, spanning most of the troposphere. From physical arguments, however, it is clear that the true physical contributions to the retrieval come from a broader range of altitudes, which, in the tropics, approximately span the full troposphere (Fu and Johanson, 2004; 2005). Although derived values of y are robust (±10%, Gillett et al., 2004, Johanson and Fu, 2005), the veracity of the retrieval for climate change has been a subject of debate (due to the accuracy and global representativeness issues mentioned above), and will be further addressed in Chapter 4.

In the following chapters, two simple retrievals will be utilized in comparison studies with the products of the observing systems. The tropospheric retrieval generated from global mean values of T_2 and T_4, is identified as T^*_G where $y = 0.143$ (Johanson and Fu, 2005), and when applied to tropical mean values is identified as T^*_T where $y = 0.100$ (Fu and Johanson, 2005).

A summary of the sources of biases and uncertainties for the datasets and other products described above is given at the end of this chapter. There are several datasets yet to be generated (or not yet at a stage sufficient for climate analysis) from other sources that have the potential to address the issue of vertical temperature distribution. A generic listing of these datasets with a characterization of their readiness is given in Table 2.1.

Table 2.1 Dataset types and readiness for high quality climate monitoring related to the vertical temperature structure of the atmosphere. “Usage of Data” indicates the level of application of the dataset to the vertical temperature issue. “Understanding” indicates the level of confidence (or readiness) in the dataset to provide accurate information on this issue.

I

🟩 : Adequate for long-term global climate variations =Green
🟨 Improvements or continued research needed for long-term global climate variations = Yellow
🟥 Problems exist or a lack of analysis to date inhibit long-term global climate variation studies = Red
P: Polar orbiter, twice per day per orbiter per ground location
G: Geostationary, many observations per day per ground location
2x Day: Twice daily at site
Hrly: Up to several times per day, many report hourly
Syn: Synoptic or generally up to 8 times per day. (Buoys continuous)

2. ANALYSIS OF CLIMATE RECORDS

Two factors can interfere with the accurate assessment of climate variations over multi-year periods and relatively large regions. First, much larger variability (weather or “atmospheric noise”) on shorter time or smaller space scales can, if inadequately sampled by the observing network, bias estimates of relatively small climate changes. For example, an extended heat wave in an un-instrumented region accompanied by a compensating cold period in a well-instrumented region may be interpreted as a “global” cold period when it was not. Such biases can result from either spatial or temporal data gaps (Agudelo and Curry, 2004). Second, instrumental errors, particularly biases that change over time, can create erroneous trends. The seriousness of each problem depends not only on the data available but also on how they are analyzed. Finally, even if global climate is known accurately at all times and places, there remains the issue of what measures to use for quantifying climate change; different choices can sometimes create different impressions, e.g., linear trends versus low frequency filtered analyses that retain some information beyond a straight line.

Upper air layers experience relatively rapid horizontal smoothing of temperature variations, so that on annual mean time scales, the atmosphere is characterized by large, coherent anomaly features, especially in the east-west direction (Wallis, 1998, Thorne et al., 2005b). As a result, a given precision for the global mean value over, say, a year, can be attained with fewer, if properly spaced, upper air measurement locations than at the surface (Hurrell et al., 2000). Thus, knowledge of global, long-term changes in upper-air temperature is limited mainly by instrumental errors. However, for some regional changes (e.g., over sparsely observed ocean areas) sampling problems may compete with or exceed instrumental ones.

a) Climate Records

Various groups have developed long time series of climate records, often referred to as Climate Data Records (CDRs) (NRC, 2000b; 2000c; 2004) from the raw measurements generated by each observing system. Essentially, climate records are time series that include estimates of error characteristics so as to enable the study of climate variation and change on decadal and longer time scales with a known precision.

Long-term temperature changes occur within the context of shorter-term variations, which are listed in Table 2.2. These shorter changes include: periodic cycles such as day-night and seasonal changes; fairly regular changes due to synoptic weather systems, the Quasi-Biennial Oscillation (QBO), and the El Niño-Southern Oscillation (ENSO); and longer-term variations due to volcanic eruptions or internal climate dynamics. These changes have different vertical temperature signatures, and the magnitude of each signal may be different at the surface, in the troposphere, and in the stratosphere. Details are given in Table 2.2. Some of these signals can complicate the identification of temperature trends in climate records.

Table 2.2 Listing of atmospheric temperature variations by time scale and their properties. (Time scales and sources of global temperature variations)

Variation	**Description**	**Dominant Period**	**Approx. Magnitude**	**Detectibility**	**Effect on Trend Estimates**
Diurnal[1]	Warmer days than nights, due to earth's rotation on its axis affecting solar heating.	Daily (outside of polar regions)	Highly variable. Surface skin T changes up to 35K. Boundary layer changes <10K. Free tropospheric changes <1K. Stratospheric changes ~0.1-1 K.	Well detected in surface data. Poorly detected globally in the troposphere and stratosphere due to infrequent sampling (once or twice daily) and potential influence of measurement errors with their own diurnal signal. A few ground-based systems	Satellite data require adjustment of drift in the local equatorial crossing time of spacecraft orbits. Inadequate quantification of the true diurnal cycle hinders this adjustment. Different diurnal adjustments by different groups may

Variation	Description	Dominant Period	Approx. Magnitude	Detectibility	Effect on Trend Estimates
				detect signal well.	partly account for differences in trend estimates.
Synoptic[2]	Temperature changes associated with weather events, such as wave and frontal passages, due to internal atmospheric dynamics.	1-4 days	Up to ~15K or more at middle latitudes, ~3K in Tropics.	Well detected by observing systems designed to observe meteorological variability.	Not significant, but contributes to noise in climate data records.
Intraseasonal[3]	Most notably, an eastward-and vertically-propagating pattern of disturbed weather in the tropical Indo-Pacific ocean region, of unknown cause. Also, atmospheric "blocking" and wet/dry land surface can cause intra-seasonal variations at mid-latitudes.	40-60 days (Tropics), < 180 days (mid-latitudes)	1-2 K at surface, less aloft (tropics), larger in mid-latitudes.	Temperature signals moderately well detected, with tropical atmosphere limited by sparse radiosonde network and IR-based surface temperature limited by cloud. Reanalysis data are useful.	Not significant due to short duration, but may be important if character of the oscillation changes over time.
Annual[4]	Warmer summers than winters, and shift in position of major precipitation	Yearly	~2-30 K; greater over land than sea, greater at high than low latitudes,	Well observed.	Trends are often computed from "anomaly" data, after the

Variation	**Description**	**Dominant Period**	**Approx. Magnitude**	**Detectibility**	**Effect on Trend Estimates**
	zones, due to tilt of the earth's axis of rotation affecting solar heating.		greater near the surface and tropopause than at other heights.		mean annual cycle has been subtracted. Changes in the nature of the annual cycle could affect annual-average trends.
Quasi-Biennial Oscillation (QBO)[5]	Nearly periodic wind and temperature changes in the equatorial stratosphere, due to internal atmospheric dynamics.	Every 23-28 months (average of 27 months because occasionally periods of up to 36 months occur.)	Up to 10 K locally, ~0.5 K averaged over the tropical stratosphere.	Fairly well observed by equatorial radiosonde stations and satellites.	Like ENSO, can influence trends in short data records, but it is relatively easy to remove this signal.
Interannual[6]	Multiannual variability due to interaction of the atmosphere with dynamic ocean and possibly land surfaces; most notably, ENSO. Can also be caused by volcanic eruptions.	ENSO events occur every 3-7 years and last 6-18 months; major volcanic eruptions, irregular but approximately every 5-20 years with effects lasting ~ 2 years.	Up to 3K in equatorial Pacific (ENSO), smaller elsewhere. Volcanic warming of stratosphere can exceed 5K in tropics cooling of surface <2K.	Fairly well observed, although the vertical structure of ENSO is not as well documented, due to sparseness of the tropical radiosonde network.	ENSO affects surface global mean temperatures by ±0.4K, and more in the tropical troposphere. Large ENSO events near the start or end of a data record can strongly affect computed trends, as was the case for the 1997-98 event. Changes in ENSO

Variation	**Description**	**Dominant Period**	**Approx. Magnitude**	**Detectibility**	**Effect on Trend Estimates**
					frequency or strength affect (and may be coupled with) long-term trends.
Decadal to interdecadal oscillations and shifts[7]	Like interannual, but longer time scales. Prominent example is the PDO/ Interdecadal Pacific Oscillation. Despite long time scale, changes can occur as abrupt shifts, for example, a warming shift around 1976. Others include regional changes in the North Atlantic, Pacific-North American, Arctic, and the Antarctic oscillations. Some changes also caused by 11-year solar cycle.	Poorly known; 50-year PDO cycle suggested by 20th-century observations; others a decade or two; solar 11-year cycle detectable also.	Not well studied. The 1976-77 shift associated with a sharp warming of at least 0.2K globally, though difficult to distinguish from anthropogenic warming. 11-year cycle leads to stratospheric temperature changes of ~2K, and interacts with the Quasi-Biennial Oscillation (QBO).	Relatively large regional changes are well observed, but global expression is subject to data consistency issues over time and possible real changes.	Can account for a significant fraction of linear trends calculated over periods of a few decades or less regionally. Such trends may differ significantly from one such period to the next.
Sub-centennial 60-80 year fluctuation or "Atlantic	Fluctuates in instrumental and paleo data at least back to c.1600. Seems to	60-80 years	~ ±0.5C in parts of the Atlantic. Apparently detectable in	Detectable globally above the noise, clear in North Atlantic SST.	Effects small globally, but probably detectable in last few

Variation	Description	Dominant Period	Approx. Magnitude	Detectibility	Effect on Trend Estimates
Multidecadal Oscillation”[8]	particularly affect Atlantic sector. Possible interhemispheric component.		global mean ~ ±0.1C		decades. Readily detectable over this period in North Atlantic Ocean where it clearly affects surface temperature trends and probably climate generally.
Centennial and longer variations[9]	Warming during 20th Century due to human influences, solar, and internal variability. Earlier changes included the “little ice age” and “medieval warm period.”	None confirmed, though 1500 year Bond cycle possible.	20th century warming of ~0.6K globally appears to be as large or larger than other changes during the late Holocene.	Surface warming during 20th century fairly well observed; proxies covering earlier times indicated 20th century warmer than the past 5 centuries	Natural temperature variations occur on the longest time scales accessible in any instrumental record.

---

[1] Christy et al., 2003; Mears et al., 2003; Vinnikov and Grody. 2003; Dai and Trenberth, 2004; Jin, 2004; Seidel et al., 2005.

[2] Palmen and Newton, 1969

[3] Duvel et al.,2004.

[4] Wallace and Hobbs, 1977

[5] Christy and Drouilhet, 1994; Randel et al., 1999; Baldwin et al., 2001

[6] Parker and Brownscombe, 1983; Pan and Oort, 1983; Christy and McNider, 1994; Parker et al., 1996; Angell,

2000; Robock, 2000; Michaels and Knappenberger, 2000; Santer et al., 2001; Free and Angell, 2002a; Trenberth et al., 2002; Seidel et al., 2004; Seidel and Lanzante, 2004

[7] Labitzke, K.,1987; Trenberth and Hurrell, 1994; Lean et al., 1995; Zhang et al., 1997; Thompson et al., 2000; Douglass and Clader, 2002; Seidel and Lanzante, 2004; Hurrell et al., 2003; Folland et al., 1999; Power et al., 1999; Folland et al. 2002.

[8] Schlesinger and Ramankutty, 1994; Mann et al., 1998; Folland et al., 1999; Andronova and Schlesinger, 2000; Goldenberg et al., 2001; Enfield et al., 2001

[9] Folland et al., 2001a.

Our survey of known atmospheric temperature variations, how well they are measured, and their impact on trend estimates suggests that most observing systems are generally able to quantify well the magnitudes of change associated with shorter time scales. For longer time scale changes, where the magnitudes of change are smaller and the stability requirements more rigorous, the observing systems face significant challenges (Seidel et al., 2004).

b) Measuring Temperature Change

Over the last three to five decades, global surface temperature records show increases of almost two tenths of a °C per decade. Explaining atmospheric and surface trends therefore demands relative accuracies of a few hundredths of a degree per decade in global time series of both surface and upper-air observations. As this and subsequent chapters will show, the effects of instrumental biases on the global time series are significantly larger than a few hundredths of a degree for the upper-air data, though the global surface temperature compilations do appear to reach this level of precision in recent decades (Folland et al., 2001b). These biases, especially those of the upper air, must therefore be understood and quantified rather precisely (see section 3 below). For this fundamental reason, reliable assessment of lapse rate changes remains a considerable challenge.

Natural modes of climate variability on regional scales are manifested in decadal fluctuations in (a) the tropical Pacific, e.g., ENSO, and (b) the northern latitudes, e.g., the North Atlantic, Pacific-North American and the Arctic atmospheric oscillations (Table 2.2). Even fluctuations on longer time scales have been proposed, e.g., the Atlantic Multidecadal Oscillation/60-80 year variation (Schlesinger and Ramankutty, 1994; Enfield et al., 2001). Each of these phenomena is associated with regions of both warming and cooling. Distinguishing slow, human-induced changes from such phenomena requires identifying the patterns and separating the influences of such modes from the warming signal (e.g., as attempted by SST by Folland et al., 1999.) In addition, these oscillations could themselves be influenced by human-induced atmospheric changes (Hasselmann, 1999).

3. LIMITATIONS

A key question addressed in this report is whether climate records built by investigators using various components of the observing system can meet the needs for assessing climate variations and trends with the accuracy and representativeness which allows any human attribution to be reliably identified. Climate record builders have usually underestimated the overall uncertainty in their products by relying on traditional sources of uncertainty that can be quantified using standard statistical methods. For example, published linear trend values exist of the same temperature product from the same observing system whose error estimates do not overlap, indicating serious issues with error determination. Thus, in 2003, three realizations of T_2 (or MSU channel 2) 1979-2002 global trends were published as +0.03 ±0.05 +0.12 ±0.02, and +0.24 ±0.02 °C per decade (Christy et al., 2003; Mears et al., 2003; and Vinnikov and Grody, 2003, respectively.) Over 40% of the difference between the first two trends is due to the treatment of a single satellite in the 1984-1986 period, with a combination of lesser differences during later satellite periods. The third dataset has more complex differences, though it is being superseded by a version whose trend is now lower (Grody et al., 2004, Vinnikov et al. 2005).

This situation illustrates that it is very challenging to determine the true error characteristics of datasets (see Chapter 4), although considerably less attention has been paid to this than to the construction of the datasets themselves. In this report, we refer to systematic errors in the climate data records as "construction errors." Such errors can be thought of as having two fundamentally different sources, *structural* and *parametric* (see Box 2.1). The human decisions that underlie the production of climate records may be thought of as forming a *structure* for separating real and artificial behavior in the raw data. Assumptions made by the experts may not be correct, or important factors may have been ignored; these possibilities lead to *structural uncertainty* (Thorne et al., 2005a) in any trend or other metric obtained from a given the climate record. Experts generally tend to underestimate structural uncertainty (Morgan, 1990). The T_2 example above shows that this type of error can considerably exceed those recognized by the climate record builders. Sorting out which decisions are better than others, given the fact many individual decisions are interdependent and often untestable, is challenging.

Structural uncertainty is difficult to quantify because this requires considering alternatives to the fundamental assumptions, rather than just to the specific sampling or bias pattern in the available data (the main source of parametric uncertainty). For example, is an apparent diurnal variation due to (a) real atmospheric temperature change, (b) diurnal solar heating of an instrument component, (c) a combination of both, or (d) something else entirely? If the answer is not known *a priori*, different working assumptions may lead to a different result when corrections are determined and applied.

There may be several ways to identify structural errors. First, it is well known in statistics that one should examine the variability that is left over when known effects are removed in a data analysis, to see whether the residuals appear as small and "random" as implied by the assumptions. Even when the residuals are examined, it is often difficult to identify the cause of any non-randomness. Second, one can compare the results with external or independent data (such as comparing SST and NMAT observations). However, one then encounters the problem of assessing the accuracy of the independent data; because, in the case of global atmospheric temperature data there are no absolute standards for any needed adjustment. Christy et al. (2000) demonstrate the use of internal and external methods for evaluating the error of their upper air time series. They assumed that where agreement of independent measurements exists, there is likely to be increased confidence in the trends. Third, one can try to assess the construction uncertainty by examining the spread of results obtained by multiple experts working independently (e.g., the T_2 example, Thorne et al., 2005a). Unfortunately, though valuable, this does not establish the uncertainties of individual efforts, nor is it necessarily an accurate measure of overall uncertainty. If all investigators make common mistakes, the estimate of construction uncertainty may be too optimistic; but if some investigators are unaware of scientifically sound progress made by others, the estimate can be too pessimistic.

A general concern regarding all of the datasets used in this analysis - land air temperature, sea surface temperature, radiosonde temperature, and satellite-derived temperature – is the level of information describing the operational characteristics and evolution of the associated observing system. As indicated above, the common factor that creates the biggest differences between analyses of the same source data is the homogeneity adjustments made to account for biases in the raw data. All homogeneity adjustments would improve with better metadata - that is, information (data) about the data (see chapter 6). For satellite-derived temperature, additional metadata such as more data points used in the pre-launch calibration would have been helpful to know, especially if done for differing solar angles to represent the changes experienced on orbit. For the in situ data sets, additional metadata of various sorts likely exist in one form or another somewhere in the world and could be acquired or created. These include the type of instrument, the observing environment, the observing practices and the exact dates for changes in any of the above.

Below we identify various known issues that led to errors in the datasets examined in this report, and which have generally been addressed by the various dataset builders. Note that reanalyses inherit the errors of their constituent observing systems, though they have the advantage of seeking a degree of consensus among the various observing systems through the constraint of model physics. The complex reanalysis procedure transforms these errors of output data into errors of construction methodology that are hard to quantify.

Errors primarily affecting *in situ* observing systems.

Spatial and temporal sampling: The main source of this error is the poor sampling of oceanic regions, particularly in the Southern Hemisphere, and some tropical and Southern Hemisphere continental regions (see Text Box 2.1). Temporal variations in radiosonde sampling can lead to biases, (e.g., switching from 00 to 12 UTC) but these are generally documented and thus potentially treatable.

Local environmental changes: Land-use changes, new instrument exposures, etc., create new localized meteorological conditions to which the sensor responds. These issues are most important for land near-surface air temperatures but can also affect the lower elevation radiosonde data. Some changes, e.g., irrigation, can act to increase nighttime minima while decreasing daytime maxima, leaving an ambiguous signal for the daily mean temperature. Such changes are sources of error only if the change in the immediate surroundings of the station is unrepresentative of changes over a larger region.

Changes in methods of observation: A change in the way in which an instrument is used, as in calibrating a radiosonde before launch, i.e., whether it is compared against a typical outdoor sensor or against a traceable standard.

Changes in data processing algorithms: A change in the way raw data are converted to atmospheric information can introduce similar problems. For radiosonde data, the raw observations are often not archived and so the effects of these changes are not easily removed.

Errors primarily affecting satellite systems

Diurnal sampling: It is common for polar orbiters to drift slowly away from their "sun-synchronous" initial equatorial crossing times (e.g., 1:30 p.m. to 5 p.m.), introducing spurious trends related to the natural diurnal cycle of daily temperature. The later polar orbiters (since 1998) have more stable station keeping. Diurnal drift adjustments for T_{2LT} and T_2 impact the trend by a few hundredths °C/decade. Changes in local observation time also significantly afflict *in situ* temperature observations, with a lesser impact on the global scale.

Orbit decay: Variations in solar activity cause expansion and contraction of the thin atmosphere at the altitudes where satellites orbit, which create variable frictional drag on spacecraft. This causes periods of altitude decay, changing the instrument's viewing geometry relative to the earth and therefore the radiation emissions observed. This issue relates most strongly to T_{2LT}, which uses data from multiple view angles, and is of order 0.1 °C/decade.

Calibration shifts/changes: For satellite instruments, the effects of launch conditions or changes in the within-orbit environment (e.g., varying solar shadowing effects on the spacecraft components as it drifts through the diurnal cycle) may require adjustments to the calibration equations. Adjustment magnitudes vary among the products analyzed in this report but are on the order of 0.1 °C/decade for T_{2LT} and T_2.

Surface emissivity effects: The intensity of surface emissions in observed satellite radiances can vary over time due to changes in surface properties, e.g. wet vs. dry ground, rough vs. calm seas, etc., and longer-term land cover changes, e.g., deforestation leading to higher daytime skin temperatures and larger diurnal temperature cycles.

Atmospheric effects: Atmospheric composition can vary over time (e.g., aerosols), affecting satellite radiances, especially the infrared.

Errors affecting all observing systems

Instrument Changes: Systematic variations of calibration between instruments will lead to time-varying biases in absolute temperature. These involve (a) changes in instruments and their related components (e.g., changes in housing can be a problem for *in situ* surface temperatures), (b) changes in instrument design or data processing (e.g., radiosondes) and (c) copies of the same instrument that are intended to be identical but are not (e.g., satellites).

Errors or differences related to analysis or interpretation

Construction Methodology: As indicated, this is often the source of the largest differences among trends from datasets and is the least quantifiable. When constructing a homogeneous, global climate record from an observing system, different investigators often make a considerable range of assumptions as to how to treat unsampled or undersampled variability and both random and systematic instrument errors. The trends and their uncertainties that are subsequently estimated are sensitive to treatment assumptions (Free et al., 2002b). For example, the trends of the latest versions of T_2 from the three satellite analyses vary from +0.044 to +0.199 °C/decade (chapter 3), reflecting the differences in the combination of individual adjustments determined and applied by each team (structural uncertainty.) Similarly, the T_2 global trends of the radiosonde-based and reanalyses datasets range from –0.036 to +0.067 °C/decade indicating noticeable differences in decisions and methodologies by which each was constructed. Thus the goal of achieving a consensus with an error range of a few hundredths °C/decade is not evidenced in these results.

Trend Methodology: Differences between analyses can arise from the methods used to determine trends. Trends shown in this report are calculated by least squares linear regression.

Representativeness: Any given measure reported by climate analysts could under- or overstate underlying climatic behavior. This is not so much a source of error as a problem of interpretation. This is often called statistical error. For example, a trend computed for one time period (say, 1979-2004) is not necessarily representative of either longer or earlier periods (e.g., 1958-1979), so caution is necessary in generalizing such a result. By the same token, large variations during portions of the record might obscure a small but important underlying trend. (See Appendix for Statistical Uncertainties.)

4. IMPLICATIONS

The observing systems deployed since the late 1950s, and the subsequent climate records derived from their data, have the capability to provide information suitable for the detection of many temperature variations in the climate system. These include temperature changes that occur with regular frequency, e.g., daily and annual cycles of temperature, as well as non-periodic events such as volcanic eruptions or serious heat and cold waves. The data from these systems also have the potential to provide accurate trends in climate over the last few decades (and over the last century for surface observations), once the raw data are successfully adjusted for changes over time in observing systems, practices, and micro-climate exposure to produce usable climate records. Measurements from all systems require such adjustments and this report relies on adjusted datasets. The details of making such adjustments when building climate records from the uncorrected observations are examined in the following chapters.

Text Box 2.1: Comparing Radiosonde and Satellite Temperatures

Attempts to compare temperatures from satellite and radiosonde measurements are hindered by a mismatch between the respective raw observations. While radiosondes measure temperatures at specific vertical levels, satellites measure radiances which can be interpreted as the temperature averaged over a deep layer. To simulate a satellite observation, the different levels of temperature in the radiosonde sounding are proportionally weighted to match the profiles shown in Figure 2.2. This can be done in one of two ways.

1. Employ a simple set of geographically and seasonally invariant coefficients or weights, called a static weighting function. These coefficients are multiplied by the corresponding set of temperatures at the radiosonde levels and the sum is the simulated satellite temperature. Over land, the surface contributes more to the layer-average than it does over the ocean, and this difference is taken into account by slightly different sets of coefficients applied to land vs. ocean calculations. This same method may be applied to the temperature level data of global reanalyses. We have applied the "static weighting function" approach in this report.

2. Take into account the variations in the air mass temperature, surface temperature and pressure, and atmospheric moisture (Spencer et al., 1990). Here, the complete radiosonde temperature and humidity profiles are ingested into a radiation model to generate the simulated satellite temperature (e.g., Christy and Norris, 2004). This takes much more computing power to calculate and requires humidity information, which for radiosondes is generally of poorer quality than temperature information or is missing entirely. For climate applications, in which the time series of large-scale anomalies is the essential information, the output from the two methods differs only slightly.

There are practical difficulties in generating long time series of simulated satellite temperatures under either approach. To produce a completely homogeneous data record, the pressure levels used in the calculation must be consistent throughout time, i.e., always starting at the surface and reaching the same designated altitude. If, for example, soundings achieved higher elevations as time went on, there would likely be a spurious trend due to the effects of having measured observations during the latter period of record, while by necessity, relying on estimates for the missing values in the earlier period. We also note that HadAT utilizes 9 pressure levels for simulating satellite profiles while RATPAC use 15, so differences can arise from these differing inputs.

An additional complication is that many radiosonde datasets and reanalyses may provide data at mandatory levels beginning with 1000 and/or 850 hPa, i.e., with no identifiable surface. Thus, the location of the material surface, and its temperature, can only be estimated so that an additional source of error to the anomaly time series may occur. There are a number of other processing choices available when producing a time series of simulated satellite data for site-by-site comparisons between actual satellite data and radiosondes (or reanalyses) and these also have the potential to introduce non-negligible biases.

Averaging of spatially incomplete radiosonde observations for comparison of global and tropical anomalies also introduces some error (Agudelo and Curry, 2004). In this report we have first zonally averaged the data, then generated satellite-equivalent measures from these data and finally calculated global and tropical averages. The spatial coverage differs markedly between the two radiosonde datasets. However, as anomalies are highly correlated in longitude the relative poor longitudinal sampling density of RATPAC (and HadAT outside of the NH mid-latitudes) is not necessarily an impediment (Hurrell et al., 2000). Comparing global averages estimated using only those zonally-averaged grids observed at RATPAC station sites by MSU versus the globally complete fields from MSU, a sampling error of less than ±0.05 °C/decade was inferred for T_{2LT}. Satellite and reanalyses are essentially globally complete and thus do not suffer from spatial subsampling.

References

Agudelo, P.A. and J.A. Curry, 2004: Analysis of spatial distribution in tropospheric temperature trends. Geophys. Res. Lett., 31, L222207.

Aires, F., C. Prigent, and W.B. Rossow, 2004: Temporal interpolation of global surface skin temperature diurnal cycle over land under clear and cloudy conditions. J. Geophys. Res., 109, doi:10.1029/2003JD003527.

Andrae, U., N. Sokka and K. Onogi, 2004: The radiosonde temperature bias corrections used in ERA-40. ERA-40 Project Series #15. 37 pp. European Centre for Medium Range Weather Forecasts. Available at http://www.ecmwf.int/publications/

Angell, J.K., 2000: Tropospheric temperature variations adjusted for El Niño, 1958-1998. J. Geophys. Res., 105, 11841-11849.

Andronova, N.G. and M.E. Schlesinger, 2000: Causes of global temperature changes during the 19th and 20th centuries. Geophys. Res. Lett., 27, 2137-2140.

Baldwin, M.P., L.J. Gray, T.J. Dunkerton, K. Hamilton, P.H. Haynes, W.J. Randel, J.R. Holton, M.J. Alexander, I. Hirota, T. Horinouchi, D.B.A. Jones, J.S. Kinnersley, C. Marquardt, K. Sato, and M. Takahashi et al., 2001: The quasi-biennial oscillation. *Rev. Geophys.*, **39**, 179-229.

Chelton, D.B., 2005: The impact of SST specification on ECMWF surface wind stress fields in the eastern tropical Pacific. J. Climate 18, 530-550.

Christy, J.R., R.W. Spencer, and W.D. Braswell, 2000: MSU Tropospheric temperatures: Data set construction and radiosonde comparisons. J. Atmos. Oceanic Tech. 17,1153-1170.

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 Tech. 20, 613-629.

Christy, J.R. and W.B. Norris, 2004: What may we conclude about tropospheric temperature trends? Geophys. Res. Lett. 31, L06211.

Christy, J.R. and R. T. McNider, 1994 Satellite greenhouse signal. Nature, 367, 325.

Christy, J.R. and S.J. Drouilhet, 1994 Variability in daily, zonal mean lower-stratospheric temperatures. J. Climate, 7, 106-120.

Dai, A., and K. E. Trenberth (2004), The diurnal cycle and its depiction in the Community Climate System model, *J. Climate*, *17*, 930-951.

Douglass, D.H. and B.D. Clader, 2002: Determination of the climate sensitivity of the earth to solar irradiance. Geophys. Res. Lett., 29, 331-334.

Duval, J.P., R. Roca and J. Vialard, 2004: Ocean mixed layer temperature variations induced by instraseasonal convective perturbations of the Indian Ocean. J. Atmos. Sci., 9, 1004-1023.

Enfield, D.B., Mestas-Nuñez, A.M. and P.J Trimble, 2001: The Atlantic Multidecadal Oscillation and its relation to rainfall and river flows in the continental US. Geophys. Res. Lett., 28, 2077-2080.

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**.**, Parker, D.E., Colman, A. and R. Washington, 1999: Large scale modes of ocean surface temperature since the late nineteenth century. Refereed book: Chapter 4, pp73-102 of *Beyond El Nino: Decadal and Interdecadal Climate Variability.* Ed: A. Navarra. Springer-Verlag, Berlin, pp 374.

Folland, C.K., T.R. Karl, J.R. Christy, R.A. Clarke, G.V. Gruza, J. Jouzel, M.E. Mann, J. Oerlemans, M.J. Salinger and S.-W. Wang, 2001a: Observed climate variability and change. In: Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change [Houghton, J.T., Y. Ding, D.J. Griggs, M. Noguer, P.J. van der Linden, X Dai, K. Maskell, and C.A. Johnson (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 881 pp.

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 and D.M.H. Sexton, 2001b: Global temperature change and its uncertainties since 1861. Geophys. Res. Lett., 28, 2621- 2624.

Folland, C.K., J.A. Renwick, M.J. Salinger and A.B. Mullan, 2002: Relative influences of the Interdecadal Pacific Oscillation and ENSO on the South Pacific Convergence Zone. *Geophys. Res. Lett.*, **29** (13): 10.1029/2001GL014201. Pages 21-1 - 21-4.

Free, M., and J. K. Angell, 2002: Effect of volcanoes on the vertical temperature profile in radiosonde data. *J. Geophys. Res.*, 10.1029/2001JD001128.

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, and L.M. McMillin, 2002: CARDS Workshop on Adjusting Radiosonde Temperature Data for Climate Monitoring: Meeting Summary. *Bull. Amer. Meteor. Soc.,* 83, 891-899.

Fu, Q., C.M. Johanson, S.G. Warren, and D.J. Seidel, 2004: Contribution of Stratospheric Cooling to Satellite-Inferred Tropspheric Temperature Trends. Nature , 429, 55-58.

Fu, Q., and C.M. Johanson, 2005: Satellite-derived vertical dependence of tropical tropospheric temperature trends. Geophys. Res. Lett. (in press).

Gillett, N. P., B. D. Santer, A. J. Weaver, 2004, Stratospheric cooling and the troposphere, Nature, doi:10.1038.

Goldenberg, S.B, Landsea, C.W., Mestas Nunez, A.M. and W.M. Gray. The recent increase in Atlantic Hurricane activity: causes and implications. Science, 293, 474- 479.

Grody, Norman C., K. Y. Vinnikov, M. D. Goldberg, J. T. Sullivan, and J. D. Tarpley, 2004. Calibration of multisatellite observations for climatic studies: Microwave Sounding Unit (MSU), J. Geophys. Res. – Atm., 109, D24104, doi:10.1029/2004JD005079, December 21, 2004.

Hasselmann, K., 1999: Linear and nonlinear signatures. Nature, 398, 755-756.

Haimberger, L., 2004: Homogenization of radiosonde temperature time series using ERA-40 analysis feedback information. ERA-40 Project Report Series No. 22. European Centre for Medium Range Weather Forecasts, Shinfield Park, Reading, RG2 9AX, England. 67 pp.

Hurrell, J., S.J. Brown, K.E. Trenberth and J.R. Christy, 2000: Comparison of tropospheric temperatures from radiosondes and satellites: 1979-1998. *Bull. Amer. Met. Soc.*, **81**, 2165-2177.

Hurrell, J.W., Kushnir, Y., Ottersen, G. and M. Visbeck, Eds, 2003: The North Atlantic Oscillation: Climatic Significance and Environmental Impacts. American Geophysical Union, pp 279.

Jin, M. (2004), Analysis of land skin temperature using AVHRR observations, *Bull. Amer. Meteorol. Soc., 85*, 587–600, doi: 10.1175/BAMS-85-4-587.

Johanson, C.M. and Q. Fu, 2005: Robustness of Tropospheric Temperature Trends from MSU channels 2 and 4. J. Climate (submitted).

Jones, P.D., T.J. Osborn, K.R. Briffa, C.K. Folland, E.B. Horton, L.V. Alexander, D.E. Parker and N.A. Rayner, 2001: Adjusting for sampling density in grid box land and ocean surface temperature time series. J. Geophys. Res., 106, 3371-3380.

Jones, P.D and Moberg, A. 2003. Hemispheric and large scale surface air temperature variations: an extensive revision and an update to 2001. J. Clim., 16, 206-223.

Kalney, E. and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. Bull Amer. Metero. Soc., 77, 437-471.

Kalnay, E. and M. Cai, 2003: Impact of urbanization and land-use change on climate. Nature, 423, 528-531.

Kent, E. C. and P. K. Taylor, 2004: Towards Estimating Climatic Trends in SST Data, Part 1: Methods of Measurement. Journal of Atmospheric and Oceanic Technology, submitted.

Kent, E. C. and P. G. Challenor, 2004: Towards Estimating Climatic Trends in SST Data, Part 2: Random Errors. Journal of Atmospheric and Oceanic Technology, submitted.

Kent, E. C. and A. Kaplan, 2004: Towards Estimating Climatic Trends in SST Data, Part 3: Systematic Biases. Journal of Atmospheric and Oceanic Technology, submitted.

Kilpatrick, K. A., G. P. Podesta, et al., 2001: Overview of the NOAA/NASA advanced very high resolution radiometer pathfinder algorithm for sea surface temperature and associated matchup database." J. Geophys. Res. **106**(C5): 9179-9198.

Kursinski E.r., G.A. Hajj, J.T. Schofiled, R.P. Linfield and K.R. Hardy, 1997: Observing the Earth's atmosphere with radio occultation measurements using the Global Positioning System. J. Geophys. Res. 102, 23429-23465.

Labitzke, K., 1987: Sunspots, the QBO, and the stratospheric temperature in the north polar region. Geophys. Res. Lett., 14, 535-537.

Lean, J., J. Beer, and R. Bradley, 1995: Reconstruction of solar irradiance since 1610: implications for climate change. Geophys. Res. Lett., 22, 3195-3198.

Mann, M.E., R.S. Bradley and M.K. Hughes, 1998: Global-scale temperature patterns and climate forcing over the past six centuries. Nature, 392, 779-787.

McPhaden, M. J., 1995: The Tropical Atmosphere Ocean array is completed. Bulletin of the American Meteorological Society. 76: 739-741.

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

Michaels, P.J., and P.C. Knappenberger, 2000: Natural signals in the MSU lower tropospheric temperature record. Geophys. Res. Lett. 27, 2905-2908.

Morgan, M. G., 1990: *Uncertainty : a guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambridge University Press, 332 pp.

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

NRC 2000b. *Ensuring the Climate Record from the NPP and NPOESS Meteorological Satellites*. National Academy Press. 51 pp.

NRC 2000c. *Issues in the Integration of Research and Operational Satellite Systems for Climate Research II: Implementation*. National Academy Press. 82 pp.

NRC 2004. *Climate Data Records from Environmental Satellites*. National Academy Press. 136 pp.

Palmen, E. and C. Newton, 1969: Atmospheric Circulation Systems: Their Structure and Interpretation. Academic Press.

Pan, Y.-H., and A.H. Oort, 1983: Global climate variations connected with sea surface temperature anomalies in the eastern equatorial Pacific Ocean for the 1958-1973 period. Mon. Weath. Rev., 111, 1244-1258.

Parker, D.E. and Brownscombe, 1983: Nature, 301, 406-408.

Parker, D.E., M. Gordon, D.P.N. Cullum, D.M.H. Sexton, C.K. Folland and N. Rayner, 1997: A new global gridded radiosonde temperature data base and recent temperature trends. Geophys. Res. Lett., 24, 1499-1502.

Parker, D.E., H. Wilson, P.D. Jones, J.R. Christy and C.K. Folland, 1996: The impact of Mount Pinatubo on world-wide temperatures. Int. J. Climotol., 16, 487-497.

Pawson, S. and M. Fiorino, A comparison of reanalyses in the tropical stratosphere. Part 3: Inclusion of the pre-satellite data era, Clim. Dyn., 1999, 15, 241-250.

Power, S., Casey, T., Folland, C.K., Colman, A and V. Mehta, 1999: Inter-decadal modulation of the impact of ENSO on Australia. *Climate Dynamics***, 15**, 319-323.

Ramaswamy, V., M.-L. Chanin, J. Angell, J. Barnett, D. Gaffen, M. Gelman, P. Kekhut, Y. Koshelkov, K. Labitzke, J.-J. R. Lin, A. O'Neill, J. Nash, W. Randel, R.Rood, K. Shine, M. Shiotani, and R. Swinbank, 2001: Stratospheric temperature trends: Observations and model simulations. *Rev. Geophys.*, *39*, 71-122.

Randel, W.J., F. Wu, R. Swinbank, J. Nash, and A. O'Neill, 1999: Global QBO circulation derived from UKMO stratospheric analyses. *J. Atmos. Sci.*, **56**, 457-474.

Rayner, N. A., D. E. Parker, et al., 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. J. Geophys. Res. 108(d14).

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

Reynolds, R.W., Rayner, N.A. Smith, T.H. Stokes, D.C. and Wang W., 2002: An improved in situ and satellite SST analysis for climate. J. Clim., 15, 1609-1625.

Robock, Alan, 2000: Volcanic eruptions and climate. Rev. Geophys., 38, 191-219.

Santer. B.D., J.J. Hnilo, T.M.L. Wigley, J.S. Boyle, C. Doutriaux, M. Fiorino, D.E. Parker, and K.E. Taylor, 1999: Uncertainties in observationally based estimates of temperature change in the free atmosphere. *J. Geophys. Res.,* **104**, 6305-6333.

Santer, B.D., T.M.L. Wigley, C. Doutriaux, J.S. Boyle and 6 others, 2001: Accounting for the effects of volcanoes and ENSO in comparisons of modeled and observed temperature trends. J. Geophys. Res., 106, 28033-28059.

Schlesinger, M.E. and N. Ramankutty, An oscillation in the global climate system of period 65-70 years. Nature, 367, 723-726.

Seidel, D.J., J.K. Angell, J.R. Christy, M. Free, S.A. Klein, J.R. Lanzante, C. Mears, D. Parker, M. Schabel, R. Spencer, A. Sterin, P. Thorne and F. Wentz, 2004: Uncertainty in signals of large-scale climate variations in radiosonde and satellite upper-air temperature datasets. J. Climate, 17, 2225-2240.

Seidel, D.J., M. Free, and J. Wang, The diurnal cycle of temperature in the free atmosphere estimated from radiosondes, J. Geophys. Res. (submitted).

Seidel, D.J., and. J.R. Lanzante, 2004: An assessment of three alternatives to linear trends for characterizing global atmospheric temperature changes, *J. Geophys. Res.* 109, XXX-XXX, doi:10.1029/2003JD004414, in press.

Sherwood, S. C., Climate signal mapping and an application to atmospheric tides, Geophys. Res. Lett., 2000, 27, 3525-3528.

Simmons, A.J., Jones, P.D., da Costa Bechtold, V., Beljaars, A.C.M., Kållberg, P., Saarinen, S., Uppala, S.M., Viterbo, P. and N. Wedi, N. 2004: Comparison of trends and variability in CRU, ERA-40 and NCEP/NCAR analyses of monthly-mean surface air temperature. *J. Geophys. Res.*, **109**, No D24 D24115 http://dx.doi.org/10.1029/2004JD005306 December 21, 2004.

Simmons, A.J., 2004: Development of the ERA-40 Data Assimilation System. 20 pp. European Centre for Medium Range Weather Forecasts. Available at http://www.ecmwf.int/publications/

Smith, T. M. and R. W. Reynolds, 2004: Improved Extended Reconstruction of SST (1854-1997). J.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, Submitted.

Spencer, R. W. , J. R. Christy and N. C. Grody, 1990: Global atmospheric temperature monitoring with satellite microwave measurements: Method and results 1979-1985. *J. Climate*, **3**, 1111-1128.

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 1979-90. *J. Climate*, **5**, 858-866.

Spencer, R.W., J.R. Christy and W.D. Braswell, 2005: On the estimation of tropospheric temperature trends from MSU channels 2 and 4. J. Atmos. Oceanic Tech., submitted.

Strategic Plan for the U.S. Climate Change Science Program, 2003: A Report by the Climate Change Science Program and the Subcommittee on Global Change Research.

Thompson, D.W.J., J.M. Wallace and G.C. Hegerl, 2000: Annual modes in the extratropical circulation Part II: trends. J. Climate, 13, 1018-1036.

Thorne, P.W., D.E. Parker, J.R. Christy and C.A. Mears, 2005: Causes of differences in observed climate trends. Bull. Amer. Meteor. Soc., in press.

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

Trenberth, K.E. and J.W. Hurrell, 1994: Decadal atmosphere-ocean variations in the Pacific. Clim. Dyn., 9, 303-319.

Trenberth, K.E., Carron, J.M., Stepaniak, D.P. and S. Worley, 2002: Evolution of the El Nino-

Southern Oscillation ad global atmospheric surface temperatures. J. Geophys. Res., 107, D8, 10.1029/2000JD000298.

Wallace, J.M. and P.V. Hobbs, 1977: Atmospheric Science: An Introductory Survey. Academic Press, New York, NY, 467 pp.

Wallis, T.W.R., 1998: A subset of core stations from the Comprehensive Aerological Reference Data Set (CARDS). J. Climate, 11, 272-282.

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, A. Robock, R.J. Stouffer, P.D. Jones and M.D. Goldberg, 2005: Temperature trends at the surface and the troposphere. J. Geophys. Res. submitted.

Zhang, Y., J.M. Wallace, and D.S. Battisti, 1997: ENSO-like interdecadal variability: 1900-93 ,J. Climate, 10, 1004-1020.