

CHAPTER 2

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

Convening Lead Author: John R. Christy, Univ. of AL in Huntsville

Lead Authors: D.J. Seidel, NOAA; S.C. Sherwood, Yale Univ.

Contributing Authors: M. Cai, FL State Univ.; E. Kalnay, Univ. of MD; C.K. Folland, U.K. Met. Office; C.A. Mears, Remote Sensing Systems; P.W. Thorne, U.K. Met. Office; J.R. Lanzante, NOAA

KEY FINDINGS

- 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 (e.g., urban heat island effect) to produce usable climate records. Measurements from all systems require such adjustments and this report relies solely on adjusted data sets.
- 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 data sets. While these show promise for trend analysis, and it is very likely that current upper-air climate records give reliable indications of directions of change (e.g., warming of the troposphere, cooling of the stratosphere), some questions remain regarding the accuracy of the measurements.
- Upper-air data sets have been subjected to less scrutiny than surface data sets.
- 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 data sets 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.

Chapter 2: Recommendations

- 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 (CCSP 2003) to significantly mitigate the limitations listed above.
- The ability to fully and accurately observe the diurnal cycle should be an important consideration in the design and implementation of new observing systems.
- When undertaking efforts to retrieve data it is important 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.
- 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.
- Some largely overlooked satellite data sets should be reexamined to try to extend, fortify, or corroborate existing microwave-based temperature records for climate research, e.g., microwave data from the Nimbus E Microwave Spectrometer (NEMS) (1972) and Scanning Microwave Spectrometer (SCAMS) (1975), infrared from the High-resolution Infrared Radiation Sounder (HIRS) suite, and radio occultation from GPS.

I. 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).

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 (including changes in nearby thermally emitting structures), effects of land-use changes (e.g., urbanization), 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 25 meters, with poorer temporal and spatial 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 of the surface. The scale of the spatial and temporal coherence of SST and MAT anomalies is greater than that of near-surface air temperatures over land; thus a lower rate of oceanic sampling, in theory, can provide an accuracy similar to the more densely monitored land area.

Incomplete geographic sampling, changing measurement methods, and land-use changes all introduce errors into surface temperature compilations (Jones *et al.*, 2001). 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 data sets for different purposes. The choice may depend on whether the interest is a particular local region, the entire globe, or use of the data set 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.

Data sets of near-surface land and ocean temperatures have traditionally been derived from in-situ thermometers. With the advent of satellites, some data sets now combine both *in-situ*

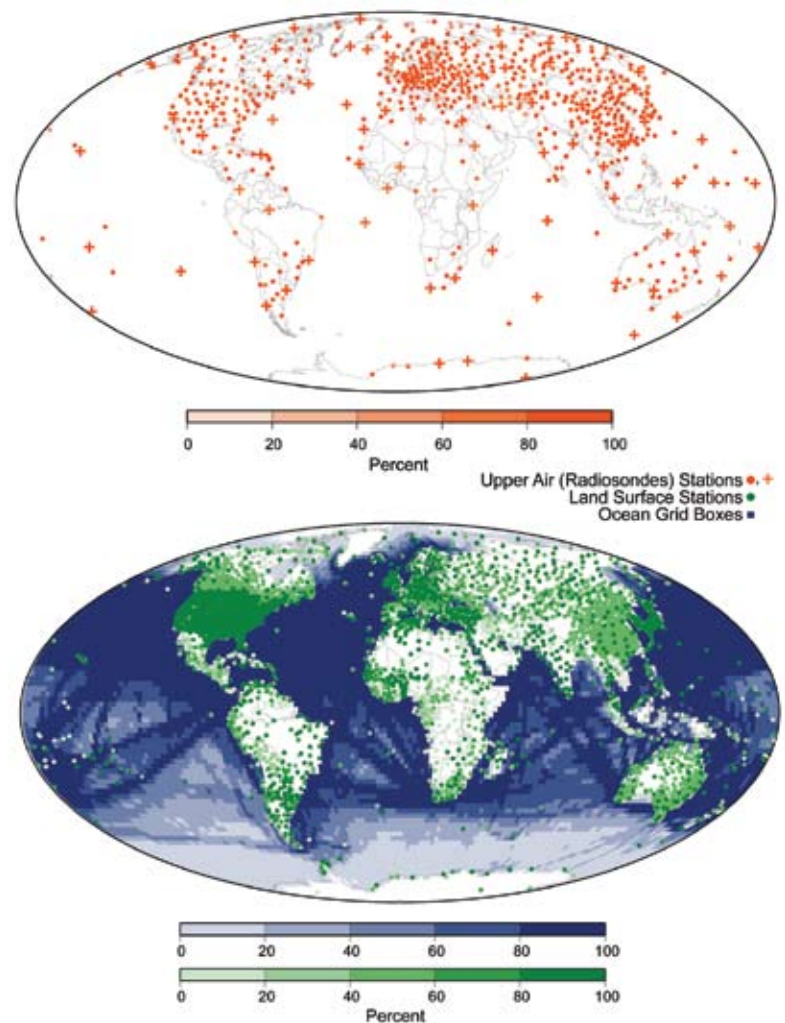


Figure 2.1 Top: Location of radiosonde stations used in the HadAT2 upper air data set with those also in the RATPAC as crosses. Bottom: Distribution of land stations (green) and SST observations (blue) reporting temperatures used in the surface temperature data sets over the period 1979-2004. Darker colors represent locations for which data were reported with greater frequency. See chapter 3 for definitions of data sets.

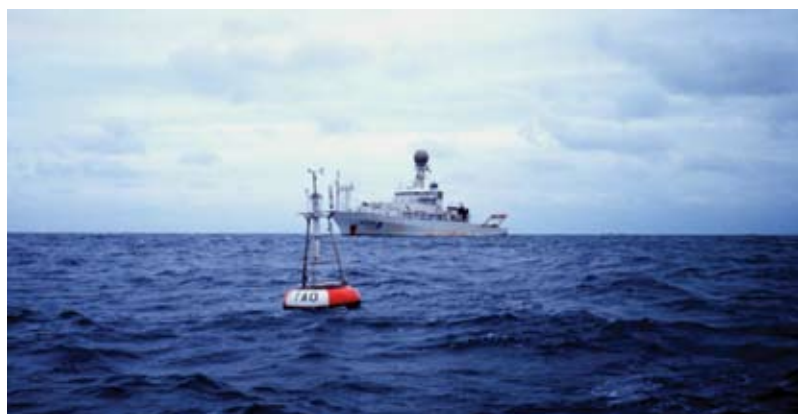
and remotely sensed data (Reynolds *et al.*, 2002; Smith and Reynolds, 2005), or use exclusively remotely sensed data (Kilpatrick *et al.*, 2001) to produce more geographically complete distributions of surface temperature. Because the satellite sensors measure infrared or microwave emission from the 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, data sets 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. Historical correction of ship data prior to 1942 is discussed by Folland and Parker (1995) and Folland (2005) and bias and random errors from ships are summarized by Kent and Taylor (2006) and Kent and Challenor (2006). They report that engine intake temperatures are typically biased 0.1-0.2°C 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, 2006) 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 (2006) also found that insulated bucket temperatures may be too cold by 0.12 to 0.16°C. 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 0.18°C. In addition, their study indicates that engine intake SSTs may have a cold bias of -0.13°C in the early 1990s. The reliability of these biases is subject to revision due to sample sizes that 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, NOAA National Climatic Data Center, and the cooperative project of the U.K. Hadley Centre and the Climate Research Unit of the University of East Anglia (HadCRUT2v). These are identified as T_{S-NASA} , T_{S-NOAA} and $T_{S-HadCRUT2v}$ respectively.

We present three global surface climate records, to a large extent based on data readily exchanged internationally.



b) Atmospheric “Upper Air” Temperatures

I. 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 extra-tropical 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 dilemma differently (see Chapters 3 and 4).

Typically, radiosonde-based data sets 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 data sets are featured in this report, the Hadley Centre Atmospheric Temperatures from the U.K. (HadAT2), and Radiosonde Atmospheric Temperatures Products for Assessing Climate (RATPAC) from NOAA. A product such as $T_{850-300}$, for example, will be identified as $T_{850-300-HadAT2}$ and $T_{850-300-RATPAC}$ for HadAT2 and RATPAC respectively.¹

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



¹ A third radiosonde data set 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 HadAT2 and RATPAC.

Unlike radiosondes, passive satellite observations sample relatively thick atmospheric layers.

and night releases differently, Sherwood *et al.*, 2005), 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 relatively 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. Furthermore, because there are far fewer instrument systems than in radiosonde data sets, 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 data sets are based on polar orbiters. These spacecraft circle the globe from near 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 Television Infrared Observation Satellite (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 that arise once the spacecraft is in orbit: diurnal drifting, orbital decay, inter-satellite biases, and calibration changes due to heating and cooling 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

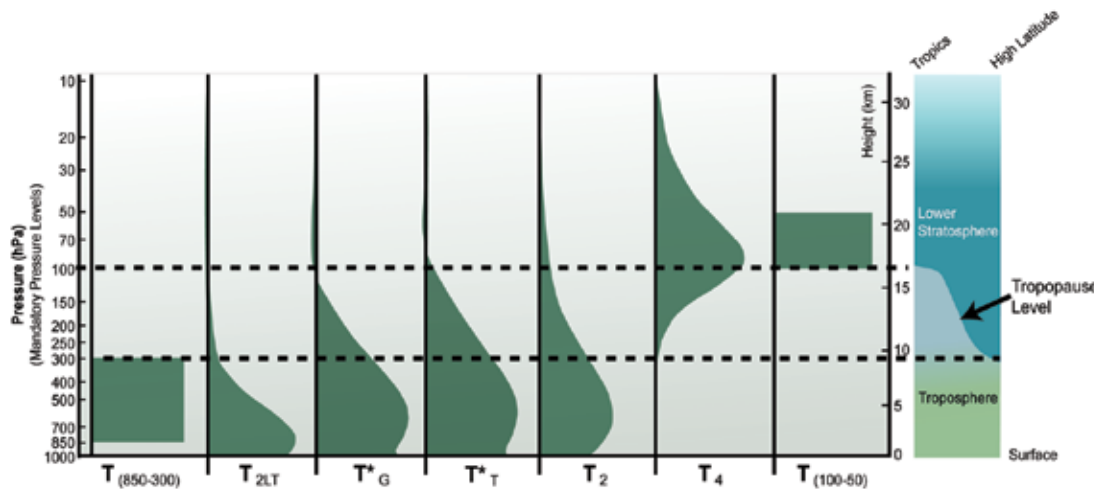


Figure 2.2 Terminology and vertical profiles for the temperature products referred to in this report. Radiosonde-based layer temperatures ($T_{850-300}$, T_{100-50}) are height-weighted averages of the temperature in those layers. Satellite-based temperatures (T_{2LT} , T_2 , and T_4) are mass-weighted 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.

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 .

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).

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 (SSM/I-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 trends derived from them are 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, which represents in a theoretical manner how the atmosphere behaves. 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

Reanalyses will be discussed here, but trends derived from them are presented only sparingly in the following chapters because of evidence that they are not always reliable.



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 Centre for Medium-Range Weather Forecasts Re-Analysis-40 (ECMWF ERA-40) beginning in 1957 (Uppala *et al.*, 2005).

Because many observational systems are employed, a change in any one will affect the time series of the final product unless flagged and corrected as ERA-40 attempts to do. 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, Bengtsson *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.

Simultaneous assimilation of radiosonde and satellite data for upper-air temperatures in reanalyses 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 data sets (*e.g.*, Pawson and Fiorino, 1999; Santer *et al.*, 1999; Randel and Wu, 2006). 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_{100-50}).

d.) Simple Statistical 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. In principle, subtracting an appropriate fraction of MSU 4 from MSU 2, as advocated by Fu *et al.* (2004), will produce a value more representative of the troposphere. The statistical retrieval has the form: Tropospheric Retrieval = $(1+y) \cdot (T_2) - (y) \cdot (T_4)$ where y is determined by regression. However this will not work exactly, because the stratospheric



Despite data adjustments, artifacts still remain in both radiosonde and satellite analyses.

contribution to a signal in either channel will depend on the nature of the signal (*e.g.*, its time scale or cause). The seriousness of this problem for trends is not large (Gillett *et al.*, 2004; Johanson and Fu, 2006; Kiehl *et al.*, 2005) but some problems have been demonstrated (Tett and Thorne, 2004; Spencer *et al.*, 2006).

Fu *et al.* (2004) used a radiosonde data set 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, 2005). In the following chapters, two simple statistical 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, 2006), 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 data sets and other products described above is given at the end of this chapter. There are several data sets 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 data sets with a characterization of their properties is given in Table 2.1.

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.

Temperature variations of upper air layers are characterized by large, coherent anomaly features on annual time scales, especially in the east-west direction (Wallis, 1998, Thorne *et al.*, 2005b). As a result, a given accuracy for the global mean value over, say, a year, can be attained with fewer, if reasonably 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 likely limited more by instrumental errors than spatial coverage. 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, 2000a; 2000b; 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 accuracy and 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

Various groups have developed long time series of climate records, often referred to as Climate Data Records, that include estimates of error characteristics.



Table 2.1 Data sources for climate monitoring related to the vertical temperature structure of the atmosphere.

DATA SOURCE	Measured Variables	Start date of main availability	Temporal Sampling	Geographic Completeness
Radiosondes (Balloons)	Upper Air Temperature	1958	2x Day	
	Upper Air Humidity	1958	2x Day	
	Upper Air Wind	1958	2x Day	
Microwave Radiometers Space-based	Upper air Temperature	1978	P	
	Sea Surface Temperature	1997	P	
	Total Column Vapor (ocean)	1987	P	
Surface-based sounders and profilers	Upper air Temperature	~1985	Hrly	
Infrared Radiometers Space-based	Upper Air Temperature	1973	P, G	
	Land Surface Temperature	1976	P, G	
	Sea Surface Temperature	1981	P, G	
	Upper Air Humidity	1973	P, G	
Visible and Infrared Radiometers	Radiative Fluxes	1979	P, G	
GPS Satellites	Temperature	2001	quasi-P	
Surface Stations Land	Land Surface Air Temperature	~1850	Hrly	
	Land Surface Air Humidity	~1880	Hrly since 1973	
Surface Instruments Ocean	Sea Surface Temperature	~1850	Syn	
	Marine Air Temperature	1856	Syn	
Reanalyses	All	1950	Syn	

- Global or near-global distribution of observations
- Large regions not regularly sampled
- Low spatial distribution of observations

P: Polar orbiter, twice per day per orbiter per ground location except in swath gaps 40°S – 40°N.

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)

Quasi-P: requires transmitter and receiver (at least one of which is satellite-mounted) to be appropriately positioned to sample atmosphere. Opportunities are not spatially/temporally systematic to date but are expected to be in the future.

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. Some of these signals can complicate the identification of temperature trends in climate records.

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 about $+0.15^{\circ}\text{C}$ per decade. Explaining atmospheric and surface trends therefore demands relative accuracies of a few hundredths of a degree C 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 accuracy 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; Knight *et al.*, 2005). 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 for 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 $+0.03 \pm 0.05$, $+0.12 \pm 0.02$, and $+0.24 \pm 0.02$ $^{\circ}\text{C}$ 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 data set has more complex differences, though it is being superseded by a version whose trend is now lower (Grody *et al.*, 2004, Vinnikov *et al.*, 2006).

This situation illustrates that it is very challenging to determine the true error characteristics of data sets (see Chapter 4), although considerably less attention has been paid to this than to the construction of the data sets 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, parametric and structural (see Appendix A). Parametric uncertainty, which results from finite sample sizes, is much less important than structural uncertainty.

Over 40% of the difference between two of the MSU data sets is due to the treatment of a single satellite in the 1984-1986 period.



Table 2.2 Global atmospheric temperature variations: their time scales, sources, and properties.

Variation	Description	Dominant Period	Approx. Magnitude	Detectability	Effect on Trend Estimates
Diurnal ² ² Christy <i>et al.</i> , 2003; Mears <i>et al.</i> , 2003; Vinogradov and Grody, 2003; Dai and Trenberth, 2004; Jin, 2004; Seidel <i>et al.</i> , 2005.	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 35°C. Boundary layer changes <10°C. Free tropospheric changes <1°C. Stratospheric changes ~0.1-1°C.	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 detect signal well.	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 partly account for differences in trend estimates.
Synoptic ³ ³ Palmen and Newton, 1969	Temperature changes associated with weather events, such as wave and frontal passages, due to internal atmospheric dynamics.	3-7 days	Up to ~15°C or more at middle latitudes, ~3°C in Tropics.	Well detected by observing systems designed to observe meteorological variability.	Not significant, but contributes to noise in climate data records.
Intraseasonal ⁴ ⁴ Duvel <i>et al.</i> , 2004.	Most notably, an eastward-and vertically-propagating pattern of disturbed weather in the tropical Indo-Pacific ocean region, related to organized convection. Also, atmospheric "blocking" and wet/dry land surface can cause intraseasonal variations at mid-latitudes.	40-60 days (Tropics), < 180 days (mid-latitudes)	1-2°C 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 ⁵ ⁵ Wallace and Hobbs, 1977	Warmer summers than winters, and shift in position of major precipitation zones, due to tilt of the Earth's axis of rotation affecting solar heating.	Yearly	~2-30°C; greater over land than sea, greater at high than low latitudes, greater near the surface and tropopause than at other heights.	Well observed.	Trends are often computed from "anomaly" data, after the mean annual cycle has been subtracted. Changes in the nature of the annual cycle could affect annual-average trends.
Quasi-Biennial Oscillation (QBO) ⁶ ⁶ Christy and Drouilhet, 1994; Randel <i>et al.</i> , 1999; Baldwin <i>et al.</i> , 2001	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°C locally, ~0.5°C 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.



Variation	Description	Dominant Period	Approx. Magnitude	Detectability	Effect on Trend Estimates
Interannual ⁷ ⁷ Parker and Brownscombe, 1983; Pan and Oort, 1983; Christy and McNider, 1994; Parker <i>et al.</i> , 1996; Angell, 2000; Robock, 2000; Michaels and Knappenberger, 2000; Santer <i>et al.</i> , 2001; Free and Angell, 2002; Trenberth <i>et al.</i> , 2002; Seidel <i>et al.</i> , 2004; Seidel and Lanzante, 2004	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 at least 2 years.	Up to 3°C in equatorial Pacific (ENSO), smaller elsewhere. Volcanic warming of stratosphere can exceed 5°C. In tropics, cooling of surface <2°C.	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 ±0.4°C, 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 frequency or strength affect (and may be coupled with) long-term trends.
Decadal to interdecadal oscillations and shifts. ⁸	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; multi-decadal PDO cycle suggested by 20 th -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 0.2°C globally, though difficult to distinguish from anthropogenic warming. 11-year cycle leads to stratospheric temperature changes of ~2°C, 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 Multidecadal Oscillation" ⁹	Fluctuates in instrumental and paleo data at least back to c.1600. Seems to particularly affect Atlantic sector. Possible interhemispheric component.	60-80 years	~ ±0.5°C in parts of the Atlantic. Apparently detectable in global mean ~ ±0.1°C	Detectable globally above the noise, clear in North Atlantic SST.	Effects small globally, but probably detectable in last few 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 ¹⁰	Warming during 20 th 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.	20 th century warming of ~0.6°C globally appears to be as large or larger than other changes during the late Holocene.	Surface warming during 20 th century fairly well observed; proxies covering earlier times indicated 20 th century warmer than the past 5 centuries	Natural temperature variations occur on the longest time scales accessible in any instrumental record.



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 climate record. Experts generally tend to underestimate structural uncertainty (Morgan, 1990). The T₂ 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 T₂ 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 data sets 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 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.

We illustrate the evolution of a data set in Table 2.3 by listing the adjustments that have been discovered and applied to the UAH T_{2LT} data set since the first version was published. The UAH satellite data set is the oldest of the satellite temperature data sets, and thus has the advantage of a traceable effort toward an improved data set. Improvements in data sets generally occur when they are used regularly to monitor climate change and are therefore more thoroughly scrutinized.



All data adjustments would improve with better metadata - that is, information about the data.

Table 2.3 Corrections in the UAH T_{2LT} data set over time. Progress occurs as data sets undergo continual and independent evaluation.

UAH Version of T _{2LT}	Main Adjustment	Net effect on T _{2LT} Trend °C/decade	Date Applied and Citation
A	Simple Bias Correction		1992, Spencer and Christy, 1992
B	Linear diurnal drift correction for NOAA-7	-0.03	1994, Christy <i>et al.</i> , 1995
C	Removal of residual annual cycle related to hot target variations	+0.03	1997, Christy <i>et al.</i> , 1998
D	Orbital Decay	+0.10	1998, Christy <i>et al.</i> , 2000
D	Removal of dependence on time variations of hot target temperature	-0.07	1998, Christy <i>et al.</i> , 2000
5.0	Non-linear diurnal correction	+0.008	2003, Christy <i>et al.</i> , 2003
5.1	Tightened criteria for data acceptance	-0.004	2004, update file at UAH
5.2	Correction of diurnal drift adjustment	+0.035	2005, Spencer <i>et al.</i> , 2006

Below we identify various known issues that led to errors in the data sets examined in this report, and which have generally been addressed by the various data set 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’s surface 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, interannual sea ice variations *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 data sets 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.*, 2002). For example, the linear trends of the latest versions of T_2 from the three satellite analyses vary from +0.04 to +0.20°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 data sets range from -0.04 to +0.07°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

Different investigators make a range of assumptions about how to treat unsampled or undersampled variability and both random and systematic instrument errors.



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 that 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 (Spencer and Christy, 1992). 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 radiosonde observations are 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 (Santer *et al.*, 1999).

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 HadAT2 utilizes 9 pressure levels for simulating satellite profiles while RATPAC uses 15, so differences can arise from these differing inputs.

An additional complication is that many radiosonde data sets 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 data sets. However, as anomalies are highly correlated in longitude the relative poor longitudinal sampling density of RATPAC (and HadAT2 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 $\pm 0.05^\circ\text{C}/\text{decade}$ was inferred for T_{2LT} . Satellite and reanalyses are essentially globally complete and thus do not suffer from spatial subsampling.




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 A for a discussion of 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 solely on adjusted data sets. The details of making such adjustments when building climate records from the uncorrected observations are examined in the following chapters.



Measurements from all systems require adjustments and this report relies solely on adjusted data sets.