



# **Climate Change & Tropospheric Temperature Trends**

**Part I - What do we know today and where is it taking us?**

By

**Scott Church**

**February 10, 2005**

## **Current Revision Level**

**Rev. 1.1** Feb, 13, 2005.

## **Acknowledgements**

I would like to thank the following for taking the time out of their already busy schedules to offer badly needed comments and suggestions regarding the content of this paper. Without their contributions, it would not have been possible. Thank you!

- Dian Seidel** (NOAA Air Resources Laboratory, Silver Spring, MD)
- Kevin Trenberth** (National Center for Atmospheric Research / Climate and Global Dynamics, Boulder, CO)
- Jerry Mahlmann** (National Center for Atmospheric Research / Climate and Global Dynamics, Boulder, CO)
- Rasmus Benestad** (Norwegian Meteorological Institute, Oslo, Norway; Contributing Editor for [www.RealClimate.com](http://www.RealClimate.com))
- Gavin Schmidt** (Goddard Institute for Space Studies, New York, NY; Contributing Editor for [www.RealClimate.com](http://www.RealClimate.com))
- David Parker** (Met Office, Bracknell, Berkshire, U.K.)

(NOTE: This paper is still being reviewed. As more comments are received and incorporated, this list will grow – and so will my gratitude.)

## **Table of Contents**

Introduction .....	5
The Dilemma .....	6
Atmospheric Temperature Monitoring with MSU & AMSU Products .....	8
a) TIROS-N Satellites and MSU Products .....	8
b) Advanced TIROS Satellites and AMSU Products .....	9
MSU and AMSU Datasets .....	10
MSU and AMSU Analysis Products .....	15
a) The University of Alabama, Huntsville (UAH) Team .....	15
b) The Prabhakara (PR) Team .....	16
c) The Remote Sensing Systems (RSS) Team .....	17
d) Vinnikov and Grody (VG) .....	18
Determining Trends from MSU/AMSU Products .....	19
a) Diurnal Corrections .....	19
b) Intersatellite Data Merging Procedures .....	20
c) The Method of Vinnikov and Grody .....	21
The Radiosonde Record .....	22
a) Angell 54 .....	25
b) HadRT .....	26
c) UAH .....	27
d) LKS .....	27
e) RIHMI .....	28
f) UAH 2004 .....	29
Comparisons of Radiosonde Analysis Methods .....	29
Discussion .....	31
a) Comparisons of Radiosonde products .....	34
b) But Should They Agree? .....	40
c) Surface-Troposphere Coupling .....	41
d) Stratospheric Signals – Fu et al. ....	43
e) Summary .....	47

Objections .....	48
a) Models and the Troposphere – Santer et al. (2003) .....	49
b) Douglass, Singer, and Michaels (2004) .....	51
c) The Fu et al. Method – A Viable Alternative to TLT? .....	54
The Road Forward .....	61
Appendix I – The Fu et al. Method .....	63
Footnotes .....	67
Glossary .....	68
References .....	70
Figures .....	80 - 141

## **Introduction**

There is general agreement among the world's climate scientists that the Earth's global average surface-air temperature is now increasing at rates that are without precedent during the last 1000 years, and that this increase is at least in part due to human activity – particularly greenhouse gas emissions and land use practices. These conclusions are based on nearly a century of temperature data from over 900 surface weather stations with close to global coverage, and a wide range of data from various proxy indicators such as tree ring cores, glacier and snow-pack change, radiosonde, rocketsonde, and satellite data, and more. These suggest that the Earth's global average temperature has risen between 0.4 and 0.8 deg C. since the early 20<sup>th</sup> century (IPCC, 2001). Even more disconcerting is the likelihood that this global warming is being driven by processes that have very long response times so that once started, it may take generations to stop even after mitigation activities are implemented around the world. Though the evidence for this warming grows stronger every day, there is still a great deal of uncertainty regarding how it will play itself out. Most climate scientists believe that by the end of the 21<sup>st</sup> century the consequences will be severe, but there is wide disagreement about the level of severity and what the actual impacts will be. There is also disagreement about the extent to which human activity is contributing to this increase. Some have argued that the observed warming is entirely natural and that we cannot do anything to mitigate it. At the more extreme end, some have even argued that the warming is *beneficial*. If indeed we are contributing to global warming, it is of the utmost importance that the remaining uncertainties about our fingerprint on the earth's climate be answered soon lest we delay too long before implementing needed changes.

One of the more important open questions involves the relationship between temperatures at the Earth's surface where we all live, and those of the troposphere and stratosphere, and how the two influence each other. Since anthropogenic (of human origin) greenhouse gases are thought to be a major contributor to this warming, and these gases are well mixed in the atmosphere, climate scientists believe that the lower and middle troposphere should warm at least as much as the surface. Even so, detecting this warming has been problematic. Many recent observations have only revealed about half as much warming as expected, and the difference is likely to be statistically significant (NRC, 2000). Climate scientists point to the many gaps in our understanding of how the surface and troposphere interact with each other as well as how they are forced by the many factors driving climate change. They also point to the many gaps and uncertainties in our data regarding the historic evolution of troposphere and stratosphere temperatures. But others who are more confident of what is already known claim that this discrepancy is a show-stopper for global warming, and proof that global warming mitigation policies are unneeded and wasteful. This perceived discrepancy between surface and troposphere temperature trends is one of the last and most significant roadblocks to a general recognition of the reality of global warming. It must be explained, one way or another, before a clear picture of the nature and extent of anthropogenic climate change can be achieved.

## The Dilemma

In situ temperature records from worldwide surface weather monitoring sites indicate that, globally averaged, surface air and sea temperatures have risen by 0.30 to 0.60 deg. C between the late 19<sup>th</sup> century and 1994, and have risen by at least another 0.10 deg. C since then (IPCC, 2001). Figure 1 shows annual anomalies of combined surface-air and sea surface temperatures (in deg. C) from 1861 to 2000 relative to 1961 to 1990 values for the northern hemisphere (Fig. 1a), the southern hemisphere (Fig. 1b), and the globe (Fig. 1c) as reported by the IPCC (2001). Annual averages are shown as red bars with  $2\sigma$  confidence intervals (twice the standard error of measurement) shown as demarcated black bars. The data are from in situ land and sea based temperature records that have been gathered and analyzed by the U.K. Met. Office (UKMO) and the Climate Research Unit (CRU) ( Jones et. al., 2001). The underlying trend is shown after averaging with a standard weighting method (dashed lines - IPCC, 1996) and after optimum averaging using variance-covariance matrices instead of correlation functions (Shen et. al., 1998; Folland et. al., 2001). Urban heat island effects (the tendency of temperatures to be artificially higher near urban centers, apart from large scale climatic trends) have been accounted for in these analyses. Using a wide variety of proxy indicators of land and sea surface temperatures, including ice cores, tree ring cores, varved lake sediments, historical records, and more, this analysis can be extended back nearly a millennium. Fig. 2 (IPCC 2001 fig. 2.20) shows the historical northern hemisphere land and sea surface temperature record from 1000 A.D. to 1998 A.D. as determined by Mann et. al. (1999). Data taken directly from in situ instruments as in the previous figure are shown in red. The blue and black curves show, respectively, a 1000 to 1980 A.D. reconstruction from this data and a 40 year smoothed representation of the underlying trend (IPCC, 2001), and the dashed purple curve shows the 1000-1900 A.D. linear trend. The shaded gray region gives  $2\sigma$  confidence intervals. Not surprisingly, the older proxy data has considerably more uncertainty than the more recent datasets. But even so, it can be clearly seen that the last century (particularly, the last few decades) show highly unusual warming trends compared to the long-term historical record. Mann et al. (1999) concluded that as of 1999, the 1990's was the warmest decade since 1000 A.D. and 1998 was the warmest year. Similar conclusions were reached using independent methods and analyses by Jones et al. (1998) and Crowley and Lowery (2000). Natural climatic variation due to solar variability, El Nino's and other interdecadal oscillations, and catastrophic events such as volcanic eruptions are contributing to these trends. But increasingly, the evidence suggests that they are largely of anthropogenic origin, and the anthropogenic contributions are likely to increase significantly over the next century unless active mitigation measures are taken (IPCC, 2001).

For the last 15 to 20 years, independent analyses of global climatic temperature trends have been made using mathematical simulations of global atmospheric climate in the hope of independently verifying the in situ and proxy temperature records, and to forecast the trends that can be expected over the next century based on current and projected human industrial and land use activities. These models range from simple models that are intended to characterize one or two particular phenomena (e.g. carbon sequestration by oceans or tropical rainforests, mass and energy transport by oceanic thermohaline cycles, or greenhouse gas emissions) to more complex three dimensional models that are intended to simulate larger regions of global climate using inputs from in situ data and the results of simpler models. The most complex of these are coupled suites of three dimensional oceanic and atmospheric general circulation models (AOGCM's) that make use of various modeled inputs (IPCC, 2001, Chap. 8). In these simulations, independent models of the world's oceans and the atmosphere are joined together and supplemented with modeled inputs such as sea-ice and fresh water fluctuations, greenhouse gases and particulates, volcanic eruptions, and more. Typically, these models are initialized at some previous point in history and integrated over time (or "spun up") until they produce a reasonable simulation of the existing global climate prior to the industrial age. Some of these models require non-physical adjustments in mass transport, energy, or other factors before they will produce stable representations of global climate. These "flux adjustments" can be viewed as corrections for various uncertainties in modeled characteristics of the global climate, and are not representative of actual climate behavior when forced. Increasingly, current state of the art AOGCM's do not require them to produce realistic climate simulations. With or without these adjustments, once an AOGCM is spun up to a reasonable and stable global climate, it is then forced with simulated natural and human induced atmospheric forcings to evaluate how the global oceanic climate will respond. AOGCM's allow for a wide range of input scenarios to be evaluated, and this allows us to test the outcomes of many possible mitigation scenarios for how the

next century will play out in regards to global warming, and how effective our various proposed solutions to it are likely to be.

Reliable AOGCM's are crucial to reasonable forecasts of how our activities can be expected to impact future global climate. Our confidence in their ability to produce these forecasts requires at least that they be able to accurately simulate the existing climate change of the last century. Results have been mixed. Current state of the art models reproduce certain global features, such as surface air temperature and large scale ocean/atmosphere heat transport, quite well. Other features, like precipitation and cloud coverage, are not well simulated (though there has been increasing success in modeling some regional monsoons). In general though, the best of these models produce global average climates that are reasonable representations of observation on many if not all important points. One of the more important areas of concern is how well these models represent the vertical structure of the lower atmosphere, particularly the troposphere (NRC, 2000). Generally, AOGCM's predict that surface-air temperatures and lower to middle troposphere temperatures should evolve at similar rates. For at least the last 25 years, this has not been observed. Though surface-air temperatures have warmed considerably during this period (at least 0.16 to 0.20 deg. K/decade), similar tropospheric trends have not been observed. Some of this is known to be due to short-term climatic fluctuations like El Nino's and volcanic eruptions (most notable El Chicon in 1982, and Mt. Pinatubo in 1991), and recent AOGCM's have enjoyed some success in correcting for these. But even after these corrections, a statistically significant difference remains that is likely to be real. This discrepancy brings into question how well current AOGCM's model the vertical structure of the atmosphere. It is generally agreed that to improve the current generation of AOGCM's and the reliability of their forecasts, we must gain a better understanding of vertical and horizontal latent heat transport and how the atmosphere is forced (IPCC, 2001; NRC, 2000). But many others have pointed to this discrepancy as proof that AOGCM's cannot be used to detect an anthropogenic "fingerprint" in recent global warming, and in some cases, even that the atmosphere has not warmed at all over the last century at all. Discussions of these discrepancies have generated much controversy over the last decade in both the scientific community and among policy makers as well. The tropospheric temperature record of the last 25 years is commonly cited as proof that anthropogenic global warming is not happening, and the United States should not adopt the Kyoto Protocol for greenhouse gas reduction targets (Douglass et al., 2004; 2004b; Ferguson and Lewis, 2003).

Though surface temperature trends are based on a wide range in situ and proxy historical data, lower and middle troposphere temperatures have come primarily from two sources – radiosonde data from selected global locations and radiative brightness temperature data from Microwave Sounding Units (MSU's) flying aboard NOAA/NASA Polar Orbiting Environmental Satellites (POES's). The radiosonde record dates from the middle of the 20th century, and the MSU satellite record goes back to 1979<sup>1</sup>. To settle the current controversy, these records must be examined, including how they were generated, where their strengths and limitations are, and how they do or do not relate to the surface record and the predictions of AOGCM's. Of the two records, the MSU record is generally thought to provide the most hope for resolving the discrepancies. Though less fine grained than the radiosonde data, it is the only extant record that provides true global coverage of the lower atmosphere, and its error characteristics are better understood. Both however have strengths and weaknesses that the other lacks, and are used to independently evaluate each other. In what follows, we will examine the MSU satellite and radiosonde records in depth, paying special attention to these strengths and weaknesses, and how they relate to each other. Then we will examine how the various datasets have been used and misused in discussions of global warming in both scientific and public policy forums.

## Atmospheric Temperature Monitoring with MSU and AMSU Products

In April of 1960, NASA and NOAA (then known as the Environmental Science Services Administration, or ESSA) began operating Polar Orbiting Environmental Satellites (POES) under a joint program for a variety of weather forecasting and short-term climate science studies. These satellites revolutionized weather forecasts by providing for the first ever space based global weather and storm observations in real-time. The first in this series was the No. 1 Television Infrared Observation Satellite (TIROS-1) which carried low and high resolution television cameras and operated for 78 days. Successive generations of TIROS class satellites expanded NOAA weather forecasting capabilities throughout the 60's and 70's. TIROS-N, which was put into service in early 1979, was the research prototype for a whole new class of TIROS satellite that carried a much wider array of sensor packages with capabilities that had been considerably enhanced from its predecessors. It provided real-time visible imagery of cloud and weather patterns, and infrared data on atmospheric and oceanic temperatures, humidity, ozone levels, snow and sea ice cover, and a variety of other climatic parameters. After TIROS-N, the following satellites were assigned letter designations during development, and then numeric designations when put into service. For instance, TIROS-N was followed by NOAA-A, which became NOAA-6 on its launch date of June 27, 1979. Between 1978 and 2001 NOAA operated 9 generations of these spacecraft, culminating in NOAA-J (14) which was placed into service in December of 1994 and operated as the POES afternoon observation satellite until March of 2001. Beginning with the launch of NOAA-K (15), NOAA began operating a newer and more advanced version of the TIROS-N Series platform - the Advanced TIROS-N (TIROS-ATN) that carries an expanded and updated sensor packages compared to TIROS-N. The latest of these spacecraft, NOAA-17 (shown in Figure 3) was launched on June 24, 2002. As of this writing, NOAA-16 and NOAA-17 are the currently designated operational spacecraft in the POES program.

### TIROS-N Satellites and MSU Products

Figure 4 shows the TIROS-N flight path with respect to the earth's surface. TIROS-N and its successors followed sun-synchronous polar orbits. An orbital plane precession rate of  $0.986^\circ$  kept the sun, earth, and the orbital plane of each spacecraft in a common plane so that a constant relative daylight illumination was maintained throughout the year as the earth orbits the sun. They relied on the earth turning under their orbital plane to achieve fully global views. Equatorial crossings were at 25 degree latitudinal separations resulting in 14.1 orbits per day and repeated Local Equatorial Crossing Times (LECT's) every 8 days. NOAA operated these satellites in a two-satellite pattern at all times so that data was gathered four times daily at any particular location. Full global coverage was achieved by each operational pair of satellites every 3 to 4 days. On-board sensor packages included the Advanced Very High Resolution Radiometer (AVHRR) and the TIROS Operational Vertical Sounder (TOVS), which was composed of three different sensors that passively measured incoming infrared and/or microwave radiation. Of the three TOVS components, the one that is most important for tropospheric temperature studies is the Microwave Sounding Unit (MSU), which measures upwelling microwave radiation from the surface to the lower stratosphere.

Figure 5 shows how the MSU's monitored the earth/atmosphere system from the POES orbital track. The MSU sensors, which are manufactured by the Jet Propulsion Laboratory (JPL) in Pasadena, CA, are four-channel Dicke radiometers that consist of two reflector antennas with two channels each operating at 50.3, 53.74, 54.96, and 57.95 GHz. Microwave energy received passively by each antenna is separated into vertical and horizontal polarization components by an orthomode transducer and sent to one of these 4 MSU channels. Nominal beam width is 7.5 deg at full width half maximum power (FWHM). Every 25.6 seconds these antennas scan a 47.4 deg (to beam centers) off-track portion of the sky to either side of the directly downward looking, or nadir, direction. Data was sampled at 11 positions separated by 9.47 degrees and recorded as digital "counts". Each count corresponds to a fixed amount of microwave energy received by the radiometer beam that is then assigned to one viewing pixel, so that count number gives the radiation density per pixel. Count number, and therefore radiation density, are proportional to the temperature of the emitting body (in this case, the atmosphere, and to a lesser degree, the earth's

surface). Positions 1 and 11 are the extreme scan angles at 47.4 deg. left and right (labeled “Scan Point” in Figure 5), and position 6 is the Nadir view which looks directly at the satellite subpoint along its orbital flight path. At each position during scan a Dicke switch connected to the incoming signal switched between the signal and a microwave load at instrument temperature. Once per Earth scan the MSU made a calibration measurement by checking the temperature of deep space (2.7 deg. K) against own onboard “hot calibration target”, of which there are two – one for Channels 1 and 2, and another for Channels 3 and 4. These hot calibration targets are quasi-blackbody emitters operating at known temperatures that were monitored by two platinum resistance thermistors according to pre-launch thermal vacuum chamber calibrations by JPL. The MSU radiometers differenced microwave emissions from these hot calibration targets and deep space (at 2.73 deg. K) to generate a calibration scale against which to compare temperature readings. The scan period of 25.6 seconds, when combined with satellite orbital speed yielded a scan to scan separation of roughly 150 km and nadir and extreme angle spatial resolutions of 110 km and 200 km respectively.

### Advanced TIROS Satellites and AMSU Products

Beginning with NOAA-15 (launched on May 18, 1998), the TIROS-N spacecraft were replaced by the Advanced TIROS-N, or TIROS-ATN spacecraft and are the current platforms for the POES program. Like their predecessors, they follow sun-synchronous polar orbits in a two-satellite profile with orbital plane precession rates that guarantee a constant sun-earth illumination profile with respect to the spacecraft throughout the year. Onboard sensor packages have been expanded and upgraded compared to their TIROS-N counterparts. The most important change for tropospheric temperature measurements was the upgrade of the TOVS package to include the new Advanced Microwave Sounding Unit (AMSU) products from JPL. The AMSU is now separated into three components – AMSU-A1, AMSU-A2, and AMSU-B – that collectively monitor 20 channels at frequencies ranging from 23.8 GHz to 183.3 GHz. With the expanded frequency range, the AMSU package now provides detailed information about temperature profiles from the surface up to an altitude of 3 mb (45 km)<sup>2</sup> and more sensitive measurements of atmospheric humidity and cloud pattern profiles. The channels of greatest interest for tropospheric temperature profiles are AMSU-A1 Channels 3, 5, 7, and 9 (respectively, 50.3, 53.6, 54.94, and 57.29 GHz), which correspond closely with the 4 TIROS-N Series MSU channels. Figure 6 shows a schematic of the TIROS-ATN platform and the new sensor packages it carries.

Like the TIROS-N Series MSU package, AMSU-A1 is a cross track scanning radiometer with an instantaneous field of view of 3.3 deg. At full width half max power. Functional performance is similar to that described above for MSU devices, but with increased sensitivity. Cross track views per scan have been increased from 11 to 31 with view separations reduced to 3.33 deg. from 9.47 deg. and one full scan spanning a 48.3 deg. view to either side of nadir. The time for a full scan, including deep space and onboard hot target calibration views has been reduced from 25.6 seconds to 8 seconds. At TIROS-ATN orbital speed, this reduces the scan to scan separation from 150 km to 47 km, and nadir and extreme angle spatial resolutions from 110 km and 200 km to 48 km and 87 km respectively, giving much higher image resolution. Temperature sensitivity (**NEΔK**) has dropped to 0.25 from 0.30 on Channels 3, 5, and 7 where tropospheric temperatures are most closely monitored – a 17 percent improvement. As with TIROS-N Series MSU products, hot target calibration temperatures are determined by monitoring an onboard quasi-blackbody radiator at microwave frequencies with platinum resistance thermistors, and calibration was done in thermal vacuum chamber tests by JPL prior to launch. The existing configuration yields a hot target temperature of 200 deg. K with an accuracy of better than 0.2 deg. K.

## **MSU and AMSU Datasets**

The four MSU channels and their AMSU counterparts collectively monitor temperatures of the troposphere and lower stratosphere, with TIROS-N based MSU Channels 1 through 4 corresponding respectively to AMSU Channels 3, 5, 7, and 9 on the TIROS-ATN platforms. Each of these channels receives its input from broad layers of the atmosphere weighted according to sensitivities that vary as a function of altitude with peak signal occurring at different altitudes for each channel (weightings as a function of altitude for each of the 4 channels are shown in Figure 7). What is measured is the total radiative brightness of that layer with most of the “weight” coming from the altitude of peak emission. This emissive profile is used to derive an average, or bulk, brightness temperature of the layer in question. Each of the 4 channels overlap to some extent. There are slight differences between the weighting functions of the MSU and AMSU products, but they are small enough to be neglected for tropospheric temperature calculations.

Channel 1 (50.3 GHz) is sensitive to the lowest two or three kilometers of the atmosphere. Data from this channel is heavily contaminated by emissions from the surface and atmospheric water and ice and is of limited utility for tropospheric temperature studies. Channel 2 (53.74 GHz) monitors a deeper portion of the atmosphere with its weighting peaked at an altitude of 500 hPa, or approximately 7 km<sup>2</sup>. Emissions monitored on this channel are relatively insensitive to humidity and thus are more representative of actual deep troposphere temperatures. This channel also receives significant input from the surface – up to 10 percent over oceans and 20 percent over mountainous regions such as the Himalayas (NRC, 2000). Channel 2 is also affected by precipitation sized ice particles in deep convective cloud regions which can contaminate data from mid-latitude squall lines. Channel 3 (54.96 GHz) measures a deeper portion of the troposphere than Channel 2 and has a weighting function that peaks at an altitude of 250 hPa. At this altitude it straddles the tropopause in extra-tropical latitudes, and this is of limited utility in generating a pure troposphere signal. Channel 4 (57.95 GHz), which has a weighting function peaking at 70 hPa, receives most of its signal from the lower stratosphere, with the remainder coming from the very uppermost regions of the troposphere. Because the stratosphere is known to be cooling with respect to the troposphere, largely as a result of ozone depletion and is also significantly impacted by the effects of aerosols and volcanic eruptions, it may respond quite differently than the troposphere to these and other forcings (Bengtsson, 1999; NRC, 2000; IPCC, 2001). Channel 4 measurements provide important ancillary data about how the lower stratosphere might be impacting measurements from other channels.

Because the MSU/AMSU Channel weighting functions overlap and the different scan angles of cross track views see the atmosphere at differing angles, information from different channels and view angles can be combined to generate “synthetic” channels that represent thinner portions of the atmosphere. For instance, a fractional portion of the channel 3 signal can be subtracted from the channel 2 signal to yield a signal representing a thinner portion of the troposphere with less contamination of stratospheric signals. Other combinations of signals from nadir and off-nadir views can be combined to isolate other layers. In the early 90’s, it was shown that by combining Channel 2 data at nadir with fractionally weighted data from 8 of the original 10 off-nadir views, a weighted measurement could be derived that emphasizes a thinner layer of the lower troposphere centered at 4 km altitude (Spencer and Christy, 1992b). This virtual channel has come to be called MSU2LT (for Channel 2 Lower Troposphere). With the advent of the AMSU products, the same philosophy can be followed using the 30 view angles of AMSU products. The newer AMSU weighted lower troposphere view is now called MSUTLT (Christy et al., 2003). Figure 7 shows the weighting function of the MSU2LT channel (labeled as MSU/AMSULT). The MSU2LT and MSUTLT channels remove most of the stratospheric influence, but they are based on a wide angle swath that neglects zonal<sup>3</sup> (that is, east to west) temperature gradients. As such, they are sensitive to temperature fluctuations that happen on timescales comparable to the east to west drift in the satellite’s equatorial crossings and are thus subject to sampling errors. They are also sensitive to direct surface emissions.

As a measure of deep atmospheric temperatures, the MSU and AMSU records far exceed that of other surface and radiosonde based temperature measuring products. Since they see much larger vertical profiles and have far more global geographic coverage than other products, they are less subject

to local surface layer noise than ground based upper air stations. With predictable orbits, they are less subject to the “geographic noise” of relocated measurement platforms that plague radiosonde and rocketsonde data. Thus, they provide a degree of consistency and coverage that cannot be replicated by other means. Because of this, most investigators today believe that the MSU and AMSU records present the best available option for resolving open questions regarding climate change and tropospheric temperatures. Even so, it must be remembered that the TIROS-N and TIROS-ATN Series of spacecraft were designed to support local weather forecasting and shorter term regional climate studies. They are lacking in many important respects for long-term climate change studies and the shortcomings must be taken into account. These can be classified into a few groups as follows.

### Poor Vertical Resolution

Unlike radiosondes, which take discreet measurements at specific altitudes, all MSU products measure weighted averages of large portions of the sky. This is both a blessing and a curse. Bulk temperature measurements of atmospheric layers are more likely to be stable than highly localized vertical temperature profiles and meaningful for evaluation of long-term climate change. But at the same time, there are many large-scale phenomena important to tropospheric temperatures that occur on vertical and geographic scales that MSU products cannot resolve. For instance, tropospheric heating is impacted by various vertical vs. horizontal heat transport mechanisms that at times can isolate the upper atmosphere from the surface (NRC 2000; Spencer and Christy, 1992b; Gaffen et al., 2000b; Trenberth and Stepaniak, 2003; Trenberth and Stepaniak, 2004). MSU measurements are too coarse to resolve these effects, yet they can significantly impact local and globally averaged tropospheric temperature trends.

MSU channels also see portions of the sky that are expected to have different long-term behavior due to global warming. Channel 2 for instance, sees the mid and upper troposphere which are expected to warm under natural and anthropogenic forcings, but it also sees the lower stratosphere which is expected to cool due to ozone depletion and other effects (we will see later that this has recently been proven to be a crucial point). Some of the stratospheric “noise” problem can be avoided by using mixed channels such as MSU2LT and MSUTLT. Since these Channels use difference methods to weight observations toward the surface, they reduce the lower stratospheric signal “pollution”. But these difference methods rely on side looking views that because of the curvature of the earth and the increased length of optical viewing paths (which are long compared to temperature gradients over the regions being viewed). As such, they significantly amplify the sampling noise between contributing channels which more than doubles the error in MSU2LT vs. MSU2 (Spencer and Christy, 1992b; NRC 2000). Similar problems plague the later MSUTLT record. MSU2LT and MSUTLT also pick up a significant percentage of their signal from the surface - roughly 10% over ocean, 20% over land (NRC 2000; Spencer and Christy, 1992b). This pollutes their lower troposphere data with surface effects that are even more problematic than the stratospheric impacts on Channel 2. Spencer and Christy have estimated the impact of this noise at approximately 0.03 deg K/decade (Spencer and Christy, 1992b) but recent estimates suggest the error may be as high as 0.1 deg K/decade, making the noise from MSU2LT and MSUTLT channels equal to or greater than that of most extant trend measurements from these products (Stendel et al., 2000). These factors complicate how MSU data are to be weighted and widely varying answers can be obtained from a few equally reasonable assumptions (NRC 2000). One other factor has come to light recently as well. It now appears that the lower troposphere MSU2LT and TLT products might be polluted at high southern latitudes by Antarctic sea-ice. Because these channels receive 10 to 20 percent of their input from the surface and sea-ice has a very high albedo compared with the ocean, an increase in sea-ice will appear to them as spurious warming, a decrease will appear as cooling. The same affect will be present at the North Pole, but to a diminished degree owing to the smaller fraction of sea-ice present at higher latitudes. There is evidence to suggest that strong cooling trends at high southern latitudes in 2LT and TLT products may be related to annual and inter-annual sea-ice variations rather than lower troposphere temperatures (Swanson, 2003).

## Short Temporal Record

MSU and/or AMSU products have been in service since 1979, making their time series about 26 to 27 years long as of this writing. This is too short for useful predictions of global change. Anthropogenic impacts on global climate (at least, as they relate to current concerns about global warming) are thought to have originated at the dawn of the Industrial age over 2 centuries ago, and can only be clearly separated from natural effects such as solar variability since the mid-20<sup>th</sup> century (NRC 2000; IPCC 2001). Against this time scale, the MSU record provides an incomplete time series at best. Many tropospheric and stratospheric phenomena which impact this record have occurred on time scales that make up a significant percentage of the MSU record as well. During the last 25 years there have been 4 ENSO events (El Niño Southern Oscillation) in 1983, 1987, 1991, and 1997, and at least 2 volcanic eruptions (El Chichón in 1982 and Mt. Pinatubo in 1991) that are known to have significantly impacted both stratospheric and tropospheric temperature time series. These impacts have ranged from one to five years so that collectively they have impacted a majority of the MSU record to at least some degree (Christy and McNider, 1994; Jones, 1994; Santer et al. 2001; Wigley and Santer, 2002). These effects must be separated from the background tropospheric record before meaningful predictions can be made. Thus, it is unlikely that the last 27 years will be representative of decades to come and care must be taken in basing predictions on them (NRC 2000).

## Onboard Instrumental Biases

MSU sensory devices also suffer from a number of systematic equipment errors. The most important of these is drift in sensor gain (i.e. – the ratio of the perceived signal to the actual signal). MSU sensors detect upwelling atmospheric microwave emissions, amplify them for processing, and convert the radiosity measurements into temperatures based on known physical laws that relate the two. In the original MSU design, microwave emissions are detected by the two MSU antennas and juxtaposed with counts from the hot target microwave load at a known instrument temperature and deep space (2.73 deg. K) by the Dicke switch. Then they are passed through a 10-110 MHz low noise mixer, reamplified, and digitized by a 12 bit analog-to-digital converter to obtain the final data that is sent to ground stations. MSU sensors record temperature as “digital counts” which correlate to a microwave radiosity that in turn correlates to a temperature. The relationship between radiometer counts and temperature is non-linear, and the non-linearity must be calibrated for each MSU package prior to launch. If this calibration is not done accurately, non-linear errors in radiometer gain will result that can be very hard to correct for later. It now appears that this has happened for some of the MSU packages, NOAA-12 in particular (Mo, 1995). Uncertainties in non-linear calibration coefficient like these cannot be removed later with hot target calibrations.

In addition to direct equipment calibration non-linearities, another non-linear source of error has been identified that varies with the temperatures of the hot target and the physical radiometer components themselves. Though the hot target is in principle fixed at a known temperature, most experience at least some drift in temperature over their service lives. As the sun illuminates radiometer components from different angles during the satellite’s orbit, the sensor components themselves will also vary in temperature. These component temperature variations will affect the gain of the radiometer components and the temperature calibration in a non-linear manner. This has come to be known as Instrument Body Effect (IBE). Some of this IBE can be measured and corrected for in post launch analysis because differences in perceived temperature between pairs of satellites will identify those satellites with non-linear gain anomalies and allow for the combining of all overlapping satellite records in ways that reduce the resulting error – a process known as “merging”. A gain correction function that varies with time can then be applied to those satellites for which a non-linear signal response is found (Christy et al., 2000). In addition, over longer time series and many measurements much of the noise from these variations will average out and the correction functions will become more reliable. In this manner, a significant portion of non-linear noise can be accounted for. But this process cannot remove all errors, particularly residual calibration errors not correlated with radiometer component temperature, and some errors will persist (NRC 2000).

## Orbital Decay and Attitude Drift

During its service life of 2 to 8 years, episodic solar wind events will cause the satellites to experience some orbital and attitude degradation. As the spacecraft orbit decays it will lose altitude, and MSU and AMSU sensors will see different portions of the sky from distances other than those expected (Wentz and Schabel, 1998). Slight variations in satellite attitude (roll orientation) will also alter MSU scan views. Nadir will be slightly off of nadir for instance, and other views will also be similarly degraded. This will have the effect of eroding the weighting functions for each scan and introducing spurious temperature variations (Mears et al., 2002). Fortunately, these two effects can be accounted for. The orbital mechanics for the POES spacecraft are well known and altitude degradation can be precisely determined. Loss of altitude can be combined with corrections to radiative transfer functions to remove orbital decay effects. Satellite attitude bias is systematic across all scan views so that when combined with known orbital decay, roll induced variations can be adjusted to the nadir view. These corrections have been applied to most recent datasets (Wentz and Schabel, 1998; Mears et al., 2002; Christy et al., 2003).

## Diurnal Sampling Variations

As already mentioned, TIROS satellite orbits are sun synchronous with a daily orbital precession rate of 0.986° that preserves orientation and LECT's throughout the year. However, for most TIROS platforms, this precession rate is not perfect, and slight variations in it introduce a drift in orbital path with respect to the sun and earth. This drift over extended periods causes the daily equatorial crossing times of each satellite to occur earlier or later in the day than expected with respect to local ephemeris time. Since temperature varies on a diurnal cycle from day to night, this introduces another spurious temperature variation into MSU products. Drifts of up to 0.5 hr/year have been observed. Diurnal sampling errors, though systematic in their effects, occur randomly and correlate imperfectly with observations, which can make them difficult to account for in datasets. One approach is to compare data from view angles to the left and right of nadir at predetermined daily times during ascending (northbound) and descending (southbound) orbital segments. Data from different view angles will correspond to different local times at different points in satellite orbit – from roughly an hour at the equator to several hours at the poles. Observations of these differences over many ascending and descending orbital cycles allow for a drift rate to be backed out of the data variations (Christy et al., 2000). The approach is straightforward, but the resulting calculation is quite involved and compounds many other uncertainties making the derivation of reliable numbers troublesome. Another approach is to use high resolution climate model simulations to recreate the diurnal brightness temperature variations expected along the satellite orbital path. The data from such models may then be used to correct each measurement to local noon and remove diurnal drifts. The modeled diurnal cycle temperature variations can in turn be validated against actual morning to evening temperature variations as observed by MSU platforms themselves (Mears et al., 2002). With proper validation of the climate modeling against observations, this methodology has the advantage of considerably reduced sampling noise compared to the cross scan methodology.

## Intersatellite Data Drift

To date, NOAA has operated 11 TIROS-N and/or TIROS-ATN Class satellites since 1978. In theory, these spacecraft and their instrument packages should all behave identically. The MSU and AMSU packages in particular all contain the exact same components (per class) and are all calibrated prior to launch according to the same guidelines and protocols. In theory! In practice however, significant differences in measured brightness temperature have been observed between pairs of satellites whose service lives overlap - up to 0.4 deg. K (Hurrell and Trenberth, 1998; Christy et al. 1998; NRC, 2000). Much of this difference is related to IBE but not all of it. It is not clear where the remainder is coming from (NRC, 2000). Analyses to date have accounted for this drift by comparing data from pairs of satellites whose service lives overlap each other. Typically, during service life overlap, data from co-orbiting pairs of satellites are smoothed with running averages to remove noise and facilitate comparison. Then, after correcting for other noise sources, a single continuous time series can be created from the independent

records by generating a single curve that minimizes the differences with the individual records – in other words, “merging” them. Typically, merge calculations of this sort are done with a least squares mathematical technique and the reliability of the merge is evaluated with a statistical analysis of the residuals (the leftover differences between the final time series and the individual satellite records).

For this to work, it is important to have a long enough overlap period between co-orbiting satellites for a statistically significant comparison of records. In most cases, overlaps are long enough for this to be done. But a few of the overlaps, particularly the NOAA-9/NOAA-10 overlap, are not. NOAA-10 was put into service in September of 1986. The NOAA-9 MSU unit failed unrecoverably in May of 1987 and produced no more data after that (the satellite was deactivated in February of 1998). As a result, only 90 days of overlapping data were generated between these two satellites - not adequate for a truly solid merge of the two. Accounting for this overlap has been the most problematic challenge facing the MSU/AMSU record. The merging procedure must also correct for the IBE and other equipment non-linearities as well. But for the reasons noted above, this can only be done to a certain extent.

Based on these considerations, differing dataset merging methodologies have been applied to the MSU/AMSU record, resulting in trend differences that at one time or another have spanned a 0.76 deg. K/decade spread (Mears et al., 2002, Christy et al., 1998; 2000; 2003; Vinnikov & Grody, 2004). In some cases the short overlaps have been omitted entirely from long-term datasets (Christy et al., 2000; 2003). In others, it has been used and evaluated against the remaining long-term trends (Mears et al., 2002). Of these, the methodologies with the lowest error characteristics and the best fit with overall long-term trends have been used and corrections have been made as the sources of error were discovered (Christy et al., 1998; Christy et al., 2003; Mears et al., 2002). But even after nearly 15 years of error removal, there are still considerable differences in the results arrived at by the latest analyses. To date, it appears that the poor overlap of the NOAA-9 record with that of adjoining satellites is the largest contributor to MSU/AMSU dataset uncertainties, accounting for at much a 65 percent of the total spread between currently existing analyses based on MSU/AMSU products (Karl et al., 2002).

### Antarctic Sea-Ice and Melt Pool Albedo Impacts

MSU and AMSU devices measure atmospheric temperatures as digital counts of radiation under the presumption that the primary source of this radiation is upwelling emissions from the atmosphere that follow the weighting functions of each channel (see Figure 7). They are not able by themselves to determine where the radiation counts they are measuring originated from. Thus, the accuracy of their measurements of troposphere temperatures and trends depend on the presumption that all digital counts are the result of atmospheric radiation and the reliability of the channel weighting functions. The high southern latitudes (above 60 deg. S) contain significant portions of sea-ice of Antarctic origin. Likewise, high northern latitudes exhibit significant sea-ice fraction of their own, though these are of considerably smaller extent than their Antarctic counterparts. During the austral summer melt season this sea-ice shrinks to an area of roughly 3 million km<sup>2</sup>, and then increases to 18 million km<sup>2</sup> during the winter and spring (Cavalieri et al., 1997). Because of its high albedo, sea-ice area increases will increase the surface generated digital counts to MSU and AMSU sensors, and will appear to them as a spurious warming. Likewise, decreases in this sea-ice and the corresponding increases in melt pond area will appear as a cooling. Sea-ice area data has been gathered since the late 70's by Scanning Multichannel Microwave Radiometer (SMMR) and Special Sensor Microwave/Imager (SSM/I) devices flying aboard NASA Nimbus-7 and Defense Dept. Meteorological satellites. These devices are passive microwave detectors similar to the MSU and AMSU devices aboard NOAA POES satellites, and measure sea-ice as brightness temperature readings. Data from these platforms shows that between 1979 and 2002 Arctic sea-ice shrunk by about 360,000 km<sup>2</sup> and overall Antarctic sea-ice increased slightly, though West-Antarctic sea-ice decreased (Cavalieri et al., 1997). However, radiosonde data from the Antarctic region also indicates that during the satellite era there may have been lower troposphere warming as well as high troposphere cooling due to ozone depletion (Thompson and Solomon, 2002). In addition, the breakup of the Larsen A and B ice shelves (in January of 1995 and March of 2002, respectively) also might have contributed to relative sea-ice and relative ocean/melt pool areas. The Antarctic peninsula has warmed considerably in recent decades, as shown by various direct and proxy measurements reveal (Vaughan et al., 2001). The resulting warmer temperatures have likely increased melt pool area on the ice that remains during austral summer. These melt pools are indistinguishable to MSU and AMSU

devices from the surrounding ocean, and thus must be considered as part of the surface radiation balance from higher latitudes. Recent comparisons of daily anomalies at southern latitudes higher than 60 deg. Has revealed a strong correlation between Antarctic ice summer melt periods and 2LT and TLT cooling trends (Swanson, 2003). Due to its higher altitude weighting, these fluctuations will exhibit only a minor impact on the Channel 2 trend and thus will have little impact on the MSU/AMSU products of the RSS, Prabhakara, and VG teams.

### **MSU/AMSU Analysis Products**

It has already been noted that the MSU and AMSU records were originally intended for regional weather forecasting and climate studies. For applications such as these, the errors described above are usually small in comparison with the data being generated and are of little consequence. But when studying long-term global climate change, small errors can add up quickly, and a great deal of data massaging is necessary to make them useful. One can get an idea of the degree of massaging involved from the fact that after nearly 15 years, the most mature of these analyses has recently undergone its 5<sup>th</sup> major revision (Christy et al. 2003) and one update of it (Christy et al., 2004). Yet its conclusions differ considerably from those of other analyses based on different data reduction methods.

As of this writing, there are 4 noteworthy analyses of tropospheric temperature trends based on MSU and AMSU products. Each makes use of a different approach to dealing with the uncertainties in MSU and AMSU records and each has its strengths and its weaknesses. All account for solar wind induced orbital decay and known calibration related instrument biases. The chief differences between them lie in how they deal with diurnal sampling errors and IBE, and how they merge independent MSU/AMSU records to create a 20 to 25 year continuous time series.

### **The University of Alabama, Huntsville (UAH) Team**

The earliest use of MSU products for climate change studies dates to the early 90's when a team led by John. R. Christy of the University of Alabama, Huntsville and Roy W. Spencer of the NASA Marshall Space Flight Center published the first revision of their MSU based lower troposphere studies. Christy and Spencer's team (hereafter referred to as UAH for University of Alabama, Huntsville) assembled a full time series of MSU data from 1978 to 1990 for the lower troposphere. It was the UAH team that derived the MSU2LT and MSUTLT synthetic channels to emphasize the lower troposphere (peak weighting at roughly 3.5 km) by differencing Channel 2 nadir and off-nadir views. They also developed another synthetic channel known as MSU2RT in their earlier analyses, though this product is seldom used any more, having been discarded in favor of the better characterizes 2LT and TLT products.

In their first analysis, UAH combined data from TIROS-N through NOAA-12 using a simple merging procedure to remove then known instrument biases and formed a consistent lower troposphere temperature time series that was designated as UAH Version A (Spencer & Christy, 1990; Spencer & Christy, 1992a; 1992b). By 1994, enough new data had been gathered to justify a reanalysis. It had been discovered that NOAA-11 had experienced eastward drifts in its diurnal cycles over its service life that increased its LECT's and introduced a spurious warming into the dataset. It was at this time that UAH developed the MSU2LT channel to replace MSU2RT. Using this new channel, and incorporating updates to corrections for missing data and satellite overlaps, Version B was released (Christy et al., 1995).

At the time, Versions A and B spurred much debate because they indicated a decadal cooling trend for MSU2R and MSU2LT. Version B showed a cooling of -.058 deg. K/decade, which was 0.03 deg. K/decade cooler than that of Version A. These trends were inconsistent with what climatologists expected for the lower troposphere based on the predictions of the best AOGCM's of the day. Shortly after the release of Version B, it was discovered that NOAA-12 was suffering from unusually large instrument biases, and that NOAA-7 had experienced a considerable drift in its LECT compared to that of NOAA-11. Correcting for these errors lead to Version C (Christy et al., 1998), which produced a trend that was 0.03 deg. K/decade more positive than that of Version B. After the release of Version C an

additional source of instrument gain error was discovered that impacted MSU2 and MSU2LT. A new set of NES-DIS non-linear calibration corrections became available for NOAA-12 which allowed for more accurate estimates of its non-linear instrument biases. Then, it was discovered that solar wind drag had caused the orbits of all NOAA POES satellites to decay, resulting in losses of altitude of up to 15 km. These altitude changes introduced spurious cooling into the MSU time series (Wentz and Schabel, 1998). Incorporation of these effects led to the release of UAH Version D, which yielded a trend of +0.06 deg. K/decade for MSU2LT and 0.04 deg. K/decade for MSU2 (Christy et al., 2000).

In later versions of their analysis products, UAH incorporated a merge procedure that models intersatellite discrepancies with a simple relationship that relates measured and actual atmospheric brightness temperature to non-linear instrument errors using pre-launch calibration coefficients, and to hot target temperature with a constant coefficient called the "hot target calibration factor" that is unique to each satellite. First, diurnal cycle errors were removed by comparing data from off-nadir views along the satellite's track. Since at any given point these views are looking at different locations along lines of latitude, they are seeing different simultaneous times of day. Comparing the evolution of these separate views with time allows for the diurnal drift of each satellite to be evaluated and removed. After removing diurnal cycle errors, data from overlapping satellite records are smoothed to better facilitate comparison. Then these data points are incorporated in a set of some 30,000 equations describing the hot target factors and the offsets. These equations are solved simultaneously using a least squares method to calculate the hot target factors and a unified time series that minimizes the discrepancies between it and the individual records. Typically, a preliminary calculation is done that optimizes calculation of the target factors. These are then used in a more comprehensive analysis to derive the final global time series and decadal atmospheric temperature trends.

UAH released Version 5.0 (as opposed to "E") in 2003 (Christy et al., 2003). In this version, data is analyzed for TIROS-N to NOAA-16 spanning from January 1979 to April 2002. Of these, NOAA-14 was the last to carry the original MSU package. With the launch of NOAA-15 in May of 1998, AMSU packages became the standard. Since these packages utilize more channels and views than their predecessors, the newer analysis was altered to incorporate the data into previous time series. The new lower troposphere channel MSUTLT was developed for this analysis from MSU Channel 2 and AMSU Channel 5, replacing the MSU2LT channel of earlier (see Figure 5). Likewise, two other channels called MSUTMT and MSUTLS were derived to replace Channels 2 and 4 respectively. With the newer data and methods incorporated, UAH Version 5.0 shows a lower troposphere (MSUTLT) temperature trend of +0.061 deg. K/decade and a middle troposphere (MSUTMT) trend of +0.013 deg. K/decade. Version 5.1 (Christy et al., 2004) updated Version 5.0 to include data from Nov. 1978 to late 2003 and an updated estimate of the NOAA-9 hot calibration target factor over that of earlier versions. UAH is updating their MSU2LT and MSUTLT products on an ongoing basis. As of October 2004, MSU2LT was showing a lower troposphere temperature trend of +0.077 deg. K/decade.

In recent years the evolution of UAH methods and analyses has been generally in the direction of increasing decadal tropospheric temperature trends. But their analyses still show less tropospheric warming than is expected from current AOGCM studies – results which continue to generate controversy.

### The Prabhakara (PR) Team

In 1998 another analysis of MSU products was published by a team led by C. Prabhakara of the NASA/Goddard Space Flight Center (Prabhakara et al., 1998). Prabhakara and his colleagues analyzed MSU records from NOAA-6 through NOAA-14 spanning the period from 1980 to 1996 to generate their time series. Satellite orbital decay was taken into account by using only nadir view data (a decaying orbit will only affect the side looking off-nadir views). Diurnal, instrument calibration, and IBE effects were all accounted for in a simple and direct manner. First, PR limit the data from each satellite to only 3 years to minimize the impact of diurnal drifts. Then, all 3 error sources are accounted for by directly analyzing differences between the AM and PM data from each satellite overlap period and qualitatively estimating the simultaneous impact of all three. Then, the individual records were unified into a single time series using the usual least squares method. Analysis was limited strictly to MSU channel 2 (middle troposphere) so as to avoid the noise inherent in MSU2LT. They obtained a middle troposphere trend of 0.109 deg. K/decade – considerably different from that of UAH, and in better agreement with the predictions of AOGCM's and theory (Prabhakara et al., 1998).

In November of 2000 PR published an updated analysis that refined their earlier work with explicit analyses of diurnal drift, instrument errors and IBE (Prabhakara et al., 2000). PR noted that for each satellite average brightness temperature was correlated with diurnal drift, indicating that the temperature changes that resulted from LECT drift impacted the hot target temperature. With this in mind they began by generating separate merged time series for the morning (AM) and Afternoon (PM) satellite data based on the corresponding overlapping records. Having separated the morning and afternoon records, they were able to compare trends with hot target temperature records. These were then used to derive an average hot target factor of 0.03 for all the satellites. With this factor, they derived a new global time series by merging independent satellite records. Their new middle troposphere trend was 0.13 deg. K/decade, also in good agreement with the predictions of AOGCM's but significantly different from the UAH trend.

### The Remote Sensing Systems (RSS) Team

In 2002 another analysis of MSU and AMSU products was published by a team of researchers from the Santa Rosa, CA based company Remote Sensing Systems (RSS). They produced time series for the middle troposphere and lower stratosphere but chose not to analyze the lower troposphere with synthetic channels like MSUTLT due to the noise inherent in these channels. Their time series, which covers tropospheric temperatures from 1979 to 2001, uses MSU Channel 2 and AMSU Channel 5 for the middle troposphere and MSU channel 4 and AMSU channel 9 for the stratosphere in the same manner as UAH analyses. It was the RSS team that first discovered the relationship between satellite orbital decay and measured brightness temperature (Wentz and Schabel, 1998). Their analysis accounts for errors from this decay, diurnal drift, satellite roll, interannual and intra-annual variations in radiometer gain, including IBE and non-linear calibration per the usual pre-launch coefficients (Mears et al, 2003) and spans the period from 1979 to 2001 (TIROS-N through NOAA-16). RSS used analysis methods similar in many respects to those of UAH, but there are a few important differences. Like UAH, they first remove diurnal errors independently of other errors. Then they remove instrument calibration errors, IBE, and satellite offsets with merge calculations that derive hot target factors and a unified time series using the usual least squares technique. Diurnal cycle errors were removed in a manner very different from that of UAH. RSS noted that the UAH method of comparing off-nadir views across the satellite flight path suffers from sampling error and other noise sources. They opted instead to model the diurnal cycle using a model of the cycle which was then directly compared with MSU observations at similar times. This method "leverages" MSU data with higher resolution and avoids the noise sources of the UAH method. Since the diurnal cycle is well understood, and as such, lends itself to modeling, they argued that this results in a better characterization of the diurnal cycle (Mears et al., 2003; 2003c). With the diurnal cycle accounted for, they evaluated overlapping satellite records using running 5-day averages (pentads) to smooth the data, and did merge calculations in a manner similar to that of the UAH team (Mears et al., 2003c), except that their hot target factor calculations are based on ocean data only. Ocean diurnal cycles differ from those of land masses in that they are much smaller and more stable due to the temperature moderating effects of water. This makes them much more predictable and easy to evaluate. Since the hot target factor is characteristic of the satellite itself, it is independent of the scene observed, RSS argued that an ocean only merge calculation would allow hot target factors to be determined more reliably prior to the final merge calculation for trends (Mears et. al., 2003; 2003b; 2003c).

The RSS analysis is noteworthy because their results, like those of PR, differ considerably from those of UAH despite a general similarity in analysis approach. Whereas UAH Version 5.0 indicates a statistically insignificant warming of 0.013 deg. C/decade for MSUTMT (middle troposphere), RSS observes a statistically significant warming of 0.097 deg. K/decade for the same layer. The difference is significant on two counts. First, given the similarities in their approach to analysis and the thoroughness and commitment brought to the datasets by both teams, the discrepancy is a testimony to the difficulties in characterizing MSU and AMSU datasets (there is a common misconception among many today that the satellite data is somehow "more reliable" than surface data – this is not true). Second, RSS's results are not only higher than those of UAH, their middle troposphere results are also higher than UAH's lower troposphere results (+0.061 deg. K/decade for Version 5.0 MSU2LT). As noted above, MSU2/AMSU5 receives a significant portion of its signal from the lower stratosphere, which is known to have cooled over the last 25 years due to ozone depletion (Bengtsson, 1999). RSS results will include this, indicating that if anything, their middle troposphere numbers are on the low side, making it likely that the discrepancy is

even larger than the numbers imply. The same is true of the PR trend estimates which despite a simpler analysis approach are in good agreement with RSS trends but not with UAH. Like UAH, RSS is updating their analysis on an ongoing basis. As of October 2004, their MSU2/AMSU5 middle troposphere temperature trend was 0.131 deg. K/decade.

Of the MSU/AMSU analysis product produced to date, the UAH and RSS products are the most mature and best characterized in terms of error correction and data inputs, and most discussion of lower atmosphere temperature trends revolves around them. As we will see, though these analyses have many similarities (the two teams have been sharing data and results with each other), the differences in their methods account for nearly all of their differing results.

### Vinnikov & Grody (VG)

In September of 2003 another analysis was published which made use of a new and radically different approach to analyzing MSU/AMSU datasets and yielded a new trend estimate that differed from all other analyses to date. Konstantin Vinnikov of the University of Maryland and Norman Grody of the NOAA NESDIS group (VG) analyzed MSU/AMSU data from 1979 to 2002 using a method that is primarily statistical and model based rather than empirical to evaluate errors. In their analysis, MSU/AMSU data is fit parametrically to a mathematical model designed specifically to represent climatic data having diurnal and seasonal cycles (Vinnikov et. al., 2002a; 2002b; 2004). The model characterizes brightness temperature as seen from a satellite platform by superposing two trends – a daily trend that represents the diurnal cycle, and another that represents long-term variations. Each term is represented by a Fourier series which includes terms for the expected variation (physically determined) and a weather dependent anomaly (random variation) which includes all noise sources, including instrument biases. Only nadir and near nadir data are used so as to avoid potential errors associated with limb corrections and decaying satellite orbits. Data between satellite tracks that would ordinarily have been measured by off-nadir views are evaluated from the parametric model. For the merging of data from different satellites VG use pentad smoothing similar to RSS. Then they do merge calculations by deriving the Fourier coefficients in their parametrized function along with a least squares fit of the satellite observations to derive a final unified time series (Vinnikov and Grody, 2003). One of the issues confronting the merging of satellite data is the fact that for any two satellites with overlapping records, the contributions of different error sources might be quite different from other overlaps. Diurnal errors for instance, might be aliased into IBE or instrumental, making it is possible to conflate these errors with each other (though diurnal errors will be linear and instrument errors – notably IBE – will not be). There is therefore, no common correction for all satellites. The VG analytical method was developed to handle such problems. With this analysis, they obtained a middle troposphere trend of 0.22 – 0.26 deg. K/decade. This trend provides the best fit to date with AOGCM predictions, but is considerably different from the trends obtained by all other teams. The VG method avoids the problems inherent in characterizing independent error sources without aliasing them into each other, but is strongly dependent on the quality of their modeling. If this is good, their results will be more reliable than those of other teams. But if it is not, the resulting errors could be considerable. It must also be noted that their analysis differs from those of the other teams in two respects. They characterize diurnal effects with a single cycle and do not separate ocean and land cycles from each other. In addition, because their model is derived based from known climate patterns, it is not clear to what extent it can account for IBE and non-linear instrument errors, as these are satellite dependent and unrelated to atmospheric effects. It is likely that some of this error will be absorbed, but not all of it. This introduces a degree of uncertainty into their analyses, VG have argued that this is not enough to account for the low trend estimates of UAH.

## Determining Trends from MSU/AMSU Products

Figure 8 summarizes the tropospheric temperature trends derived from the most current MSU/AMSU analysis products. Despite the dedication of all teams, the spread of values shown is striking. For the preceding 25 years, UAH finds relatively little troposphere warming, but Vinnikov and Grody find warming consistent with the predictions of most current climate models and even greater than the observed surface warming. The results of the RSS and Prabhakara teams are in between.

It is evident that the uncertainties and differences in methodology of these analysis products have taken their toll and must be sorted out before a reliable picture of troposphere temperatures can be obtained. With the exception of Vinnikov and Grody, who used a radically different analytical method, each of the teams used basically the same approach, but with different methods of treating the individual sources of error in data collection. Of these methodological differences, 3 in particular, stand out: Satellite data merging methods, diurnal corrections, and Antarctic sea-ice contamination.

### Diurnal Corrections

Removal of the diurnal sampling effect may be done in one of two ways - empirically using direct satellite measurements, or with physically based math models in conjunction with measurements. The UAH team adopted the former approach. They compared nadir and off-nadir  $T_b$  measurements to estimate the rate of drift. At any given equatorial crossing, the nadir view gives the true  $T_b$  at LECT, and off-nadir views give  $T_b$  at earlier and later times. View 1 represents a  $T_b$  that is 80 minutes earlier than View 11. By comparing these differences for millions of cross track scans throughout the lives of all satellites, UAH was able to estimate the temperature shifts that would result from any given satellite's drift in LECT. These estimates were made separately for ascending and descending satellites. A constant bias due to satellite roll is removed by comparing the two. The final estimates of diurnal temperature shift were then applied to the MSU/AMSU record with appropriate adjustments for the differences between land and ocean diurnal cycles<sup>4</sup> (Christy et al. 2000; Christy et al. 2003). The advantage of this method is that it is based strictly on satellite observations with few extraneous assumptions. But because it is dependent on a comparison of nadir and off-nadir views, it is subject to sampling noise so that zonal averages must be used to calculate the diurnal cycle drift. The diurnal cycle cannot be removed separately for each measurement grid point (Mears et al., 2002; 2003c).

Given this inherent noisiness, the RSS team adopted the second approach. Diurnal cycles are well understood, particularly over oceanic areas where the cycles are moderated by the large thermal capacity of the world's oceans, so it is possible to construct highly accurate models of them that can be calibrated by direct MSU observations. RSS calculated a climatology of local diurnal  $T_b$  anomalies using the NCAR Community Climate Model (CCM3) (Hack, 1995; Kiehl, 1996; 1998; 1998; Hurrell et al., 1998). This model, which is robust and has been tested for its ability to replicate broad features of global climate, produces surface and atmospheric temperature profiles on a 128 x 64 grid of the earth. Output from this model is then used in models of surface emissivity and radiative heat transfer to calculate observed brightness temperature at MSU/AMSU sensors (Wentz, 1998). Corrections for oceanic surface roughness and variations in emissivity with sea surface temperature (SST) were included, and an emissivity of 0.95 was assumed for land cycles. Using 5 year's worth of hourly output from these models, a 5 year time series was constructed for each view angle at LECT. Results were tested by comparing the model predictions with direct MSU/AMSU measurements for ascending and descending cycles at the same location. To reduce noise, comparisons were made only for the 5 central MSU/AMSU fields of view, with the 4 off-nadir views corrected to nadir (Mears et al., 2003; 2003c). Figure 9 shows global MSU/AMSU brightness temperature differences between ascending and descending cycles for the month of June as measured by the center 5 fields of view (top) and as calculated by CCM3 (bottom). It can be readily seen that the agreement between the two is quite good. Since the calibration of this model is done on a global scale using MSU/AMSU near-nadir data for which noise levels are small compared to the magnitude of the diurnal cycle, this method avoids the uncertainties of the noisier UAH method, and the agreement between MSU/AMSU observation and the modeled cycle validates the calculated cycle as well (Mears et al., 2002; 2003c).

## Intersatellite Data Merging Procedures

The MSU/AMSU record spans a period of nearly 25 years and uses data gathered from 9 satellites whose service lives overlap at various points during this period. Once the appropriate diurnal corrections have been made to the data of each, their records must be merged into one continuous time series. But each of these satellites will have an offset relative to the others due to the various linear and non-linear instrument errors already described. The most notable of these is the Instrument Body Effect (IBE) discussed previously. The exact cause of IBE is not known, but it is thought to result from a combination of uncertainties in the calibration of MSU and AMSU non-linear radiometer gain prior to flight, and a drift in radiometer gain that is proportional to the hot calibration target temperature (Christy et al., 2000; 2003) and has been observed in all TIROS-N class satellites. IBE is likely to be hardware related, it will therefore be unique to each satellite and independent of the flight path and the scene being viewed at any time. In addition to it there will be remaining residual errors not accounted for already in other corrections.

Because of these errors, no two MSU/AMSU records will be measuring temperature from the same reference point – all will be somewhat “out of sync” with each other. Before a consistent time series can be produced, the independent records for all satellites must be unified – that is, merged – to a single averaged record that minimizes the discrepancies between it and the individual records (e.g. the “residuals”). The standard procedure for doing this begins by smoothing each satellite’s record with “running averages” to remove sampling noise and short-term fluctuations, after which each record is expressed as a function of the actual measured brightness temperature, the product of the hot calibration target temperature and a constant target temperature factor, and a term for any remaining non-linearities. Then, for each period where any two satellites are co-orbiting and there are valid overlaps in their data, the differences between each set of simultaneous measurements can be expressed as a difference equation in terms of these functions. The full set of these for all overlapping records provides a large set of simultaneous equations in 17 unknowns (9 Hot Calibration Target Factors and 8 offsets for the full set of overlapping records). Since there is no independent reference temperature for calibration, one of the satellite’s observed temperature records (typically, NOAA-10) is arbitrarily set to zero as the datum. Solving these equations simultaneously using multiple regression techniques determines the target factors. Then, using these Target Factors and least squares methods to minimize the residual differences between the solution curve and the individual satellite records, a unified time series for the entire MSU/AMSU history can be constructed. The number of equations that will need to be solved depends on the type of averaging used. The effectiveness of this procedure is dependent on the lengths of the overlapping records – the longer these overlaps are, the more effectively differences can be characterized and removed. Figure 10b shows the service lives of the first 9 TIROS-N series satellites (up to NOAA-14) and their respective periods of service life overlap. It can be seen that of all overlapping periods of service, the NOAA-9/NOAA-10 and NOAA-7/NOAA-9 overlaps are unusually short – NOAA-9/NOAA-10 in particular being only about 90 days. The shortness of these overlaps has proven to be one of the thornier problems facing MSU/AMSU analyses. How they were handled by each team accounts for most of the differences between their results.

The UAH team began by applying 60 to 110 day smoothing to the individual records. Short overlaps were dealt with by setting a minimum acceptable overlap time for any dataset to be used in their analysis. This led them to discard the TIROS-N/NOAA-6, NOAA-7/NOAA-9, NOAA-8/NOAA-9, NOAA-9/NOAA-10, and NOAA-10/NOAA-12 overlaps from their analysis. Figure 10a shows the selection of overlaps used in their analysis against the full record. The remaining overlaps were then used to for the derivation of hot target factors. Land and ocean data from each record were used to provide fully global coverage. Once the target factors were determined, the merging of records was accomplished using the selected overlaps to create a “backbone” against which all satellite records were compared to generate the final averaged time series (Christy et al., 2000; Christy et al., 2003). With this method, UAH avoided the noise and sampling problems presented by the shorter overlaps and had a more global basis for their derived hot target factors. According to the UAH team, their choice of overlaps to omit was guided by a decision to give top priority to minimizing the intersatellite trend differences – a goal which they appear to have achieved, reporting an RMS of all intersatellite annual trend differences of only 0.008 deg. C/yr (Christy et al., 2004). But a price was paid for this. While this is a desirable objective for intersatellite datasets, and one they have arguably achieved, the selection of overlaps used was based on a somewhat arbitrary choice of minimum required overlap,

and the resulting decadal trend is fairly sensitive to which overlaps are omitted. This raises the possibility that the UAH merge minimized the trend itself rather than the uncertainty in the overall trend measurement (minimizing the actual offset residuals in a merge does not necessarily minimize the uncertainty in trend, especially where some record overlaps have been excluded from analysis). In addition, though the use of land and ocean data for calculating hot target factors gives them an observationally based global evaluation, diurnal temperature fluctuations over land are far greater both temporally and geographically, leading to increased noise in the hot target factor calculation.

By contrast, the RSS team used a more unified approach. Their datasets were smoothed using running 5-day averages (pentads) rather than the longer period smoothing of the UAH team, allowing them to include all the data equally weighted. Shorter overlaps (e.g. – NOAA-9/NOAA-10) are de-emphasized because of they are a smaller portion of the overall time series, but they are not ignored altogether. Target factors and intersatellite offsets are simultaneously calculated for all satellite record overlaps from the same set of equations, and the unified time series and trend are produced from this (Mears and Wentz, 2003b). The use of all satellite datasets is an important advantage of this method. No dataset is arbitrarily ignored, making their analysis insensitive to any arbitrariness in what data is used to calculate a backbone for their time series. Yet because they use of pentad smoothing of their data rather than longer period smoothing, they can still account for the problems of shorter overlaps without introducing any arbitrariness in their time series associated with ignoring them. A comparison of overlaps used in the analyses of UAH and RSS is shown in Figure 10. Another important difference is that the RSS team calculated their target factors using ocean only data. Due to its large heat capacity, water retains temperature very well making oceanic diurnal temperature fluctuation more consistent and far less subject to geographical noise (it is no coincidence that extreme seasonal temperature differences are more characteristic of central continent regions such as the American Midwest, and coastal regions like the Pacific Northwest are more seasonally moderate). Also, unlike land regions, oceanic brightness temperature fluctuations are typically small on diurnal and seasonal scales. This relative stability allows for target factors to be calculated with more precision. This is important because the diurnal fluctuation is part of the data sampled when evaluating target factors, and this introduces noise in the target factor calculation that is not accounted for directly by the diurnal corrections themselves. Since hot target factors are ultimately a property of the satellite hardware itself rather than the scene being viewed, once determined, they may be used for all analysis. Thus the RSS team has been able to determine hot target factors with more precision and make use of a more complete and less arbitrary dataset when evaluating their decadal trends.

Figure 11 gives the 1979 to 2001 decadal trend in global Channel 2 (AMSU Channel 5) brightness temperature as determined by UAH Ver. 5.0 (Christy et al., 2003) and RSS Ver. 1.0 (Mears et. al., 2003). Figure 11a gives the RSS regional profile, Figure 11b shows the UAH trend, and the differences between the two is in Figure 11c. It can be seen from Figure 11c that despite the differences between UAH and RSS global trends, the two teams are in good agreement with each other over most of the northern hemisphere above 30 deg. North latitude. They differ markedly though for the tropics and high southern hemisphere (particularly the southern Pacific Ocean), where the UAH team sees much more cooling than RSS. Other notable differences occur in northern Africa where RSS observes more warming, and the Himalayas where UAH observes more warming. Thus, the differences between the global trends observed by these teams has a strong regional component weighted heavily toward the tropics and southern hemisphere. Furthermore, the most significant regional differences between the two analyses are larger than global trends derived.

### The Method of Vinnikov and Grody

In September of 2003, Konstantin Vinnikov of the University of Maryland and Norman Grody of NOAA (hereafter referred to as VG) provided an independent analysis MSU/AMSU data from 1979 to 2002 using a methodology that that differs considerably from that of UAH and RSS. VG raised three concerns with previous analyses, UAH's in particular.

- In situ corrections for diurnal cycle variations cannot be corrected for reliably due to an insufficient number of observations.
- Off-nadir brightness temperature views must be corrected to nadir before they can be used, and this introduces error – particularly when they are used in composite Channels such as UAH's 2LT and TLT Channels.
- Data merging procedures based on comparisons of overlapping satellite records cannot account for drifts in hot target calibration factors and IBE because the effects of each cannot be reliably separated from diurnal cycle variations.

To address these concerns, VG analyzed the MSU/AMSU using a model they had developed elsewhere specifically for analyzing data from multiple satellites with seasonal and diurnal cycles and differing LECT's (Vinnikov et al., 2002a, 2002b). In this model the long-term brightness temperature trend from each satellite is given as a set of Fourier components with terms representing the seasonal and diurnal cycles, the underlying climatic temperature trend, and a term for underlying constant bias intended to capture random noise and IBE. The satellite data is then fitted to this model and the Fourier components and bias constants are solved for using a least squares analysis. To account for orbital decay, they used only near nadir views. With this method they obtain a middle troposphere temperature trend of 0.22 – 0.26 deg. K/decade from MSU Channel 2 and AMSU Channel 5 (Vinnikov and Grody, 2003).

The Vinnikov and Grody method avoids the noise sources inherent in the analyses of earlier work and is flexible enough to account for a wide range of observed climatic effects. But because they have restricted their analysis to the low noise nadir views, their observations only cover a narrow earth scene directly under the satellite path. The regions between each orbital path that would have been observed by off-nadir views are missed. In addition, a single diurnal cycle is assumed and there is no discrimination between ocean and land. Because of these two omissions, the method relies heavily on zonal averaging of the data to account for unobserved regions and land/ocean variations. If VG have done their zonal averaging and statistics well, their numbers will be better than anyone else's. If not, their calculated decadal trends could differ markedly from reality. These points are still being hotly debated.

## **The Radiosonde Record**

Though the MSU/AMSU record provides the only truly global tropospheric temperature characterizations currently available, the issues discussed above present significant roadblocks to getting tropospheric temperature trends that are accurate enough to detect any potential anthropogenic fingerprint. Even after more than 15 years of investigations, many questions remain and the different teams, all equally committed, are obtaining widely varying results. In light of this, many researchers have sought independent verification of the MSU/AMSU record from alternate products. The one promising option for this is the radiosonde record.

For the last 50 – 60 years balloon borne instrument packages called radiosondes, whose sensory devices track temperature, humidity, and pressure have been launched at least once daily from many weather stations around the globe. The data from these devices provide a detailed temperature, pressure, and humidity history of the lower and middle troposphere at discrete altitudes that have been used for regional weather forecasts and meteorological studies. Once released, they ascend at known rates gathering data as they climb until they reach their maximum altitude. Data regarding wind patterns can also be obtained by monitoring radiosonde movements after release. Though not specifically designed for the purpose of assessing global change, in the absence of other in-situ data sources, they have been used for this purpose since then. As of 2000, there was a global network of some 900 upper air radiosonde launching stations located at points all around the world including polar regions. This network is predominately land based and weighted toward Northern Hemisphere locations. At least two thirds take twice daily measurements at 0000 and 1200 Coordinated Universal Time, or UTC (NRC, 2000). There are several datasets that have been compiled from this network spanning a period dating from the late 50's to the present that have been used to evaluate global tropospheric temperature trends

(Angell and Korshover, 1975; Angell, 1998; Parker et al., 1997; 2000; Brown et al., 2000; Gaffen, et al., 2000a; Lanzante et al., 2003). These datasets have been analyzed in different ways yielding different results depending on which data sources were used, sampling methods, and analytical comparisons were used (Angell, 1999; Parker et al., 1997; Santer et al., 2000; Lanzante et al., 2000; Hegerl and Wallace, 2002; Lanzante et al. 2003; Seidel et al., 2003).

The greatest strengths of the radiosonde record are the length of time over which reliable measurements have been made (which is considerably longer than the MSU/AMSU record), the maturity of the temperature sensing technology used, and the fine resolution with which sondes measure vertical temperature profiles and lapse rates<sup>4</sup>. Even so, the sonde record faces a number of challenges. How these are dealt with bears directly on the usefulness of this data as an independent check of the MSU/AMSU record. Sonde record limitations can be broadly described as follows.

- 1) Instrumental Variation: Over the years a number of makes and model of radiosonde have been put into service around the world. Instrument packages on these platforms have varied according to the manufacturer and model of sonde involved, and have evolved considerably in their sophistication and reliability since the post-WW II period. Newer devices utilize temperature thermocapacitor - wire resistor – thermistors based designs, or bimetallic sensor designs (NRC 2000). Variations in the calibration, reliability and general accuracy of these devices introduce errors which must be corrected for. In addition, a “technology growth” correction is necessary to account for the fact that advances in sensor technology since the birth of the sonde network have introduced variations in accuracy and life of sonde based temperature sensors in the last 50 years. An overview of the evolution of sonde based instrumentation and the potential sources of error associated with it was published by the World Meteorological Organization in 1996 (WMO 1996), and it is as relevant today as it was then.
- 2) Sampling Variation: The 900 odd upper air radiosonde stations around the world are at discrete, locations that tend to favor the Northern Hemisphere. The large majority are at continental land locations. There has never been a truly global coverage pattern of the sort that reliable estimates of global change require. This has led to the need to either extrapolate sonde data to global scales with the aid of models, or limit their use to partial comparisons with other datasets. Furthermore, radiosonde stations have come and gone over the years due to budget cuts, policy changes, and regional conflicts - particularly in regions of political unrest like sub-Saharan Africa or the former Soviet Union after the fall of communism. Thus, the number and location of many sonde stations has not been constant since their records began and temporal corrections and/or weightings must be introduced into these datasets to account for this “geographic noise”. Variations in land surface characteristics have also affected data from some stations. Heavy development or land use changes could lead to increased desertification or forestation for some stations over longer time periods. These can in turn impact the local surface emissivity of regions monitored by these stations. Growing urbanization will do the same. Some stations have varied the times at which daily sonde reading are taken throughout their history. These variations will create spurious temperature variations in a manner not unlike that of the diurnal cycle variations experienced by MSU products when their LECT’s vary over their service lives (NRC, 2000). There are a variety of ways to account for such effects, but each has weaknesses as well as strengths.
- 3) Vertical Data Homogeneity: Though sonde datasets have far better vertical resolution than satellite products, they vary widely in the extent of their vertical coverage. At their best, radiosondes can achieve altitudes of up to 50 km, but older or less well designed balloons commonly burst at much lower altitudes creating inhomogeneities in the altitude range covered by some datasets. Differing sonde instrument packages have also varied over the years in the number of readings taken per ascent and the specific altitude levels of each. Measurements have been taken and transmitted with anywhere from 10 m to 50 m resolution depending on sonde rise and/or data sampling rates. Older archived datasets for many global stations may contain only 20 data levels per sounding - far less than most modern records (NRC, 2000). Some datasets are based on all available data, with variations in these factors (Angell, 1988). Others, such as CLIMAT TEMP reports are based on monthly normalized data from a global subset of upper air stations and from preset pressure determined altitudes (Parker et al. 1997). Missing data and temporal variations in how they were gathered can introduce significant errors

in these monthly estimates (NRC, 2000). In addition, sonde datasets have been built over the years under widely varying surface and boundary layer conditions reflecting differing locations and local weather conditions. Variations in the daily launch times mentioned previously can also affect vertical profile readings, as these can vary throughout any given day. These factors were never considered an issue before given that, like the original MSU packages, sonde records were produced to support local meteorological studies and/or weather forecasts. But for global change studies they cannot be ignored, particularly when they are being used to independently validate other data sources with differing issues.

- 4) Onboard Instrumental Biases: Like MSU and AMSU packages, sonde based temperature sensors are also subject to instrumental biases from various sources including incident solar and infrared radiation (which is not related to ambient atmospheric temperature), drifts in calibration, and noise driven variations in measurement accuracy. Device designs have evolved in their reliability over the years as well, and these factors are not necessarily constant over the lifetime of measurements from any given station (NRC, 2000). Variations in how instrumental factors have been accounted for over the years in analyses have impacted results, particularly for the upper troposphere (Gaffen, 1994). Toward the end of the 20<sup>th</sup> century for instance, there was a growing shift toward the use of sonde instrument packages built by the Vaisala company (Parker and Cox, 1995). These packages have measurement characteristics which differ from those of earlier types and the differences can be aliased into the record of the sonde stations that implemented them.
- 5) Data Reduction and Analysis Variations: Over the years, dataset problems like those discussed above have been dealt with in varying ways. Analytical techniques that account for onboard instrument biases have evolved as understanding of them grew. There have also been variations in the way monthly mean data have been evaluated, leading many stations to use differing methods for reporting monthly means at varying locations worldwide and over the years records have been kept. This has been a problem for the CLIMAT TEMP monthly reports most commonly used for MSU/AMSU trend comparisons. Analysis of years worth of monthly averaged data from globally distributed stations using consistent rules has shown that significant errors can result from these variations (Gaffen et al. 2000b). Another issue plagues the sonde record that is of particular relevance to MSU/AMSU comparisons. MSU and AMSU packages measure the weighted average temperature of broad layers of the troposphere and stratosphere, but sondes generate specific temperature vs. altitude profiles. These profiles vary widely with location and date and bear little resemblance to MSU/AMSU weighting functions. As such, sonde data cannot be directly compared to MSU/AMSU datasets until they have been processed so that they better represent what these sensors would have observed had they been co-located. There are differing ways this can be done. The simplest approach is to average the sonde data over the same atmospheric layer seen by the corresponding MSU/AMSU channel. This neglects the altitude determined weighting function these channels see (Figure 7). More commonly, sonde datasets are weighted in a manner that reflects the MSU/AMSU weighting functions, yielding more meaningful comparisons. This however does not account for the impact of atmospheric moisture, which MSU and AMSU sensors are relatively insensitive to but sondes are not. This can introduce further error, though typically errors of this sort are typically small in comparison to the decadal trends being measured (NRC, 2000).

Effectively addressing these uncertainties requires that we examine the operational history of each reporting station as well as its data. But for a variety of reasons weather stations around the world have varied widely in the quality and completeness of their records. This has been particularly problematic for MSU/AMSU comparison studies because many of the most important regions for comparison (e.g. the tropics and northern Africa) lie in regions of chronic political instability and weather stations in these regions have had difficulty keeping their funding base and maintaining consistent locations and operational records. Other areas that are even more important for MSU/AMSU comparisons (e.g. the Southern Pacific Ocean) are under-sampled due to a scarcity of stations. Where datasets and operational records are incomplete, it is necessary to interpolate data and/or metadata from the station's past records or from the nearest stations with continuous simultaneous records.

There are two basic dataset products that have been used in radiosonde analyses relevant to MSU/AMSU comparisons. The first is based on individual soundings containing all gathered data at each level of ascent for all stations reporting data in this manner (Angell, 1988). The National Climatic Data Center (NCDC) has consolidated numerous such datasets from around the world, along with all relevant metadata about each station's operational history into a single database called the Comprehensive Aerological Reference Data Set, or CARDS database (Eskridge et al., 1995). The second is based on monthly averaged data taken from a globally distributed subset of weather stations at 9 mandatory pressure levels between 850 and 30 hPa. These are known as CLIMAT TEMP reports (Parker et al., 1997). Of all weather stations worldwide, roughly 45 percent provide both CARDS and CLIMAT TEMP datasets (NRC, 2000). Both have suffered from discontinuities in data and metadata and must be adjusted accordingly before they can be used. Typically there are 3 steps involved. First, anomalous discontinuities in the record must be identified and separated from real climatic effects. Second, the size of these discontinuities must be determined. Last, appropriate corrections must be made that will remove their impact on the trend. In doing all this there are 2 pitfalls to be avoided – natural variations in regional climate might be confused with changes in a station's operational practices, and changing operating practices might be mistaken for real climate variations. Both will introduce spurious warming or cooling into the long term record from that station that will impact a global time series. To date, several radiosonde analysis products have been developed. They differ from each other in their use of either CARDS or CLIMAT TEMP data and their methods for identifying anomalous events and correcting for them. All have strengths and weaknesses and none stands out as uniquely more reliable than the others. The ones that have been of most interest for MSU/AMSU comparisons are the following.

#### Angell 54

For over 3 decades many tropospheric radiosonde studies used an analysis product prepared by Jim Angell, currently of the NOAA Air Resources Laboratory in Silver Springs, MD. This product was based on data taken from a selection of 63 global radiosonde stations taken at mandatory layers in the troposphere (850-300 hPa), tropopause (300-100 hPa), and lower stratosphere (100-50 hPa) layers. Layer mean temperatures were ascribed to each using the hydrostatic equation and temperatures measured at each bounding pressure surface. Trends for each layer were then derived for various global regions by averaging the annual temperature anomalies using data from all stations within each (Angell and Korshover, 1983; Angell, 1988). The resulting analysis has been referred to as Angell 63. While the coverage of this network has been considered to be good and the record relatively complete, the consistency and quality of the data were not explored in much depth and the resulting trends were taken at face value. Recently, other investigators have examined individual Angell 63 station datasets more closely and determined that changes in instrument packages and operating procedures have contaminated the data from several of them (Gaffen, 1994; Gaffen et al., 2000a; Lanzante et al., 2003). To quantify the impact of this, Angell reanalyzed the 1958-2000 datasets for the 63 station network focusing on the 300-100 hPa layer. This layer, which reflects the tropopause, was selected because other layers had less complete datasets, and the full time period was chosen so as to make characterization of anomalies more apparent. A least squares regression analysis was performed for the network and the confidence intervals of each stations record was then compared to the network trend to determine which ones deviated most from the regression. Anomalous stations were defined as those whose 2 standard error of regression values were larger than 0.2 deg. K/decade. Angell identified 9 of the original 63 stations, located mainly in the tropics, as providing anomalous data. These 9 were removed and the remaining 54 station network was reanalyzed. This analysis product has come to be known as Angell 54 (Angell, 2003).

Figure 12 shows the original Angell 63 station network with the 9 anomalous stations highlighted in red. Note that all 9 of the anomalous stations removed are located in the tropics. The remaining 54 stations show reasonable global coverage emphasizing land based locations, though the density of the network is quite low given the regional variability of tropospheric temperature trends and lapse rates. Coverage of ocean regions is somewhat sparse, particularly in the Southern Pacific. Figure 13 shows the upper air trends obtained from this network for the 850-300 hPa, 300-100 hPa, and 100-50 hPa layers in 4 regions – the northern and southern hemispheres, the tropics, and the globe as compared with several other radiosonde, radiosonde-satellite re-analysis, and surface products, and UAH Version D (Angell,

2003). Here, the tropics are defined as the region from 30 deg. N to 30 deg. S Latitude, whereas the IPCC WG1 Year 2001 Report defines the tropics as between 20 deg. N to 20 deg. S Latitude (IPCC, 2001). Figure 14 shows just the Angell 54 upper air trends for the same altitudes and regions for 1958-2000 (Left), 1979-2000 (Center), and the change from the former to the latter (Right). As in Figure 13, the horizontal bars indicate 2-sigma confidence intervals. The most noticeable feature is the stronger cooling of the low stratosphere (100-50 hPa) for the 1979-2000 period vs. 1958-2000, the difference amounting to nearly 0.4 deg. K/decade. Also noticeable is the cooling in all regions of the tropopause (300-100 hPa) in the later period relative to the whole record, and possible even a shift from warming to cooling. Based on the mean trends, the lower to middle troposphere (850-300 hPa) appears to have shifted from warming over the longer record to less warming or possibly cooling everywhere except in the northern hemisphere. It also appears that lapse rates have increased during the latter period indicating that the surface appears to have warmed with respect to the lower troposphere. Figure 13 also shows that Angell 54 is in good agreement with other sonde and surface-air products, and UAH Version D. But the confidence intervals shown, as determined by twice the standard error of measurement and the observed discrepancies between other products, indicate spreads that accommodate both UAH and RSS analyses. It can also be seen that a significant range of variation is possible for all regions, particularly the tropics.

## HadRT

Another sonde analysis product has been developed by David Parker and Margaret Gordon of the Hadley Centre, U.K. MET Office. This one uses monthly gridded CLIMAT TEMP data from a globally distributed network of 400 weather stations, considerably larger than Angell 54. The methodology is best described in Parker et al. (1997). Several HadRT analysis products are currently available. HadRT2.0 uses data from 1958 to the present on a 5-degree latitude by 10 degree longitude grid. No bias corrections are applied to station data for this release. Versions 2.1 (troposphere and stratosphere) and 2.1s (stratosphere only) identify anomalous changes in station records using comparisons with MSU data from UAH Version C (Christy et al., 1998) for Version 2.1, and Version D (Christy et al., 2000) for Version 2.1s (stratosphere MSU corrections only). Later versions add additional corrections to data from 1958 to 2000, the most recent as of this writing being HadRT2.3 (troposphere) and HadRT2.3s (stratosphere) which use Laplacian methods applied to the second derivative of the corresponding NCEP reanalysis<sup>8</sup> temperature fields to fill in gaps in the data (Reynolds, 1988). Of the available products, HadRT2.1(s) is the one that to date has been most frequently used for MSU comparisons.

The HadRT method uses sonde temperature data taken at 9 standard altitudes between 850 and 30 hPa. These subjected to certain quality control measures (Parker & Cox, 1995), and given a weighting derived from the MSU Channel 2 weighting function so as to simulate a "bulk" layer temperature that would have been measured by MSU devices looking directly at the same vertical column of the atmosphere. Then for each global grid location, this weighted data is compared on a case by case basis with collocated monthly MSU anomalies and known changes in instrumentation and analysis methods taken from Gaffen (1996). Differences between the sonde and MSU anomalies are evaluated before ( $\Delta 1$ ) and after ( $\Delta 2$ ) the most recent known changes in operating procedure and/or equipment, excluding data prior to known previous changes. If  $\Delta 2 - \Delta 1$  was found to differ significantly from zero at the 95 percent confidence level, a seasonally invariant correction equal to  $\Delta 2 - \Delta 1$  was applied to the earlier data for the layer being evaluated as far back as the next previous identified instrument change. If no earlier changes are identified the correction is applied back to January 1979 when the MSU record began.  $\Delta 2$  is evaluated for the entire record following the most recent instrument change and incorporates bias corrections already evaluated in previous steps. In this manner, time series for each global grid are evaluated one step at a time going back as far as 1979. Each adjustment is then apportioned to the individual mandatory altitude layers according to average bias adjustments estimated at each layer in the tropics and extra-tropics for different classes of instrument and/or analysis method changes (Parker et al., 1997). This method benefits from the leveraging of the MSU data which is consistent and truly global in a way that raw sonde data is not. But CLIMAT TEMP data were gathered at observation times (0000 UTC or 1200 UTC) that may differ from location to location, and often the time series for a given gridded location may combine both. This can result in inconsistencies and complicate comparisons with other similar methods. It is also presumed that natural events such as volcanic eruptions will affect sonde and MSU products in a similar manner. The MSU flight history has seen at least 2 major volcanic eruptions

which are known to have impacted global temperature profiles in both the troposphere and lower stratosphere. MSU devices rely on radiative measurements from space while sondes utilize local direct temperature measurements using thermocaps (dielectric temperature sensors) which may be impacted quite differently. It is not at all clear that both products have responded to these events in a similar manner. Since corrections are based on the MSU record, the two are not entirely independent. To some degree this compromises the usefulness of this product for evaluation of MSU analyses, although the corrections were applied only to a relatively small portion of the data, so the impact is not severe.

## UAH

The UAH team has developed their own method of preparing sonde analysis products. Their method compares monthly anomalies of sonde simulated MSU data ( $RaobT_b$ ) to actual MSU data in a manner similar to that of the HadRT method (Christy et al., 2000). There are a few key differences though. Whereas HadRT uses monthly gridded CLIMAT TEMP data, the UAH method uses raw CARDS data taken directly from specific stations and wherever possible, considers data gathered at 0000 UTC separately from that taken at 1200 UTC. In addition, this method evaluates record discontinuities based only on bulk temperatures derived from collocated MSU and  $RaobT_b$  data, as opposed to HadRT which applies corrections on a layer by layer basis. First, a time series is formed by differencing  $RaobT_b$  and MSU data. Then, a 30 month running average time series is formed from this one and used to identify anomalous discontinuities in record. It has been shown that using this method, differences of about 0.3 K in the tropics to 0.6 K at higher latitudes are significant (Free et al., 2002).

Because it relies only on bulk temperature comparisons, the UAH method does not make use of individual station metadata so it is less likely to be impacted by changes in record keeping methods that plague many sonde station histories. But as such, it goes without the station metadata that is consistent and reliable and thus lacks important input from independent sources. Like HadRT, this method is strongly tied to the UAH Version D MSU data and is therefore subject to all the same limitations, including its limited usefulness for independently evaluating MSU products. A more serious problem is that the UAH sonde product for Version D was based on a network of 97 U.S. controlled stations covering North America, Bermuda, Iceland, and the Western Pacific. This network was chosen because it has an unusually high degree of consistency in equipment and methods, making these stations more reliable than most (Christy et al., 2000; Luers & Eskridge, 1998). But a price is paid in limited global coverage. We will see later on that coverage of this network is sparse to non-existent in those global regions that are most important for comparing the relative reliability of different MSU/AMSU products, particularly UAH vs. RSS.

## LKS

Despite good global coverage and robust methods, the HadRT and UAH products depend on MSU datasets for detection and correction of sonde time series. This limits their usefulness as independent validation of MSU products, particularly UAH MSU products. In addition, because the MSU record only goes back to 1979, these methods offer no insight into the sonde record prior to that time – nearly half of the extant record. Angell 54 does not rely on MSU comparisons and is thus truly independent. But with only 54 stations globally, it has less coverage than HadRT products and also relies on data from many weather stations that are less reliable than those used by UAH. For independent validation of the MSU record, what is needed is a truly reliable sonde analysis product that has truly global coverage, particularly in the Southern Hemisphere and tropics, and effectively identifies and corrects for anomalous discontinuities in historical sonde data without referencing any MSU product.

In 2003, John Lanzante and Steven Klein of the NOAA Geophysical Fluid Dynamics Laboratory in Princeton, NJ, and Dian Seidel of the NOAA Air Resources Laboratory in Silver Spring, MD took the first major step in this direction when they published a sonde analysis product based on a global network of 87 stations that had been selected and given needed data adjustments based on criteria that did not use MSU products (Lanzante et al., 2003). It has been shown that statistical methods alone can identify many abrupt discontinuities in sonde datasets and historical weather station records are often not adequate to identify the causes of these discontinuities (Lanzante, 1996; Gaffen et al., 2000a). Lanzante,

Klein, and Seidel (LKS) capitalized on this and other data to create a sonde analysis product that is independent of MSU products. Their analysis uses a global network of 87 sonde stations with global coverage similar to that of the Angell 54 network but more density. Figure 15 shows the global station network used. Within this network, each station's record was painstakingly examined in detail with special attention being paid to the following,

- Statistics and station metadata
- 0000 UTC temperature measurements minus 1200 UTC measurements (where both were taken)
- Temperatures measured at nearby levels at the same station
- Temperatures predicted using statistical regression of existing measured temperatures and winds
- Historical records of sonde launch times
- The Southern Oscillation Index
- Volcanic eruption history
- Comparable temperature data from nearby stations

Each member of the team evaluated the records of all 87 stations on a case by case basis and made recommendations as to where data adjustments were needed and to what degree. Afterwards they met together, compared results, and reached a consensus from which the final analysis product was created (Lanzante et al., 2003). Adjustments for identified discontinuities in each record were made at specific pressure levels (altitudes) by extrapolating from time series taken at nearby reference pressure levels at the same station. The reference levels were chosen so that they either required no correction, or had already been corrected to guarantee their reliability as a reference for other changes. If no suitable reference data was available, an adjustment was made by interpolating from data before and after the discontinuity (Lanzante et al., 2003). This method, though thorough, is quite labor intensive and so does not lend itself to really large networks, and as such, it lacks the degree of coverage of HadRT. But it is also the only record currently used for independent MSU comparison studies that is truly independent of MSU data.

## RIHMI

Alexander Sterin (1999) of the All-Russian Research Institute of Hydrometeorological Information (RIHMI) in Obninsk, Russia has prepared another upper air radiosonde analysis product that is based on a much broader base of data than those considered do far. Sterin used CARDS and telecommunicated data from a network of over 800 radiosonde stations worldwide and reanalyzed it into a monthly gridded temperatures at two pressure levels - 850-300hPa (troposphere) and 100-50hPa (lower stratosphere). Initial quality control checks were done using the Complex Quality Check (CQC). Data that passed this test was then spatially processed into anomalies for the globe and for 3 latitude zones: the Southern Extra-tropics (-90°S to -20°S), the Tropics (-20°S to +20°N), and the Northern Extra-tropics (+20°N to +90°N). Processing was done using an algorithm based on adaptive polynomial interpolation and sequential corrections (Reitenbach and Sterin 1996) with mass weighting for the vertical layers. Latitudinal Adjustments were made for unobserved regions using direct interpolation from locations with verified data. The initial version of this product used only gross spatial and temporal consistency checks for quality control. Later versions used improved quality control checks on raw data and expanded the network of stations used to over 2500 stations worldwide (Sterin, 2001). This method has the advantage of a much larger network and better coverage than other sonde analysis products. Normalizing from a much larger dataset will help eliminate the effects of data clustering as well. But RIHMI also suffers from having limited checking for inconsistencies in station metadata and historical discontinuities as well as the added uncertainty introduced by interpolating in global regions where there is no data.

## UAH 2004

In March of this year the UAH team published an updated radiosonde/MSU comparison that addressed many of the difficulties associated with other comparison studies (Christy and Norris, 2004). Christy and Norris noted the various methodological differences between their MSU/AMSU products and those of RSS, as well as the southern hemisphere coverage issues with independent radiosonde analyses that have been sought for MSU intercomparison studies. They also discussed the other issues that have plagued these studies, in particular solar heating, lag, incompleteness of record, and equipment variations such as the switch from Phillips or VIZ-B to Vaisala RS-80 radiosondes at many stations (Parker et al., 1997; NRC, 2000; Christy et al., 2000; 2003; Seidel et al., 2003; 2004; Angell, 2003). Many of these problems become more severe with increasing altitude (Parker et al., 1997; Gaffen et al., 2000a; Lanzante et al., 2003). The VIZ-B/Phillips to Vaisala evolution alone for instance, can account for corrections of up to 1 to 3 deg. K per station in the lower stratosphere, which is larger than the trends being measured in this layer. In addition, variations in tropopause height and altitude have also affected radiosonde records at many stations (Angell, 2003). Though corrections have been necessary in the lower troposphere as well, they have generally been smaller.

Realizing this, Christy and Norris sought an independent MSU-radiosonde intercomparison study that focused on the lower troposphere alone and attempted to redress the southern hemisphere coverage issue. They constructed a time series of monthly lower troposphere temperature anomalies for the 271 month period of Jan. 1979 to July 2001 using a global network of 89 stations, which is shown in Figure 42. These stations were selected based on their meeting a minimum requirement of at least 60 percent of the monthly records could be generated from daily soundings. A subset of these 89 stations met a further requirement that 75 percent of their monthly records could be similarly generated. To determine the necessary adjustments to this network, a method similar to the UAH method described above was used, and the results were checked against independent records of Durre et al. (2002) who constructed a similar record for the northern hemisphere, and Peter Thorne of the Hadley Centre, who constructed an independent product based only on the records of neighboring stations.

Christy and Norris's Year 2004 intercomparison product goes a long way to resolving many of the issues of earlier products, including an increased emphasis on southern hemisphere stations where other products have been weak, and a concentration on lower altitude data that avoid many of the problems of more complete high altitude datasets. But even so, an examination of Figure 42 reveals that the increased southern hemisphere coverage is still primarily over land, emphasizing South America and Australia, but remains scant in the southern Pacific where the largest RSS-UAH differences remain. In addition, while much has been done to enhance the completeness of record for the requisite 89 stations, incompleteness of record is still an issue, as are the difficulties in assuring that record discontinuities are all captured.

## Comparisons of Radiosonde Analysis Methods

In addition to accurate data, a reliable radiosonde time series requires an accurate assessment of the data's history and data gathering methods. It has already been noted that this is no easy task. Incomplete records, variations in equipment and methods which may or may not have been documented over the last 50 years and a host of other complications make historical reconstructions challenging. Apart from their data reduction methods, the main differences in the radiosonde analysis product described above lie in the way that they detect and correct for anomalies and/or gaps in datasets. There are two problems to be avoided. First, changes in the record unrelated to climate must be detected and accurately corrected for. Second, real climate signals must not be mistakenly identified as spurious. How well each of the current sonde analysis products does either is a matter of intense research.

In October of 2000 Dian Seidel of the NOAA Air Resources Laboratory and Tom Pederson of the National Climatic Data Center convened the CARDS Workshop on Adjusting Radiosonde Temperature Data for Monitoring at NCDC in Ashville, NC. At this workshop representatives from several research centers discussed how well the current generation of sonde analysis products achieve these goals. Attendees divided up into 7 groups to analyze the historical upper air sonde data from 12 stations

worldwide. The 12 stations used in this exercise were chosen for their emphasis on countries with large networks and their high quality of metadata. Two of the stations, located in Australia, were also included because of known discrepancies between their results and those of the UAH Version D results at the same location. Each team prepared an analysis of this network employing one of the established methods currently in use, including the methods discussed here. Some teams used daily data and others used monthly gridded data, as differing methods required. The objective was to determine which methods were most reliable at capturing discrete changes in tropospheric and stratospheric temperature records, characterizing them as natural or anomalous, and accurately correcting for them if necessary. The methods and results of this exercise are discussed in Free et al. (2002). Figure 16 shows the results of each team's characterization of temperature change events from the 12 station record. Included are the methods of LKS (denoted as GFDL), HadRT (denoted as Met Office), and UAH (results from 3 other methods tested at the Workshop are also presented, but are not considered here as they have not been used extensively for MSU comparisons). Identified change points are given per station and per investigating team/method. Figure 17 shows in tabular form the percentage of instances where any 2 methods agreed on an event within a 6 month window (top) and the total number of events detected by each team/methodology.

It is evident from these results that agreement between the various methods was the exception rather than the rule. The average number of record change points captured per decade, real and anomalous, was highest for the UAH method and lowest for the LKS method. The total number of changes identified was considerably less than the number of known metadata events based on extant records. For instance, the National Climatic Data Center (NCDC) has record of 21 known changes in method and/or equipment after 1979 at 10 of the 12 stations that are significant for the data gathered. Yet out of these 21 changes, there were only 4 occasions where changes at any given station were reliably captured by all teams and methods. Agreement between any 2 methods averages only about 50 percent. There is also little agreement on the pressure levels (altitudes) where temperatures needed adjusting, and the size of the needed correction. In many cases, different methods yielded different signs for their adjustments. One case in point was Darwin, Australia, where there is a fairly complete record. Figure 18 shows the change points identified by each team at all altitudes from the surface to 10 hPa for the Darwin record. The corrections applied by each team are given as colored triangles pointing upward for positive corrections and downward for negative ones. The size of each triangle is proportional to the magnitude of the correction it denotes. Figure 19 shows the uncorrected and corrected temperature trends for Darwin at 50 hPa (stratosphere) and 200 hPa (tropopause) for 4 of the methods studied. Despite the completeness of the metadata and records from Darwin, there is no point at which all 4 methods detect the same event. A significant discontinuity at 50 hPa that gives a large correction for HadRT and LKS yields only a small correction in NCDC - and is not even detected by UAH. Likewise, UAH detects at least 2 events that do not appear to have been accounted for in HadRT, LKS, or NCDC. Another large event in 1953 at 200 hPa is given a significant correction in LKS, but is not accounted for by either UAH or HadRT because it occurs prior to the beginning of the MSU record, which is required by both methods. It can also be seen that the magnitudes of the corrections for the troposphere, though not unduly large, are significant (at least 20 percent of the value), and those needed for the stratosphere are quite large. The HadRT stratospheric correction (1.94 +/- 1.42 deg. K) is over 3 times the size of the adjusted trend (while stratospheric trends are not directly relevant to troposphere temperature trends, they do impact the MSU Channel 2 signal, and must be considered).

These results demonstrate the difficulties associated with preparing radiosonde time series, and are in agreement with the results of similar studies (Santer et. al., 1999; Gaffen et. al., 2000a). Furthermore, it has also been shown that adjustments to sonde datasets can have a significant impact on confidence intervals as well as the derived trend. Even one level shift in a dataset can increase by as much as 50 percent the length of time series necessary to accurately derive a trend (Weatherhead et. al., 1998). Each of the methods discussed here has proven useful for independent checks of MSU records, even those like the HadRT and UAH methods that are not entirely independent of this record. But all have yielded mixed results, emphasizing the importance of ongoing efforts to improve access to weather station records and datasets. The National Research Council recently convened a Panel on Reconciling Temperature Observations to address these concerns as well as the issues already discussed regarding the MSU record. Their report (NRC, 2000) recommends that station metadata be updated and expanded for all stations worldwide, not just for a few as has been done to date. They also recommended that policies be put in place to regulate future changes in equipment and methods to guarantee continuity of

records in the future, and efforts are under way to put these recommendations into practice. For today, caution must be exercised when comparing separate radiosonde analyses with each other and the MSU record (NRC, 2000; Free et al., 2002).

## **Discussion**

What then, can be made of all this? How do the satellite and radiosonde trends relate to each other? Are they truly measuring the same quantities? If so, how do these datasets relate to the climatic variability that has been observed in the upper atmosphere in the last 25 years? For that matter, are the last 25 years representative of the century to come, and do they allow us to either detect or refute an anthropogenic impact on climate change? These are the questions that must be addressed with these records as they currently exist.

To address the length of the record, we must examine the last 25 years in light of the most significant climatic events that have impacted it. During this period, there have been at least two major volcanic eruptions that are known to have significantly reduced solar radiative heating of the troposphere causing cooling over 3 to 4 year periods (El Chicon in 1982 and Mt. Pinatubo in 1991). There were also 5 ENSO (El Nino Southern Oscillation) events and 4 La Nina events which resulted in warming and cooling, respectively. At least two of the ENSO events (1982-83 and 1997-89) and one La Nina event (1988-89) were particularly strong. Figures 20 and 21 show global temperature anomalies for the middle troposphere and lower stratosphere respectively, from 3 MSU/AMSU products and 2 radiosonde datasets (Seidel et al., 2003). The MSU/AMSU curves represent UAH Versions D and 5.0 (Christy et al., 2000; 2003) and RSS (Mears et al., 2003). The sonde curves are for HadRT2.1 and HadRT2.1s for the troposphere and stratosphere respectively (Parker et al., 1997), and LKS (Lanzante et al., 2003). In both figures, the bottom curve gives the average of all 5 datasets and the curves for each dataset above it give deviations from this average. Figure 22 shows 5 globally averaged and one tropically averaged time series from these datasets compared with the Quasi-Biennial Oscillation Index (QBI) and the Southern Oscillation Index (SOI), also from Seidel et al. (2003). In all 5 datasets the short-term variations in temperature anomalies (1-2 years) are larger than the resulting decadal trends and none of the trends shown are monotonic. The signatures of the El Chicon and Pinatubo eruptions are clearly visible as stratosphere warming and troposphere cooling in the years following each, as are the larger El Nino events since 1980. Also visible are the impacts of the QBI and SOI oscillations.

There is significant geographical variance as well, with fluctuations in the tropics being larger than the corresponding global trends particularly during the late 80's and the late 90's. With this much variance in the data over time scales representing significant portions of the record, the measurement of long-term trends becomes sensitive to how the trends are computed, and it is not clear how meaningful these trends are. This is particularly true of the radiosonde record, which is far more sensitive to variations in metadata than the satellite record and has suffered extensively from poorly kept station records. As was noted above, a shift of one level in a dataset due to corrections for metadata detected events can increase by as much as 50 percent the amount of time necessary to accurately derive a trend (Weatherhead et al., 1998).

The derived trends from MSU/AMSU also show much variation. Figure 8 shows the calculated tropospheric temperature trends as determined by the latest analyses of each investigating team. The values shown for RSS reflect their first version (Mears et al., 2003) and their most up-to-date figures as of this writing (Oct. 2004). UAH figures reflect Version 5.0 (Christy et al., 2003) and also their latest figures. Despite years of uncompromising dedication - in the case of the RSS and UAH teams, over 8 and 15 years of effort respectively - one is struck by the divergence in these trends. UAH observes little if any lower and middle tropospheric warming, but VG observe warming that equals, and in some places even exceeds that of the well documented surface warming. A closer look at these results reveals much about these differences impacted the measured trends.

Figure 23 shows the global time series of Channel 2 brightness temperature for the UAH and RSS teams, and Figure 24 shows their hot target temperature factors. The constant value of 0.03 used by Prabhakara's team to characterize all sources of error (except diurnal drift and orbital decay) including

IBE is shown for comparison. RSS target factors were derived using ocean-only data and are shown with and without the diurnal corrections derived this way. The UAH target factors were derived from land and ocean data. An examination of this data makes several things readily apparent.

- The diurnally corrected and uncorrected RSS target factors differ only slightly from each other, the largest differences being that for TIROS-N. This demonstrates the relatively weak contribution of oceanic diurnal cycles to target factor and the RSS method of using ocean-only data for target factor calculations.
- Comparison of each team's target factors shows that despite the differences in their merging methods, there is generally good agreement among them. The one obvious exception is NOAA-09, where the UAH team obtains a value that is much larger than anyone else's. The RSS value for NOAA-09 agrees well not only with their target factors for other satellites. It also agrees with the other UAH team values, and these are in turn in good agreement with the corresponding RSS values. Prabhakara's generalized constant target factor is generally similar. Given that all the JPL built MSU packages were nearly identical at launch time, this is not surprising. The only value that is a clear outlier is the UAH determined NOAA-09 factor.
- The UAH and RSS global time series' are nearly identical at the beginning of the record, but begin to diverge from each other at a more or less constant rate during 1986, when NOAA-09 was in service and very near to the NOAA-09/NOAA-10 service life overlap. The nearly step function change in trend variation that happened here appears to account for most of the difference between the global decadal trends calculated by the UAH and RSS teams.

It is noteworthy that the divergence between the UAH and RSS time series' begins during 1986, at or very near the NOAA-09/NOAA-10 overlap. This implies that the overall trend is very sensitive to how this overlap is handled. As part of their analysis, the RSS team extracted the residuals (remainders) left over for each dataset after they were combined to fit a final time series curve. The covariance matrix generated for the standard deviations of these residuals during merge calculations can be used to derive an estimate of the statistical uncertainties in each satellite's dataset. Using a Monte-Carlo method, they superposed noise onto the after-the-fit residuals from their Ocean-Only merged time series and generated an ensemble of 30,000 sets of "noisy" deviations from the covariance matrix in 17 merging parameters and added them to the intersatellite differences from the fitted trend. These were then used to create a set of 30,000 "noisy" merging calculations that were used to derive the error estimates in the resultant trends. Performing the same analysis on the temperature differences between all pairs of co-orbiting satellites during overlaps in service, they were able to characterize the sensitivity of the overall trends to each service life overlap.

Figure 25 shows a graphical representation of the results. It can be seen that compared to the others, the NOAA-09 offset from the fitted trend, and its target factor, are both poorly constrained and largely dependent on each other, so that errors in one will contribute directly to errors in the other. It can also be seen that the NOAA\_07/NOAA-09, NOAA\_09/NOAA-09, and NOAA-09/NOAA-10 overlap overlaps have significantly more uncertainty than the others, and the overall decadal trend is particularly sensitive to how the NOAA-09/NOAA-10 overlap is handled. This can be demonstrated further by using the UAH target factors in the RSS ocean-only merging analysis. When the RSS team did this they obtained a 1979-2001 global MSU2 trend of 0.014 deg K/decade for the first case, which is much closer to the corresponding UAH value of -0.011 deg K/decade than to their measured trend of 0.099 deg K/decade. Using their target factors and substituting only the UAH NOAA-09 factor gives a trend of 0.022 deg K/decade, showing that the large majority of the RSS-UAH difference is accounted for by the NOAA-09 factor only (Mears et al., 2003c). Likewise, the RSS team also applied the UAH merging method and diurnal correction to their own pentad averaged datasets and obtained a NOAA-09 target factor of 0.075, which compares more favorably to the UAH value of 0.095 than to all other factors determined by both teams and Prabhakara's team (Mears et al., 2003).

Thus, the overall trend differences between UAH and RSS can be attributed to how each team smoothed their data and conducted their dataset merge and target factor derivations, particularly for the NOAA-09 target factor. The inclusion or omission of NOAA-09/NOAA-10 overlap and/or the other NOAA-09 overlaps, allows for a great deal of variability in outcome and makes it possible to minimize either the trend or the uncertainty depending on what is chosen or ignored. This introduces a degree of arbitrariness into the UAH analysis that renders their results questionable. By including all of the datasets with appropriate weighting for short and/or noisy overlaps, as the RSS team did, this arbitrariness can be avoided while still accounting for the related uncertainties. It is significant that by neglecting several overlaps, including NOAA-09/NOAA-10, UAH obtained a NOAA-09 target factor that is anomalously high not only compared to all other target factors found by other teams, but also to *their own values* for all other target factors. In light of this, it now appears that the UAH team selected their overlaps and merging methods to minimize the global trend in tropospheric temperatures rather than the overall uncertainty. Their stated objective was to minimize the intersatellite trend differences, and this they largely succeeded in doing (Christy et al., 2004). But by selectively omitting some satellite overlaps, they have obtained a NOAA-09 target factor that is inconsistent with their own results for other satellites and the target factors derived by other teams. It must be noted that the hot target calibration factor is a property of the hot target itself and not the scene viewed or the data gathered. As such, a large factor is indicative of an issue with the hot target itself. The hot targets carried by MSU and AMSU packages on POES satellites are little more than simple blackbody emitters regulated by simple electronic devices and monitored with two PRT thermocouples. There will always be at least some variability in the manufacturing of such devices, as there would in the manufacture of any product. But a variance of 300 percent for one of these devices as compared with all other such devices in service is unconvincing. Add to this the noisiness associated with the nadir/off-nadir differencing derivation of their composite TLT numbers for the lower troposphere, and the associated pollution of these numbers with data from the surface, and it is not surprising that many researchers today question the UAH lower and middle troposphere temperature trends. In their favor, UAH trends to compare favorably with some radiosonde datasets, a fact that continues to give credibility to their derived trends. But the issues with radiosonde analyses limit their viability as an independent check and cannot be used to conclusively validate UAH in light of the other problems. There is also the issue of Antarctic sea-ice and melt-pool area impacts on UAH 2LT and TLT products. Because these factors will not affect the products of other teams who have based their results on Channel 2 alone, it is not surprising that UAH lower troposphere products show smaller trends for this layer than would be expected from Channel 2 and surface temperatures alone (Swanson, 2003), particularly as there is independent reason to believe that many regions of the Antarctic have shown surface and lower troposphere warming in recent decades (Vaughan et al., 2001; Thompson and Solomon, 2002).

It remains to consider Vinnikov and Grody (2003). As noted, their analysis avoids much of the noise associated with other methods, but leaves us having to account for IBE, discrepancies between land and oceanic diurnal cycles, and the way their data is zonally averaged between satellite tracks. Given that they use pentad smoothed datasets, as did the RSS team, and only low noise nadir views, it is reasonable to estimate the impact each of these effects has on their analysis by comparing them with the corresponding RSS estimates. Figure 26 shows RSS Version 1.0 and UAH Version 5.0 decadal trends for land and ocean data, with and without diurnal corrections. It can be seen that the RSS diurnal correction contributes about 0.007 deg K/decade to the oceanic trend and 0.064 deg K/decade to the land trend. It is reasonable to conclude that depending on how the VG diurnal cycle is zonally averaged, their numbers would be corrected by up to half of this amount, or around 0.03 deg K/decade. Interestingly, VG observed a trend in their diurnal maxima and minima that contradicts what has been observed (Vinnikov and Grody, 2003). Their analysis gives a diurnal cycle of the warming trend that yields a daytime maximum and a nighttime minimum, indicating that the diurnal cycle is increasing in amplitude, which is not observed to be the case (Karl et al., 1993). When VG correct for this in their analysis, they get a downward shift of 0.04 deg K/decade, not unlike this 0.03 deg K/decade interpolation, yielding their lower bound of 0.22 deg K/decade. Strictly speaking, the potential error they addressed in warming trend diurnal cycle is a correction of temporal phase, not geographical location. If VG's numbers were corrected for both diurnal effects, it would be reasonable to expect them to drop to something like 0.22 – 0.03, or 0.19 deg K/decade. For IBE, VG rolled all error estimates into a single linear correction which could be expected to absorb part, but not all of the IBE accounted for by the target factors in the UAH and RSS analyses. Since IBE is determined by the product of hot target factor and target temperature variation, we can estimate the impact of it by considering the observed variations in target temperature for

selected MSU packages. Figure 27 shows MSU2 monthly hot target temperatures for a few MSU packages. Inspection of these variations reveals that while the amplitude of variation is in some cases quite large (e.g. NOAA-12), the normalized average fluctuations are on the order of 2 to 7 deg K/decade with the average somewhere in between. The UAH and RSS target factors shown in Figure 24 average to around 0.018, neglecting the UAH NOAA-09 factor. From these values we can make a very rough estimate of the impact of IBE corrections to the VG trend as being on the order of -0.036 to -0.126 deg. K/decade at the very largest. With their Version D release, UAH produced a more thorough estimate of IBE impacts by estimating the dependence of the MSU calibrated antenna digital count errors with variations in Hot Target temperature (Christy et al., 2000). They obtained an estimated impact of 0.026 deg. K impact per deg. K temperature variation in the Hot Target. Based on these estimates, it appears that if Vinnikov and Grody's zonal averages of observed temperatures between satellite paths are reasonable, correcting their analysis to include IBE and both land and ocean diurnal cycles will likely cause their numbers to converge with those of the RSS or Prabhakara teams rather than the UAH team.

### Comparisons of Radiosonde Products

Given the issues with MSU/AMSU analysis products, several teams have produced comparison studies of the two. Some have sought independent validation of MSU results (Christy et al., 2000; 2003). Others have sought to refine the reliability of radiosonde analyses using MSU products as an independent check, or to clarify the correlations and interdependencies between sonde and MSU products (Angell, 2000; 2003; Seidel et al., 2003). These studies have yielded differing results. Some found no differences in trend between the surface and tropospheric temperature records since 1958 (Angell, 2000). Others indicate that the troposphere has warmed relative to the surface before 1978 and then cooled with respect to it thereafter (Gaffen et al., 2000a). For many of these results, variance in the data is similar in magnitude to the trends being measured, particularly in regions such as the tropics and high southern latitudes that are particularly important for comparisons with different MSU/AMSU analysis products. This variance in results has impacted attempts to validate MSU/AMSU analysis products using sonde analyses.

The UAH team in particular has compared their results with selected networks of radiosonde sites as part of their Version D and 5.0 releases, and more recently, a lower troposphere intercomparison with their Version 5.1 release (Christy and Norris, 2004). In Version D (Christy et al., 2000) UAH derived a time series for the 97 radiosonde stations they used covering North America, Bermuda, Iceland, and the Western Pacific. These time series were then compared to Channel 2LT data from UAH's 2.5 degree monthly maps at the same locations. Care was taken to construct the selected sonde network prior to evaluation of the corresponding MSU data points to insure that the two were independent. Direct radiosonde records, which give a vertical temperature profile with specific readings being taken at regular height intervals as the device ascends (once for every 5 meter rise is common), were converted into their corresponding weighted averages as they would have been read by the MSU with a constant surface emissivity being assumed (the MSU would directly detect averaged fluctuations of actual surface emissivity as it influenced the 2LT data). Good agreement was found between the two. Both datasets agreed on lower tropospheric trend to within 0.005 deg. K/decade with an annual correlation of 0.97. UAH notes however that the decadal trend observed for these sites was +0.16 deg. K/decade – considerably higher than the Version D global trend of -0.01 deg. K/decade demonstrating that the sites chosen, though noteworthy for their reliability as sonde sites go, were not representative of the whole earth. Seeking sonde data with more global coverage, they also compared their Version C and D datasets with Angell 63 and HadRT2.0 analyses (Angell, 1988; Parker, et. al., 1997) after checking the data for internal consistency and binning it into a global grid to facilitate comparison with their own globally gridded results. Figure 28 shows how the 4 datasets compared. It can be seen that the overall agreement is fairly good, though there are noticeable differences in the 1984-85 and 1990-95 timeframes and lower correlations with MSU (0.91 for HadRT2.0 and 0.93 for Angell). It should also be noted that the Angell 63 network predates Angell 54 and has not been corrected for the 9 stations known to have anomalous records (Angell, 2003), and the HadRT2.0 dataset has not been corrected for anomalous record changes either, as were later HadRT products such as HadRT2.1.

For Version 5.0 (Christy et al., 2003), UAH generated 3 composite simulated radiosonde temperatures designated as RLT, RMT, and RSL (to facilitate comparison with their TLT, TMT, and TLS Channels, respectively) and compared them with MSU results on a range of scales. Analysis was done using CARDS data for a single station and a small network of 28 stations chosen for their internal consistency of method and equipment, and an additional comparison was made using a more global network with fewer controls. For their single station comparison, Minqin, China (38.6 deg. N, 103.1 deg. E, 1367 m elevation) was used. Minqin was chosen for its unique consistency in equipment, methods, and continuity of record. The 28 stations listed in their appendix, were chosen to optimize the same levels of consistency to the greatest degree possible. Excellent agreement was found for Minqin with a monthly correlation between the two datasets of 0.90 for the lower troposphere and 0.95 for the middle troposphere, monthly standard deviation differences of less than 0.53 deg. K, and only 0.01 deg. K/decade difference in overall trend. Good agreement was also found for their 28 station network, with the corresponding figures being 0.94 and 0.93 for monthly correlation, 0.15 and 0.12 deg. K for monthly standard deviation variance, and an overall trend difference of -0.05 deg. K/decade.

For their larger global analyses, UAH compared TLT data with RLT data generated from 4 other sonde analysis products – HadRT2.0, Angell 63 (as with Version D), RIHMI, and a temperature reanalysis product from the National Centers for Environmental prediction (NCEP) generated from in situ sonde and satellite observations (other than MSU) <sup>8</sup>. Comparison was also made with results from a general circulation climate model (Kalnay et al., 1996). In each case, temperature vs. altitude data for a vertical column of 850 to 300 hPa altitudes (corresponding to the lower troposphere) was assigned a weighting similar to that observed by MSU devices to generate an RLT simulated MSU measurement to facilitate comparison with TLT data. Good agreement was found for all products, with trend differences of less than 0.02 deg. K/decade for all products. Figure 29 shows time series of annual lower troposphere temperature anomalies from UAH's Version 5.0 TLT data and the corresponding products from NCEP and HadRT. NCEP is a reanalysis product that uses data from in situ sources, satellite data (other than MSU, but using NOAA sounding profiles similar to those used in MSU products), and results from a global circulation model (Kalnay, 1996). HadRT data for this comparison was taken from HadRT 2.1 which corrects for discrete temperature changes using comparisons with MSU Ver. D data, and thus is not entirely independent of MSU results. Likewise, NCEP data is dependent to some extent on radiosonde products and is not entirely independent from them either.

Figure 30 gives a summary of 95 percent confidence intervals for monthly anomalies between their TLT (lower troposphere) and TMT (middle troposphere) data, and their Minqin, 28-station network, HadRT, and NCEP analyses. Despite problems with instrument and method consistency for the larger scale analyses and gaps in data (Christy et al., 2003) there is generally good agreement between all datasets. Since the publication of UAH Ver. D, the Angel 63 product has been updated to Angell 54 as described above. Removal of the anomalous 9 stations resulted in a significant warming of the 1958-2000 record compared to the original up to an altitude of 100 hPa, particularly in the tropics. The global changes in trend are smaller, but also show an increase. Most of the change however occurs before 1979 and the beginning of the MSU record. There is general consensus from this data that globally, the troposphere and the surface have warmed at roughly the same rate since 1958, but that prior to 1979 the troposphere warmed more than the surface, and since 1979 has cooled with respect to it. Removal of the anomalous sonde stations improves the agreement between most sonde datasets and the UAH Version D MSU record (Angell, 2003). Christy and Norris' Year 2004 intercomparison study also found good agreement between the MSUTLT record and their 89 station radiosonde record, with the global trend differences generally falling within +/- 0.04 K/decade of each other depending on the subsample chosen and whether the data had been adjusted for discontinuities or not (Christy and Norris, 2004). They also found consistently larger differences between their Version 5.1 MSUTMT product and the corresponding product from RSS, though it is not at all clear that RSS-UAH discrepancies can shed any light on the MSUTLT record (for which RSS has no comparable product).

Independent radiosonde analysis products have yielded varying results depending on the methods and networks used. Globally, Brown et al. (2000) find that since 1958 the surface to lower troposphere has warmed by about 0.20 deg. K/decade. Lanzante et al. (2003) find a warming of around 0.15 deg. K/decade between 1958 and 1997 for the same layers using datasets adjusted via the LKS methodology. Gaffen et al. (2000a) find a warming of 0.08 deg. K/decade. With the update of Angell 63 to Angell 54, Angell (2003) finds 850-300 hPa tropospheric warming equal to that of the surface and a tropical warming

of 0.13 deg. K/decade for 1958 to 2000. Figures 31 and 32 respectively show 1958-1997 temperature trends for 3 atmospheric layers from 5 radiosonde analysis products, and similar 1979-2001 trends for 3 MSU products and one radiosonde product. Each figure also shows a typical confidence interval that generally represents each of these datasets (Seidel et al., 2003). The HadRT trends represented here are from HadRT2.1s. For the 850 – 300 hPa layer, Angell 54 yields 0.10 deg. K/decade while HadRT and LKS yield 0.08 and 0.125 deg. K/decade respectively. RIHMI yields about 0.45 deg. K/decade, and with respect to the typical  $2\sigma$  confidence interval shown, is a clear outlier with respect to the other sonde products.

Over this longer period, these results are more or less consistent with the predictions of many AOGCM's and traditional theories of global atmospheric vertical energy and temperature transport. But the shorter periods of 1958-1979 and 1979-2000 (the non-satellite and satellite era's respectively) reveal a complex evolution of regionally and global lapse rates. For 1979-2000, Gaffen et al. (2000a) and Brown et al. (2000) find that in the tropics, the lower troposphere warmed with respect to the surface between 1958 and 1978 (just prior to the beginning of the MSU record) and then cooled with respect to it afterward. Angell (2003) finds that between 1958 and 1979 the global 850-300 hPa layer warmed with respect to the surface by nearly 0.13 deg. K/decade, and then cooled with respect to it between 1979 and 2000. Brown et al. (2000) and Hegerl and Wallace (2002) observed similar trends. For the shorter period of 1979-1997, LKS shows a global trend of 0.07 deg. K/decade for the 850-300 hPa layer, and -0.05 deg. K/decade for the 1015-10 hPa layer weighted to simulate MSU2 (lower and middle troposphere), and HadRT2.0 yields 0.09 deg. K/decade and -0.11 deg. K/decade respectively. For weightings simulating MSU2LT, the corresponding figures from these products are 0.075 deg. K/decade for LKS and -0.01 deg. K/decade for HadRT2.0. Extending the satellite era record from 1997 to 2001 yields good regional and global agreement between UAH and HadRT2.0 for MSU2LT, but considerable disagreement for MSU2, where the HadRT2.0 data appears to diverge from both UAH and RSS both globally and in the northern hemisphere and tropics (Seidel et al., 2003). The LKS dataset does not extend past 1997.

Global stratospheric trends have been less variable and are characterized by better agreement between radiosonde and MSU products. While stratospheric trends are not directly related to tropospheric temperatures and considerations of anthropogenic greenhouse warming, they are of indirect importance because of their impact on MSU2 data. Strong stratospheric cooling trends (due to ozone depletion) will introduce some spurious cooling into MSU2 trends, which are otherwise generally considered to be synonymous with the lower and middle troposphere. For the period of 1958-1997, Angell finds cooling trends of -0.48 deg. K/decade and -0.41 deg. K/decade for the globe and the tropics, respectively. HadRT shows comparable cooling of -0.39 deg. K/decade globally and -0.37 deg. K/decade, and LKS finds cooling of -0.38 and -0.36 deg. K/decade respectively for the same regions (Seidel et al., 2003). RIHMI shows far smaller cooling trends than the other sonde products with global and tropical trends of -0.20 deg. K/decade and -0.21 deg. K/decade respectively. Once again, RIHMI is an outlier compared to the other sonde products. Similar figures are obtained for the satellite era, indicating that during the last half century stratospheric lapse rates have been less variable than their tropospheric counterparts. Regional data reveal some interesting observations. Angell 54 finds much more long-term cooling for the southern hemisphere than other products (-0.79 deg. K/decade vs. -0.42 deg. K/decade for HadRT2.0, the next closest result). For 1979-1997, all analysis products are more regionally self-consistent, but LKS shows much more cooling than the other products. In all cases, the radiosonde products appear to show more cooling than MSU. This might be due to the fact that MSU4 is fairly strong at 100 hPa which is quite near the tropopause, a region characterized by high lapse rates and rapid transitions from warming to cooling, particularly in the tropics. In addition, some sonde datasets such as Angell 54 show more cooling than others (e.g. LKS) due to differences in southern hemisphere coverage that will cause some datasets to selectively emphasize the south pole ozone hole more than others (Angell, 2003). This may also explain part of the increased cooling observed between sonde datasets and MSU. It is also known that the shift toward Vaisala radiosondes, and the further evolution of these packages during the 1990's (e.g. the evolution of designs from the RS11 to RS90 Series of devices) has also introduced a spurious apparent cooling, particularly in the stratospheric record. Differences in how this evolution was corrected for may result in differences in apparent cooling trends. For the 300-100 hPa layer, which globally straddles the tropopause, there is considerable disagreement on the 1958-1997 trend between all products at statistically significant levels, with some products disagreeing even on the sign of this trend (Seidel et al., 2003).

Though there are statistically significant differences between various radiosonde and satellite products, confidence intervals are quite large, particularly for radiosonde simulated MSU2 and MSU2LT data in the southern hemisphere and tropics. In many cases the trends show  $2\text{-}\sigma$  confidence intervals that are larger than the trends being measured. Stratospheric confidence intervals tend to be much larger than their tropospheric counterparts, reflecting the many problems associated with incomplete datasets due to failed equipment at higher altitudes, particularly the bursting of inferior balloons at random stratospheric altitudes before datasets could be completed. In general, there is much more confidence in data from the northern hemisphere than the southern hemisphere and tropics, reflecting the better quality of data and coverage from that region (Seidel et al., 2003). But all current radiosonde datasets agree that globally, over the longer term (1958 to 2000) the surface and 850-300 hPa layers have warmed at comparable rates, but since 1979 the surface has warmed relative to the 850-300 hPa layer with the estimates ranging from 0.04 to 0.14 deg. K/decade for the various datasets (Angell, 2003). There appears to have been a step function change in lapse rate around 1976-77 that is not yet fully understood. Thus, the long-term trend in tropospheric temperatures are consistent with the predictions of the best extant AOGCM's, but over shorter intervals these temperatures have evolved in more complex ways. It is not at all clear that trend estimates for the last 25 years can be either generalized to model predictions or extrapolated into the future.

So how do radiosonde and MSU analysis products compare with each other? Are they truly complementary, and does one vindicate the other as so many people hope? Perhaps, but with so many factors influencing both, and given the complexities of how the upper atmosphere has evolved in the last 50 years, the devil is in the details. It must be remembered that while the various tropospheric analyses from MSU differ in their resulting trends, these datasets are truly global in their coverage while their radiosonde counterparts are not. Regional temperature trends can vary widely, despite the superficial appearance of global agreement. Similar global trends often mask large variations and/or consistencies in regional trends. Sonde analyses will be useful in MSU comparison studies only if they have adequate coverage and confidence levels in regions where MSU trends from different teams disagree. Figure 11A shows 1979-2001 MSU global decadal trends as determined by the RSS team (a) the UAH team (b), and the difference between the two (c) from Mears et al. (2003). Figure 11B shows a similar comparison for 1979-2002 trends (Mears et al., 2003b). It can be seen that the two analyses are in relatively good agreement with each other over most of the northern hemisphere. The most significant differences between them are in the tropics, northern Africa, parts of Siberia, and in the high southern latitudes (Mears et al., 2003). It is these regions that discriminate between RSS and UAH global trends. Comparing Figures 11A(c) and 11B(c) with the regions covered by the networks used in various radiosonde analysis products reveals that the bulk of the sonde coverage is elsewhere. For instance, the Minqin sonde site (38.6 deg. N, 103.1 deg. E, 1367 m elevation) used in UAH Version 5.0 for their single station comparison, is at a location where there is very little difference between the UAH and RSS trends (Christy et al., 2003; Mears et al., 2003). The Angell 54 station network has only 6 stations representing both the south temperate and south polar zones where many of the largest differences are. Several of the southern hemisphere sites are in a mid-latitude band where UAH and RSS again largely agree (e.g. Hedland in Australia, Amsterdam in the Indian Ocean, and Pascua in the Western Pacific). The LKS network has better coverage than Angell 54, as does HadRT and RIHMI, but all are still relatively sparse in regions that are of greatest interest for MSU/AMSU comparisons, particularly the South Pacific. Furthermore, the LKS dataset only covers the period up to 1997 and thus covers only  $\frac{3}{4}$  of the extant MSU/AMSU record, and in particular, it omits much of the significant warming that characterizes the end of the 20<sup>th</sup> century and the beginning of the 21<sup>st</sup>. The impact of these differences can be seen in regional trends and confidence levels.

For 1978-1997, UAH Versions D and 5.0 show excellent agreement with LKS both regionally and globally for MSU2. But there is much less agreement for MSU2LT where LKS shows higher trends both globally and in all regions. The opposite is true for HadRT, which shows relatively good agreement with both UAH products for MSU2LT, but more global and regional cooling for MSU2. RSS consistently shows more warming UAH and LKS, but confidence intervals largely overlap except for the southern hemisphere, which may well over-represent cooling due to the sparse network sampling problems already discussed. HadRT2.0 consistently shows more cooling than UAH and RSS for what is likely the same

reasons, though agreement is better with UAH (a fact that might possibly be related to the two not being fully independent). It can also be seen that while LKS shows excellent agreement with UAH for MSU2, it shows considerably more stratospheric cooling than other sonde and MSU products, which suggests that it may be under-representing MSU2. If so, correcting for this may well bring LKS into better MSU2 agreement with RSS, which shows less MSU4 cooling than UAH, HadRT, and LKS (Seidel et al., 2003).

Extending the comparisons to 2001 changes the picture yet again. For the longer period, LKS data cannot be used, as it only extends to 1997. For this period, HadRT again shows good agreement with both UAH products both regionally and globally, but confidence intervals are high with respect to the observed trends, particularly in the tropics. For MSU2, there is less agreement, though again UAH is closer to HadRT than RSS. But confidence intervals are large enough to include all 3 MSU products, and HadRT differs considerably from MSU even in the northern hemisphere where UAH and RSS data are best characterized. Note also that once again, HadRT shows considerably more stratospheric cooling (MSU4) than UAH and RSS, raising the likelihood that it may be under-representing MSU2 (Seidel et al., 2003).

UAH has acknowledged the difficulties these considerations present for their sonde validation studies, and have sought to address many of them in their most recent work. Since the release of UAH Ver. 5.1, they have expanded the number of stations used in their sonde validations, particularly in the southern hemisphere where comparisons are most important and datasets are weakest. The expanded network of 89 stations in the southern hemisphere and tropics still shows good agreement with their Version 5.1 MSUTLT analyses (Christy and Norris, 2004; Christy et al., 2004). Though many of their southern hemisphere stations have only 60 to 75 percent availability of monthly data for the 1979-2001 period, against this concern is the fact that there is relatively good agreement between their 60 and 75 percent availability data, and between their adjusted and unadjusted data as well, suggesting that the impact of these discontinuities on the TLT record has not been large. But the issue of coverage remains. Once again, a comparison of Figures 11A(c) and 11B(c) with Figure 41 shows that even with its expanded southern hemisphere coverage, the 89 stations are concentrated mainly on land in regions such as Australia and the South American continent. The southern oceans, the most critical regions for comparison, are still under-represented. Furthermore, though Christy and Norris have pointed out that their MSUTLT comparisons with this network are quite favorable while their MSUTMT to RSS comparisons at the same locations are not (Christy and Norris, 2004), it is not at all clear that MSUTMT trends between differing products can be evaluated based on MSUTLT comparisons. As such, agreement on this point is hardly conclusive regarding the middle troposphere.

Another problematic aspect to these sonde comparisons can be seen by comparing Figures 11A(c) and 11B(c). Figure 11A(c) shows the global discrepancy between UAH Ver. 5.0 and RSS Ver. 1.0 for the period of 1979-2001 (Mears et al., 2003). It can be seen that the largest disagreement is in the southern hemisphere and tropics, and that the disagreement is largest at the higher southern latitude oceanic regions and in the tropical Pacific Ocean (the yellow regions) – as has been previously stated. Figure 11B(c) shows the same data for updated UAH and RSS products giving trends for 1979-2002 (Mears et al., 2004). The addition of just one year's worth of data has changed the situation dramatically. Now we see that the agreement between UAH and RSS has improved considerably over the tropics and most of the southern latitudes. The principal area of disagreement south of the equator is now concentrated mainly at the high southern latitudes and over oceanic regions. UAH and RSS show little difference in trends over the South American continent and Australia where UAH has most expanded their southern hemisphere sonde data. Thus, even though the latest global tropospheric trends from both teams still show marked differences, the areas of agreement actually expanded between 2001 and 2002, and the regions that are most crucial for discriminating between the two products has actually retreated even further into the regions of poorest sonde representation, suggesting that there is enough variability in the record to significantly alter the validity of intercomparison studies with just one or two additional years to the record.

These considerations readily show the difficulties encountered when attempting to validate MSU analysis products with radiosonde datasets. Certainly, much can be learned from such studies. Any independent dataset is likely to clarify things that other types of data might obscure and radiosonde data is no different. The increased vertical resolution of these products, and the fact that they use different methods to measure temperature can add robustness to atmospheric temperature studies. There are currently several attempts under way to expand radiosonde products. The LKS product is currently being

expanded beyond 1997 to the present and refined to include better characterized data records for the 1957-1997 period. The resulting product will be the NOAA Radiosonde Atmospheric Temperature Products for Assessing Climate (RATPAC). HadRT2.1 is also being upgraded using various spatial consistency checks and additional data. The work being done to develop the Global Climate Observing System Upper Air, or GUAN network also has done much to improve the quality and coverage of radiosonde data since the National Research Council published their landmark study calling for such improvements (NRC, 2000). Clearly, care must be taken when interpreting the existing MSU and radiosonde analyses, particularly as independent validation of each other. As of this writing, what can be said with reasonably high confidence is the following;

- Of the radiosonde products most used for MSU comparisons, all have strengths and weaknesses, and no one product stands out as exceptionally well characterized compared to the others. There are significant differences in the methodologies each uses to detect and correct for spurious changes in historical data records, and little agreement among them as to what changes are necessary and to what degree. Yet these changes can have a significant effect on derived trends and there are statistically significant differences among the trends produced by each.
- It is generally agreed that all the radiosonde analyses considered here are currently necessary for a complete characterization of tropospheric and stratospheric temperatures. Selecting any one of them for comparisons emphasizing and one particular period (such as 1979-1997 or 1979-2001 for instance) is inappropriate. Of the sonde products most used for MSU comparisons, only the LKS product (Lanzante et al., 2003) can be said to be truly independent of MSU datasets at this time.
- None of these radiosonde analyses has truly global coverage. In general, the northern hemisphere is well covered geographically and temporally, and metadata records are relatively complete. But other regions are not so well characterized. Particularly problematic are the tropics, the high southern latitudes, the southern oceans, and northern Africa.
- Despite the differences, there is general agreement among radiosonde products that the long-term record (1958 to 2001) shows little difference between surface and tropospheric warming rates, but the shorter records are more complex. The troposphere warmed with respect to the surface between 1958 and 1978, and cooled with respect to it thereafter during the satellite era. There appears to have been a step function change in lapse rate trends some time between 1977 and 1979.
- Radiosonde analysis products generally show more stratospheric cooling than do MSU products. Some of this may be due to sampling errors associated with the southern hemisphere being under-represented, and some may be due to shifts toward the use of Vaisala sonde devices during the 1990's.
- Differences in trends from UAH and RSS analysis MSU2 products (RSS does not have an MSU2LT product) are mainly the result of regional variations in their data – most notably, the tropics, South Pacific, Atlantic, and Indian Oceans (in particular, the very high oceanic southern latitudes), and Northern Africa. These areas are poorly characterized in radiosonde analysis products compared to other global regions and do not clearly discriminate between UAH and RSS analyses. UAH values for MSU2LT and MSU2 tend to be closer to some radiosonde analysis products for certain regions and time periods (e.g. LKS for MSU2 from 1979-1997, and HadRT2.0 for MSU2LT from 1979-2001), but confidence intervals are usually large enough to accommodate RSS numbers also. Agreement is not preserved for other products and timeframes.
- The evolution of tropospheric temperature trends from MSU/AMSU products since 2000 has been in the direction of increased warming, gradually closing the gap between the predictions of AOGCM's and observation for the last 25 years. The long-term radiosonde record, which predates the MSU/AMSU record, shows good agreement with the predictions of AOGCM's. Of the extant MSU/AMSU analysis

products, only UAH Version C (Christy et al., 1998), now over 6 years old, and earlier UAH analyses have shown tropospheric cooling. This is now known to have been due to uncorrected spurious cooling in the MSU record.

- Of the MSU/AMSU analysis products to date, UAH and RSS are the most mature and best characterized. UAH products show the least amount to lower and middle tropospheric warming, to a degree that cannot be accounted for by the most up-to-date versions of AOGCM's. While there is much to be said for this product, it suffers from an analysis method that is somewhat arbitrary with its merge methodology, and yields anomalously high target factors for the NOAA-09 satellite that are not observed by other teams. The RSS methodology avoids these problems and shows more consistency and lower noise overall from various sources. While the observed means of radiosonde derived trends are closer to the UAH trends than to those of RSS products (which consistently show more warming than their UAH counterparts), the range of uncertainty in these analyses is far less than what is required to discriminate between the two products. As such, many of not most climate scientists today have more confidence in the RSS analysis products than those of UAH. Extant AOGCM's can comfortably reproduce the RSS derived middle troposphere and stratosphere temperature trends.

In summary, radiosonde analyses can be considered as complimentary to MSU products. Where they are continuous and well characterized, they provide a higher level of vertical resolution and a longer record than is available from MSU products, allowing for better characterization of lapse rates and an independent evaluation of MSU data at isolated locations. But their issues with continuity of record, sampling error due to geographical limitations, and incomplete metadata limit their ability to act as a truly reliable independent validation of MSU products.

### But Should They Agree?

It was noted earlier that under most global warming scenarios, the troposphere is expected to warm at a rate equal to or greater than the surface (IPCC, 2001). But as we have seen, for at least the last 20 to 25 years both the satellite and radiosonde records show much less warming than expected. The surface record is well characterized by a wide variety of direct and proxy indicators and can be established independent of the troposphere record. But the relationship between the two is less well understood. To date, the belief that their temperature trends should be similar has been largely due to the belief that they are strongly coupled and will exchange heat readily. AOGCM studies have tended to support this. In general, most have done a good job of reproducing observed global surface temperature trends, though they have been less reliable in regards to humidity and sea level pressure (IPCC, 2001, Chap. 8). Recent work has improved the record, but many questions remain. What role does the surface play in the radiative forcing of the troposphere vs. direct solar forcing? How much of a role does the vertical or horizontal advection of latent heat, moisture, or air mass play in surface-troposphere interactions? Have tropospheric humidity and cloud cover significantly impacted the relationship between the two over the last 25 years? These and many more questions are still up for grabs.

Also in question is whether the two trends really have been all that different since the late 70's. In their landmark Year 2000 Reconciling Observations of Global Temperature Change (NRC, 2000) the National Research Council examined the surface and troposphere records at length, including the uncertainties inherent in MSU and radiosonde products, the issues surrounding AOGCM simulations of the surface and upper atmosphere, and the degree to which natural climate variations might be playing a role. They concluded that while many open questions remain, there is a residual discrepancy surface and troposphere temperature trends that is statistically significant, and is not accounted for by state of the art AOGCM's – the two may well be exchanging heat, mass, or radiative forcing in ways that these models are not accounting for. This has led some to claim that global warming is not happening (Singer, 1999; Douglass et al., 2004; 2004b).

Some of the difference can be attributed to natural climate variations. Interannual cycles such as ENSO (El Nino Southern Oscillation), PDO (Pacific Decadal Oscillation), and NAO (North Atlantic

Oscillation) can significantly affect the surface temperature record for up to 3 years, independent of the long-term trend. Volcanic eruptions such as El Chicon in 1982 and Mt. Pinatubo in 1991 have also disturbed the troposphere and stratosphere with large injections of aerosols and particulates and caused significant climatic variation for several years after their occurrence. In fact, interannual cycles and catastrophic events have impacted more than a third of the MSU/AMSU history. While events such as these explain much of the difference between surface and upper air records, they have not explained all of it (Christy & McNider, 1994; Santer et al., 2000). Since the release of the NRC Year 2000 Report, AOGCM's have improved considerably. Those that are forced by both natural and anthropogenic climate inputs, including volcanic forcings (a relatively recent development) can reproduce observations to a fair extent. The ECHAM4/OPYC model developed jointly by the Max Plank Institute for Meteorology (Hamburg) and the European Centre for Medium-Range Weather Forecasts is a case in point. This model, in which an atmospheric model is coupled to an isopycnal ocean model using a 2.8 deg. latitude/longitude grid, 18 atmospheric layers, and flux corrections for heat transfer and fresh water mass flow, can reproduce many of the observed surface/troposphere discrepancies including the effects on ENSO and the eruptions of Pinatubo and El Chicon. But the fit with UAH Versions D and 5.0 is near, and in some cases, beyond the limit of confidence intervals for a significant portion of the record (Santer et al., 2000). It must also be noted that the NRC report was released prior to the publication of RSS Version 1.0 (Mears et al., 2003), which has changed the picture dramatically. As of this writing, most state of the art models that include all observationally known forcings can comfortably reproduce the RSS record. Therefore, the case for a clear discrepancy between surface and troposphere trends and the validity of AOGCM representations of both, boils down to evaluations of UAH vs. RSS, and the extent to which data from various radiosonde products can be used as independent validation of either.

It must also be remembered that microwave and radiosonde temperature data are not the only indications we have of a warming troposphere. There are also independent proxy measurements of climate changes that are consistent with increasing troposphere temperatures. Santer et al. (2003b) used reanalysis data from the National Centers for Environmental Prediction (NCEP) and the European Centre for Medium-Range Weather Forecasts (ECMWF), in conjunction with several runs of both atmosphere only, and ocean-atmosphere coupled circulation models to study variations in tropopause height as characterized by tropopause lapse rate between 1979 and 2000. They found that after volcanic effects and other known natural forcings were included, the observed changes in tropopause height over this period are a robust zero-order response of the climate system to forcing by well-mixed greenhouse gases and stratospheric ozone depletion. Increasing tropospheric temperatures imply an increasing atmospheric energy content below 100 hPa, and this in turn leads to decreasing tropopause lapse rates and an increase in tropopause height that is observed. Observations of this sort are difficult to reconcile with a troposphere that is not warming.

### Surface-Troposphere Coupling

If the surface and troposphere are indeed strongly coupled to each other thermally, then discrepancies between their temperature trends are indeed puzzling. At the very least, we would have to say either that uncertainties in MSU and radiosonde records are more uncertain than we imagined, or that there are regional and/or global interactions between the two that AOGCM's are not getting. Most of the climate simulation models used over the last decade or so assume thermal coupling across atmospheric vertical layers and have been less well characterized regarding things might interfere with this coupling (e.g. water vapor, sea level pressure, or deep convection cells). So it is not surprising that they predict similar surface and upper air temperature trends. But if the troposphere is even partially decoupled from the surface, either regionally or globally, then surface and upper air trends may well diverge (NRC, 2000). Recently, several lines of research have emerged suggesting that this may well be the case. One of the most promising has been the work of Kevin Trenberth and David Stepaniak of the National Center for Atmospheric Research (Boulder, CO) on the earth's global radiation budget. Trenberth and Stepaniak studied the earth's energy budget and the way solar energy input to the atmosphere and surface are redistributed globally. Among other things, they found that important zonal and poleward energy transports occur in the tropics and extra-tropics that redistribute latent heat much more strongly in these directions than vertically, decoupling the surface from the troposphere in these regions. The findings are particularly significant because it is primarily in these regions that lapse rates are much higher than expected from models, and the surface and troposphere trends are most noticeably different, and

uncertain, in the various datasets. There are two mechanisms at work here which strongly couple vertical and poleward heat transport providing an almost seamless energy balance that connects outgoing long-wave radiative cooling with annual variation of solar atmospheric heating. Radiative cooling of the earth at the top of the atmosphere is globally uniform. But because the earth's rotational orbital plane is tilted with respect to its solar orbital path (the ecliptic plane), the weighting of solar heating will shift in a meridional (north – south) direction annually – which is, of course, why there are seasons at higher latitudes. This requires a poleward energy transfer that must balance. Trenberth and Stepaniak showed that this balance has two components which favor a poleward transfer of latent heat that largely decouples the surface from the troposphere, particularly in the tropics and extra-tropics (Trenberth & Stepaniak, 2003a,b). They found that in lower latitudes the dominant mechanism of latent heat transport is the overturning of Hadley and Walker cells. In the upward cycle of these cells the dominant diabatic heat transfer occurs from the convergence of moisture driven by the cell motion itself. This results in a poleward transport of dry static energy that is partially, but not completely balanced by an equatorial transport of latent heat, leaving a net poleward transport of moist static energy. In the subtropics, the subsidence warming in the downward branch of these cells is balanced by cooling that arises from the poleward transport of energy by transient baroclinic eddies. These eddies are broadly organized into storm tracks that covary with global stationary atmospheric waves in a symbiotic relationship where one feeds the other. The relatively clear skies in the subtropics feed this cycle by allowing for strong solar absorption at the surface which feeds the latent heat transport cycle through evaporation, and in return, this is compensated by subsurface ocean heat transport that is itself driven by the Hadley circulation winds. The relationship between these cycles and how they exchange energy is shown in Figure 35.

For their analysis of the magnitudes of these effects, Trenberth and Stepaniak used overall energy transports derived from reanalysis products for the period 1979–2001 from the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) as derived by Kalnay et al. (1996) and used in Trenberth et al. (2001). These were deemed to be most consistent with the overall heat budget as determined from Top of Atmosphere (TOA) and ocean measurements (Trenberth and Caron 2001; Trenberth & Stepaniak, 2003a). Other complimentary heat budget data from the Southampton Oceanographic Centre (SOC) heat budget atlas was also used to characterize ocean surface heat transfer (Josey et al. 1998, 1999). Trenberth and Stepaniak noted that this data had considerable uncertainties due to sampling error and systematic biases from bulk flux parameterizations, but they were careful to use them only with relevant physical constraints that limited the impact of these uncertainties on their results (Trenberth et al., 2001; Trenberth and Stepaniak, 2003b). TOA data was taken mainly from Earth Radiation Budget Experiment (ERBE) satellite measurements of TOA radiation (Trenberth 1997). Precipitation estimates were taken from the Climate Prediction Center (CPC) Merged Analysis of Precipitation (CMAP) precipitation estimates (Xie and Arkin, 1997).

Figures 36 and 37 show typical zonally average annual magnitudes of the energy transfers involved in these various processes in the tropics and extra-tropics for the North Pacific (Fig. 34), and the South Pacific (Fig. 35) for the ERBE period February 1985–April 1989. It can be seen that the net effect is to give the earth's energy budget has a strong poleward component in the tropics and extra-tropics that redistributes a significant portion of surface reradiated, convective, and latent heat poleward rather than vertically. This should at least partially decouple surface temperature trends from upper troposphere trends in these regions in ways not accounted for in previous AOGCM's. Given that this effect is most evident in the tropics and extra-tropics, we should expect that heat transfer processes that would ordinarily bring the troposphere up to the same temperature as the surface will be at least partially diverted, leaving the troposphere cooler (or perhaps under some circumstances, warmer) in these regions than would otherwise be expected. The fact that it is the tropics and extra-tropics that display the largest discrepancies between UAH and RSS analyses lends further support to this theory. There are still considerable uncertainties in the magnitudes of some of the heat transfer budgets in this process, and more work needs to be done to fully characterize it (Trenberth & Stepaniak, 2003a,b), so the degree to which this process contributes to discrepancies between various MSU analyses and the surface record needs further examination.

The important point here is that the existence of such a mechanism means that *we should expect at least some disconnect between surface and troposphere warming rates in these regions*. Even if this disconnect proves to be of considerable magnitude, it would not present any issues for the long-term surface record, which we must remember, is robust and well characterized *independent of the*

*troposphere record* (NRC, 2000; IPCC, 2001). As it is today, MSU products, and to a lesser extent radiosonde products, vary between those that predict little if any disconnect and can be comfortably reproduced by state-of-the-art AOGCM's (Mears et al., 2003; Prabhakara et al., 2000; Vinnikov and Grody, 2003) and those that show relatively large, statistically significant disconnects (Christy et al., 2003). The truth is likely to be somewhere in-between. For our purposes, it is enough to emphasize that demonstrable differences between surface and tropospheric temperature trends do *not* invalidate either record.

### Stratospheric Signals – Fu et al.

There is one more question that still needs to be asked – *How well do MSU observations represent actual troposphere air temperatures?* It was noted earlier that MSU sensors detect upwelling radiation emissions from atmospheric oxygen on discreet frequencies according to weighting functions that give radiated intensity as a function of altitude. Figure 7 shows the weighting functions of Channels 2 and 4 from 1000 to 10 hPa altitudes. The region of interest for studies of anthropogenic greenhouse gas warming is the lower to middle troposphere covering the 850 – 300 hPa layer. The radiation detected by MSU Channel 2 and AMSU Channel 5 peaks at roughly 700 hPa (about 7 km) in the lower region of this layer, and receives most of its input from it making it, as we saw, most representative of the 850 – 300 hPa layer. Globally, the tropopause falls within the 300-100 hPa layer, and the 100-50 hPa layer crosses the lower stratosphere. The MSU Channel 4 and AMSU Channel 9 signals peak at about 90 hPa and receive most of their signal from the 200-10 hPa layer making them most representative of the lower stratosphere. Even though these channels are mainly sensitive to different layers, it can be seen that Channel 2 receives a non-negligible portion of its input from the 300 hPa layer up to nearly 30 hPa altitude, amounting to roughly 15-20 percent of its signal (NRC, 2000; Fu et al., 2004). Figures 20 through 22 show why this is problematic. Figure 22 shows averaged trends for several upper-air products from both radiosonde and MSU for MSU Channels 2 and 4, and the 850-300 and 100-50 hPa layers that respectively correspond to them. Figure 20 shows MSU2 in detail for UAH and RSS products along with 2 radiosonde products. Figure 21 show similar data for MSU4. For the 1979-2002 period, the 850-300 layer has generally trended upward while the 100-50 hPa layer has trended downward. Here we can clearly see the stratospheric cooling that was discussed earlier – due mainly to ozone depletion (Bengtsson, 1999; NRC, 2000; IPCC, 2001). Therefore, the lower stratosphere will alias a spurious cooling in the MSU2/AMSU5 trend making it less representative of that layer's actual trends. UAH chose to account for this with their synthetic 2LT and TLT Channels, which earlier we saw were weighted to lower altitudes, avoiding significant input from above 100 hPa. We also saw how this approach, while elegant and effective to a fair degree, also significantly amplified sampling noise and other signal pollution from surface emissions.

In May of 2004, a team lead by Qiang Fu of the University of Washington published a study in Nature that presented another approach to removing this “trend contamination”. Because MSU4 provides a near direct measure of the 100-50 hPa layer, it can be used to estimate the lower stratospheric cooling that is being aliased into MSU2 and remove it (Fu et al., 2004). What Fu and his colleagues did was to derive explicit vertical temperature profiles from global radiosonde data and use these to derive effective weighting functions for MSU2 and MSU4 that remove the lower stratospheric influence from MSU2, leaving behind a pure 850-300 hPa layer brightness temperature. One of Fu's co-authors, Dian Seidel, was also a co-author of the LKS radiosonde analysis (she is the “S” in LKS). Bringing her considerable expertise with radiosonde networks to Fu's team, Ms. Seidel used the 87 station LKS network (Lanzante et al., 2003) to provide monthly temperature profiles for the surface and 15 pressure levels from 1000 to 15 hPa for the globe, the northern and southern hemispheres, and the tropics for the period of 1958 to 1997. Known weighting profiles for MSU2 and MSU4 (Christy et al., 2003) were then superposed on this data and used to derive a new weighting function  $\mathbf{W}(h)_{FT}$  that is positive up to roughly 100 hPa and negative above. The net effect is a weighting that averages to zero above 100 hPa and equal to the MSU2 weighting below. This effectively removes the stratospheric input and leaves only the desired 850-300 hPa layer brightness temperature. This can be expressed as follows,

$$\mathbf{W}(h)_{FT} = \alpha_2(h)\mathbf{W}_2(h) + \alpha_4(h)_{200-0}\mathbf{W}_4(h) \quad (1)$$

where  $\mathbf{W}_2(h)$  and  $\mathbf{W}_4(h)$  are the MSU2 and MSU4 weightings respectively, as functions of altitude and  $\alpha_2(h)$  and  $\alpha_4(h)$  are additional factors that correct out the lower stratospheric influence on MSU2 indicated by the MSU4 signal below 100 hPa.

Apart from the obvious El Chicon and Pinatubo volcanic events, the stratosphere has displayed a more regionally and temporally stable trend history during the satellite era than the troposphere. Because of this, and the fact that MSU2 and MSU4 weighting functions have been temporally stable as well, it is possible to use equation 1 to express the 850-300 hPa layer brightness temperature *trend*,  $\mathbf{T}^{tr}_{850-300}$ , as a linear sum of the observed MSU Channel 2 and Channel 4 brightness temperatures  $\mathbf{T}^{tr}_2$  and  $\mathbf{T}^{tr}_4$  factored respectively by constant coefficients  $\mathbf{a}_2$  and  $\mathbf{a}_4$ , plus a constant offset term to reflect the contribution of surface emissions. Thus,  $\mathbf{T}_{850-300}$  can be expressed as,

$$\mathbf{T}^{tr}_{850-300} = \mathbf{a}_0 + \mathbf{a}_2\mathbf{T}^{tr}_2 + \mathbf{a}_4\mathbf{T}^{tr}_4 \quad (2)$$

where  $\mathbf{a}_2$  and  $\mathbf{a}_4$  reflect the corresponding weighting function terms in equation 1. Fu's team used monthly averaged  $\mathbf{T}^{tr}_2$  and  $\mathbf{T}^{tr}_4$  data taken directly from MSU/AMSU data from Mears et al. (2003) and derived the coefficients  $\mathbf{a}_0$ ,  $\mathbf{a}_2$  and  $\mathbf{a}_4$  from these temperatures and their LKS based synthetic weighting profiles. Then, they derived the satellite era trend in  $\mathbf{T}^{tr}_{850-300}$  from equation 2 using least squares methods. Note that in equation 1,  $\mathbf{W}(h)_{FT}$  is *not* a real weighting function as the requirement that it remove the stratospheric signal makes it go negative above some critical altitude. This is in contrast to actual brightness temperature weighting functions for  $\mathbf{T}_2$  and  $\mathbf{T}_4$  such as those shown in Figure 7, which must be everywhere positive (real heat sources cannot have negative emissivities). Fu's team found  $\mathbf{a}_2$  and  $\mathbf{a}_4$  to be equal to 1.156 and -0.153 respectively for globally averaged anomalies. For global trends,  $\mathbf{W}(h)_{FT}$  generally peaks in the middle troposphere at the same level as  $\mathbf{W}(h)_2$  but 15 percent higher, and goes negative at the top of the tropopause (100 hPa), exactly as we would expect for removal of the Channel 4 influence.

To check the validity of their derived values for  $\mathbf{a}_2$  and  $\mathbf{a}_4$ , the team applied simulated MSU2 and MSU4 weighting functions to the LKS dataset (Lanzante et al., 2003) and derived effective globally averaged  $\mathbf{T}^{tr}_2$  and  $\mathbf{T}^{tr}_4$  values for the full period of 1958 to 1997. These were used to calculate an effective global  $\mathbf{T}^{tr}_{FT}$  from equation 2 for the for the same period, which was then evaluated against the actual LKS radiosonde determined temperature profiles for the 850-300 hPa layer. They found excellent agreement between the two, with RMS error and trend difference of 0.065 deg. K and 0.001 deg. K/decade respectively, and a correlation coefficient of 0.984. This shows that globally, the vertical structure of stratospheric temperatures are very coherent and equation 2 can be used to remove nearly all of the Channel 4 signal from Channel 2, giving a well characterized measurement of  $\mathbf{T}_{850-300}$  (Fu et al., 2004). See the Appendix for more details on this method.

Figure 39 shows uncorrected MSU Channel 2 troposphere temperature trends from UAH Version 5.0 (Christy et al., 2003) and RSS Version 1.0 (Mears et al., 2003), and the Fu team's corrected troposphere temperature trends (Fu et al., 2004). Fu and his colleagues found that the stratospheric contribution to Channel 2 during the MSU service record was about -0.08 deg. K/decade globally. Correcting the RSS Channel 2 record yields a global trend is 0.18 deg. K/decade. This represents a tropospheric warming of roughly 1.1 times the surface record for the same period. The stratospheric contribution in the tropics was somewhat less, -0.05 deg. K/decade, due to a lower tropopause and smaller stratospheric contributions there, consistent with what is expected for the moist adiabatic lapse rates which characterize these regions. These trends also bring the relative tropospheric and surface trends into the range predicted by the most up-to-date general circulation models. Santer et al. (2003) reported that their control run of the ECHAM4/OPYC replicated a 20 year trend differential of 0.08 deg. K/decade, in agreement with the observed stratospheric correction found by Fu's team. Other AOGCM's forced with the appropriate balance of natural and anthropogenic inputs yield similar results. AOGCM's still differ widely in their quality and their ability to reproduce observed results, and multiple runs of the same models can produce varying results also when forced with slightly different, but equally possible

scenarios. But it now seems clear that the best of these models can comfortably reproduce not only the corrected and uncorrected RSS values, but also Fu's corrected UAH trends. Applying these corrections to UAH's Version 5.0 TMT and TLT records yields a global tropospheric trend of 0.09 deg. K/decade, comparable to the uncorrected RSS Channel 2 trend, and a tropical trend of 0.08 deg. K/decade. While these trends represent better agreement with observed surface trends than UAH's uncorrected numbers, they appear to be inconsistent with the UAH TLT trends. These are expected to have minimal influence from the stratosphere and receive 10 to 20 percent of their signal from the surface. But in the tropics for instance, the UAH TLT trend of -0.01 deg. K/decade bears little resemblance to the 850-300 hPa corrected trend of 0.08 deg. K/decade. Furthermore, the UAH TLT trend is cooling with respect to their Channel 2 trend by -0.04 deg. K/decade which is contrary to expectations. It is possible that this might be related to the noise sources discussed previously for TLT retrieval and differences in the UAH merging procedures with respect to those of other teams – in particular, the contributions of Antarctic sea-ice and summer melt pools may contribute to these discrepancies (Swanson, 2003). The low southern latitudes and the tropics are where we expect at least some degree of troposphere decoupling from the surface due to the possible meridional advection of heat (Trenberth and Stepaniak, 2003; 2003b).

The success of the Fu et al. method requires accurate comparisons of high resolution vertical temperature trend profiles that can be compared with MSU sensed layer trends for derivation of the appropriate corrections. For the comparisons to be meaningful, the vertical temperature profiles need to be as independent of the MSU record as possible. The most viable options available for this task are radiosonde analyses. We saw earlier that those problems face a host of coverage and accuracy issues, and these will be a concern for the Fu et al. method. Fortunately, the coverage issue is less problematic because stratospheric temperatures and trends have been less variable during the satellite era than their tropospheric counterparts. But the trend uncertainties are less easily avoided and tend to increase with altitude, making it all the more important to independently check the method across multiple products (NRC, 2000; Seidel et al., 2004). Indeed, this was one of the main concerns voiced by many critics of the Fu et al. method after it was initially published (Spencer, 2004b; Whipple, 2004). Fu realized that these criticisms should be taken seriously and with co-author Celeste Johanson he redid his analysis using other upper-air products. The new analysis was published as a letter in the December 15, 2004 issue of *Journal of Climate* (Fu and Johanson, 2004).

As noted earlier, the original study Fu et al. (2004) based their weighting function derivation on vertical temperature profiles derived from the LKS radiosonde analysis (Lanzante et al., 2003) which was chosen because in addition to being one of the better characterized radiosonde products in terms of its corrections for spurious changes in record and data gathering methods, it is one of the few that is completely independent of the MSU record. But no single radiosonde product today, including this one, is without its issues. Therefore conclusions based on multiple complementary radiosonde products are more robust (Seidel et al., 2004). In their revised analysis Fu and Johanson recognized the need for multiple radiosonde products in their method and they broadened their radiosonde product base. For their new analysis they used a multi-dataset vertical stratospheric trend profile that had been derived as part of the Stratospheric Processes and their Role in Climate initiative's Stratospheric Temperature Trends Assessment program (SPARC-STTA). This profile, which was derived by Ramaswamy et al. (2001), gives 1979-1994 trends above 120 hPa at 45 deg. N. Latitude as derived from a mix of radiosonde, lidar, satellite, and reanalysis products. It is shown in their Figure 30 - reproduced here in Figure 50, and the various analysis products used to derive it are listed in Figure 51. The radiosonde datasets used included the Angell 63 network (Angell, 1988), UK RAOB (Parker and Cox, 1995) and Oort and Liu (1993) products, among others. Like the radiosonde analyses we reviewed earlier, these vary in their degrees of global coverage, number of stations per network, and the number of vertical layers sampled (and the corresponding depth of stratospheric layer covered as well). A few have been updated since this publication, Angell 63 being one (Angell, 2003). Lidar data for 44 deg. N. Latitude and altitudes above 10 hPa (Hauchecorne et al., 1991) are included, although this layer will have little relevance to MSU trends. Satellite records included MSU4 data from UAH Version C (Christy et al., 1995) through 1994, and data from 3 direct and 5 synthetic (derived) channels of the Stratospheric Sounding Unit (SSU) and High-Resolution Infrared Sounder (HIRS) detectors. These devices are passive detectors similar to the MSU's but operating in the infrared with weighting functions that emphasize the middle and upper stratosphere. Like the MSU's, they are carried on the NOAA TIROS-N/ATN series of satellites (see Figure 6). The lowest peak signal (SSU15X) is centered near 40 hPa (Nash and Forrester, 1986; WMO, 1990). The reanalysis piece was provided by the NCEP/NCAR Reanalysis (Kalnay et al., 1996), and

analyzed stratospheric datasets from the NASA Goddard Space Flight Center (GSFC) and the UKMO/SSUANAL (Schubert et al., 1993; Bailey et al., 1993). These products use various combinations of SSU, HIRS-2, and MSU4 data in conjunction with atmospheric GCM modeling to fill in gaps in the observational record. In many ways this suite of datasets is comparable to those we have reviewed so far. The SSU and HIRS records need to be merged in much the same manner as the complete MSU record. They also have their accompanying burden of sampling and data characterization uncertainties as well, and these are not unlike those plaguing the MSU's. The issues with the radiosonde record have already been discussed, as has the fact that these uncertainties increase with altitude. Many, but not all, of these issues have been corrected for prior to their analysis by Ramaswamy's team. The remaining uncertainties were accommodated by weighting the various datasets according to their confidence intervals when contrasting the vertical stratospheric trend profile in Figure 50 (Ramaswamy et al., 2001).

Fu and Johanson rescaled the 1979-1994 SPARC-STTA vertical profile of Ramaswamy's team for 45 deg. N. Latitude to the 1979-2001 global MSU4 trend of UAH Version 5.0 (Christy et al., 2003), making it more representative of global means for the longer period. Profiles were derived using distance and pressure altitudes, and trend profiles for the 200-120 hPa layer were linearly extrapolated from 120 hPa trends after rescaling. The resulting curves are shown in Figure 52 where  $R_H$  gives the distance derived profile,  $R_P$  the pressure derived one, and  $R_{Had}$  the one based on HadRT2.1s. Confidence intervals will be more or less similar to those shown in Figure 50 with a slight increase due to the rescaled UAH Version 5.0 MSU4 confidence interval. It can be seen that the altitude trends derived by distance and pressure are quite similar above 100 hPa, and diverge below it due to differences in the respective 120 hPa trends, with the distance altitude derived trend being in better agreement with the collective radiosonde record than the pressure altitude profile. But all are in reasonably good agreement with 100-300 hPa global mean trends from other radiosonde analyses (Seidel et al., 2004) and may be taken as representative of typical other multi-dataset profiles for these layers. Fu and Johanson used the 2 rescaled SPARC-STTA profiles and the HadRT2.1s profile to derive 3 new sets of synthetic weighting coefficients for the MSU2 and MSU4 covered layers. These were then used to derive 3 new stratospheric "contamination" signals, one for each product. The resulting contaminations in  $\Delta T^{tr}_2$  were,

-0.073 +/- 0.004 deg. K/decade ( $R_H$  - Distance altitude rescaled SPARC-STTA)

-0.066 +/- 0.004 deg. K/decade ( $R_P$  - Pressure altitude rescaled SPARC-STTA)

-0.083 +/- 0.006 deg. K/decade ( $R_{Had}$  - HadRT2.1s)

(Fu and Johanson, 2004)

These figures are to be compared with the original figure of -0.08 deg. K/decade (Fu et al., 2004). It is evident that the agreement between the various products is quite good, as are the confidence intervals on associated with the derivations. The small confidence intervals are largely attributable to the fact that the vertical portion of the weighting function above 200 hPa (equation 1) integrates to zero, removing most of the noise associated with these layers. We should also note that these corrections are *global* rather than regional. There will likely be regional differences, particularly in the tropics where highly variable lapse rates have been observed and various physical processes, such as the Hadley cell driven meridional advection of latent heat, might be at least partially decoupling troposphere and surface trends (Trenberth and Stepaniak, 2003). In fact, in the tropics the Fu et al. synthetic brightness temperature trends and weighting function (equations 1 and 2) are likely to be more representative of the entire troposphere and tropopause below 100 hPa, rather than the 850-300 hPa layer (Fu et al., 2004b). Thus, the Fu et al. method is fairly insensitive to multiple dataset differences and proves to be a robust method for removing the stratospheric contamination in MSU2.

Based on these values for  $\Delta T^{tr}_2$ , the corrected UAH Version 5.0 TMT trends range from 0.076 to 0.093 deg. K/decade. The corresponding TLT trend for the same period is 0.045 deg. K/decade after removal of an estimated 0.01 deg. K/decade stratospheric interference on that product (Fu and Johanson, 2004). UAH reported TLT trends of 0.061 deg. K/decade for the period 1979-2002 (Christy et al., 2003) and 0.06 deg. K/decade for 1979-1998 (Christy et al., 2000). Though they are not precisely the same, the

Fu derived  $T_{FT}^{tr}$  trends should be similar to TLT trends for the same period. Furthermore, we saw earlier that the TLT channel receives a significant amount of noise from surface emissions (NRC, 2000). So given the known surface trend, we might expect TLT to show spurious warming rather than cooling. In particular, we might expect warming of the TLT record with respect to the corrected TMT or  $\Delta T_2^{tr}$  records. TLT vs.  $\Delta T_2^{tr}$  trend discrepancies will need further investigation. But apart from uncertainties in the Fu et al. method, there are 2 most likely candidates for spurious cooling in TLT trends are,

- The UAH “backbone” based merge method, which along with the stated goal of minimizing intersatellite trend differences, may be minimizing *trends* as well and not reducing RMS scatter in the residuals, which would likely provide a better overall trend estimate (Mears et al., 2003; Christy and Norris, 2004).
- High latitude contamination for sea-ice and summer melt pool emissions to which the TLT record will be vulnerable, particularly in the high southern latitudes, but the MSU2 and MSU4 records will not (Swanson, 2003).

Whatever issues remain with the UAH TLT global trends, the corrected UAH MSU2 trends, regional and global, are still within the range of what can be reproduced by GCM’s supplemented by considerations of tropospheric decoupling mechanisms.

## Summary

In light of the previous discussions, we can summarize the current state of knowledge regarding surface and tropospheric temperature trends and their relationship to anthropogenic greenhouse gases as follows;

- The global surface temperature trend is well characterized by a wide range of in situ and proxy data apart from the tropospheric record (NRC, 2000; IPCC, 2001). State-of-the-art AOGCM’s can comfortably reproduce this trend, but also predict a similar long-term trend for the upper atmosphere. The apparent disagreement between the surface and upper air records hinges on these AOGCM predictions, and the belief that the two should be well coupled (NRC, 2000; IPCC, 2001). Over the longer period for which upper air data is available (1958 to the present), the two are in excellent agreement as expected. But short-term trends have shown much variation indicating that the troposphere and surface interact in more complex ways than previously thought. There is evidence that for certain regions of the globe (particularly the tropics), the surface and troposphere may well be decoupled to some extent so that short-term differences in trend are to be expected (Trenberth and Stepaniak, 2003). Because of this, the last 25 years are highly unlikely to be representative of long-term trends in the surface-troposphere temperature differences (NRC, 2000). Because of this, short-term upper air trends cannot be considered to be an indicator of surface global warming or an anthropogenic global warming fingerprint.
- The disagreement between the tropospheric trends of the RSS and UAH teams is likely related to how each team handled the merging of intersatellite datasets and the derivation of hot target factors – particularly the NOAA-9 target factor and the shorter overlaps (e.g. NOAA-9/NOAA-10). The UAH derived value for the NOAA-9 target factor appears to be an outlier compared not only with the RSS value, but with all other RSS and UAH target factors. This anomalous factor appears to be related to UAH’s choice of a “backbone” method of merging that neglects shorter overlaps. The RSS analysis which uses all overlaps appears to result in a set of target factors and trend residuals for the merged time series, resulting in less trend error and a more consistent set of target factors. The difference between these two methods accounts for at least 65 percent of the difference between the trends of each team (Mears et al., 2003; Christy et al., 2003; Santer et al., 2003).

- Though many uncertainties still remain, the best current estimates of troposphere temperature trends for the 850-300 hPa approach the RSS team value of 0.10 deg. C/decade uncorrected for spurious stratospheric cooling, and corrected trends approach 0.18 deg. K/decade (Mears et al., 2003; Fu et al., 2004).
- UAH MSU2LT and MSUTLT products are more strongly influenced by surface radiation emissions than the MSU Channel 2 products of all teams. As such, they will be much more strongly influenced by annual and inter-annual Arctic and Antarctic sea-ice and melt pool areas, particularly the latter. These impart a distinct cooling trend to high southern latitudes (above 60 deg. S) and as such are likely to contribute to the lower trends observed by these products in the southern oceans.
- Vinnikov and Grody have independently shown tropospheric warming rates that agree with expectations, but have not been corrected for differences between land and ocean diurnal cycles or instrument body effect. If corrected for these effects, their trends appear to be more likely to approach those of the RSS team rather than the UAH team (Vinnikov and Grody, 2004).
- Radiosonde analyses are in reasonable agreement with both RSS and UAH trends for most regions where the three overlap. But the noise inherent in these datasets is large enough that they cannot be reliably used to discriminate between RSS and UAH products. In regions where the RSS and UAH products diverge (e.g. the tropical Pacific and northern Africa) radiosonde coverage is too sparse and poorly characterized to be useful. This situation has not changed appreciably in the 4 years since the National Research Council released their year 2000 report on satellite derived troposphere temperatures and global change.
- Tropospheric trends corrected for a spurious stratospheric cooling signal are within the range of what can be comfortably reproduced by the best extant general circulation models with the appropriate natural and anthropogenic forcings (Fu et al., 2004; 2004b; Fu and Johanson, 2004). These models not only capture the signal of the observed troposphere warming, they also capture natural climate variations related to ENSO, PDO, and volcanic eruptions such as those of Mt. Pinatubo and El Chicon. The remaining discrepancies between models and observation can be explained by various mechanisms of poleward energy transport in the tropics and extra-tropics that decouple the troposphere from the surface (Trenberth and Stepaniak, 2003a,b; Santer et al., 2003).

Many issues still remain regarding troposphere and surface temperature trends and their relationship to anthropogenic greenhouse gas emissions and land use activities. More work needs to be done to improve the data quality, particularly that from radiosondes. The ongoing development of the GUAN and RATPAC radiosonde products are a very positive step in this direction as is the ongoing work of the UAH and RSS teams to better characterize and expand their own datasets. But the greatest mysteries surrounding the apparent disagreement between surface and troposphere temperature trends during the last 25 years have largely been resolved. There is no longer any valid reason to dispute global warming at the earth's surface based on tropospheric temperature trends.

## **Objections**

It is evident from the previous discussions that since the Year 2000 report of the National Research Council on upper-air trends (NRC, 2000) much progress has been made toward addressing the problems discussed in this paper. The last few years have seen a steady erosion of support for a long-term disparity between surface and troposphere temperature trends that is both robust and unexplainable in terms of known mechanisms. Support has also diminished for the belief that such a disparity, even if it does exist, disproves anthropogenic climate change. Yet many questions remain and the issue of troposphere temperature trends from satellite and radiosonde products continues to be controversial. Some have challenged the more recent developments in this field, arguing that the upper-air record still

shows an irreconcilable disparity with the surface that argues against anthropogenic climate change. The large majority of these criticisms can be summarized in one or more of the following claims;

- 1) UAH MSU/AMSU products, which continue to show lower trends than other MSU products, have been independently confirmed by the radiosonde record and are therefore more reliable other satellite products – particularly those of RSS.
- 2) Apart from their agreement with the predictions of state-of-the-art AOGCM's, there is no valid reason to favor RSS MSU/AMSU analysis products over those of UAH.
- 3) State-of-the-art AOGCM's continue to disagree substantially with the best characterizations of the upper-air record and are therefore unable to demonstrate an anthropogenic “fingerprint” on climate change.
- 4) The Fu et al. method cannot account for perceived surface-troposphere disparities because its adjusted MSU2 weighting function overcorrects for the stratospheric and aliases a spurious warming into the free troposphere trend. Critics have argued that this is largely because the method relies on statistical trend evaluation to characterize stratospheric MSU2 contributions in terms of MSU4, and (it is claimed) this cannot be done reliably on a regional or global scale for the entire satellite era.

While these criticisms have received most of their coverage from forums outside of the peer-review process (e.g. think tank and advocacy group publications), they have received noteworthy attention within the scientific community as well. Douglass, Singer, and Michaels (2004; 2004b) argue for the first three points in two papers published by *Geophysical Research Letters* in July of 2004. The last point has been argued mainly in scientific conferences, and now most recently in a paper published by *Nature* in December, 2004 (Tett and Thorne, 2004). As they relate to the strengths and weaknesses of the upper-air record in general, these points have already been addressed. But the specific criticisms that have appeared in recent journal publication and/or conference settings will now be addressed.

### Models and the Troposphere - Santer et al. (2003)

When RSS Version 1.0 was first made public in early 2003 it attracted immediate attention because it was the first new MSU analysis product produced that treated that record in the same level of detail as the pioneering UAH products. Like those products, it addressed all currently known sources of error, improving on the characterization of some of them, and incorporated more recent data than the extant UAH product at that time (Version D – Version 5.0 was published later that year). But unlike UAH products, it predicted satellite era troposphere temperature trends that were noticeably higher, and roughly consistent with those of Prabhakara et al. (2000). RSS published their full analysis product later that same year in *Journal of Climate* (Mears et al., 2003). These results were consistent with the predictions of state-of-the-art AOGCM's. Ever since, there has been lively, and at times heated debate as to whose analysis product is more accurate. Some have even gone so far as to claim that the RSS team had cooked their analysis to justify the surface record and AOGCM predictions. RSS published their full analysis product later that same year in *Journal of Climate* (Mears et al., 2003).

In spring of 2003, a team led by Ben Santer of the Lawrence Livermore Laboratory that included Carl Mears, Frank Wentz, and Matthias Schabel of RSS, published a paper in the journal *Science* that compared results from a state-of-the-art AOGCM with RSS and UAH analysis products to see how well the results of either could be accounted for by the latest model improvements. Santer's team compared four runs of the Dept. of Energy's Parallel Climate Model (PCM) with MSU data from RSS Version 1.0 and UAH Version D (Christy et al., 2000) to see if either MSU product could reproduce an anthropogenic “fingerprint” that was visible in PCM. This model, which is described in Washington et al. (2000) is a coupled land, ocean, atmosphere, and sea-ice model that does not use flux corrections at interfaces. The atmospheric and land components are taken from the NCAR's Version 3 Community Climate Model (CCM3) and Land Surface Model (LSM). CCM3 is the same atmospheric model that RSS used to characterize their diurnal correction. The reliability of CCM3 for diurnal behavior has already been seen

(Figure 9). The ocean and sea-ice components are taken from the Los Alamos National Laboratory Parallel Ocean Program (POP) and a sea-ice model from the Naval Postgraduate School. In PCM, these various components are tied together with a flux coupler that uses interpolations between the component model grids in a manner similar to that used in the NCAR Climate System Model (CSM). Grid resolution varies from ½ deg. at the equator to 2/3 deg. near the North Atlantic. The atmospheric component (CCM3) uses 32 vertical layers from the surface to the top of the atmosphere. In various experiments PCM has been very reliable in reproducing observed global surface temperature behavior (see Figures 40 and 41) and stable, well characterized results for a broad range of forcings, and has done an excellent job of capturing ENSO and volcanic effects as well.

Santer's team ran four realizations of the "ALL" PCM experiment which makes use of well-mixed greenhouse gases (including anthropogenic greenhouse gas emissions), tropospheric and stratospheric ozone, direct scattering and radiative effects of sulfate and volcanic aerosols, and solar forcing (Ammann et al., 2003; Meehl et al., 2003). All used identical forcings but differing start times. Simulated MSU temperatures were derived from global model results by applying MSU Channel 2 and 4 weighting functions to the PCM output across its 32 vertical layers, and these were then compared with UAH and RSS analysis products. The goal was to see if an anthropogenic fingerprint on global tropospheric temperature trends could be detected in either of the two MSU products. First, the model was "fingerprinted" using standard techniques (Hasselmann, 1979; Santer et al., 1995) to see if observational uncertainties had a significant impact on PCM's consistency. Internal climate noise estimates (which are necessary for fingerprint detection experiments) were obtained from PCM and the ECHAM/OPYC model of the Max-Planck Institute for Meteorology. The anthropogenic fingerprint on climate change was taken to be the first Empirical Orthogonal Function (EOF),  $\Phi$ , of the mean of the four ALL runs of PCM. Then, increasing expressions of  $\Phi$  were sought in UAH and RSS analyses in an attempt to determine the length of time necessary for it to be detected at a 5 percent statistical significance level in both observational records (Santer et al., 2003).

They found that a clear MSU Channel 2 anthropogenic fingerprint was consistently found only in the RSS dataset. This is not surprising as the RSS team found consistently warmer Channel 2 trends than UAH. What is more noteworthy, is that this was true only for the mean-included comparisons. When the means are removed from both datasets, the fingerprint was clearly visible at the 5 percent level in 6 out of 8 cases for the RSS and UAH analyses – a consequence of the fact that PCM captures the observed equator to pole temperature and trend gradients quite well, and these are in turn manifested in  $\Phi$ . The team concluded that the main differences in the ability of the RSS and UAH products was due to the large global mean and trend differences between the two, and these were in turn likely to be due to uncertainties in how each was analyzed. Santer's team correctly concluded that,

"Our findings show that claimed inconsistencies between model predictions and satellite tropospheric temperature data (and between the latter and surface data) may be an artifact of data uncertainties."

(Santer et al., 2003)

This is exactly what we would expect. We saw earlier that nearly two thirds of the trend discrepancy between the UAH and RSS analyses is related to the differing methods each team used to characterize IBE, do their merge calculations, and to a lesser extent, their differing methods of smoothing and diurnal drift correction. Since detection of the anthropogenic fingerprint in PCM as characterized by  $\Phi$ , depends on this difference, it would not be surprising if the difference between detection and non-detection is the result of data and/or data processing uncertainties. The fact that the mean-removed analyses of both teams do capture the fingerprint demonstrates the ability of PCM and its component models to capture real tropospheric and surface effects.

Even so, some have claimed that these results are a self fulfilling prophecy. PCM and RSS Version 1.0 it is argued, were used to justify each other and the RSS product has been preferred because of the agreement rather than its own merits as an observational analysis - Point 2 above (Christy, 2003). But the criticism does not bear scrutiny. In fact, Santer's team did not use PCM to determine the accuracy of UAH or RSS products. They compared simulated MSU Channel 2 observations from PCM with the corresponding records from UAH Version D and RSS Version 1.0 to see if an anthropogenic fingerprint

on global warming could be detected in either. They found that an anthropogenic fingerprint, as characterized by the first empirical orthogonal function in the PCM runs Santer et al. used, *can* be detected in both products, but that it is observable in UAH Version D only after removing global mean values from the dataset. From this observation they concluded that both products likely capture an anthropogenic fingerprint, and the difference between the two products is largely a matter of how each team handled data uncertainties. Examinations of the merge methodologies of each team, their smoothing methods, and their characterizations of IBE and diurnal correction have already verified this independent PCM. Furthermore, the fact that an anthropogenic fingerprint can be found in each product by the Santer team methodology demonstrates the ability of PCM to reproduce many of the temporal and geographical patterns inherent in real temperature trends, demonstrating that one MSU product is not likely to be more physically consistent with PCM than the other.

### Douglass, Singer & Michaels (2004)

In July of 2004, David Douglass<sup>5</sup>, S. Fred Singer<sup>6</sup>, and Patrick Michaels<sup>7</sup> lead teams that published two papers in Geophysical Research Letter in which they claim to have demonstrated that there is a clear disparity between surface and lower troposphere temperature trends (Douglass et al., 2004), and that current state-of-the-art AOGCM's cannot accommodate it (Douglass et al., 2004b). In the first of these papers Douglas et al. (hereafter, DEA) use MSU data, radiosonde data, and a reanalysis product applied to the period of 1979 to the present to argue that the disparity exists and that it cannot be accounted for by any known tropospheric dynamics. To do this, they start with global surface temperature data from Jones et al. (2001). These are monthly anomalies with respect to the 1961-1990 average of global surface air temperatures over land, and below-surface water temperatures for oceanic regions, as represented within a 5 deg. by 5 deg. grid cells. This record is then compared with lower troposphere trends taken from UAH Version D MSU2LT Data (Christy et al., 2000) and data from a new "2-meter" temperature product (R2-2m) derived from an updated version of the National Centers for Environmental Prediction - National Center for Atmospheric Research (NCEP/NCAR) Reanalysis<sup>8</sup> (Kanamitsu et al., 2002; Kalnay et al., 1996). The latter is selected for its consistency and completeness between the surface and 850 hPa layers, and because it is (they argue) a dataset that is independent of both the MSU record and the radiosonde products that have been used to date for tropospheric intercomparison studies (Christy et al., 2000; 2003; 2004; Seidel et al., 2003, 2004; Angell, 2003). In the second (2004b), they compare results from 3 AOGCM's with surface temperature trends similar to those used in the first paper (but taken from Jones et al., 1999 rather than 2001), MSU2LT data from UAH Version D (Christy et al., 2000), radiosonde data from HadRT2.0 (Parker et al., 1997), and 50 year results from the NCEP/NCAR Reanalysis (Kisteler et al., 2001). From these datasets they argue that the models, which represent the current state of the art in AOGCM's, cannot account for the observed troposphere and surface temperature trends.

DEA based their conclusions on several claims. But upon closer inspection, each one is supported with problematic treatments of the data they cite. First, they argue that the trend from satellite and radiosonde products is significantly less than that of the surface, with exact values depending on both the choice of dataset and analysis methodology (Douglass et al., 2004). This is only true of UAH products. Trends from RSS, Prabhakara, and Vinnikov and Grody differ considerably spanning the range from full agreement to disagreement. They do mention Vinnikov and Grody (2003) and Fu et al. (2004) in reference to the MSU record (RSS Version 1.0 was not addressed) but base their MSU trends only on UAH Version D, which they claim is the only extant MSU product that is validated by the radiosonde record (Douglass et al., 2004).

Closer inspection reveals that the support for this claim is driven entirely by the datasets and time frames they have chosen for their comparison. The MSU and radiosonde records chosen just happen to be the ones that are closet in agreement for the period 1979 to 1996 *and* low in trend – UAH Version 5.0 truncated to 19967 (Christy et al., 2003) and the LKS radiosonde product (Lanzante et al., 2003). Figure 33 shows tropospheric temperature trends for UAH, RSS, LKS and HadRT2.1 for 3 layers and by global region (Seidel et al., 2004). It can be seen that there is very good agreement between LKS and both UAH products for MSU2, though regionally the confidence intervals are large enough to accommodate RSS outside of the southern hemisphere for which the UAH-RSS discrepancy is largest. So not

surprisingly, the southern hemisphere contributes most to the discrepancy. HadRT2.1 shows significant disagreement with both. For MSU2LT however, the LKS dataset shows noticeable discrepancies with UAH products, but agreement with HadRT2.1 is improved. In this case the largest regional discrepancy is again with the southern hemisphere, where now LKS shows more warming. This is particularly significant, as it is in this region that we expect the 2LT product to be most impacted by Antarctic sea-ice and summer melt pools (Swanson, 2003). Thus, even though UAH median trend estimates tend to be closer to comparably adjusted radiosonde products, agreement varies significantly by layer and region, and confidence intervals tend to be large.

When we extend the record another 4 years the picture changes yet again. Figure 34 shows the same troposphere temperature trends by layer and region as Figure 33, but for 1979-2001. In this case, UAH and RSS products are compared with HadRT2.1 (the LKS record ends in 1997). Now we see that both UAH and RSS products are in relatively good agreement with each other, and both disagree with HadRT2.1 globally and in all regions except the southern hemisphere, where UAH products are closer the HadRT2.1 than RSS. For the MSU2LT layer, both UAH products and HadRT2.1 agree well, but the confidence intervals for each are as large as the trends being measured (Seidel et al., 2004). It is worth noting that until 1997, LKS trends in all regions and globally were consistently warmer than their HadRT2.1 counterparts. Given the 1997-98 El Nino and its impact on all trends, it would have been surprising if this had not continued if LKS had been extended to 2001.

Once again, we see that agreement depends on layer and region, and confidence intervals tend to be large in comparison to the trends being measured. This is particularly true of the 2LT layer that is of most interest to DEA. Furthermore, which layer is in agreement and to what degree appears to be strongly driven by the length of record being examined. DEA did not address the issue of limited radiosonde coverage, particularly in regions such as the southern oceans that have the most impact on differences between UAH and RSS trends. Nor did they address the issue of Antarctic sea-ice and melt pool impacts which will be of particular importance for the lower troposphere 2LT trends that they are most concerned with. These factors are significant and they cannot be ignored in MSU/radiosonde comparisons.

The chosen time frame for their study (1979-1996) raises questions as well. DEA state that,

“Since we wish to examine the disparity in the temperature trends among these three datasets, we limit our analysis to a common observational time series. The starting point in our analysis will be 1979, which is the beginning year in both the R2-2m and MSU data. We truncate the analysis at December 1996 which avoids the snow cover issue in R2-2m. This also avoids the anomalously large 1997 El Nino event in the tropical Pacific which *Douglass and Clader* [2002] showed can severely affect the trend-line. We will show later in this paper that it is likely that our conclusions would change little had we been able to use data though 2003.”

(Douglass et al., 2004)

In other words, even though the extant MSU records from both UAH and RSS extend to the present, DEA considered only the first 17 years, leaving out nearly one third of it. Their stated reasons were to exclude a known issue with snow cover contamination in R2-2m and the ENSO event of 1997, but these arguments are unconvincing. There were at least 4 other ENSO events during the satellite era (1982-83, 1986-87, 1991-92 and 1994-95). The 1982-83 event was one of the largest of the 20<sup>th</sup> century and occurred during the tropospheric/stratospheric impact of the El Chicon eruption (see Figures 20-22). These were not omitted even though the 1982-83 event was almost as large as the 1997 event. Furthermore, there is at least some evidence that a relationship may exist between global warming and ENSO events, particularly their frequency (Meehl and Washington, 1996; Knutson et al., 1997; Timmermann et al., 1999; Collins, 2000). Though the jury is still out on this (Zhang et al., 1997; Knutson et al., 1997; Boer et al., 2000), there is enough evidence of a possible relationship between the two that they cannot be avoided *prima facie* in upper-air climate change studies. Likewise, avoiding the snow cover issue is also unconvincing as the MSU2LT record is impacted by this as well, particularly in those regions where UAH and RSS products differ significantly (Swanson, 2003). Even if neither of these things were an issue, we are still left with an analysis of only 2/3 of the relevant upper-air record being used to evaluate products that cover the entire period.

The truncated time period is also noteworthy in one other respect. DEA specifically compare lower troposphere trends as determined by UAH MSU2LT products with surface and upper-air trends from other records. The online community encyclopedia Wikipedia ([www.wikipedia.com](http://www.wikipedia.com)) discusses the MSU record and presents a table that shows troposphere trends from UAH products vs. record ending year from 1992 through 2003 (Wikipedia, 2004). A check of this table reveals that the year at which DEA's analysis ends, 1996, is the last year for which UAH 2LT products show a negative lower troposphere temperature trend. This is interesting because the claim that the troposphere has cooled during the satellite era has been a popular one in many forums. At face value, DEA's choice of record length supports this claim. But any and all record extensions beyond 1996 yield a warming trend that in the last several years has progressed toward a restoration of long-term agreement with the surface record. Thus, by limiting the period of their analysis DEA has,

- Omitted a full third of the MSU record, including only that portion for which a negative lower troposphere temperature trend can be derived. Longer 2LT records show warming trends that are moving in the direction of restoring long-term agreement with the surface record.
- Allowed themselves to directly compare the UAH MSU2 record with the one record which is truly independent of MSU products and shows the best agreement with their chosen MSU product for that period, LKS (Lanzante et al., 2003). The LKS record does not extend beyond 1997.
- Allowed themselves to directly compare another radiosonde product, HadRT2.0, with the UAH 2LT record over a period where there is very good agreement between the two, yet avoid a longer period over which the agreement is much worse (see their second paper cited here, Douglass et al., 2004b).

DEA argued that using the entire record has little impact on their conclusions. To demonstrate this, they did a repeat of their analyses for the 1979-2002 over ocean regions only (which they say avoids snow cover problems) produces similar trends. But the comparison is not valid. First, land regions contribute significantly to the overall trend and cannot be ignored regardless of oceanic response. DEA's reasoning on this point assumes that snow cover is one of the most dominant features of land based trends, if not *the* most important, which is incorrect. Indeed, it is enlightening to compare this argument with the MSU regional trends shown in Figure 11B. Remember that UAH Version D is cited as their only trusted authority for this record. MSU 2LT trends by global region are shown in the middle map. Note that the large majority of lower trend areas for this period are over the world's oceans. This is not surprising, as we expect oceanic regions to have a mediating effect (we have already seen this at work in the Ocean Only vs. Ocean + Land diurnal cycles discussed earlier). Similar land-ocean trend differences can also be seen in the RSS regional trends (top map), though with higher overall values. Note also that many of the warmer regions occur in tropical or extra-tropical areas like the southeast United States and the Arabian Peninsula. The idea that snow cover could be polluting tropospheric trends over Florida and Saudi Arabia is not compelling. The agreement with their earlier results appears to be a result of their choice of oceanic regions only for comparison, which due to the moderating effects of oceanic climates are expected to produce lower trends.

The regional data and figures DEA use have issues as well. Figure 43 shows their Figure 1 (Douglass et al., 2004) which presents their regional 1979-1996 trends as determined by the surface record (Jones et al., 2001), the UAH Version D MSU record (Christy et al., 2000), and the NCEP/NCAR 2-Meter Reanalysis (Kanamitsu et al., 2002). For the period they analyzed, the surface record contained many gaps, so DEA wisely conducted their study only for areas where there were consistent records for all 3 products. However, in this figure where they report regional trends, they show cells with missing data in the same color (dark blue) as those with the minimum regional cooling rates. Though the caption mentions this in passing, the casual reader is left with an inability to discriminate between regions with observed cooling and those with no data, making the figure misleading. Likewise, Figure 44 shows their Figure 2 which presents their 1979-1996 trends for the Surface Record (Jones et al., 2001), the UAH Version D MSU 2LT Record (Christy et al., 2000), and the NCEP/NCAR 2-Meter Reanalysis (Kanamitsu et al., 2002) plotted by latitude. The first thing to notice is that the plot is not symmetric about the equator. DEA extend their trends northward beyond 60 deg. N. Latitude, stopping just short of the Arctic Circle.

Yet in the Southern Hemisphere they truncate it at about 35 deg. S. Latitude without explanation. A comparison of Figure 44 with Figures 11A and 11B reveals that ending the geographic trend record here avoids the region where UAH and RSS products are most different. The region from 60 deg. S. Latitude to the South Pole is where Antarctic sea-ice and summer melt pools have the most impact on the MSU 2LT and TLT records (Swanson, 2003). These regions also significantly impact the NCEP/NCAR R2-2m record as well. Figure 45 shows zonally averaged oceanic albedo as a function of latitude in both the original NCEP/NCAR Reanalysis (Kalnay et al., 1996) and the R2-2m product used by DEA. Sharp increases beyond 60 deg. latitude at either pole reflect the heavy influence of sea-ice. The austral summer cycling of these albedos can be readily seen in the R2-2m product at higher latitudes than 60 deg. S. Note also that the R2-2m product will not reflect the effect of summer melt pools on this albedo (which will have the effect of lowering it to open ocean values). These high albedos will appear as warming trends to the UAH 2LT record, and their interaction with summer melt pools correlate strongly with lower UAH 2LT trends. The effect is much stronger in the Southern Hemisphere than in the North (Swanson, 2003). By avoiding the polar regions, DEA avoid the impact of these influences on their trends, and they avoid the regions of largest difference between UAH and RSS for MSU Channel 2.

Thus the conclusions of DEA's troposphere disparity paper (Douglass et al., 2004) are highly sensitive to their choice of region, temporal period, and analysis product. Its conclusions do not survive broader comparisons and are thus not robust. Their paper, in which they compare the upper-air record to the predictions of AOGCM's (Douglass et al., 2004b) suffers from similar difficulties. Here, DEA shift their attention from claims of a surface/upper-air discrepancy to an attempt to show that state-of-the-art AOGCM's cannot account for it. They examine results from 3 AOGCM's and compare them to the 1979-1997 surface temperature record as determined by Jones et al. (1999) and resolved to a 5 deg. by 5 deg. (latitude vs. longitude) grid, the MSU 2LT lower troposphere temperature record as determined by UAH Version D (Christy et al., 2000), the same as determined by HadRT2.0 (Parker et al., 1997), and the NCEP/NCAR 2-Meter Reanalysis (Kisteler et al., 2001). The models they choose are Hadley CM3 (Tett et al., 2002), the Goddard Institute for Space Studies GISS SI2000 atmospheric model (Hansen et al., 2002), and the Dept. of Energy Parallel Coupled Model, or PCM (Meehl et al., 2003; 2003b).

Hadley CM3 is run for the period 1985-1995 and forced with greenhouse gas emissions, sulfates, and tropospheric and stratospheric ozone. The 1961-1980 portion of this run was removed. Once again we see a truncated record – this time one that removes certain portions of both the beginning and the end of the MSU record. An examination of the upper-air history during the satellite era reveals that the portion of the record DEA omitted in their Hadley CM3 run contains the El Chicon eruption (1982) and a large El Nino event. Hadley CM3 has the ability to capture both events and in fact, results from runs with solar and volcanic forcing were available to DEA at the time they published (Tett et al., 2002; Braganza et al., 2004). An examination of Figures 20 and 22 reveals that the combined impact of these two events was a boost in tropospheric temperatures below 300 hPa for a year or two followed by a cooling period of comparable length prior to 1985 (when their run began). The impact of including these events might well have boosted the early end of the record in this model and resulted in a lower overall trend for the period they examined, which would likely have improved the agreement between their Hadley CM3 run and the MSU record. Thus, the disparity obtained in this run is unlikely to be robust.

Similar problems limit the usefulness of their GISS SI2000 run. SI2000 is a coupled ocean-atmosphere model with several alternative oceanic components and a 4 deg. x 5 deg. gridded atmospheric portion. The atmospheric portion is an update of the earlier GISS SI95 model where the number of vertical layers has been increased from 9 to 12, and the higher layers have been made higher resolution to allow for more accurate modeling of ozone and stratospheric aerosols from volcanic eruptions. Several other refinements were used to improve the performance of this model. Its higher tropopause level resolution of results in a lower 2 X CO<sub>2</sub> forcing compared to SI95 and its climate sensitivity falls within the range of 3.5-4.1 W/m<sup>2</sup> reported by IPCC WG I (2001). SI95 also contained a programming error that caused it to misrepresents sea-ice and summer melt pool absorptivity, and SI2000 contains an update that corrects for this by fixing the Antarctic and Greenland interiors at an albedo of 0.80 (Hansen et al., 2002).

Regarding DEA's use of SI2000, the most relevant piece is the oceanic component. In SI2000 the atmospheric component model is coupled at a common interface grid to any one of 5 oceanic component models denoted Ocean A through Ocean E. Some of these are more realistic than others and responses from them vary accordingly. Ocean A is based on the Sea Surface Temperatures (SSTs) and sea ice representations from the HadISST1 ocean surface model (Rayner et al., 2003) and is strictly an ocean surface and surface heat flux model with no representation of deep ocean dynamics. Ocean D is a deep ocean model based on the Geophysical Fluid Dynamics Laboratory (GFDL) Modular Ocean Model (MOM), and Ocean E is taken from the isopycnic coordinate based Hybrid Coordinate Ocean Model (HYCOM) as described in Bleck (1998). The most commonly used oceanic component in SI2000 runs is Ocean B (a Q-flux ocean) which is based on a deep ocean with diffusive penetration of heat anomalies and a diffusion coefficient that varies geographically (Hansen et al., 1984). Based on observed rates of ocean mixing of tracers, Ocean B provides a good approximation of oceanic global heat uptake for climate forcing scenarios that do not fundamentally alter the deep ocean circulation (true of most multi-decadal simulations such as those done by DEA), and has proven useful for characterizing the efficacy of each of SI2000's radiative forcings when only limited dynamical interactions are permitted.

The crucial difference to note here is that Ocean's B through E model deep ocean heat uptake and the transport of oceanic heat anomalies and Ocean A does not. The moderating effect produced by the heat absorption of the world's oceans is one of the biggest drivers of the long response time associated with climate change. Yet this moderating effect is not driven by climate change alone. The heat content of the world's ocean shows significant regional and global variability from year to year and on decadal time scales, but this variability is not in turn driven by interactions between SST and the atmosphere alone. As such, it is not possible to provide truly realistic climate modeling using SST and surface heat fluxes alone to represent the world's oceans. Because it is based on SST and surface flux only, Ocean A does not adequately capture these effects and its use in climate change studies is severely limited (Hansen et al., 2002). In fact, for some large scale effect such as the North Atlantic Oscillation, Ocean A can even yield the wrong sign for the resultant heat flux anomalies (Bretherton and Battisti, 2000). It is true that Ocean A includes heat fluxes in conjunction with directly measured SST history, so in theory deep ocean behavior would be embedded in it. In this sense, it can be said to "pass information" about this to whatever modeled atmosphere it was coupled to. SST records are commonly used for certain modeled scenarios with good results, so this approach is not de-facto a bad one. But it effectively treats the world's oceans as having infinite heat capacity, making them more of a *boundary condition* on modeled atmospheric response than part of the response itself. While oceanic heat capacity is significantly larger than that of the atmosphere, it is certainly not infinitely larger, and it is unrealistic to assume that the world's oceans are not displaying a response of their own to the very forcings that models like SI2000 are using. Deep ocean models like Ocean B will capture these effects, but SST/flux models like Ocean A will not.

For their simulation, DEA used the 6 forcing case employed by Hansen et al. (2002) for the period 1979-1998 with Ocean A for their oceanic component. As might be expected, because Ocean A neglects deep ocean moderating effects and heat transports, it yields anomalously high atmospheric temperature trends. In describing the SI2000 results they use, DEA reference Figure 16 from Hansen et al. (2002). That figure is reproduced here for the period 1979-1998 as Figures 46A and 46B and shows the change in annual-mean temperature profile vs. pressure altitude for the period 1979-1998 (assuming linear trends) as determined by SI2000 and as observed by radiosonde and MSU. The left-side plot is for Ocean A and the right-side is for Ocean B. The case used by DEA is the Ocean A 6-forcing run and their observational comparison is with the HadRT2.0 radiosonde product. Figure 46A shows the model run used by DEA (compared with HadRT2.0 as they did). It is clear that not only is this run the least representative of the available options with regard to oceanic thermal behavior, it is also the one that produces the largest discrepancy between model and observation. Both regionally and globally, Ocean B provides a better fit the both the radiosonde and MSU data. Furthermore, the MSU data shown in these figures is taken from UAH Version D (Christy et al., 2000), not RSS Version 1.0, and yet the Ocean B 6-forcing case global response consistently falls within the confidence intervals given for this MSU product, even for the lower altitude MSU 2LT data. Regionally, confidence intervals overlap. For the middle troposphere layer (850-300 hPa) RSS Version 1.0 can be expected to run roughly 0.18 deg. K higher than the MSU trends shown for the same period and would be a better fit still across all regions. It is clear from this data that even though it is not perfect, SI2000 run with Ocean B gives a very good overall representation of regional and global temperature trends for the surfaced and troposphere when forced

by well known effects. Yet DEA make no mention of it and choose instead to present only the Ocean A results that yield the largest surface-troposphere discrepancy. While there are pros as well as cons to this approach, in the very least DEA should have presented the Ocean B results and offered a compelling demonstration as to why we should prefer the Ocean A runs – particularly when their cited source spoke directly to its limitations in modeled scenarios like the one they were using (Hansen et al., 2002).

Of the 3 AOGCM's evaluated, the Dept. of Energy PCM model is the only one DEA consider with a full suite of realistic forcings and oceanic and atmospheric components, in combination with a time period that includes all the most significant ENSO and volcanic episodes for the satellite era. They present results for the "ALL" case which includes greenhouse gases, sulfate aerosols (direct effect only), stratospheric and tropospheric ozone, solar, and volcanic forcings. This is the same run that Santer et al. (2003) considered in their evaluation of the detectability of an anthropogenic fingerprint in a modeled climate. We have already seen that Santer's team did in fact, detect an anthropogenic fingerprint in that model. While PCM does not yield a good overall fit with UAH Versions D and 5.0, it does provide a good fit with RSS Version 1.0, and it has already been shown that the differences are largely a matter of analysis method (Mears et al., 2003; 2003b; Santer et al, 2003).

DEA present results from the PCM "ALL" case and their chosen runs for GISS SI2000 and Hadley CM3 as zonally averaged trends vs. latitude (their Figure 1) and as decadal trends vs. altitude (their Figure 2) compared with observational data (Douglass et al., 2004b). As with their first paper, the disparities they observe are largely dependent on the time frame, region, and datasets chosen. The limitations of the radiosonde datasets and the Reanalysis product have already been discussed. But it is noteworthy that for their radiosonde comparison they choose HadRT2.0 when HadRT2.1 was available. We saw earlier that the latter had improved considerably on the former with updated corrections for anomalous data and discontinuous records (Free et al., 2002; Seidel et al., 2003; 2004). Note also that with the exception of the northern hemisphere, their Figure 2 shows all negative trends at 800 hPa for the MSU record (MSU 2LT). Yet their cited source is UAH Version D (Christy et al., 2000) which reports an MSU 2LT trend of 0.06 deg. K/decade for 1979-2001. This is a direct result of the fact that DEA report the value through 1996 omitting the latter third of the extant MSU record. For the NCEP/NCAR Reanalysis they use the original version (R1) in this paper rather than the later version (R2-2m) used in the first. It has already been noted that the updated version of this product corrected many problems present in the first. These included corrections for bogus data in the southern hemisphere, snow and ice cover problems for the 1974-1994 period, and snowmelt pool and oceanic albedo problems for the entire record (Kanamitsu et al., 2002) – all problems that will be of importance to MSU and model comparisons. DEA do not state why the updated Reanalysis product was not used for this comparison study, though the problems with the earlier product were known to them.

In summary, DEA's conclusions are dependent on,

- A neglect of one third of the extant record, including a significant ENSO event of the late 1990's.
- A validation of the shorter record that is heavily dependent on the choice of global region that is most likely to produce minimal trend differences for both periods.
- A neglect of 3 other upper-air MSU products in their study, at least one of which overall is arguably as well characterized and the one they chose, and in a few respects, better.
- A neglect of the most recent, and improved, analyses of the MSU product they did use (other than passing remarks) – most likely because the later products (Christy et al., 2003; 2004) show higher MSU TLT trends than the one they chose (Christy et al., 2000) and that one covers a time frame closer to the truncated period they analyzed.

- A selection of only those AOGCM run periods and parameters that produce large discrepancies between troposphere and surface trends, including a choice of ocean component model for GISS SI2000 that neglects the mediating effects of deep oceanic heat transports and produces tropospheric temperatures that are considerably higher than those of other oceanic components that better represent ocean-atmosphere coupled effects – despite the fact that their cited source discusses at length the limitations of this component model for analyses like theirs.

These are not robust results. It must be concluded that DEA have not demonstrated either a surface-troposphere disparity or an inconsistency between modeled and observed results.

### The Fu et al. Method - A Viable Alternative to TLT?

Because MSU Channel 2 signal receives up to 15 percent of its signal (its raw digital counts) from the lower stratosphere (the 100-50 hPa layer), it very likely underestimates temperature trends in the lower to middle troposphere (the 850-300 hPa layer). Traditionally, this was accounted for by using MSU2 and the TLT as complementary lower troposphere products. We saw earlier that while TLT reduces the stratospheric Channel 2 “footprint”, it pays a price in sampling error and contaminating inputs from other sources such as Antarctic sea-ice and melt pools. Chiang Fu and his co-authors developed their method to avoid these problems. By using direct MSU4 temperature and trend data to correct MSU2 they avoid sampling errors associated with off-nadir MSU views and greatly minimize signal contamination from the surface. Even so, there has been much discussion as to whether their method is more reliable than data from synthetic channels like TLT. After the method was first published (Fu et al., 2004) discussion of its strengths and weaknesses occurred mainly in conference settings and some popular forums. The criticisms fell chiefly into two groups – concerns about the functional form of the corrected weighting function  $\mathbf{W}_{FT}$ , and concerns about the reliability of using statistical methods to derive the  $\mathbf{T}_2$  and  $\mathbf{T}_4$  data used with it.

Roy Spencer of the UAH team expressed concerns about the first point. He argued that because  $\mathbf{W}_{FT}$  goes negative above 100 hPa it will inevitably alias spurious warming into the troposphere trend. Spencer argued that the method might work, but only if trends are constant with altitude from the upper troposphere to the lower stratosphere (roughly 300 -50 hPa) – which they are not (Spencer, 2004). This would be a valid criticism if the method used  $\mathbf{W}_{FT}$  strictly for the derivation of MSU2 brightness temperature with the layers above 100 hPa removed. This is not the case. What Fu and his colleagues actually did can be seen more clearly in Figures 47 to 49. Figure 47 shows  $\mathbf{W}_{FT}$  compared with the weighting functions for MSU2 and MSU4, and Figure 48 shows the same information with MSU2 color banded according to the layers it detects. The region shown in light orange reflects the uncorrected free troposphere contribution to MSU2. The region shown in light blue reflects the tropopause and lower stratosphere, where 300 hPa can be considered the “lowest approach” altitude for the tropopause and 200 hPa a global mean. Figure 13 (right side) shows 1979-2001 upper-air trends as a function of altitude for several radiosonde products and single point trends for UAH Version D (Angell, 2003). Similar data is reproduced in Figures 31 and 33 as broad layer bar graph data for the longer 1958-1997 period using a different set of radiosonde products. It can be seen that the satellite era trends decrease with altitude. Within the uncertainty ranges shown, they go negative above altitudes of roughly 7 to 9 km with the global average being around 8 km (the 300-100 hPa layer). Comparing these trends with Figure 48 reveals that for the satellite era, the light blue layer has an overall negative trend and the orange layer a positive one. Because MSU2 sees the full weighting function of both, it will alias the cooling trends above 300 hPa into the warming trends below. Figure 49 shows Figure 47 shaded to reflect the layer coverage of the Fu et al. weighting function in comparison to its uncorrected MSU2 and MSU4 counterparts. The region shown in dark blue can be expressed in terms of MSU4 and is chosen so that its weighting will integrate to zero with altitude above 300 hPa. Below, the Fu et al. function will have the same weighting that MSU2 would have seen below 300 hPa if the stratosphere were not contributing to its signal (the combined light and dark orange regions). The characterization of this weighting function allows for these two regions to be separately expressed as multiples of  $\mathbf{T}_2$  and  $\mathbf{T}_4$  from which the actual free troposphere brightness temperature trend can be derived. Now it can be

seen that Spencer misunderstood the Fu et al. method.  $W_{FT}$  goes negative above 90-100 hPa because it *must* do so to prevent a stratospheric cooling from being aliased into the free troposphere trend.

Another challenge to the Fu et al. method was published in December of 2004 by the journal *Nature*. Simon Tett and Peter Thorne (hereafter, TT) of the UK Met Office used the Fu et al. method to derive new coefficients and free troposphere trends for the tropics (30 deg. S. to 30 deg. N Latitude) during the period 1978-2002 using the HadRT2.1s radiosonde analysis, the ERA-40 reanalysis (Uppala, 2003), and an ensemble of model runs (Tett and Thorne, 2004). These trends, which they denote as  $T_{fjws}$  in contrast with the  $T^{tr}_{850-300}$  derived by other methods, were then compared to corrected MSU2 trends from UAH Version 5.0 (Christy et al., 2003), RSS Version 1.0, and surface trends. A comparison of their results is given in Figure 53. For non-satellite analyses, surface temperatures were derived from the products indicated. Satellite products were compared to surface trends from the HadCRUT2v dataset. ERA-40 reanalysis based surface trends were derived using zonal averages of 2-meter temperatures over land and SST's over ocean regions. For their model comparisons TT used an ensemble of 6 runs of the atmosphere-only HadAM3 (Pope et al., 2000) and 4 runs of the coupled ocean-atmosphere HadCM3 model (Stott et al., 2000). Their HadAM3 and HadCM3 modeled results were forced with a suite of natural and anthropogenic inputs as described in the cited sources, and were identical with the exception of two corrections in HadAM3 – one for errors in ozone depletion and one for changes in sulfur cycle forcing (Tett and Thorne, 2004). Based on these results they concluded that,

- Fu et al. “trained” and tested their MSU2 and MSU4 coefficients ( $a_2$  and  $a_4$ , respectively) using the same radiosonde dataset (Lanzante et al., 2003), obtaining false agreement and overfitting of the data. Their resulting corrections are overly small and result in overly warm free troposphere trends.
- For the Fu et al. methods to work, stratospheric trends must be relatively stable over the period analyzed, but in fact they are not. In particular, they claim that the lower stratospheric impact of the quasi-biennial oscillation (QBO) will be aliased into Fu et al. derived trends.
- With the exception of HadRT2.1s, free troposphere temperature trends as derived using the Fu et al. method applied to a suite of other upper-air products, show worse agreement with observation and larger confidence intervals than does the UAH Version 5.0 TLT product.
- Agreement between model run derived trends and those based on Fu et al. derived observations show good agreement only between the HadAM3 atmosphere-only run and RSS Version 1.0.

From a review of their methods and results, several comments can be made.

First, it is odd that TT base their comparison study on the tropics only. This is precisely the latitude band for which lapse rates are largest and trends are most variable for the period they studied. Extant radiosonde and reanalysis products are poorly characterized in this region as well. It is not clear why they did not extend their analysis to include a study of global and high latitude trends, and they offered no explanation for this. Such a study would have been particularly useful because the northern latitudes in particular are where their chosen radiosonde and reanalysis products are relatively well characterized and have good coverage. Furthermore, the high southern latitudes are where we expect the biggest differences between UAH and RSS products prior to correction by the Fu et al. method, and where we expect the largest contamination of the TLT record from sea-ice and summer melt pools signals. A test of the Fu et al. method in these regions would have been far more revealing than the region they chose. TT use ERA-40 for their reanalysis product, and this analysis has made great strides over the earlier ERA-15 product in dealing with issues like sea-ice and snow cover, particularly during the satellite era (Bromwich and Fogt, 2004). Comparisons with this product in these regions might have shed light on potential problems with the TLT record, but were not investigated.

Even so, their criticisms of the tropical record are flawed as well. TT rightly point out that the Fu et al. method is most reliable when stratospheric trends are relatively stationary by region and period. But then they point to the Quasi-Biennial Oscillation (QBO) as evidence that they are not and claim that Fu et al. are aliasing QBO trends into their MSU2 correction. Figure 22 shows the stratospheric QBO signal compared to monthly anomaly time series for 6 vertical layers averaged over several upper-air products. Included in this comparison are LKS, HadRT, RIHMI, Angell 63, Angell 54, and UAH Versions D and 5.0. All six time series shown are global, and the QBO signal was determined using 50-hPa altitude zonal wind patterns from radiosonde data at Singapore (Seidel et al., 2004). Because these time series’ draw upon a variety of products including both radiosonde and MSU, they are less subject to the idiosyncracies of any particular dataset, and as they are *global* in nature they present a better comparison to the Fu et al. methods than the tropical data used by TT. Three things are apparent. First, it can be seen that apart from a slight upturn prior to 1981 (at the beginning of the satellite era) and the upward punctuations of the El Chicon and Pinatubo eruptions, the MSU4 and 50-100 hPa time series’ are fairly monotonic and stable for the entire period TT examine, so this requirement is met. It may be argued that the two volcanic events destroy this continuity, but they are also reflected in the tropospheric MSU2 and 850-300 hPa records as proportionally large dips in those records shortly after the stratospheric spikes. Therefore, both layers will reflect this activity in comparable proportions with regard to trend comparisons. Second, a close examination of the stratospheric global MSU4 and 50-100 hPa layer records reveals that at best, the QBO impact on them is barely noticeable. The tropics where TT chose to do their analysis, is the one region where we expect the most significant QBO impact, but this region tells us the *least* about the applicability of the Fu et al. method to the global trends it was used for (Seidel et al., 2004). Finally, the QBO time series is highly periodic, and therefore self-canceling. Even if it did alias a significant signal

into the tropospheric record, that signal would be largely removed by the trending process (Fu et al., 2004b). Furthermore, TT's criticisms assume that the Fu et al. weighting function goes negative above 100 hPa and will therefore alias QBO effects into the free troposphere record that are not there currently. In fact, this is true only of the Fu et al. global weighting function. . The revised Fu and Johanson tropical weighting function does not go negative until around 75 hPa. Figure 54 shows this function compared to its MSU2 counterpart. It is evident that the MSU2 weighting receives more signal from this layer than the Fu et al. weighting, and for the latter the layer above 100 hPa will cancel out while the MSU2 contribution will not (Fu and Johanson, 2004; Fu et al., 2004b). This can even be seen in the global Fu et al. and MSU2 weightings shown in Figure 49.

TT's reported disparities between  $T_{fjws}$  and  $T_{850-300}$  trends in the tropics are also less revealing than they believe. In addition to the large lapse rates and temporal variability characterizing this region, the tropical tropopause often dips as low as 300 hPa. Tropopause trends are poorly characterized across all upper-air products and can significantly affect lower altitude trends if it is not excluded from the sampling (Seidel et al., 2004). For this region,  $T_{fjws}$  is representative of the entire troposphere from the surface up to 100 hPa rather than the 850-300 hPa layer alone. This can be seen clearly in TT's own dataset. Their 1000-100 hPa layer trends agree quite well with their reported  $T_{fjws}$  trends (Fu et al., 2004b). Their ERA-40 derived trend for  $T_{850-300}$  is 0.03 deg. K/decade. The ERA-40 vertical trend profile in this region is revealing. It is positive below 775 hPa, negative between 700 and 400 hPa, and strongly positive between 300 hPa and the tropopause – which itself may occur anywhere between 300 and 100 hPa in this region for the period TT analyze. Therefore, for this region the  $T_{850-300}^{tr}$  trend may be much smaller than its  $T_{fjws}$  counterpart simply because of the vertical variability of this region (Fu et al., 2004b). But the global record will not reflect this.

TT state that modeled results agree with  $T_{fjws}$  trends only for the atmosphere-only runs, but once again there are serious omissions. They use HadAM3 (atmosphere-only) and HadCM3 (coupled ocean-atmosphere) forced with natural and anthropogenic inputs for their model comparisons. But like Douglass et al. (2004b) they did so using model components and regional constraints that are not representative of the method they are testing. TT report that while their atmosphere-only and atmosphere-ocean coupled model runs gave similar results, their couple model runs (HadCM3) yielded a higher range of trends and only were consistent only with the corrected  $T_{fjws}$  trend of RSS Version 1.0. However, TT's HadCM3 coupled model run was not based on a true deep ocean model, but on HadISST which is a simple analysis of observed SST's (Tett, 2004). As such, like the Ocean A component of the GISS SI2000 model, it neglects the moderating effects of deep ocean latent heat advection and will thus overestimate atmospheric trends. In light of this, it is instructive to compare TT's use of HadCM3 with that of Douglass et al. (2004b) and the corresponding runs of Ocean A and Ocean B forced GISS SI2000 (Hansen et al., 2002). DEA obtained SST forced HadCM3 data (Tett et al., 2002) directly from Tett and used it to generate vertical trend profiles for the tropics (30 deg. S to 30 deg. N Latitude). These vertical profiles are directly comparable to the data used by TT, the sole exception being that whereas TT report 1979-2001 trends by layer, DEA truncate their analysis to 1979-1996 so as to create the surface-troposphere trend disparity that their case depends on.

Figure 55 shows DEA's 1979-1996 vertical trend profile for the same tropical region analyzed by TT (Douglass et al., 2004b). Included is a direct comparison of HadCM3 for the period 1975-1995 and GISS SI2000 forced with Ocean A for a similar period (1979-1998). Like TT, DEA use HadCM3 runs that were forced with natural and anthropogenic inputs, and did the same for their GISS SI2000 results. For this region and these periods, it is evident that both SST forced models give strikingly similar results indicating that they are largely comparable for this region and period. Extending the record to 2001 would not be likely to change this result significantly, as both models can be expected to capture the large 1997 ENSO event which dominates this portion of the record. Given the similarities between the two models, the impact of replacing the SST driven Ocean A component of GISS SI2000 that was used by DEA with a true deep ocean component like Ocean B. Figure 46B shows the difference. The left side figure shows the 1979-1998 vertical trend profile for the tropics and extra-tropics (40 deg. S to 40 deg. N Latitude - a region slightly wider than that used by TT) that is obtained using Ocean A SST forcing. The right side shows the comparable trend profile obtained from Ocean B forcing. A clear moderating effect of roughly -0.07 to 0.015 deg. K/decade can be seen in the Ocean B run, demonstrating what we saw earlier when

examining DEA's results – the neglect of deep ocean latent heat advection leads to a model induced overestimation of atmospheric trends.

It is reasonable to expect similar behavior in HadCM3. Had TT used a true deep ocean component in their HadCM3 runs we would expect their modeled tropical trends to be lower by a similar spread. This would have put them in a range where given the uncertainties in forcing and model component responses, they would be adequate representations of either UAH or RSS corrected tropospheric trends. With regard to the uncertainties inherent in an exercise like this one, note also that neither Had CM3 or GISS SI2000 with either forcing scenario reproduces the positive-negative-positive vertical trend variability that is observed in the tropics as we saw earlier (Fu et al., 2004b). Something approaching this behavior is somewhat noticeable in the high southern latitudes (Figure 46B, lower right), but is not captured in the tropics and extra-tropics. This alone should lead us to exercise caution when using model runs for comparisons with observation in localized latitude bands like the tropics that are highly variable and not well characterized in upper-air products. TT raise many important question regarding the use of the Fu et al. method and have shown that care must be used in applying it. But their specific criticisms are at best a poor representation of how the method is used, and thus they do not stand up to scrutiny.

With regard to tests of the Fu et al. method, there is one more that needs to be considered. In the same December 2004 issue of Nature, alongside of TT, Nathan Gillett and Andrew Weaver of the University of Victoria, BC and Ben Santer (hereafter, GWS) published the results of their application of the Fu et al. method to AOGCM derived upper-air temperatures for the period 1958-1997 (Gillett et al., 2004). GWS used global upper-air temperatures from a four-run ensemble of the DOE PCM coupled ocean-atmosphere model forced with natural and anthropogenic inputs (Santer et al., 2003c; Washington et al., 2000) and used the Fu et al. method to derive values for the  $\mathbf{a}_0$ ,  $\mathbf{a}_2$ , and  $\mathbf{a}_4$  coefficients. These were then applied to MSU2 and MSU4 brightness temperature trends that had been obtained by applying the respective weighting functions to PCM temperatures and using least squares methods to obtain the corresponding layer trends. The resulting  $\mathbf{T}_{FT}$  (free troposphere) trends were then compared with the equivalent  $\mathbf{T}_{850-300}$  and TLT trends that had been derived from PCM. Results are shown in Figure 56.

GWS found that the Fu et al. derived  $\mathbf{T}_{FT}$  trends agree with the model “observed”  $\mathbf{T}_{850-300}$  trends to within +/- 0.016 deg. K/decade. Similar agreement was found for the northern and southern hemispheres and the tropics (Gillett et al., 2004). It is interesting to note that GWS's  $\mathbf{T}_{FT}$  trends also agree with their simulated TLT trends for the same period and regions, indicating that the two do reflect similar upper-air layers. The significance of this test as compared to others is that *the PCM modeled climatology is precisely known, and is therefore not subject to the observational uncertainties that plague the existing satellite, radiosonde, and reanalysis records* (e.g. sampling noise, incomplete coverage and temporal record, differences in merge method, etc.). Whether or not it is accurate in its finest details compared to observations is beside the point. PCM does in fact reproduce the large scale behavior of the surface and troposphere and captures most of its more significant features. Therefore, it represents a valid “upper-air” environment against which the Fu et al. method itself can be tested. Because the objective of the GWS study was to test this method rather than reproduce observed climate variables, the robustness of the Fu et al. technique was demonstrated.

Thus, though a number of challenges have been made to the Fu et al. method, none of them withstands scrutiny. Given the relative stability of the stratospheric record and the independent test of the method's robustness using modeled and multi-dataset applications, what criticisms remain regarding the statistical characterization of the method's trend analysis are not likely to stand the test of time. As the quality of radiosonde, rawinsonde, and AMSU products grows, better characterizations of  $\mathbf{W}_{FT}$  and  $\mathbf{T}_{FT}$  will emerge that will allow for more complete investigations of the Fu et al. Weighting function and TLT products. Until then debate regarding the two methods will likely continue, and both will be treated as complementary approaches to stratospheric trend removal.

## **The Road Forward**

Despite the progress that has been made, many unanswered questions regarding upper-air trends remain and much work remains to be done. In their year 2000 report on upper-air datasets and global change the NRC identified four areas where changes needed to be made before significant progress could be made to expand upper-air datasets and reduce the existing uncertainties (NRC, 2000). Specifically, they called for,

- An international program to expand and update existing networks of upper-air temperature monitoring products to improve their quality and consistency. The existing products were not designed for monitoring global change and they lack the accuracy required to properly characterize long-term small trends and/or the needed level of global coverage. Improvements should include new temperature processing algorithms, better archiving of data, improved record keeping at upper-air stations, more consistent data retrieval methodologies, and expanded access to datasets by all research teams.
- A more comprehensive analysis of the uncertainties inherent in upper-air temperature retrievals and the data reduction and analysis methods applied to them. This should cover all the errors described in this paper, and an expanded search for further potential uncertainties. In regard to the satellite data, the uncertainties in the NOAA-09 record and inter-satellite merge methods are of particular concern. Lack of consistency in station records and data retrieval methods and are high priority problems for radiosonde products. All of this will need to be addressed.
- Improvements to the way current generation of state-of-the-art AOGCM's model the effects of combined natural and anthropogenic forcings of the troposphere, and a rerun of all the best upper-air simulations prior to the year 2000. In particular, the vertical structure of the troposphere and stratosphere needs better characterization in these models as does the 3-dimensional structure of greenhouse gas and ozone distribution and its evolution with time.
- The implementation of newer and more sophisticated protocols to insure the quality and consistency of data used in operational numerical weather prediction models. These newer protocols would add considerable reliability improvements to the upper-air datasets used in predictions of global change.

Since then, many important strides have been made toward these goals. AOGCM's have improved considerably since the NRC report. The latest and best ones do far better than their predecessors at reproducing observed 20th century climate changes. Many have done so without resorting to flux corrections. The UAH and RSS teams continue the painstaking work of refining their own MSU/AMSU troposphere and stratosphere products. Slowly, but surely, their analyses are growing in accuracy as more uncertainties are discovered and corrected for and with each passing day their datasets chip away at the short record length that has plagued satellite dataset analyses of global change. Radiosonde datasets are improving also. At the time of this writing, the U.K. Met Office has released Version 2.3/2.3s of their HadRT product and are working on an even newer product designated HadAT. HadAT will be more spacio-temporally consistent than HadRT products and will rely on near-neighbor station comparisons rather than MSU comparisons for the detection of anomalous record changes. As such, it will be independent of the MSU record. The NOAA Air Resources and Geophysical Fluid Dynamics Laboratories continue their work expanding the original LKS radiosonde analysis (Lanzante et al., 2003). The new product, designated the Radiosonde Atmospheric Temperature Products for Assessing Climate (RATPAC) has to date extended the original product to 2003 using a new semi-automated first difference method to append 1997-present radiosonde data to data from the original 81 station network. The method avoids the tedious, labor intensive work of the original and is designed to provide optimal use in estimating long-term small trends, which is exactly what is needed for global change studies.

Proposals have also been made that would allow the use of new data that was not previously suitable for long-term upper-air studies. MSU Channel 1 for instance, views the lower troposphere and could provide much needed data about the surface to 850 hPa layer. To date, it has not been used for these studies due to the high degree of contamination it receives from surface emissivity and the problems encountered in correcting for it. There are also plans to develop improved characterizations of surface emissivity that would make accurate surface corrections possible for this channel and allow it to be used. This would result in valuable ancillary data as well as improved surface corrections for MSU2 and MSU TLT. Likewise the High Resolution Infrared Sounder (HIRS) which operates in the infrared, is also more sensitive to the lower troposphere than MSU2 and would also be usable if accurate surface emissivity corrections were available. This device would also provide high resolution detection of clouds, and thus better characterizations of their impact on incoming and outgoing long-wavelength radiation (Karl et al., 2002). As these proposals are implemented and the length of upper-air datasets grows, so will our understanding of the troposphere and our ability to detect any and all anthropogenic "fingerprints. As these methods are refined and our knowledge of troposphere and stratosphere dynamics grows, the remaining uncertainties should diminish. Given the independent evidence for an anthropogenic fingerprint on global surface temperature trends, we have every reason to believe that the remaining questions about the resulting impact on the troposphere will be resolved as well.

## **Appendix I - The Fu et al. Method**

The Microwave Sounding Unit (MSU) and Advanced Microwave Sounding Unit (AMSU) packages carried by NOAA's Polar Orbiting Environmental Satellites (POES) measure upwelling atmospheric radiation at specific microwave frequencies and use these to derive a bulk brightness temperature for the atmospheric "column" being viewed at any given moment. The total amount of radiation received is determined by surface temperature, atmospheric temperature, and the optical depth of the atmosphere at the frequency being monitored (which is actually quite large at the microwave frequencies used by MSU detectors, and therefore not likely to be a significant factor in the weighting function profile). For each MSU observation, the intensity measured will be given by the sum total of these emissions, the atmospheric contribution increasing with altitude until a peak is reached and falling off to zero from there. Thus, after correction for surface effects, the total intensity measured will correspond to a bulk Brightness Temperature  $T_b$  of this vertical layer with the lion's share of the signal corresponding to the altitude of peak emission. For a perfect blackbody emitting at Frequency  $\nu$  with no dispersion, Brightness Temperature  $T_b$  is defined as,

$$I_\nu = B_\nu(T_b) \quad (A-1)$$

Where  $I_\nu$  is the Specific Intensity of the detected radiation at Frequency  $\nu$ , given in  $J \text{ sec}^{-1} \text{ m}^{-2} \text{ ster}^{-1} \text{ Hz}^{-1}$  and  $B_\nu(T)$  is the corresponding Blackbody Intensity as determined from Planck's Law. For a perfect signal with no dispersion, this is given by,

$$I_\nu = \frac{2h}{c^2} \left[ \frac{\nu^2}{e^{h\nu/kT_b} - 1} \right] \quad (A-2)$$

Where  $h$  is the Planck Constant,  $k$  is the Boltzmann Constant, and  $c$  is the Speed of Light. In reality, MSU and AMSU radiometers detect radiation with nominal beam widths of 7.5 deg (MSU) and 3.3 degrees (AMSU) at full width half maximum power (FWHM). In addition, they also experience input from side lobe emissions, reflectance off of spacecraft surfaces, and occasional "light pollution" from the moon. These are corrected out so that after calibration against the hot target and deep space temperatures,  $I_\nu$  measured as raw "digital counts" will correlate directly to a  $T_b$  value.

Because the measured  $I_\nu$  is the sum total of upwelling radiation emissions from broad layers of the atmosphere,  $T_b$  will be given by the integrated sum of the actual vertical temperature of the atmosphere factored by a weighting function that reflects both the emissive strengths at the frequency being measured and the optical depth of the atmosphere to that frequency. Thus, the observed brightness temperature corresponding to MSU/AMSU observation at nadir is given by,

$$T_b = W_s T_s + \int_{850}^0 W(z) T(z) dz \quad (A-3)$$

where  $W_s$  is the Surface Emissivity factor,  $T_s$  is Surface Temperature,  $W(z)$  is the Weighting Function giving emitted intensity at Frequency  $\nu$  as a function of Altitude  $z$ , and  $T(z)$  is the actual Vertical Temperature Profile of the atmospheric column within the MSU/AMSU beam view. As a weighting, or "emission density" function of altitude,  $W(z)$  must also meet the requirement of,

$$\int_{850}^0 W(z)dz = 1.0 \quad (\text{A-4})$$

The specific atmospheric layer corresponding to  $T_b$  will be dependent on the frequency being detected and its associated weighting function. MSU Channel 2 (MSU2/AMSU5) detects at a frequency of 53.74 GHz and receives most of its input from the lower and middle troposphere. MSU Channel 4 (MSU4/AMSU9) detects at 57.95 GHz and receives mainly from the lower stratosphere. Figure 7 shows the Weighting Functions for MSU2 and MSU4, respectively designated as  $W_2(z)$  and  $W_4(z)$ . It can be seen that whereas MSU2 receives most of its signal from the lower and middle troposphere, it still receives significant input from above 300-100 hPa, which corresponds globally to the tropopause. During the satellite era stratospheric trends have been quite different than those of the troposphere. Because MSU2 sees these differences above 300-100 hPa, it will alias those trends into the troposphere signal and will therefore not accurately reflect tropospheric trends.

With regard to comparisons of surface and troposphere temperature trends, the layer that is of most interest is the 850-300 hPa layer, often referred to as the "Free Troposphere". To effectively measure trends for this layer using MSU2 we need to remove the signal contributions it receives from the surface and the atmosphere above 300 hPa. Because MSU4 receives nearly all of its signal from above 200 hPa and has a temperature trend history that is relatively stable compared to that of the troposphere, it can be used to do this. The Brightness temperature  $T_2$  observed by MSU2 is derived from equation A-3 as,

$$T_2 = W_s T_s + \int_{850}^0 W_2(z)T(z)dz \quad (\text{A-5})$$

Above the surface affected layer, this can be separated into vertical layer components above and below the tropopause giving,

$$T_2 = W_s T_s + \int_{850}^{300} W_2(z)T(z)dz + \int_{300}^0 W_2(z)T(z)dz \quad (\text{A-6})$$

where 200 hPa is taken to be the global average tropopause altitude. It is evident from Figures 7 and 52 that both MSU2 and MSU4 receive significant portions of their signal from the 300-100 hPa layer, and MSU2 is still receiving non-negligible contributions from above it. MSU4 however, receives virtually all of its signal from above 300 hPa and in fact peaks near 100 hPa where stratospheric brightness temperatures and trends will be clearest. Because of this, if the MSU2 signal above 300 hPa is expressed as a function of MSU4, it will be possible to account explicitly for both the temperature and trend impacts on MSU2 and explicitly separate them from the desired free troposphere trend. We now define the Free Troposphere Brightness Temperature  $T_{FT}$  for the 850-300 hPa layer as,

$$T_{FT} = \int_{850}^{300} W_2(z)T(z)dz = T_2 - W_s T_s - \int_{300}^0 W_2(z)T(z)dz \quad (\text{A-7})$$

This can in turn be expressed in terms of a free troposphere Weighting Function  $\mathbf{W}_{FT}$  in the usual manner as,

$$\mathbf{T}_{FT} = \mathbf{W}_s \mathbf{T}_s + \int_{850}^0 \mathbf{W}_{FT}(z) T(z) dz \quad (\text{A-8})$$

Taking the 200 hPa level as the global average tropopause height, we may express  $\mathbf{W}_{FT}$  in terms of the MSU2 and MSU4 signals by respectively defining the Channel Correction Functions  $\alpha_2(z)$  and  $\alpha_4(z)$  with the following conditions,

$$\mathbf{T}_{FT} = \mathbf{W}_s \mathbf{T}_s + \int_{850}^0 \alpha_2(z) \mathbf{W}_2(z) T(z) dz + \int_{850}^0 \alpha_4(z) \mathbf{W}_4(z) T(z) dz \quad (\text{A-9.1})$$

$$\int_{200}^0 \alpha_2(z) \mathbf{W}_2(z) T(z) dz + \int_{200}^0 \alpha_4(z) \mathbf{W}_4(z) T(z) dz = \mathbf{0} \quad (\text{A-9.2})$$

$$\int_{850}^0 \mathbf{W}_{FT}(z) dz = \int_{850}^0 \alpha_2(z) \mathbf{W}_2(z) dz + \int_{850}^0 \alpha_4(z) \mathbf{W}_4(z) dz = \mathbf{1.0} \quad (\text{A-9.3})$$

The resulting form of  $\mathbf{W}_{FT}$  is shown in Figure 48 alongside of  $\mathbf{W}_2$  and  $\mathbf{W}_4$ . The function peaks at the same location as  $\mathbf{W}_2$ , but roughly 15 percent larger, and goes negative above 100 hPa. Note that it *must* do this to remove the stratospheric component in the MSU2 signal (equation A-9.2). When Fu's team originally published (Fu et al., 2004) this was a widely voiced criticism. Because a real weighting function must be everywhere positive (i.e. - a real microwave emitter is by definition *not* an absorber), some argued that the negative portion of this weighting function proved that Fu's team had overcorrected for the stratosphere. In fact, concern about this was voiced at Nature during the review process prior to publication until the methodology was clarified (Seidel, 2004). Yet because  $\mathbf{W}_{FT}$  is a *modified* weighting function with "real" and "virtual" components, negative values are allowed.

MSU Weighting Functions are non-dimensional and approximately independent of time. Therefore, they may be decoupled from  $\mathbf{T}(z)$  allowing us to express equation A-9.1 in the form,

$$\mathbf{T}_{FT} = \mathbf{a}_0 + \mathbf{a}_2 \mathbf{T}_2 + \mathbf{a}_4 \mathbf{T}_4 \quad (\text{A-10})$$

where,

$$\mathbf{a}_0 = \mathbf{W}_s \mathbf{T}_s \quad (\text{A-10.1})$$

$$\mathbf{a}_2 = \int_{850}^{200} \alpha_2(z) \mathbf{W}_2(z) dz + \int_{200}^0 \mathbf{W}_2(z) dz \quad (\text{A-10.2})$$

$$\mathbf{a}_4 = \int_{200}^0 \alpha_4(z) W_4(z) dz \quad (\text{A-10.3})$$

and  $\mathbf{T}_2$  and  $\mathbf{T}_4$  are the observed MSU2 and MSU4 Brightness Temperatures as taken directly from MSU digital counts. Free Troposphere trends can now be derived from equation A-10 as,

$$\frac{\partial}{\partial t} \mathbf{T}_{\text{FT}} = \mathbf{a}_0 + \mathbf{a}_2 \frac{\partial}{\partial t} \mathbf{T}_2 + \mathbf{a}_4 \frac{\partial}{\partial t} \mathbf{T}_4 \quad (\text{A-11})$$

Fu's team used global monthly average temperature anomalies from the surface to the 10 hPa layer derived from the LKS radiosonde dataset (Lanzante et al., 2003) applied to equations A-9 to derive  $\mathbf{a}_0$ ,  $\mathbf{a}_2$ , and  $\mathbf{a}_4$ . Then, regional and global average trends for  $\mathbf{T}_2$  and  $\mathbf{T}_4$  were derived using least squares methods applied to the MSU products of UAH (Christy et al., 2003), RSS (Mears et al., 2003), and Vinnikov and Grody (2004). From these, equation A-11 yields the desired  $\mathbf{T}_{\text{FT}}$  trend. The later work of Fu and Johanson (2004) used the same method to derive  $\mathbf{a}_0$ ,  $\mathbf{a}_2$ , and  $\mathbf{a}_4$  from the HadRT2.1 radiosonde dataset (Parker et al., 1997) and the SPARC-STTA stratospheric temperature profile dataset of Ramaswamy et al. (2001) as an independent check of the original method. The key to this method's success is the relative regional stability of form and trend in  $\mathbf{T}_4$  during the satellite era. This allows for the meaningful use of least squares methods applied to  $\mathbf{T}_4$  as a means of using regional and global stratospheric trends to correct for that layer's impact on  $\mathbf{T}_2$ . Some have criticized the Fu et al. method on the ground that this cannot meaningfully be done. This would be true if stratospheric trends were not monotonic, and regionally and temporally consistent with respect to  $\mathbf{T}_2$  trends during the satellite era. But to at least first order, they are. The specific criticisms made of the method are dealt with in the main body of this paper.

## Footnotes

- 1) To a lesser extent, rocketsonde and lidar studies have also been done, but these have not proven practical for regional or global monitoring.
- 2) Ambient pressure generally drops as altitude increases and can therefore be used as a measure of altitude. The International Standard Atmosphere relates altitude with a global average pressure at a fixed temperature. Local atmospheric pressures often vary from that predicted for a given altitude by the International Standard Atmosphere. Since most atmospheric phenomena are sensitive to pressure rather than actual altitude, it is common in climate change literature to speak of altitude in pressures, generally millibars (mb) or hectopascals (hPa) rather than actual distances.
- 3) In climate change parlance, the term *Zonal* refers to the East to West direction, along lines of latitude. The term *Meridional*, or poleward, refers to North to South, along lines of longitude.
- 4) *Lapse rate* is the rate of change of temperature with altitude, expressed in degrees per unit of distance. It is customary to discriminate between moist and dry lapse rates, as the moisture content of the atmosphere is important for how lapse rates affect weather phenomena.
- 5) Dept. of Physics, University of Rochester, NY.
- 6) University of Virginia, retired. Science and Environmental Policy Project (SEPP).
- 7) Environmental Sciences Dept., University of Virginia.
- 8) The NCEP/NCAR Reanalysis is a composite upper-air analysis product containing several meteorological parameters combined in a global spatial grid of 2.5° x 2.5° (latitude x longitude) resolution from the surface up to the 10 hPa level. It uses data from land and ship based measurements of temperature, wind and humidity, weather forecasts, MSU satellite data, and rawinsonde data (that is, data from radiosondes that have been tracked by radar or radio-theodolite to obtain wind speed and direction). These data sources are tied together by an AGCM (Atmospheric General Circulation Model – no ocean coupling) run in a “frozen” state to evaluate upper-air temperature, pressure, wind, and humidity from 1948 to the present. The MSU data are used to provide weekly raw “soundings” for the Reanalysis. They are not actual weighted brightness temperature measurements of the sort used in upper-air MSU products like those of UAH and RSS, and are independent of those products. Because it is heavily dependent on model based extrapolations from global rawinsonde data, the NCEP/NCAR Reanalysis cannot be considered as independent of the radiosonde record.

The NCEP/NCAR Reanalysis has proven to be a valuable tool in many upper-air studies because of its reliance on multiple datasets and the stability of the AGCM that ties them together, minimizing the impact of flaws in any one dataset. However, like other upper-air products it too has difficulties that limit its usefulness for studies of the troposphere and lower stratosphere. These include changes in synoptic land station and ship observations records, contamination of some of its data by surface snow and sea-ice albedo, problems accounting for some regional weather patterns such as the annual Indian monsoon season, and all the same limitations of coverage and record continuity that plague the radiosonde record. It is also subject to many of the same issues facing AOGCM's as well, which can be more problematic in that it is being used for a fine detail extrapolation of in situ data, whereas AOGCM's are typically used only for large scale predictions of regional and global upper-air trends. The particular version used by Douglass et al. (2004) for their intercomparison study is a recent update of the Reanalysis that is based on 2-meter resolution vertical layer rawinsonde readings. The original NCEP/NCAR Reanalysis product is best described in Kalnay et al. (1996), and the 2-meter update used by Douglass et al. (2004) is described in Kanamitsu et al. (2002).

## Glossary

<b>AOGCM</b>	-	Oceanic and Atmospheric General Circulation Model.
<b>AMSU</b>	-	Advanced Microwave Sounding Unit.
<b>AVHRR</b>	-	Advanced Very High Resolution Radiometer.
<b>CARDS</b>	-	Comprehensive Aerological Reference Data Set.
<b>CMAP</b>	-	Climate Prediction center Merged Analysis of Precipitation.
<b>CPC</b>	-	Climate Prediction center. NOAA/National Weather Service.
<b>CRU</b>	-	The Climate Research Unit. University of East Anglia, Norwich, U.K.
<b>DEA</b>	-	Douglass et al.
<b>ECMWF</b>	-	European Centre for Medium-Range Weather Forecasts.
<b>ENSO</b>	-	El Nino Southern Oscillation.
<b>ERBE</b>	-	Earth Radiation Budget Experiment.
<b>FWHM</b>	-	Full Width Half Maximum power.
<b>GES</b>	-	Greening Earth Society.
<b>GFDL</b>	-	Geophysical Fluid Dynamics Laboratory. Princeton, NJ.
<b>GISS</b>	-	Goddard Institute for Space Studies. NASA. New York, NY.
<b>GSFC</b>	-	Goddard Space Flight Center. NASA. Greenbelt, MD.
<b>GUAN</b>	-	Global Climate Observing System Upper Air Network.
<b>GWS</b>	-	Gillette, Weaver, and Santer.
<b>HIRS</b>	-	High-Resolution Infrared Sounder.
<b>hPa</b>	-	Hectopascal. SI unit of pressure equivalent to one millibar.
<b>IBE</b>	-	Instrument Body Effect.
<b>IPCC</b>	-	Intergovernmental Panel on Climate Change.
<b>JPL</b>	-	The Jet Propulsion Laboratory, Pasadena, CA.
<b>LECT</b>	-	Local Equatorial Crossing Time.
<b>LKS</b>	-	Lanzante, Klein, Seidel. (Lanzante et al., 2003).
<b>MM</b>	-	McKittrick and Michaels.
<b>MSU</b>	-	Microwave Sounding Unit.
<b>MSU2LT</b>	-	Microwave Sounding Unit Channel 2 – Lower Troposphere.
<b>MSU2MT</b>	-	Microwave Sounding Unit Channel 2 – Middle Troposphere.
<b>MSU2RT</b>	-	Microwave Sounding Unit Channel 2 – Lower Troposphere (UAH Version B).
<b>MSUTLT</b>	-	MSU Channel 2/AMSU Channel 5 – Lower Troposphere.
<b>MSUTMT</b>	-	MSU Channel 2/AMSU Channel 5 – Middle Troposphere.
<b>MSUTST</b>	-	MSU Channel 4/AMSU Channel 9 – Lower Stratosphere.
<b>NAO</b>	-	North Atlantic Oscillation.
<b>NASA</b>	-	National Aeronautics and Space Administration.
<b>NCAR</b>	-	National Center for Atmospheric Research. Boulder, CO.
<b>NCDC</b>	-	National Climatic Data Center. NOAA.

<b>NCEP</b>	- National Centers for Environmental Prediction.
<b>NCPFR</b>	- National Center for Public Policy research.
<b>NOAA</b>	- National Oceanographic and Atmospheric Administration.
<b>NRC</b>	- National Research Council. National Academy of Sciences, Washington D.C.
<b>PDO</b>	- Pacific Decadal Oscillation.
<b>POES</b>	- Polar Orbiting Environmental Satellite.
<b>PRT</b>	- Platinum Resistance Thermocouple.
<b>PR</b>	- Prabhakara et al.
<b>QBI</b>	- Quasi Biennial Oscillation Index.
<b>RMS</b>	- Root Mean Square.
<b>RAOB</b>	- RAwinsonde OBservation.
<b>RATPAC</b>	- Radiosonde Atmospheric Temperature Products for Assessing Climate.
<b>RSS</b>	- Remote Sensing Systems. Santa Rosa, CA.
<b>SEPP</b>	- The Science and Environmental Policy Project.
<b>SOI</b>	- Southern Oscillation Index.
<b>SPARC-STTA</b>	- Stratospheric Processes and their Role in Climate - Stratospheric Temperature Trends Assessment Program.
<b>SST</b>	- Sea Surface Temperature.
<b>SSU</b>	- Stratospheric Sounding Unit.
<b>TOA</b>	- Top of Atmosphere.
<b>TIROS</b>	- Television Infrared Observation Satellite.
<b>TIROS-ATN</b>	- Advanced Television Infrared Observation Satellite.
<b>TOVS</b>	- TIROS Operational Vertical Sounder.
<b>TT</b>	- Tett and Thorne.
<b>UAH</b>	- University of Alabama, Huntsville, AB.
<b>UCAR</b>	- University Corporation for Atmospheric Research. Boulder, CO.
<b>UKMO</b>	- The U.K. Met. Office. Exeter and London, U.K.
<b>VG</b>	- Vinnikov and Grody.

## References

- Ammann, C.M., G.A. Meehl, W.M. Washington and C.S. Zender. 2003. A monthly and latitudinally varying volcanic forcing dataset in simulations of 20th century climate. *Geophys. Res. Lett.*, **30**, doi:10.1029/2003GL016875RR.
- Angell J.K. 1988. Variations and trends in tropospheric and stratospheric global temperatures, 1958 – 87. *J. Climate*, **1**, 1296-1313.
- Angell J.K. 1999. Comparison of surface and tropospheric trends estimated from a 63-station radiosonde network, 1958 - 1998. *Geophys. Res. Lett.*, **26**, 2761-2764.
- Angell J.K. 2000. Difference in radiosonde temperature trend for the period 1979-1998 of MSU data and the period 1959-1998 twice as long. *Geophys. Res. Lett.*, **27**, 2177-2180.
- Angell J.K. 2003. Effect of exclusion of anomalous tropical stations on temperature trends from a 63-station radiosonde network, and comparison with other analyses. *J. Climate*, **16**, 2288-2295.
- Angell J.K. and J. Korshover. 1983. Global temperature variations in the troposphere and low stratosphere, 1958-1982. *Mon. Wea. Rev.*, **111**, 901-921.
- Bailey, M. J., A. O'Neill, and V. D. Pope. 1993. Stratospheric analyses produced by the United Kingdom Meteorological Office. *J. Appl. Meteorol.*, **32** (9), 1472–1483.
- Baliunas, S.L. 2003. The Air Up There – Is It Hotter? *Tech Central Station Online*. Published on Nov. 21, 2003. Available online at <http://www.techcentralstation.com/112103D.html>. Accessed Nov. 30, 2004.
- Bassist A. N., and M. Chelliah, 1997. Comparison of tropospheric temperatures derived from the NCEP/NCAR reanalysis, NCEP operational analysis, and the Microwave Sounding Unit. *Bull. Amer. Meteor. Soc.*, **78**, 1431–1447.
- Beder, S. 1998. *Global Spin: The Corporate Assault on Environmentalism*. Chelsea Green Publishing Company. 288pp. ISBN 1 870098 67 6.
- Beder, S. 1999. Climatic Confusion and Corporate Collusion: Hijacking the Greenhouse Debate. *The Ecologist*. March/April 1999. 119-122. Available online at <http://www.uow.edu.au/arts/sts/sbeder/ecologist2.html>. Accessed on Jan. 9, 2005.
- Benestad, R.E. (2004). Are temperature trends affected by economic activity? Comment on McKittrick & Michaels. *Climate Research*, **27**, 171-173. Available online at [http://www.realclimate.org/Benestad04\\_CR\\_c027p171.pdf](http://www.realclimate.org/Benestad04_CR_c027p171.pdf). Accessed on Dec. 19, 2004.
- Benestad, R.E. (2004b). Are Temperature Trends affected by Economic Activity? *Realclimate.org*. Available online at <http://www.realclimate.org/index.php?p=41>. Accessed on Dec. 19, 2004.
- Bleck, R., Ocean modeling in isopycnic coordinates, in *Ocean Modeling and Parameterization*, edited by E. P. Chassignet and J. Verron, pp. 423–448, Kluwer Acad., Norwell, Mass., 1998.
- Boer, G.J., G. Flato, and D. Ramsden, 2000b: A transient climate change simulation with greenhouse gas and aerosol forcing: projected climate for the 21st century. *Clim. Dyn.*, **16**, 427-450.
- Bretherton, C. S., and D. S. Battisti. 2000. An interpretation of the results from atmospheric general circulation models forced by the time history of the observed sea surface temperature distribution. *Geophys. Res. Lett.*, **27**, 767–770.
- Bromwich, D.H., and R.L. Fogt. 2004. Strong trends in the skill of the ERA-40 and NCEP/NCAR Reanalyses in the high and middle latitudes of the Southern Hemisphere, 1958-2001. *J. Climate*, **17**, 4603-4619.
- Brown, S.J., D.E., Parker, C.K. Folland, and I. Macadam. 2000: Decadal variability in the lower-tropospheric lapse rate *Geophys. Res. Lett.*, **27**, 997-1000.
- Bengtsson, L., Roeckner, E. and Stendel, M. 1999: Why is the global warming proceeding much slower than expected? *J. Geophys. Res. - Atmospheres*, **104**, D4, 3865-3876.

- Braganza, K., D.J. Karoly, A.C. Hirst, P. Stott, R. Stouffer, and S.F.B. Tett. 2004. Simple indices of global climate variability and change Part II: attribution of climate change during the twentieth century. *Climate Dynamics*, **22**, 823-838.
- Cavalieri, D. J., P. Gloersen, C. L. Parkinson, J. C. Comiso, and H. J. Zwally. 1997. Observed Hemispheric Asymmetry in Global Sea Ice Changes. *Science*, **278**, 1104–1106.
- Christy, J.R. and McNider, N.T. 1994: Satellite greenhouse signal. *Nature*, **367**, 325.
- Christy, J.R., R.W. Spencer, and R.T. McNider. 1995: Reducing noise in the MSU daily lower-tropospheric global temperature dataset. *J. Climate*, **8**, 888-896.
- Christy, J.R., and W.D. Braswell. 1997: How accurate are satellite 'thermometers'? *Nature*, **389**, 342.
- Christy, J.R., R.W. Spencer, and E.S. Lobl. 1998: Analysis of the merging procedure for the MSU daily temperature series. *J. Climate*, **11**, 2016-2041.
- Christy, J.R., R.W. Spencer, and W.D. Braswell. 2000: MSU tropospheric temperatures: Dataset construction and radiosonde comparisons. *J. Atmos. And Oc. Tech.*, **17**, 1153-1170.
- Christy, J.R., R.W. Spencer, W.B. Norris, and W.D. Braswell. 2003: Error estimates of Version 5.0 of MSU-AMSU bulk atmospheric temperatures. *J. Atmos. And Oc. Tech.*, **20**, 613-629.
- Christy, J.R. May 13, 2003. Testimony before the U.S. House of Representatives' Committee on Resources. Available online from CO2 Science magazine, Volume 6, Number 22: 28 May 2003 at [http://www.co2science.org/edit/v6\\_edit/v6n22edit.htm](http://www.co2science.org/edit/v6_edit/v6n22edit.htm). Accessed Nov. 28, 2004.
- Christy, J.R., R.W. Spencer, and W.D. Braswell. 2004: Update on microwave-based atmospheric temperatures from UAH. *15<sup>th</sup> Symposium on Global Change and Climate Variations*. 84<sup>th</sup> AMS Annual Meeting. Seattle, WA. Jan. 11, 2004. Extended abstract available online at <http://ams.confex.com/ams/pdfpapers/68728.pdf>.
- Christy, J.R. and W.B. Norris. 2004. What may we conclude about global tropospheric temperature trends? *Geophys. Res. Lett.*, **31**, L06211, doi:10.1029/2003GL019361.
- Collins, M., 2000. The El-Niño Southern Oscillation in the second Hadley Centre coupled model and its response to greenhouse warming. *J. Climate*, **13**, 1299-1312.
- Covey, C., A. Abe-Ouchi, G.J. Boer, G.M. Flato, B.A. Boville, G.A. Meehl, U. Cubasch, E. Roeckner, H. Gordon, E. Guilyardi, L. Terray, X. Jiang, R. Miller, G. Russell, T.C. Johns, H. Le Treut, L. Fairhead, G. Madec, A. Noda, S.B. Power, E.K. Schneider, R.J. Stouffer and J.S. von Storch. 2000. The Seasonal Cycle in Coupled Ocean-Atmosphere General Circulation Models. *Clim. Dyn.*, **16**, 775-787.
- Crowley, T.J. and T. Lowery. 2000. How warm was the Medieval warm period? *Ambio*, **29**, 51-54.
- Douglass, D. H., B. D. Pearson, S. F. Singer, P. C. Knappenberger, and P. J. Michaels (2004). Disparity of tropospheric and surface temperature trends: New evidence. *Geophys. Res. Lett.*, **31**,
- Douglass, D. H., B. D. Pearson, and S. F. Singer (2004b). Altitude dependence of atmospheric temperature trends: Climate models versus observation. *Geophys. Res. Lett.*, **31**,
- Douglass, D. H., S. F. Singer, and P. J. Michaels 2004c. Settling Global Warming Science. *Tech Central Station Online*. Published on Aug. 12, 2004. Available online at <http://www.techcentralstation.com/081204D.html>. Accessed Dec. 5, 2004.
- Durre, I., T. C. Peterson, and R. S. Vose. 2002. Evaluation of the effect of the Luers-Eskridge radiation adjustments on radiosonde temperature homogeneity. *J. Climate*, **15**, 1335–1347.
- Eskridge, R.E., O.A. Alduchov, I.V. Chernykh, P. Zhai, A.C. Polansky, and S.R. Doty. 1995. A Comprehensive Aerological Reference Data Set (CARDS): Rough and systematic errors. *Bull. Amer. Meteor. Soc.*, **76**, 1759-1775.
- Ferguson, B. and M. Lewis. 2003. Specific Comments on the Menendez Climate Change Amendment Findings. May 6, 2003. *Competitive Enterprise Institute*. Available online at [http://www.cei.org/dyn/view\\_expert.cfm?expert=224](http://www.cei.org/dyn/view_expert.cfm?expert=224). Accessed Nov. 28, 2004.
- Free M., I. Durre, E. Aguilar, D. Seidel, T.C. Peterson, R.E. Eskridge, J.K. Luers, D. Parker, M. Gordon, J. Lanzante, S. Klein, J. Christy, S. Schroeder, B. Soden, L.M. McMillan, and E. Weatherhead. 2002:

- Creating Climate Reference Datasets. CARDS Workshop on Adjusting Radiosonde Temperature Data for Climate Monitoring. *Bull. Amer. Meteor. Soc.*, June 2002, 891-899.
- Folland, C.K., Rayner, N.A., Brown, S.J., Smith, T.M., Shen, S.S.P., Parker, D.E., Macadam, I., Jones, P.D. and others. 2001. "Global temperature change and its uncertainties since 1861." *Geophys. Res. Lett.*, **28**, 2621-2624.
- Fu, Q., and C.M. Johanson. 2004: Stratospheric Influences on MSU-Derived Tropospheric Temperature Trends. *J. Climate*, To be published Dec. 15, 2004.
- Fu, Q., C.M. Johanson, S.G. Warren, D.J. Seidel. 2004. Contribution of stratospheric cooling to satellite-inferred tropospheric temperature trends. *Nature*, **429**, (6987), 55-58.
- Fu, Q., C.M. Johanson, S.G. Warren, D.J. Seidel. 2004b. Atmospheric science: Stratospheric cooling and the troposphere (reply). *Nature*, **432**, doi:10.1038/nature03210 Brief Communications. Dec. 2, 2004. Abstract available at [http://www.nature.com/cgi-taf/DynaPage.taf?file=/nature/journal/v432/n7017/abs/nature03210\\_fs.html](http://www.nature.com/cgi-taf/DynaPage.taf?file=/nature/journal/v432/n7017/abs/nature03210_fs.html). Accessed on Dec. 27, 2004. Subscription required for access to the full article.
- Gaffen, D.J. 1994. Temporal inhomogeneities in radiosonde temperature records. *J. Geophys. Res.*, **99**, 3667-3676.
- Gaffen, D.J. 1996. A digitized metadata set of global upper-air station histories. *NOAA Tech. Memo ERL ARL 211*, 38 pp.
- Gaffen, D.J., M.A. Sargent, R.E. Habermann, and J.R. Lanzante, 2000a: Sensitivity of tropospheric and stratospheric temperature trends to radiosonde data quality. *J. Climate*, **13**, 1776-1796.
- Gaffen, D.J., Santer, B.D., Boyle, J.S., Christy, J.R., Graham, N.E. and Ross, R.J. 2000b: Multidecadal changes in the vertical temperature structure of the tropical troposphere. *Science*, **287**, 1242-1245.
- Gelbspan, R. 1998. The Heat Is on: The Climate Crisis, the Cover-Up, the Prescription. *Perseus Publishing Company*. 288pp. ISBN 0738200255.
- Gelbspan, R. 2004. Boiling Point: How Politicians, Big Oil and Coal, Journalists and Activists Are Fueling the Climate Crisis--And What We Can Do to Avert Disaster. *Basic Books*. 254pp. ISBN 046502761X.
- Gelman, M. E., A. J. Miller, R. N. Nagatani, and C. S. Long. 1994. Use of UARS data in the NOAA stratospheric monitoring program. *Adv. Space Res.*, **14** (9), 21-31.
- Gillett, N.P., B.D. Santer, and A.J. Weaver. 2004. Atmospheric science: Stratospheric cooling and the troposphere. Arising from: Q. Fu et al., *Nature* 429, 55-58 (2004). *Nature*, **432**, doi:10.1038/nature03209. Brief Communications. Dec. 2, 2004. Abstract available at [http://www.nature.com/cgi-taf/DynaPage.taf?file=/nature/journal/v432/n7017/abs/nature03209\\_fs.html](http://www.nature.com/cgi-taf/DynaPage.taf?file=/nature/journal/v432/n7017/abs/nature03209_fs.html). Accessed on Dec. 27, 2004. Subscription required for access to the full article.
- Greening Earth Society (GES). May 6, 2003. *Virtual Climate Alert*, **4**, 9. Available online at <http://www.co2andclimate.org/Articles/2003/vca9.htm>. Accessed Nov. 21, 2004.
- Hansen, J., A. Lacis, D. Rind, G. Russell, P. Stone, I. Fung, R. Ruedy, and J. Lerner. 1984. Climate sensitivity: Analysis of feedback mechanisms, in *Climate Processes and Climate Sensitivity*. *Geophys. Monogr. Ser.*, **29**, edited by J. E. Hansen and T. Takahashi, pp. 130-163, AGU, Washington, D.C.
- Hansen, J., M. Sato, R. Ruedy, A. Lacis, K. Asamoah, K. Beckford, S. Borenstein, E. Brown, B. Cairns, B. Carlson, B. Curran, S. de Castro, L. Druryan, P. Etwarrow, T. Ferde, M. Fox, D. Gaffen, J. Glascoe, H. Gordon, S. Hollandsworth, X. Jiang, C. Johnson, N. Lawrence, J. Lean, J. Lerner, K. Lo, J. Logan, A. Lueckett, M.P. McCormick, R. McPeters, R.L. Miller, P. Minnis, I. Ramberran, G. Russell, P. Russell, P. Stone, I. Tegen, S. Thomas, L. Thomason, A. Thompson, J. Wilder, R. Willson, and J. Zawodny. 1997. Forcings and chaos in interannual to decadal climate change. *J. Geophys. Res.* **102**, 25679-25720.
- Hansen, J., R. Ruedy, J. Glascoe, and M. Sato. 1999. GISS analysis of surface temperature change. *J. Geophys. Res.* **104**, 30997-31022. Data from this study is available online from the Goddard Institute for Space Studies at <http://www.giss.nasa.gov/data/>. Accessed on Dec. 18, 2004.
- Hansen, J., M. Sato, L. Nazarenko, R. Ruedy, A. Lacis, D. Koch, I. Tegen, T. Hall, D. Shindell, B. Santer, P. Stone, T. Novakov, L. Thomason, R. Wang, Y. Wang, D. Jacob, S. Hollandsworth, L. Bishop, J.

Logan, A. Thompson, R. Stolarski, J. Lean, R. Willson, S. Levitus, J. Antonov, N. Rayner, D. Parker, and J. Christy. 2002. Climate forcings in Goddard Institute for Space Studies SI2000 simulations. *J. Geophys. Res.*, **107** (D17), 10.1029/2001JD001143.

Hasselmann, K. 1979. In *Meteorology of Tropical Oceans*. D.B. Shaw (Ed.) Royal Meteorological Society of London. pp. 251-259.

Hauchecorne, A., M.-L. Chanin, and P. Keckhut. 1991. Climatology and trends of the middle atmospheric temperature (33–87km) as seen by Rayleigh lidar over the south of France. *J. Geophys. Res.*, **96**, 15,297–15,309.

Hegerl, G.C. and J.M. Wallace. 2002: Influence of patterns of climate variability on the difference between satellite and surface temperature trends. *J. Climate*, **15**, 2412-2428.

Intergovernmental Panel on Climate Change (IPCC) – Working Group I, (1996): *Climate Change 1995 – The Science of Climate Change*, Houghton, J.T., F.G. Meira Filho, B.A. Callander, N. Haarris, A. Kattenberg, and K. Maskell eds. 1996. Cambridge University Press.

Intergovernmental Panel on Climate Change (IPCC) – Working Group I, (2001): *Climate change 2001 – The Scientific Basis*, Houghton, J.T., Ding, Y., Griggs, D.J., Noguer, M. van der Linden, P.J., Dai, X., Maskell, K., Johnson, C.C. eds. 2001. Cambridge University Press.

Jones, P.D. 1994: Recent warming in global temperature series. *Geophys. Res. Lett.*, **21**, 12, 1149-1152.

Jones, P.D., K.R. Briffa, T.P. Barnett, and S.F.B. Tett. 1998: High-resolution paleoclimatic records for the last millennium: interpretation, integration, and comparison with General Circulation Model control run temperatures. *The Holocene*. **8**, 455-471.

Jones, P.D., M. New, D.E. Parker, S. Martin, and I.G. Rigor. 1999: Surface air temperature and its changes over the past 150 years. *Rev. Geophys.* **74**, 173-199.

Jones, P.D., T.J. Osborne, K.R. Briffa, C.K. Folland, E.B. Horton, L.V. Alexander, D.E. Parker, and N.A. Raynor. 2001: Adjusting for sampling density in grid box land and ocean surface temperature time series. *J. Geophys Res.* **103**, 3371-3380.

Hack, J.J. Rosinski, J.M., Williamson, D.L., Boville, B.A. and Truesdale, J.E. 1995. Computational Design of the NCAR Community Climate Model. *Parallel Comput.*, **21**, 1545-1569.

Hurrell, J., Hack, J.J. Boville, B.A., Williamson, D., and Kiehl, J.T. 1998. The Dynamical Simulation of the NCAR Community Climate Model Version 3 (CCM3). *J. Climate*, **11**, 1207-1236.

Kalnay, E. et. al. 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteor. Soc.* **77**, 437-471.

Karl, T.R. et. al. 1993: *Bull. Am. Meteorol. Soc.* **37**, 173-200.

Karl T.R., J.R. Christy, B. Santer, F. Wentz, D. Seidel, J. Lanzante, K. Trenberth, D. Easterling, M. Goldberg, J. Bates, C. Mears 2002: Draft White Paper: Understanding recent atmospheric temperature trends and reducing uncertainties. Prepared in support of the Strategic Plan for the U.S. Climate Change Science Program. Available online at <http://www.climate-science.gov/Library/stratplan2003/temptrends-wp-26nov2002.pdf>. Accessed Jan. 9, 2005.

Kiehl J.T., J.J. Hack, G.B. Bonan, B.A. Boville, B.A. Briegleb, D.L. Williamson, and P.J. Rasch. 1996: Description of the NCAR Community Climate Model (CCM3). Boulder: National Center for Atmospheric Research.

Kiehl J.T., J.J. Hack, G.B. Bonan, B.A. Boville, B.A. Briegleb, D.L. Williamson, and P.J. Rasch. 1998: The National Center for Atmospheric Research Community Climate Model: CCM3. *J. Climate*, **11**, 1131-1149.

Kinne, O. 2003. Climate Research: an article unleashed worldwide storms. *Climate Research*, **24**, 197-198.

Kistler, R., E. Kalnay, W. Collins, S. Saha, G. White, J. Woollen, M.I Chelliah, W. Ebisuzaki, M. Kanamitsu, V. Kousky, H. van den Dool, R. Jenne and M. Fiorino. (2001). The NCEP\_NCAR 50-year

- Reanalysis: monthly means. *Bull. Amer. Meteor. Soc.*, **82**, 247-267. Dataset available online at <http://dss.ucar.edu/datasets/ds090.2/data/monthly/TMP.2m/>. Accessed on Dec. 5, 2004.
- Knutson, T.R., S. Manabe and D. Gu, 1997. Simulated ENSO in a global coupled ocean-atmosphere model: multidecadal amplitude modulation and CO<sub>2</sub>-sensitivity. *J. Climate*, **10**, 138-161.
- Koshelkov, Y. P., and G. R. Zakharov. 1998. On temperature trends in the Arctic lower stratosphere. *Meteorol. Gidrol.*, **5**, 45-54.
- Labitzke, K., and H. van Loon. 1994. Trends of temperature and geopotential height between 100 and 10 hPa in the Northern Hemisphere. *J. Meteorol. Soc. Jpn.*, **72**, 643-652.
- Lanzante, J.R. 1996: Resistant, robust, and nonparametric techniques for the analysis of climate data: Theory and examples, including applications to historical radiosonde station data. *Int. J. Climatol.*, **16**, 1197-1226.
- Lanzante, J.R., S.A. Klein, D.J. Seidel. 2003: Temporal homogenization of monthly radiosonde temperature data. Part I: Methodology. *J. Climate*, **16(2)**, 224-240.
- Lanzante, J.R., S.A. Klein, D.J. Seidel. 2003: Temporal homogenization of monthly radiosonde temperature data. Part II: Trends, sensitivities and MSU comparison. *J. Climate*, **16(2)**, 241-262.
- Lelieveld, J., G.J. Roelofs, L. Ganzeveld, J. Feichter, and H. Rodhe. 1997. Terrestrial sources and distribution of atmospheric sulphur. *Phil. Trans. R. Soc. Lond. B.*, **352**, 149-158.
- Luers, J.K., and R.E. Eskridge. 1998: Use of radiosonde temperature data in climate studies. *J. Climate*, **11**, 1002-1019.
- Mann, M.E., R.S. Bradley, and M.K. Hughes. 1999. Northern Hemisphere Temperatures During the Past Millennium: Inferences, Uncertainties, and Limitations. *Geophys. Res. Lett.*, **26**, 759-762.
- McKittrick, R. and P.J. Michaels. (2004). "A Test of Corrections for Extraneous Signals in Gridded Surface Temperature Data". *Climate Research*, **26** pp. 159-173.
- McKittrick, R. and P.J. Michaels. (2004b). "Correction: A Test of Corrections for Extraneous Signals in Gridded Surface Temperature Data". Published Sept. 13, 2004. Available online at <http://www.uoquelph.ca/~rmckitri/research/MM04.Correction.pdf>. Accessed on Dec. 19, 2004.
- McKittrick, R., and P.J. Michaels. 2004c. Are temperature trends affected by economic activity? Reply to Benestad. *Climate Research*, **27**, 175-176.
- Mears, C.A., M.C. Schabel et al. 2002: Correcting the MSU Middle Tropospheric Temperature for Diurnal Drifts. *Proceedings of the International Geophysics and Remote Sensing Symposium*, Volume III, pg. 1839-1841, 2002.
- Mears, C.A., Schabel, M.C., Wentz, F.J. 2003: A Reanalysis of the MSU Channel 2 Tropospheric Temperature Record. *J. Climate*, **16** (22), 3650-3664.
- Mears, C.A., Schabel, M.C., Wentz, F.J. 2003b: Understanding the difference between the UAH and RSS retrievals of satellite-based tropospheric temperature estimate. *Workshop on Reconciling Vertical Temperature Trends*, NCDC, Oct. 27-29, 2003. Available online at [http://cimss.ssec.wisc.edu/itwg/itsc/itsc13/session9/9\\_mears.pdf](http://cimss.ssec.wisc.edu/itwg/itsc/itsc13/session9/9_mears.pdf).
- Mears, C.A., Schabel, M.C., Wentz, F.J. 2003c: A New Tropospheric temperature Dataset From MSU. *14<sup>th</sup> Symposium on Global Change and Climate Variations*, Long Beach, CA., Feb. 11, 2003. Available online at <http://ams.confex.com/ams/pdfpapers/55016.pdf>.
- Meehl, G.A. and W.M. Washington, 1996: El Nino-like climate change in a model with increased atmospheric CO<sub>2</sub>-concentrations. *Nature*, **382**, 56-60.
- Meehl, G. A., Washington, W. M., Wigley, T. M. L., Arblaster, J. M., Dai, A., 2003. Solar and greenhouse gas forcing and climate response in the twentieth century *J. Clim.* **16**, 426-444.
- Meehl, G. A., Washington, W. M., Arblaster, J. M., 2003b. Factors affecting climate sensitivity in global coupled climate models. Paper presented at the American Meteorological Society 83<sup>rd</sup> Annual Meeting, Long Beach, CA.

- Meehl, G.A. W.M. Washington, T.M.L. Wigley, J.M. Arblaster and A. Dai. 2003. Solar and Greenhouse Gas Forcing and Climate Response in the Twentieth Century. *J. Climate*, **16**, 426-444.
- Meehl, G.A., G.J. Boer, C. Covey, M. Latif and R.J. Stouffer, 2000. The Coupled Model Intercomparison Project (CMIP). *Bull. Am. Met. Soc.*, **81**, 313-318.
- Michaels, P.J., R. McKittrick, and P.C. Knappenberger. 2004. Economic Signals in Global Temperature Histories. *15th Symposium on Global Change and Climate Variations; The 84th AMS Annual Meeting*. Seattle, WA, Jan 2004. Available online at <http://ams.confex.com/ams/pdfpapers/72436.pdf>. Accessed on Dec. 18, 2004.
- Milloy, S. Aug. 1, 2003. Global Warming is not a WMD. Fox News Channel. Available online at <http://www.foxnews.com/story/0,2933,93466,00.html>. Accessed on Nov. 21, 2004.
- Milmoe, P.M. 1999. Evaluation of the Environmental Impacts from APCA/CW Partnership. The 1999 American Council for an Energy Efficient Economy (ACEEE) Summer Study. Available online at [http://yosemite.epa.gov/oar/globalwarming.nsf/uniqueKeyLookup/ESTU5LVMC8/\\$file/ACEEEevaluat.doc?OpenElement](http://yosemite.epa.gov/oar/globalwarming.nsf/uniqueKeyLookup/ESTU5LVMC8/$file/ACEEEevaluat.doc?OpenElement). Accessed on Dec. 19, 2004.
- Mo, T. 1995: A Study of the Microwave Sounding Unit on the NOAA-12 satellite. *IEEE Trans. Geosci. Remote Sens.*, **33**, 1141-1152.
- Nash, J., and G. F. Forrester. 1986. Long-term monitoring of stratospheric temperature trends using radiance measurements obtained by the TIROS-N series of NOAA spacecraft. *Adv. Space Res.*, **6** (10), 37-44.
- National Center for Public Policy Research (NCPPr). Aug. 13, 1998. Study Challenging Validity of Satellite Data Flawed. Satellite Data Still the Most Reliable Means of Measuring Planet's Temperature. NCPPr Press Release. Ridenour, D. (ed.). Available online at
- National Research Council (NRC). Panel on Reconciling Temperature Observations. 2000. Reconciling Observations of Global Temperature Change. Wallace, J.M. et. al. (eds.). *National Academy Press*, Washington DC. ISBN 0-309-06891-6. Available online at <http://www.nap.edu/books/0309068916/html/index.html>.
- National Research Council (NRC). Committee on Abrupt Climate Change. 2002. Abrupt Climate Change: Inevitable Surprises. Alley, R.B. et. al. (eds.). *National Academy Press*, Washington DC. ISBN 0-309-07434-7. Available online at <http://books.nap.edu/catalog/10136.html>.
- Oort, A. H., and H. Liu. 1993. Upper-air temperature trends over the globe, 1956-1989. *J. Climate*, **6**, 292-307.
- Parker, D.E., and D.I. Cox. 1995. Toward a consistent global climatological rawinsonde database. *Internat. J. Climatol.*, **15**, 473-496.
- 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., C.K. Folland, and I. Macadam. 2000: Why is the global warming proceeding much slower than expected? *J. Geophys. Res.*, **104**, 3865-3876.
- Pope, V. D., Gallani, M. L., Rowntree, P. R., and Stratton, R. A. (2000). The impact of new physical parametrizations in the Hadley Centre Climate Model - HadAM3. *Climate Dynamics*, **16**, 123-146.
- Prabhakara, C., J.R. Iacovoazzi, J.M. Yoo and G. Dalu. 1998: Global Warming deduced from MSU. *Geophys. Res. Lett.*, **25**(11), 1927-1930.
- Prabhakara, C., J.R. Iacovoazzi, J.M. Yoo and G. Dalu. 2000: Global Warming: Evidence from satellite observations. *Geophys. Res. Lett.*, **27**(21), 3517-3520.
- Ramaswamy, V., M.L. Chanin, J. Angell, J. Barnett, D. Gaffen, M. Gelman, P. Keckhut, 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** (1), 71-122.

- Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent and A. Kaplan. 2003. Global Analyses of SST, Sea Ice and Night Marine Air Temperature Since the Late Nineteenth Century. *Journal of Geophysical Research*, **108** (D14), 4407, doi:10.1029/2002JD002670.
- Reitenbach, R.H., and A.M. Sterin. 1996. An objective analysis scheme for climatological parameters. Thirteenth Conference on Probability and Statistics in Atmospheric Sciences. American Meteorological Society. pp. 334-338.
- Reynolds, R.W. 1988. A real-time Global Sea Surface Temperature Analysis. *J. Climate*, **1**, 75-86.
- Reynolds, R. W., and T. M. Smith. 1994. Improved global sea surface temperature analyses. *J. Climate*, **7**, 929– 948.
- Robinson, A.B., S.L. Baliunas, W. Soon, and Z.W. Robinson. 1998. Environmental Effects of Increased Atmospheric Carbon Dioxide. *Unpublished*. Available online at <http://www.oism.org/pproject/s33p36.htm>. Accessed on Nov. 30, 2004.
- Spencer, R.W. and J.R. Christy. 1990. Precision monitoring of global temperature trends from satellites. *Science*, **247**, 1558-1562.
- Santer, B.D., Mikolajewicz, U., Brüggemann, W., Cubasch, U., Hasselmann, K., Höck, H., Maier-Reimer, E. and Wigley, T.M.L. 1995. Ocean variability and its influence on the detectability of greenhouse warming signals. *Journal of Geophysical Research* **100**, 10693-10725.
- 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, D.J. Gaffen, L. Bengtsson, C. Doutriaux, J.S. Boyle, M. Esch, J.J. Hnilo. G.A. Meehl, E. Roeckner, K.E. Taylor, M.F. Wehner. 2000. Interpreting Differential Temperature Trends at the Surface and in the Lower Troposphere. *Science*, **287**, 5456, 1227-1232.
- Santer, B. D., T. M. L. Wigley, J. S. Boyle, D. J. Gaffen, J. J. Hnilo, D. Nychka, D. E. Parker, and K. E. Taylor. 2000b. Statistical significance of trends and trend differences in layer-average atmospheric temperature time series. *J. Geophys. Res.*, **105**, 7337– 7356.
- Santer, B.D., T.M.L. Wigley, C. Doutriaux, J.S. Boyle, J.E. Hansen, P.D. Jones, G.A. Meehl, E. Roeckner, S. Sengupta, and K.E. Taylor. 2001. Accounting for the effects of volcanoes and ENSO in comparisons of modeled and observed temperature trends. *J. Geophys. Res.*, **106**, 28033-28059.
- Santer, B.D., T.M.L. Wigley, G.A. Meehl, M.F. Wehner, C. Mears, M. Schabel, F.J. Wentz, C. Ammann, J. Arblaster, T. Bettge, W.M. Washington, K.E. Taylor, J.S. Boyle, W. Bruggemann, C. Doutriaux. 2003. Influence of Satellite Data Uncertainties on the Detection of Externally Forced Climate Change. *Science*, **300**, 1280-1284.
- Santer, B.D., R. Sausen, T. M. L. Wigley, J. S. Boyle, K. AchutaRao, C. Doutriaux, J. E. Hansen, G. A. Meehl, E. Roeckner, R. Ruedy, G. Schmidt, and K. E. Taylor. 2003b. Behavior of tropopause height and atmospheric temperature in models, reanalyses, and observations: Decadal changes. *J. Geophys. Res.*, **108** (D1), 4002, doi:10.1029/2002JD002258.
- Santer B.D., M.F. Wehner, T.M.L. Wigley, R. Sausen, G.A. Meehl, K.E. Taylor, C. Ammann, J.M. Arblaster, W.M. Washington, J.S. Boyle and W. Bruggemann. 2003c. Contributions of anthropogenic and natural forcing to recent to recent tropopause height changes. *Science*, **301**, 479-483.
- Schiermeier, Q. 2004. Global warming anomaly may succumb to microwave study. *Nature, Nature News*, **429**, 7, doi:10.1038/429007a. Available online at [http://www.nature.com/news/2004/040503/pf/429007a\\_pf.html](http://www.nature.com/news/2004/040503/pf/429007a_pf.html). Accessed on Dec. 27, 2004.
- Schubert, S. R., R. Rood, and J. Pfaendtner. 1993. An assimilated data set for Earth science applications. *Bull. Am. Meteorol. Soc.*, **74**, 2331–2342.
- Seidel, D.J. 2004. Personal communication, Sept. 2004.
- Seidel, D.J., J. Angell, J. Christy, M. Free, S. Klein, J. Lanzante, C. Mears, D. Parker, M. Schabel, R. Spencer, A. Sterin, P. Thorne, and F. Wentz. 2003. Intercomparison of Global Upper-Air Temperature Datasets from Radiosondes and Satellites. *AMS Annual Meeting, 2003*.

- Seidel, D.J., J.K. Angell, J. Christy, M. Free, S.A. Klein, J.R. Lanzante, C. Mears, D. Parker, M. Schabel, R. Spencer, A. Sterin, P. Thorne, F. Wentz. 2004. Uncertainties in Signals of Large-Scale Climate Variations in Radiosonde and Satellite Upper-Air Temperature Datasets. *J. Clim.*, **17**, 2225-2240.
- Shen, S.S., M. Thomas, C.F. Ropelewski, and R.E. Livezey. 1998. An optimal regional averaging method with error estimates and a test using tropical Pacific SST data. *J. Clim.*, **11**, 2340-2350.
- Singer, S.F. 1999. *EOS*, **80**, 183.
- Soon, W. and S. Baliunas. 2003. Proxy climatic and environmental changes of the past 1000 years. *Climate Research*, **23**, 89-110.
- Spencer, R.W. 2004. When Is Global Warming Really a Cooling? *Tech Central Station Online*. Published on May 5, 2004. Available online at <http://www.techcentralstation.com/050504H.html>. Accessed Dec. 26, 2004.
- Spencer, R.W. 2004b. Global Warming: The Satellite Saga Continues. *Tech Central Station Online*. Published on Dec. 3, 2004. Available online at <http://www.techcentralstation.com/120304F.html>. Accessed Dec. 27, 2004.
- Spencer, R.W. and J.R. Christy. 1990. Precision monitoring of global temperature trends from satellites. *Science*, **247**, 1558-1562.
- Spencer, R.W. and J.R. Christy. 1992a. Precision and radiosonde validation of satellite gridpoint temperature anomalies. Part I: MSU Channel 2. *J. Clim.*, **5**, 847-857.
- Spencer, R.W. and J.R. Christy. 1992b. Precision and radiosonde validation of satellite gridpoint temperature anomalies. Part II: A tropospheric retrieval and trends during 1979-90. *J. Clim.*, **5**, 858-866.
- Spencer, R. W., and J. R. Christy. 1993. Precision lower stratospheric temperature monitoring with the MSU technique: Validation and results, 1979–1991. *J. Climate*, **6**, 1194–1204.
- Stendel, M., Christy, J.R. and Bengtsson, L. 2000. Assessing levels of uncertainty in recent temperature time series. *Clim. Dyn.*, **16**, 1405-1423.
- Sterin, A.M., 1999. An analysis of linear trends in the free atmosphere temperature series for 1958-1997. *Meteorologiai Hidrologia*. **5**, 52-68.
- Sterin, A.M., 2000. Variations of upper-air temperature in 1998-1999 and their effect on long period trends. Proc. 24th Annual Climate Diagnostics and Prediction Workshop. NOAA. pp. 222-225.
- Sterin, A.M., 2001. Tropospheric and Lower Stratospheric Temperature Anomalies Based on Global Radiosonde Network Data . In Trends Online: A Compendium of Data on Global Change. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee, U.S.A.
- Stott, P.A., Tett, S.F.B., Jones, G.S., Allen, M.R., Mitchell, J.F.B., and Jenkins, G.J. 2000. External control of 20th century temperature by natural and anthropogenic forcings. *Science* **290** (5499): 2133-2137.
- Tett, S. 2004. Personal communication, Dec. 21, 2004.
- Tett, S.,F. B., G.S. Jones, P.A. Stott, D.C. Hill, J.F.B. Mitchell, M.R. Allen, W.J. Ingram, T.C. Johns, C.E. Johnson, A. Jones, D.L. Roberts, D.M.H. Sexton, and M.J. Woodage. 2002. Estimation of natural and anthropogenic contributions to twentieth century temperature change. *J. Geophys. Res.*, **107** (D16), 10.1029/2000JD000028.
- Tett, S., and P. Thorne. 2004. Tropospheric temperature series from satellites. Arising from: Q. Fu et al., *Nature* 429, 55–58 (2004). *Nature*, **432**, doi:10.1038/nature03208. Brief Communications. December 2, 2004. Abstract available at [http://www.nature.com/cgi-taf/DynaPage.taf?file=/nature/journal/v432/n7017/abs/nature03208\\_fs.html](http://www.nature.com/cgi-taf/DynaPage.taf?file=/nature/journal/v432/n7017/abs/nature03208_fs.html). Accessed on Dec. 27, 2004. Subscription required for access to the full article.
- Thompson, J., and S. Solomon, Interpretation of Recent Southern Hemisphere Climate Change. *Science*, **296**, 895–899.

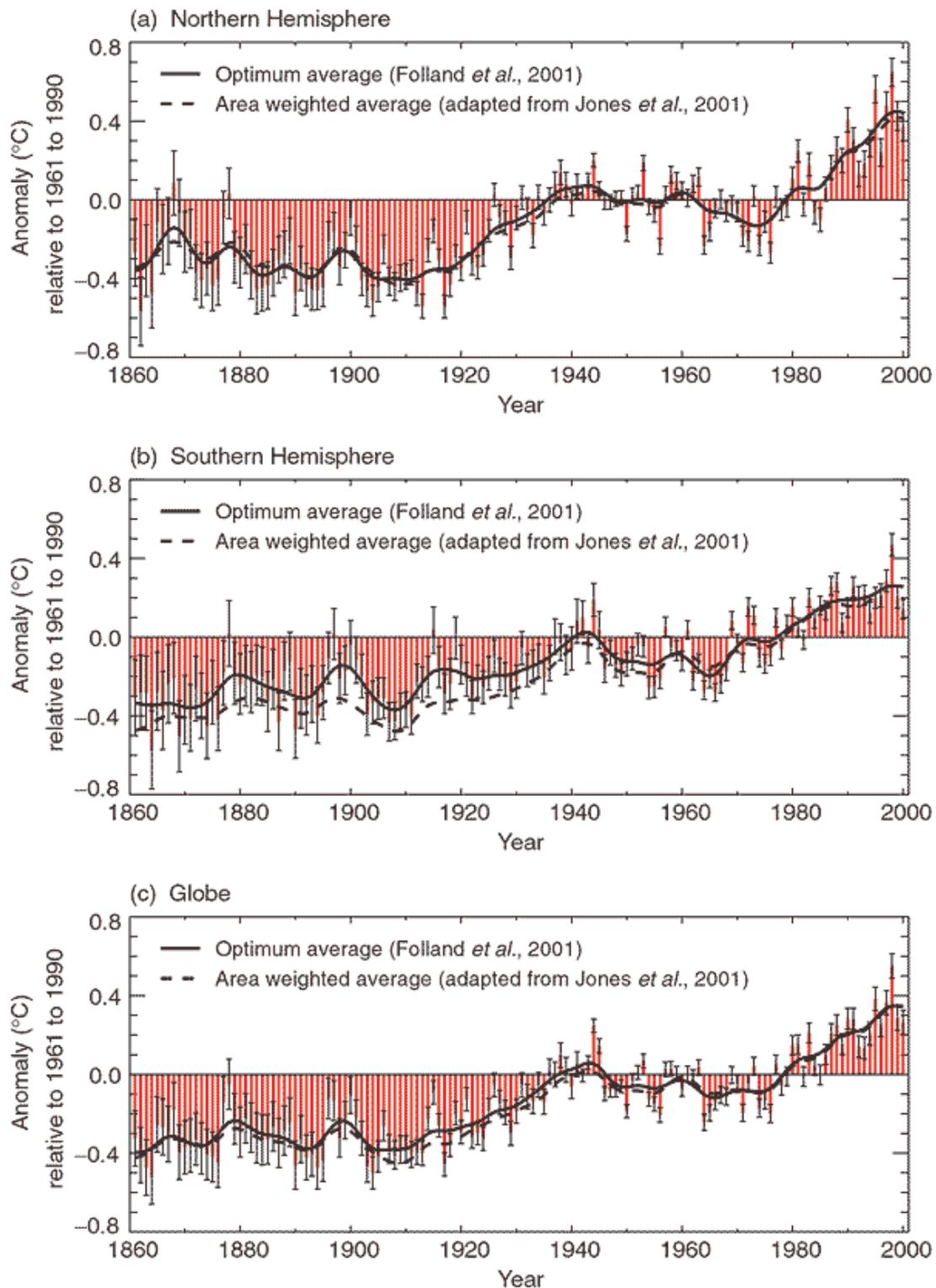
- Timmerman, A., J. Oberhuber, A. Bacher, M. Esch, M. Latif and E. Roeckner, 1999. Increased El Nino frequency in a climate model forced by future greenhouse warming. *Nature*, **398**, 694-696.
- Trenberth, K.E. 1984. Signal vs. noise in the Southern Oscillation. *Mon. Wea. Rev.*, **112**, 326-332.
- Trenberth, K.E. and J. M. Caron. 2001. Estimates of meridional atmosphere and ocean heat transports. *J. Climate*, **14**, 3433–3443.
- Trenberth, K.E., J. M. Caron, and D. P. Stepaniak, 2001: The atmospheric energy budget and implications for surface fluxes and ocean heat transports. *Climate Dyn.*, **17**, 259–276.
- Trenberth, K.E. and D.P. Stepaniak. 2003a. Covariability of components of poleward atmospheric energy transports on seasonal and interannual timescales. *J. Clim.*, **16**, 3691-3705.
- Trenberth, K.E. and D.P. Stepaniak. 2003b. Seamless poleward atmospheric energy transports and implications for the Hadley circulation. *J. Clim.*, **16**, 3706-3722.
- Trenberth, K.E. 2004. Personal communication, Jan. 2004.
- Uppala, S. 2003. In *Proc. Workshop Reanalysis 5–9 November 2001*, 1–10 (European Centre for Medium-range Weather Forecasting, Reading).
- Vaughan, G.J. Marshal, W.M. Connolley, J.C. King and D.P. R. Mulvaney. 2001. Climate Change: Devil in the Detail. *Science*, **293**, 1777-1779.
- Vinnikov, K.Y, A. Robock, D.J. Cavalieri, C.L. Parkinson. 2002a. Analysis of seasonal cycles in climatic trends with application to satellite observations of sea ice extent. *Geophys. Res. Lett.*, **29 (9)** doi: 10.1029/2001GL014481.
- Vinnikov, K.Y, A. Robock, A. Basist. 2002b. Diurnal and seasonal cycles of trends of surface air temperature. *J. Geophys. Res.*, **107 (D22)**, 4641, doi:10.1029/2001JD002007.
- 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, A. Robock, N.C. Grody and A. Basist. 2004. Analysis of Diurnal and Seasonal Cycles and Trends in Climatic Records with Arbitrary Observation Times. *Geophys. Res. Lett.*, In Press.
- W.M. Washington, J.W. Weatherly, G.A. Meehl, A.J. Semtner Jr., T.W. Bettge, A.P. Craig, W.G. Strand Jr., J.M. Arblaster, V.B. Wayland, R. James, Y. Zhang. 2000. Parallel Climate Model (PCM) control and transient simulations. *Climate Dynamics*, **16**, 755-774.
- Weatherhead, E.C. and Coauthors. 1998. Factors affecting the detection of trends: Statistical considerations and applications to environmental data. *J. Geophys. Res.*, **103**, 17 149-17 161.
- Wentz, F.J. 1998. Algorithm Theoretical Basis Document: AMSR Ocean Algorithm. Remote Sensing Systems. Santa Rosa, CA. *RSS Tech. Report 110398*. Nov. 3, 1998.
- Wentz, F.J. and M. Schabel. 1998. Effects of orbital decay on satellite-derived lower tropospheric temperature trends. *Nature*, **394**, 661-664.
- Whipple, D. 2004. Climate: The Tropospheric Data Do Conform. *UPI Newswire*. Dec. 67, 2004. Available online from Space Daily at <http://www.spacedaily.com/news/climate-04zzzza.html>. Accessed on Dec. 27, 2004.
- Wigley. T.M.L. and B.D. Santer. 2003. Differential ENSO and volcanic effects on surface and tropospheric temperatures. *J. Clim.*, Submitted.
- Wikipedia. 2004. Satellite temperature measurements. *Wikipedia*. Available online at [http://en.wikipedia.org/wiki/Satellite\\_temperature\\_measurements](http://en.wikipedia.org/wiki/Satellite_temperature_measurements). Accessed Dec. 12, 2004.
- Wilks, D. S. 1995. Statistical Methods in the Atmospheric Sciences: An Introduction. *Academic Press, New York*, 467 pp.
- World Meteorological Organization (WMO). Report of the International Ozone Trends Panel: 1988, Global Ozone Res. and Monit. Proj., *Rep. 18*, chap. 6, pp. 443–498, Geneva, 1990.

World Meteorological Organization (WMO). 1996. Measurements of upper air temperature, pressure, and humidity. *Guide to Meteorological Instruments and Methods of Observation*, Chapter 12. WMO-No. 8, Sixth Edition, Geneva, I.12-I.12-32.

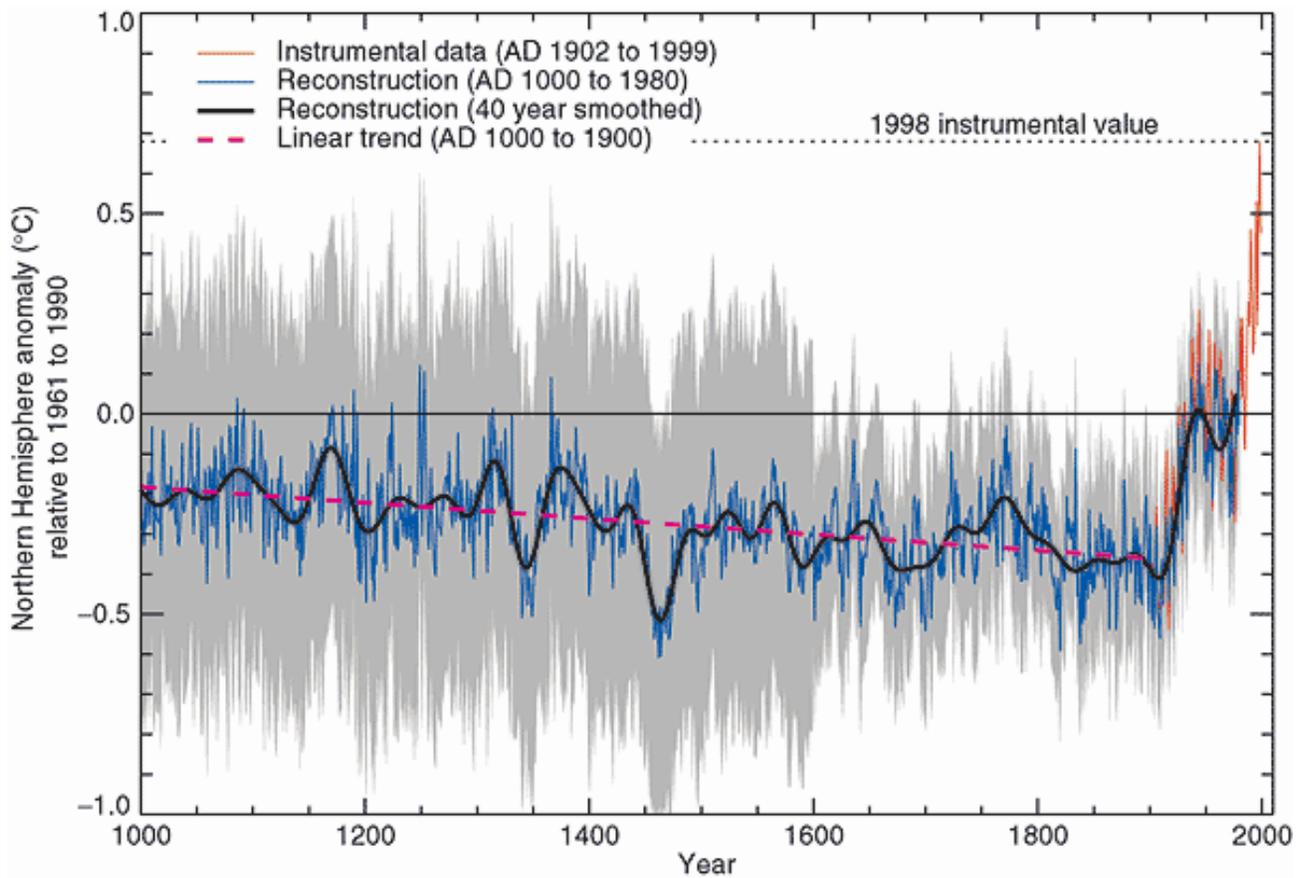
Xie, P. and P.A. Arkin, 1996. Analyses of Global Monthly Precipitation Using Gauge Observations, Satellite Estimates, and Numerical Model Predictions. *J. Climate*, **9**, 840-858.

Xie, P. and A. Arkin. 1997. Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates and numerical model outputs. *Bull. Amer. Meteor. Soc.*, **78**, 2539–2558.

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



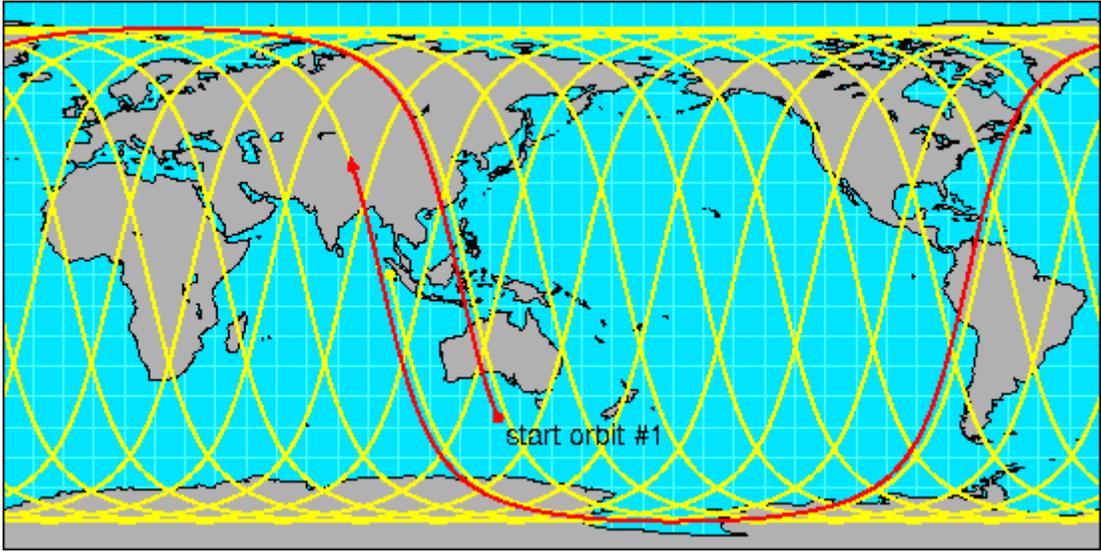
**Figure 1:** Smoothed annual anomalies of combined land-surface air and sea surface temperatures ( $^{\circ}\text{C}$ ), 1861 to 2000, relative to the 1961 to 1990 average for (a) Northern Hemisphere; (b) Southern Hemisphere; and (c) Globe. The curves shown use running near-decadal averages. (Folland et al., 2001) The solid curves are optimally averaged (Folland et al., 2001), and the dashed curves are standard area weighted (adapted from Jones et al., 2001). The red bars are annual averages and are shown with twice their standard errors (demarcated black bars). From IPCC (2001, Chapter 2.2.2.3).



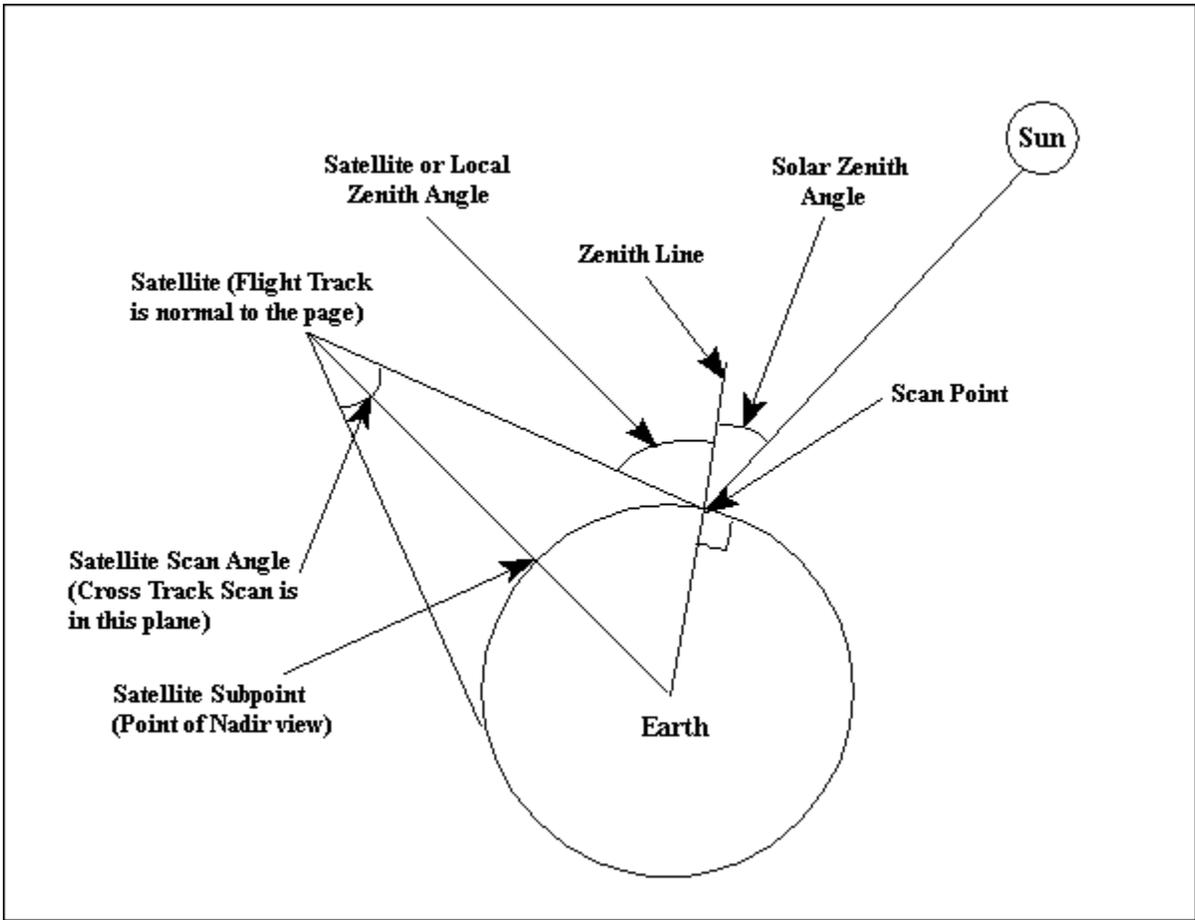
**Figure 2:** Millennial Northern Hemisphere surface air temperature reconstruction (shown in blue) and directly measured data from various instrumental records (shown in red) from AD 1000 to 1999 (adapted from Mann et al. 1999). The black curve shows the underlying trend smoothed by 40-year running averages, the purple dashed line shows the linear trend from AD 1000 to 1850. The shaded gray area gives confidence intervals as two standard error limits. From IPCC (2001, Chapter 2.3.2.2).



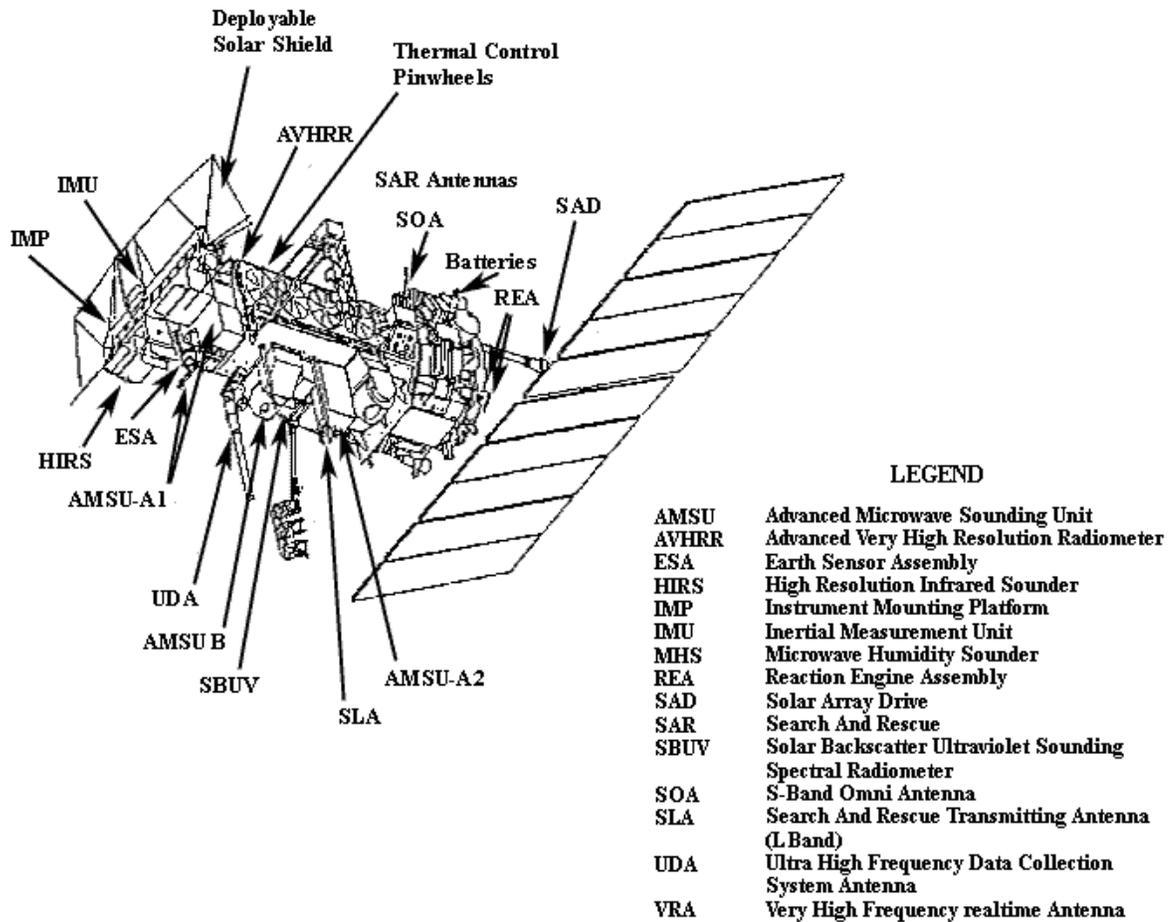
**Figure 3:** NOAA-17 – The latest in the NOAA/NASA Polar Orbiting Environmental Satellite TIROS-ATN Program.



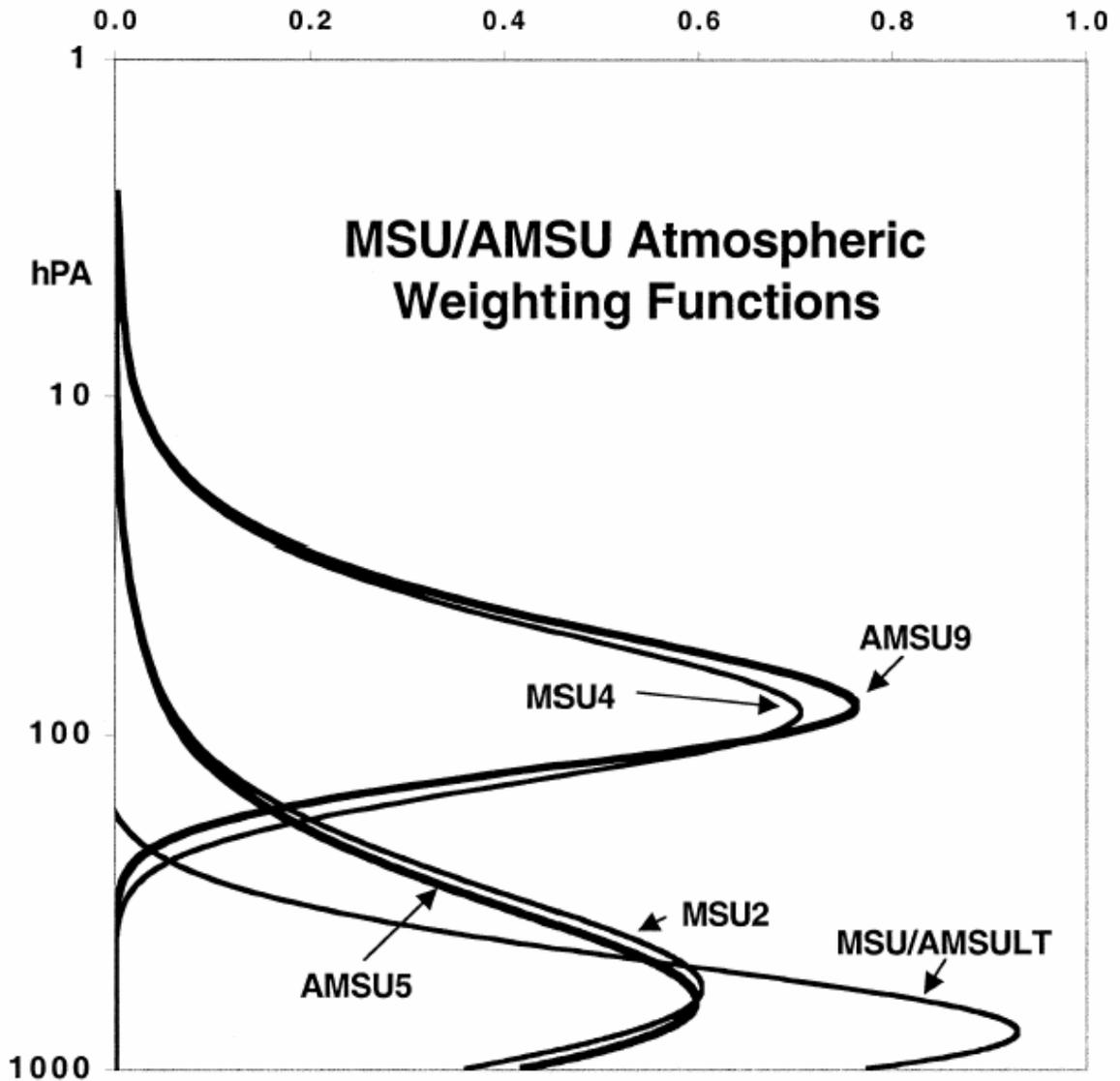
**Figure 4:** POES Satellite Flight Path with respect to the earth's surface.



**Figure 5:** POES Satellite Flight Path and Scan Geometry in the ecliptic (Sun/Earth) plane. The north pole is normal to the page.



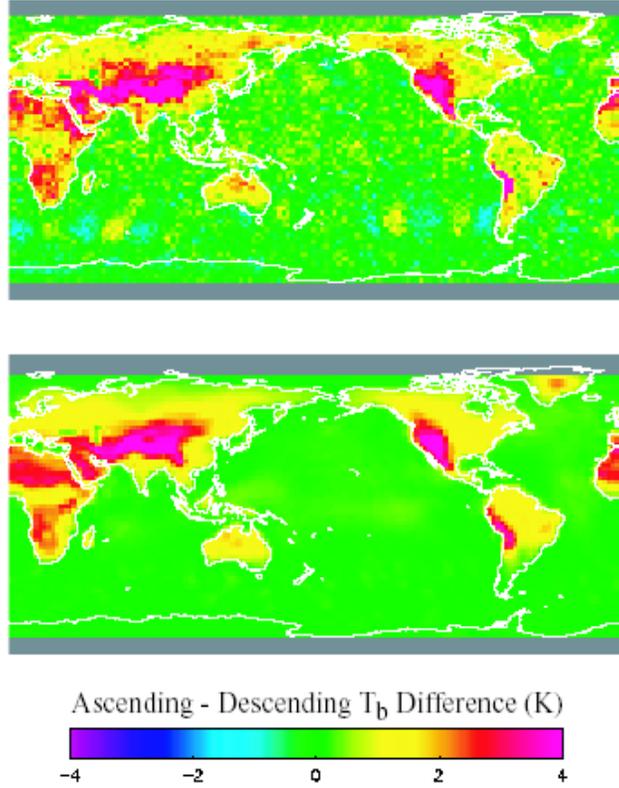
**Figure 6:** The TIROS-ATN Spacecraft and its components. The A1, A2, and B Advanced Microwave Sounding Units are the packages used for troposphere and stratosphere temperature studies.



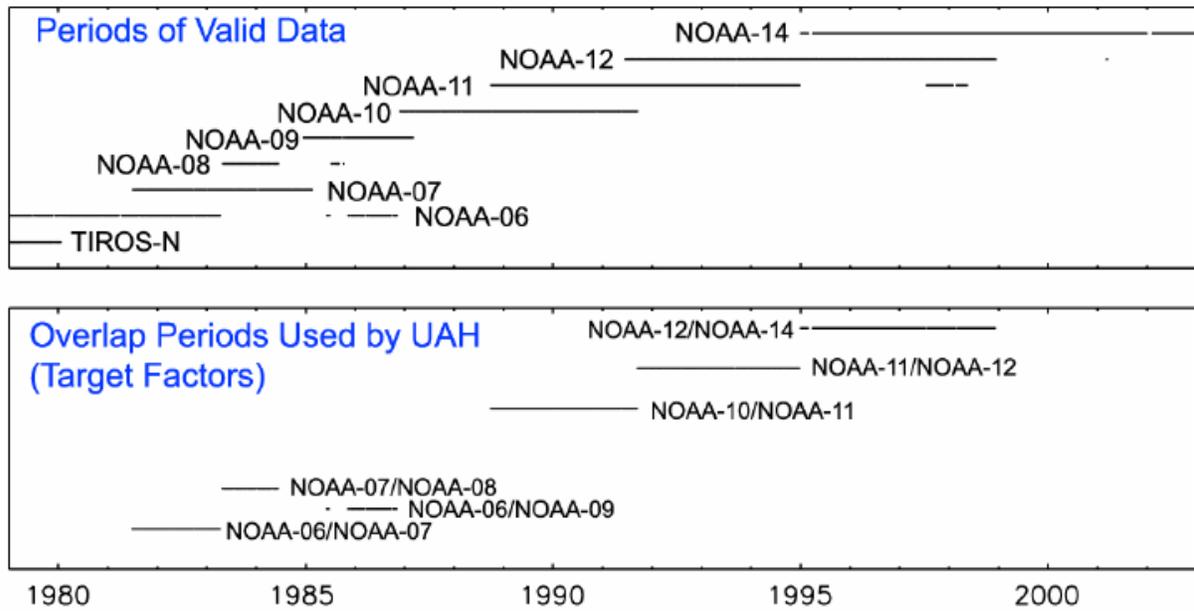
**Figure 7:** Static atmospheric weighting function profiles as a function of altitude (in pressure units <sup>3</sup>) for MSU and AMSU products. Not shown are the surface contribution factors, which for land (ocean) for TLT are 0.20 (0.10) and for TMT are 0.10 (0.05) of the total weighted profile. The land surface contribution increases for higher surface altitudes. The MSU/AMSULT profile is calculated from views of Nadir and Off-Nadir views of Channel 2. From Christy et. al. (2003).

<b>Global Average Atmospheric Temperature Trends from MSU/AMSU (Deg. K/Decade)</b>				
	<b>TLT</b>	<b>2</b>	<b>4</b>	<b>Ref.</b>
<b>UAH</b>	<b>0.061 (+/- 0.050)</b>	<b>0.013 (+/- 0.050)</b>	<b>-0.510 (+/- 0.100)</b>	Christy et. al., 2003 1979 – 2002 data
	<b>0.08 (+/- 0.05)</b>			Christy et. al., 2004 1978 – 2003 data
	<b>0.06 (+/- 0.06)</b>	<b>0.04 (+/- 0.06)</b>	<b>-0.49 (+/- 0.10)</b>	Christy et. al., 2000 1979 – 19984 data
<b>RSS</b>	<b>--</b>	<b>0.097 (+/- 0.020)</b>	<b>-0.400</b>	Mears et. al., 2003; Seidel et. al., 2003 1979 – 2001 data
		<b>0.129</b>		Mears et. al., 2004 1979 – 2004 data
<b>Vinnikov &amp; Grody</b>	<b>--</b>	<b>0.240 (+/- 0.020)</b>	<b>--</b>	Vinnikov & Grody, 2003 1979 – 2002 data
<b>Prabhakara</b>	<b>--</b>	<b>0.13 (+/- 0.05)</b>	<b>--</b>	Prabhakara et. al., 2000 1980 - 1999 data

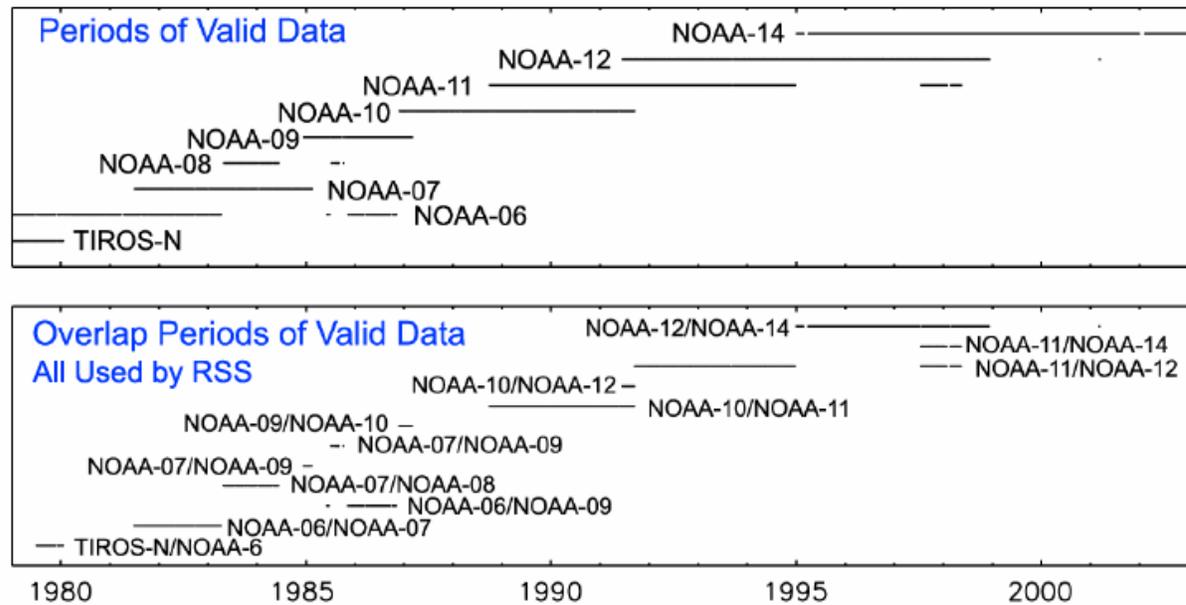
**Figure 8:** Global average tropospheric temperature results from MSU and AMSU records. TLT results are representative of the lower troposphere. Channels 2 and 4 give the middle troposphere and lower stratosphere respectively.



**Figure 9:** Ascending-Descending Channel 2 brightness temperature differences for the entire MSU dataset for the central 5 fields of view, the month of June, and for ascending node Local Equatorial Crossing Times (LECT's) of 15:00 to 16:00 (Top), and the same as simulated by CCM3 diurnal climatology by the RSS Team (from Mears et. al., 2002).

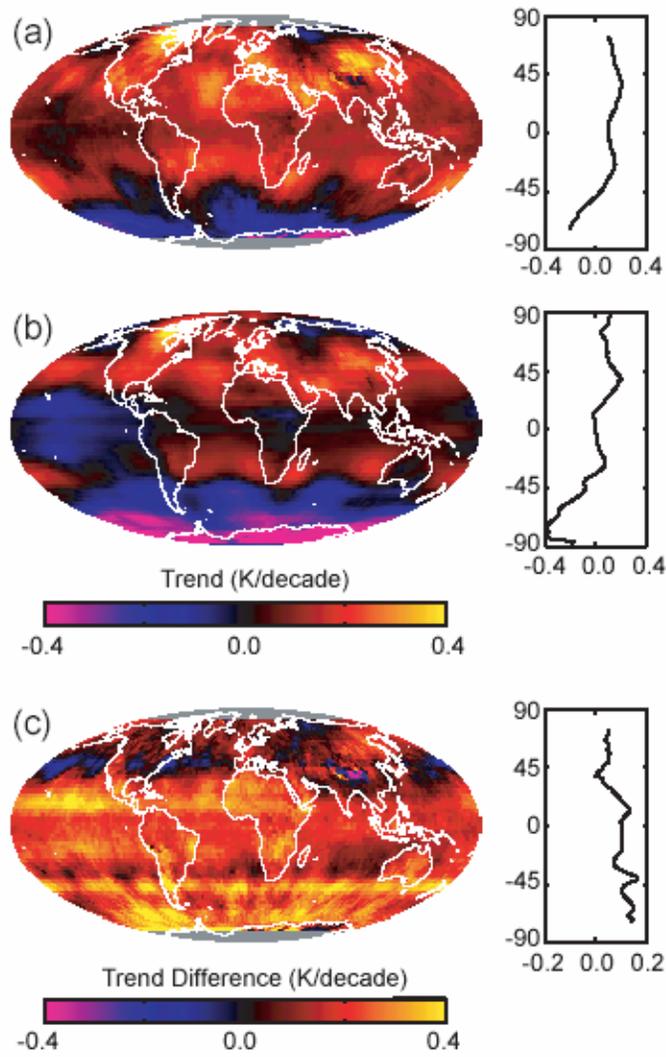


a) Service Life Overlaps used by UAH.



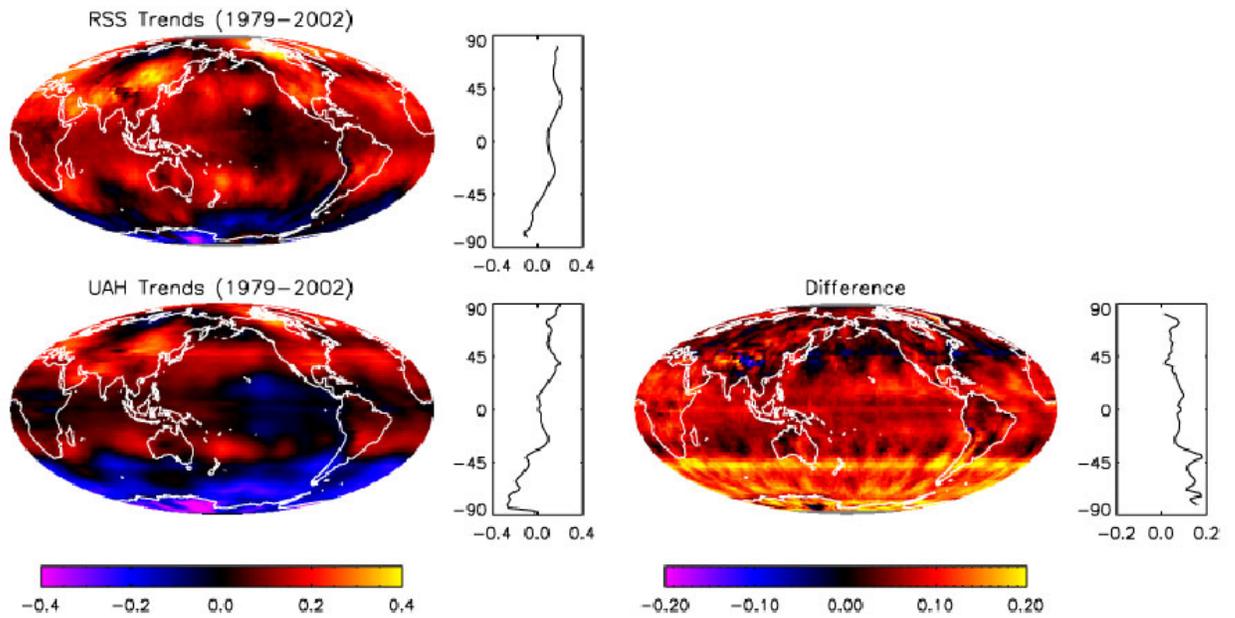
b) Service Life Overlaps used by RSS.

**Figure 10:** Service lives of NOAA POES Satellites from 1979 to 2003 and the overlaps in service used for Hot Target Coefficient derivation and merge calculations by, a) UAH and, b) RSS. Taken from Mears et. al. (2003).

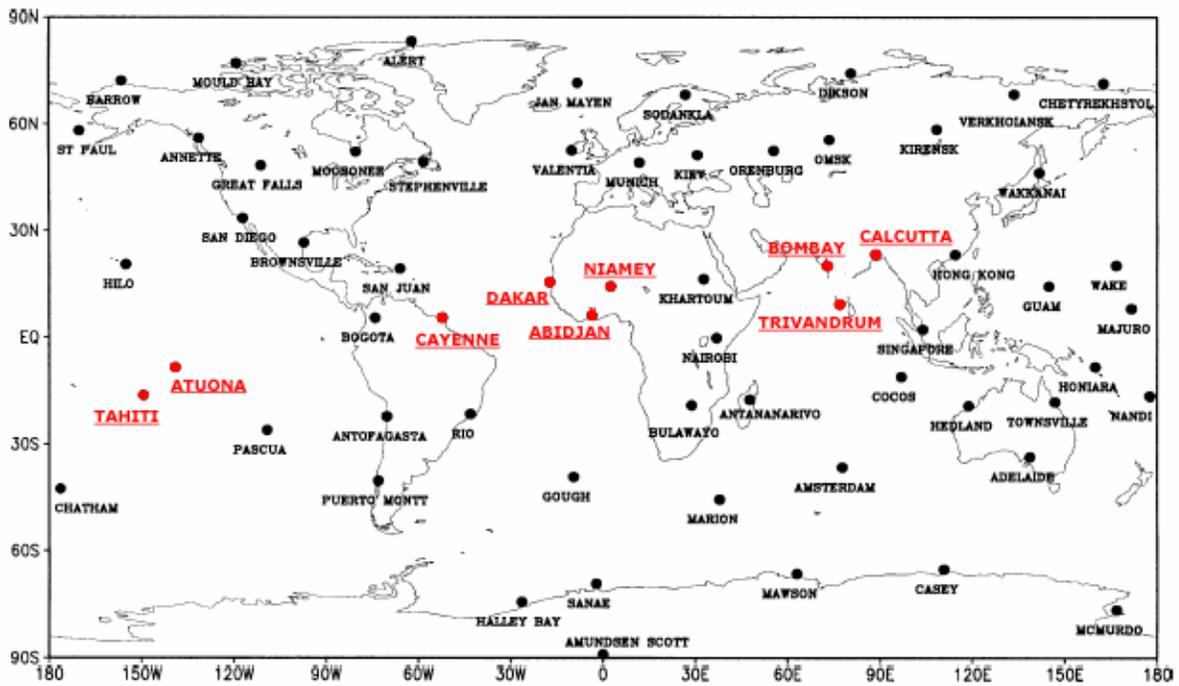


**Figure 11A:** MSU Channel 2 brightness temperatures for 1979 to 2001 as determined by, a) RSS Ver. 1.0 (Mears et. al., 2003), b) UAH Ver. 5.0 (Christy et. al., 2003), and, c) the difference between the two. Taken from Mears et. al. (2003).

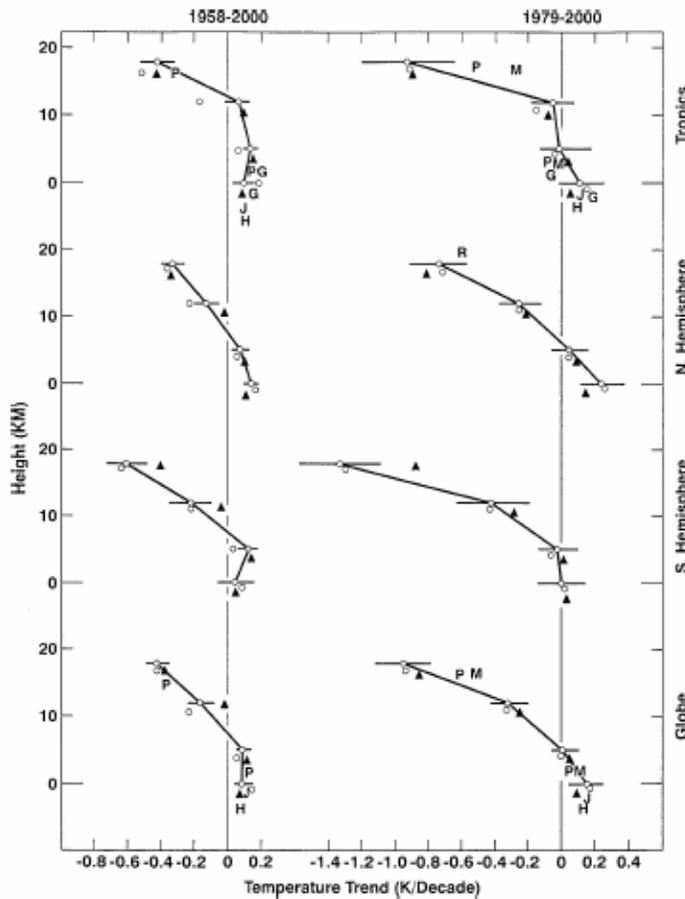
## Decadal Trend Map Comparison



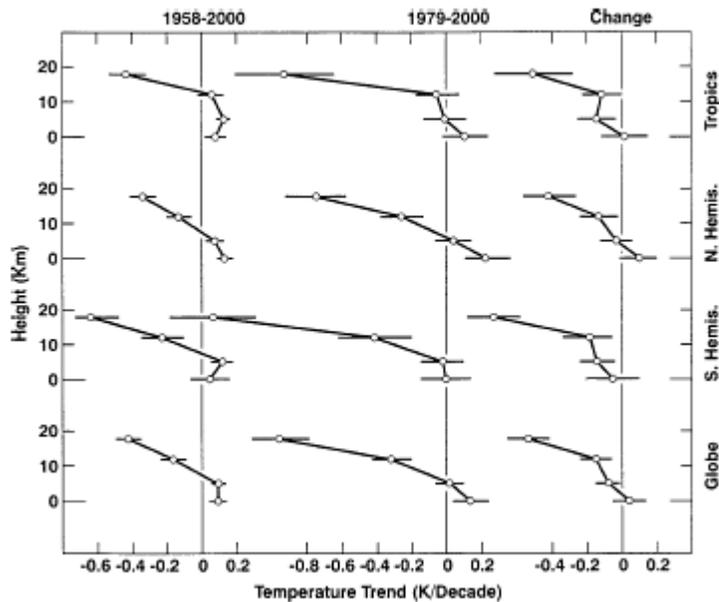
**Figure 11B:** Same as Figure 11A but for 1979 to 2002. Taken from Mears et. al. (2003b).



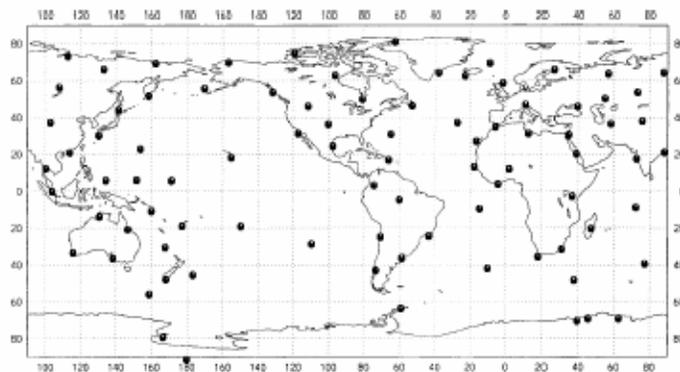
**Figure 12:** The Angell 63 Network with the 9 anomalous stations identified by Angell shown in red. These stations were removed in the Angell 54 network (from Angell, 2003).



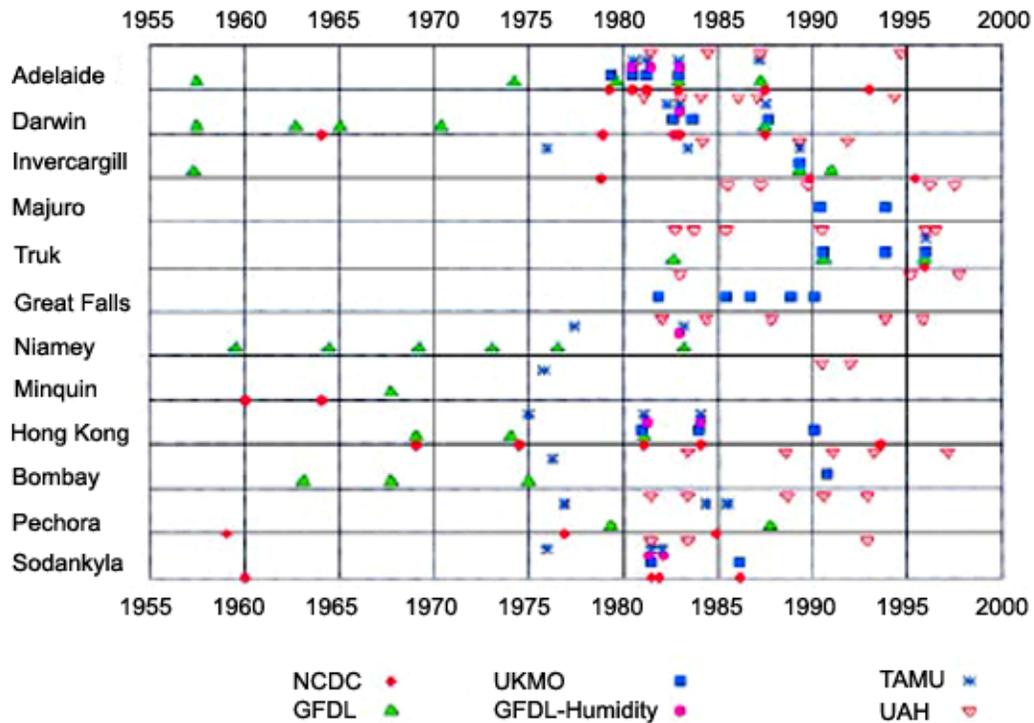
**Figure 13:** Upper air temperature trends in deg. K/decade from Angell 54 at various troposphere and lower stratosphere altitudes for the Northern Hemisphere, the Southern Hemisphere, the Tropics, and the globe, compared with those from MSU, other radiosonde analysis and re-analysis products, and surface-air data, for 1958-2000 (Left) and 1979-2000 (Right). MSU data (M) are from UAH Ver. D (Christy et. al., 2000). Alternate sonde products are from Lanzante et. al. (2003: solid triangles), Parker et. al. (1997: P), and Gaffen et. al. (2000b: G). The re-analysis product is a radiosonde-satellite product from Ramaswamy et. al. (2000: R). Surface temperature trends are from Jones et. al. (2001: J) and Hansen et. al. (1999: H). Trends shown for Lanzante et. al. (2003) are for 1959-1997 (Left) and 1979-1997 (Right), and data for Gaffen et. al. (2000b) are for 1960-1997 (Left) and 1979-1997 (Right). The small circles unconnected by straight lines show trends for the original Angell 63 network (Angell, 1988). The horizontal bars show 2-sigma confidence intervals for each trend indicated. Figure taken from Angell, 2003.



**Figure 14:** Upper air temperature trends in deg. K/decade from Angell 54 at various troposphere and lower stratosphere altitudes for the Northern Hemisphere, the Southern Hemisphere, the Tropics, and the globe for 1958-2000 (Left), 1979-2000 (Center), and the change from the former to the latter (Right). The horizontal bars show 2-sigma confidence intervals for each trend indicated. Figure taken from Angell, 2003.



**Figure 15:** The 87 station network used in the LKS radiosonde analysis (Lanzante et. al., 2003).



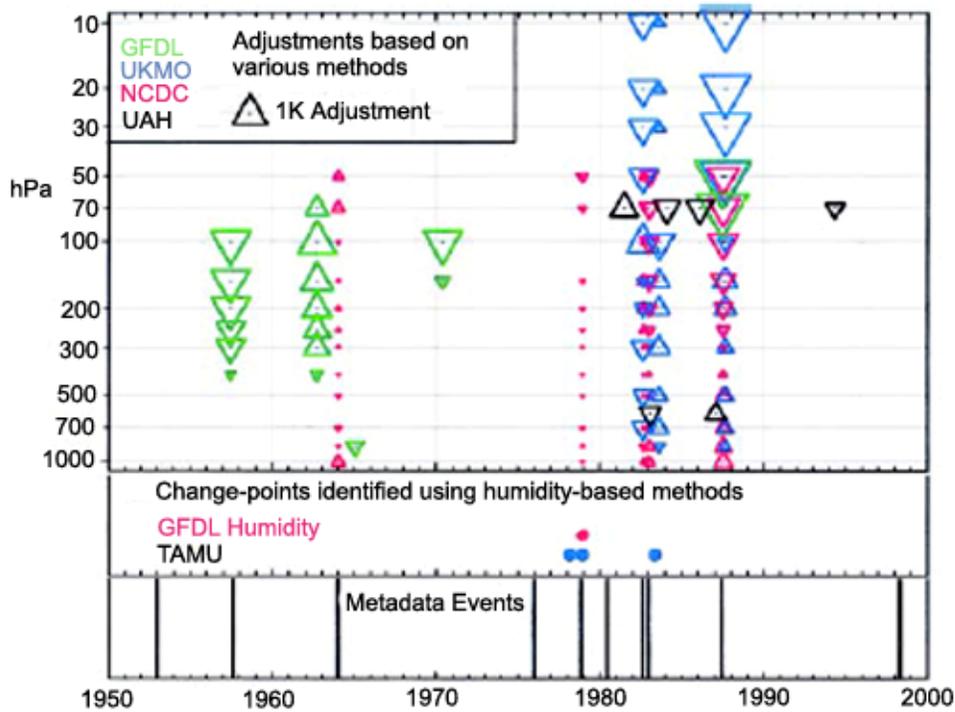
**Figure 16:** Discrete, and possibly anomalous temperature change points found at any level by the six participating groups participating in the 2002 CARDS Workshop. Each team used a different standard method for detecting possibly anomalous change points in the historical temperature records of the 12 stations shown and suggested corrections for all that were considered anomalous. The LKS methodology (Lanzante et. al., 2003) is denoted by (GFDL), and the HadRT methodology (Parker et. al., 1997) is denoted by UKMO. Symbols on or above the grid line for each station show change points for that station. Countries for each station are shown in Figure 17b. Taken from Free et. al., 2002).

	NCDC	Met Office	GFDL	GFDL-Humidity	TAMU	UAH	Average
<b>NCDC</b>	X	56	55	53	70	21	50.9
<b>Met Office</b>	56	X	39	58	45	23	44.0
<b>GFDL</b>	55	39	X	44	49	12	39.9
<b>GFDL-Humidity</b>	53	58	44	X	82	25	52.6
<b>TAMU</b>	57	45	49	82	X	13	51.6
<b>UAH</b>	21	23	12	25	13	X	18.5

	NCDC	Met Office	GFDL	GFDL-Humidity	TAMU	UAH
<b>Adelaide, Australia</b>	6	4	3	3	4	4
<b>Darwin, Australia</b>	3	3	1	1	3	6
<b>Invercargill, New Zealand</b>	2	1	2	X	2	3
<b>Majuro, Marshall Islands</b>	0	2	0	0	0	5
<b>Truk, Micronesia</b>	1	3	3	X	1	6
<b>Great Falls, MT, United States</b>	0	5	0	X	0	3
<b>Niamey, Niger</b>	X	X	1	1	1	5
<b>Minquin, China</b>	0	0	0	X	0	2
<b>Hong Kong, China</b>	3	3	1	2	2	X
<b>Bombay, India</b>	X	1	0	X	0	5
<b>Pechora, Russia</b>	1	X	2	X	2	5
<b>Sodankyla, Finland</b>	3	2	X	2	2	3

**Figure 17:** Comparison of change points identified by the different methods used by each team participating in the 2002 CARDS Workshop (top). The values shown are the percentages of change points that occurred within 6 months of each other for each pair of methods. Where one or both teams found change points, the percent agreement is the number of common change points divided by the total number of change points found by the two teams in question. Stations where neither of them found any change points are considered to be in 100% agreement. Each entry in the last column is the average of the percentages shown in that row. Comparisons with the Met Office and UAH are limited to 1979–97. The stations may not constitute a representative sample of all radiosonde data, and not all groups produced results for all stations. The bottom chart shows the number of change points identified by each team for 1979–97. An X denotes stations for which a group did not provide data, and

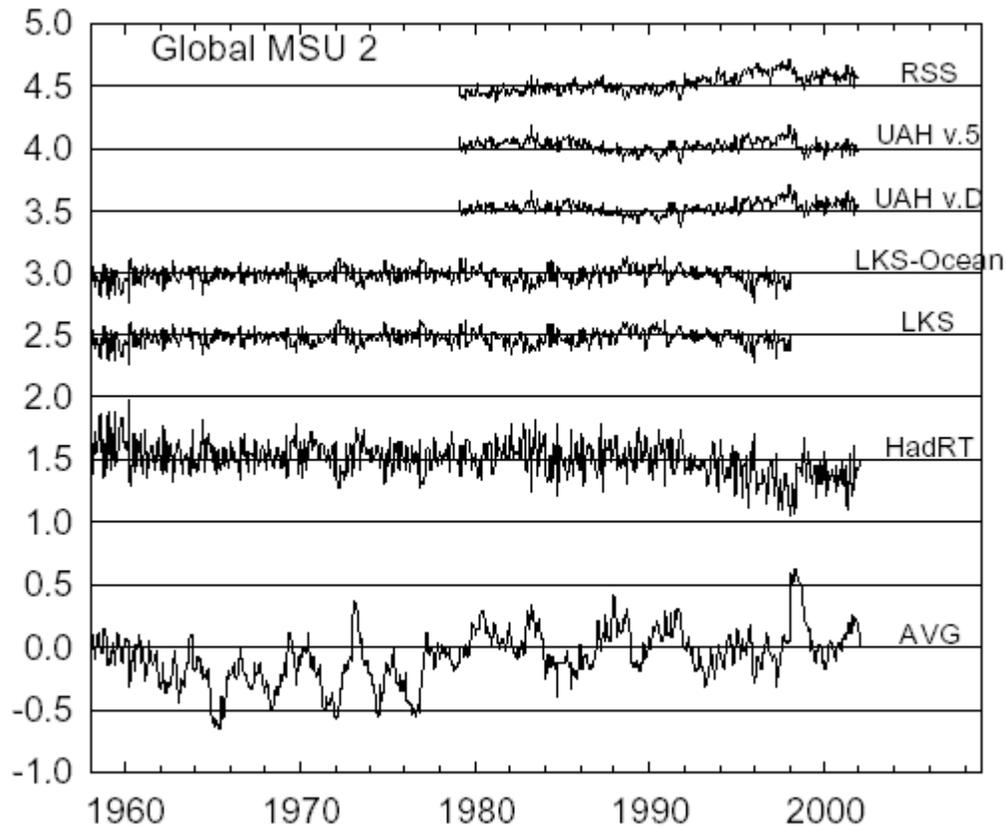
an 0 denotes stations where a group examined the record but found no break points from 1979 through 1997. Taken from Free et. al., 2002.



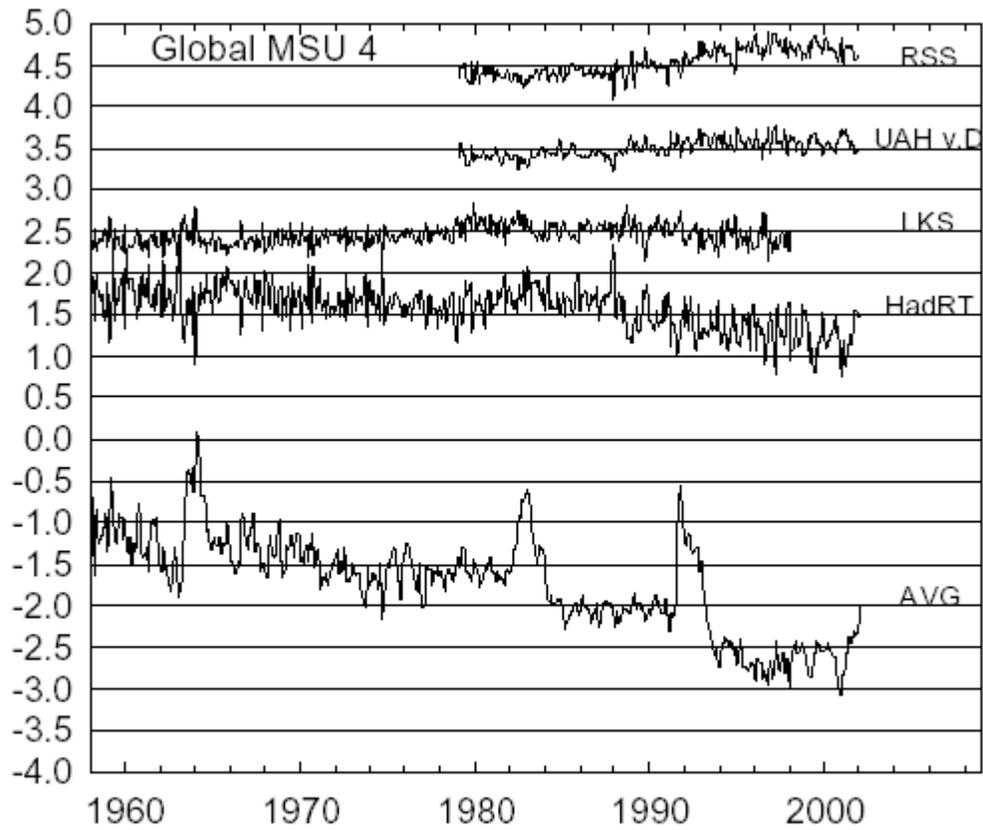
**Figure 18:** Comparison of change points identified by the different methods used by each team participating in the 2022 CARDS Workshop (top). The values shown are the percentages of change points that occurred within 6 months of each other for each pair of methods. Where one or both teams found change points, the percent agreement is the number of common change points divided by the total number of change points found by the two teams in question. Stations where neither of them found any change points are considered to be in 100% agreement. Each entry in the last column is the average of the percentages shown in that row. Comparisons with the Met Office and UAH are limited to 1979–97. The stations may not constitute a representative sample of all radiosonde data, and not all groups produced results for all stations. The bottom chart shows the number of change points identified by each team for 1979–97. An X denotes stations for which a group did not provide data, and an 0 denotes stations where a group examined the record but found no break points from 1979 through 1997. Taken from Free et. al., 2002.

Unadjusted				
	GFDL	NCDC	Met Office	UAH
50 hPa	$-2.58 \pm 1.30$	$-2.44 \pm 1.24$	$-2.41 \pm 1.12$	$-0.89 \pm 0.74$
850 hPa	$-0.12 \pm 0.24$	$-0.07 \pm 0.20$	$-0.07 \pm 0.20$	
Adjusted				
	GFDL	NCDC	Met Office	UAH
50 hPa	$-0.62 \pm 0.96$	$-1.60 \pm 1.04$	$-0.47 \pm 0.86$	$-0.24 \pm 0.48$
850 hPa	$-0.10 \pm 0.22$	$-0.27 \pm 0.22$	$-0.09 \pm 0.20$	
Difference				
	GFDL	NCDC	Met Office	UAH
50 hPa	$1.96 \pm 1.61$	$0.83 \pm 1.62$	$1.94 \pm 1.41$	$0.65 \pm 0.88$
850 hPa	$0.02 \pm 0.32$	$-0.20 \pm 0.30$	$-0.02 \pm 0.28$	

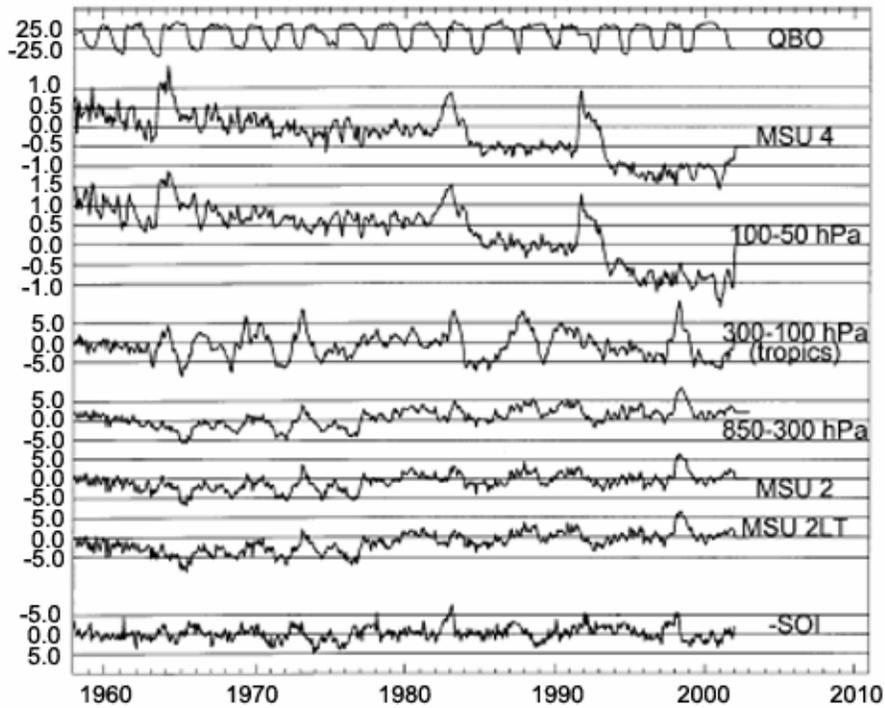
**Figure 19:** Tropospheric (850 hPa) and stratospheric (50 hPa) temperature trends in deg. K/decade at Darwin, Australia as determined by 4 of the 6 teams participating in the 2002 CARDS Workshop. UAH trends represent Darwin data weighted to simulate MSU Channel 4. The difference uncertainties are twice the square root of the sum of the squares of the individual time series standard errors, and represent the 95 percent confidence interval. Taken from Free et. al., 2002.



**Figure 20:** Global temperature anomalies for the middle troposphere from MSU/AMSU and 2 radiosonde datasets. The HadRT sonde dataset represents monthly CLIMAT TEMP reports and the LKS sonde dataset is from an 87 station network corrected for temporal inhomogeneities. The bottom curve gives the average trend for all products and the individual product curves give deviations from the average (from Seidel et. al., 2003).

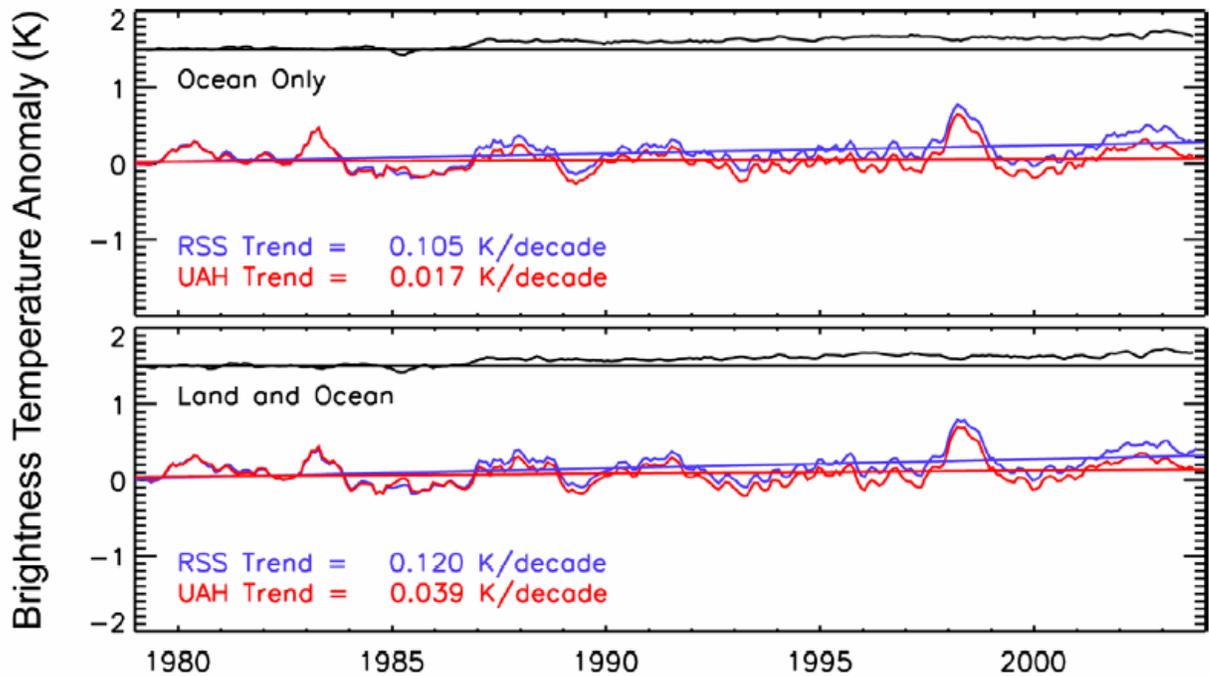


**Figure 21:** Global temperature anomalies for the lower stratosphere from MSU/AMSU and 2 radiosonde datasets. The HadRT sonde dataset represents monthly CLIMAT TEMP reports and the LKS sonde dataset is from an 87 station network corrected for temporal inhomogeneities. The bottom curve gives the average trend for all products and the individual product curves give deviations from the average (from Seidel et. al., 2003).

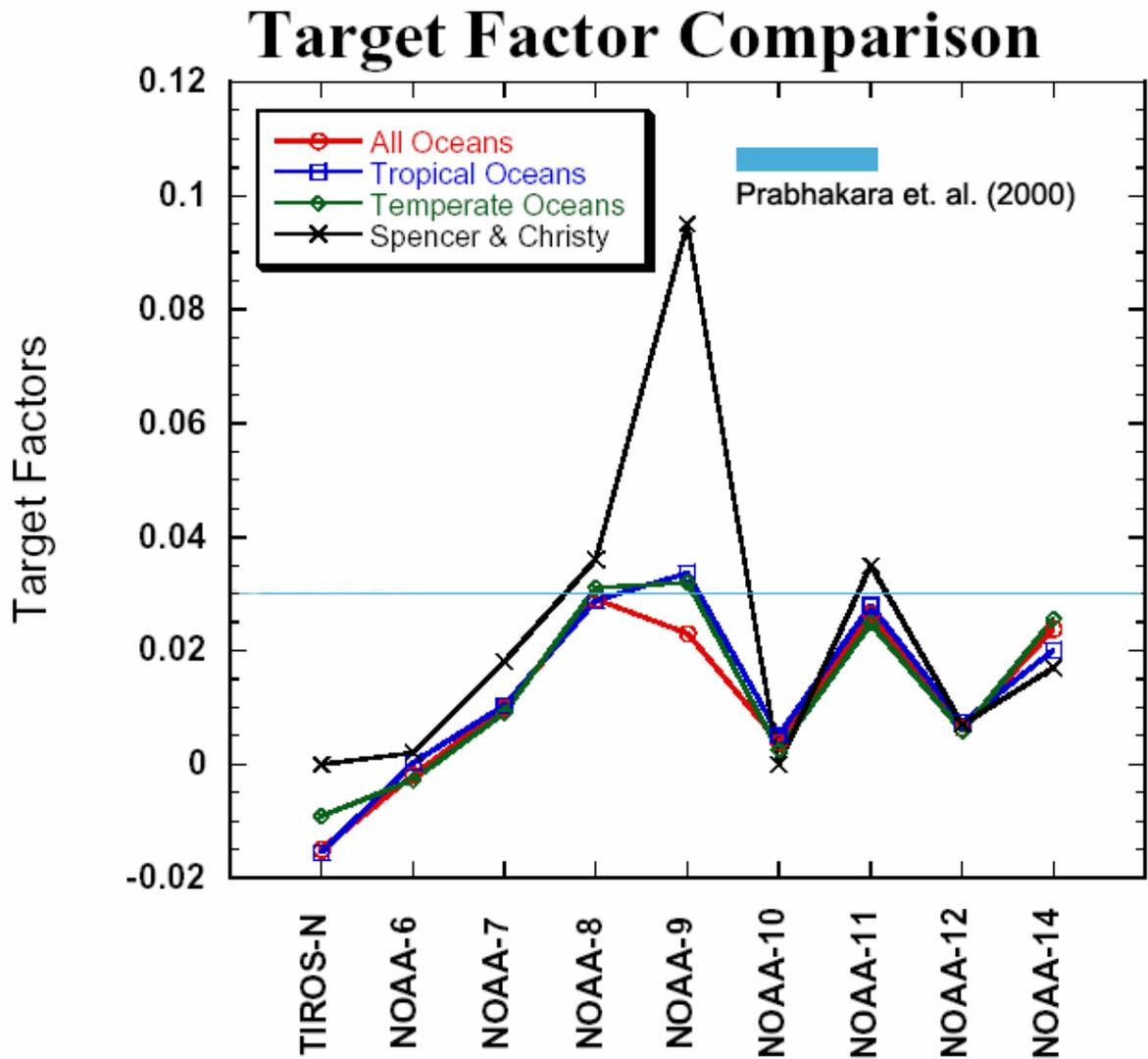


**Figure 22:** Multidataset-average monthly anomaly time series for 6 vertical layers compared with time series for the Quasi-Biennial oscillation (QBO) as determined by 50-hPa altitude zonal wind patterns from radiosonde data at Singapore, and the Southern Oscillation Index (SOI) as determined by Trenberth (1984). The datasets shown are global averages of data from LKS, HadRT, RIHMI, Angell 63, Angell 54, and UAH Vers. D and 5.0. All are global average time series except for the 300-100 hPa (tropopause) time series which is for the Tropics only. Taken from Seidel et. al., 2003.

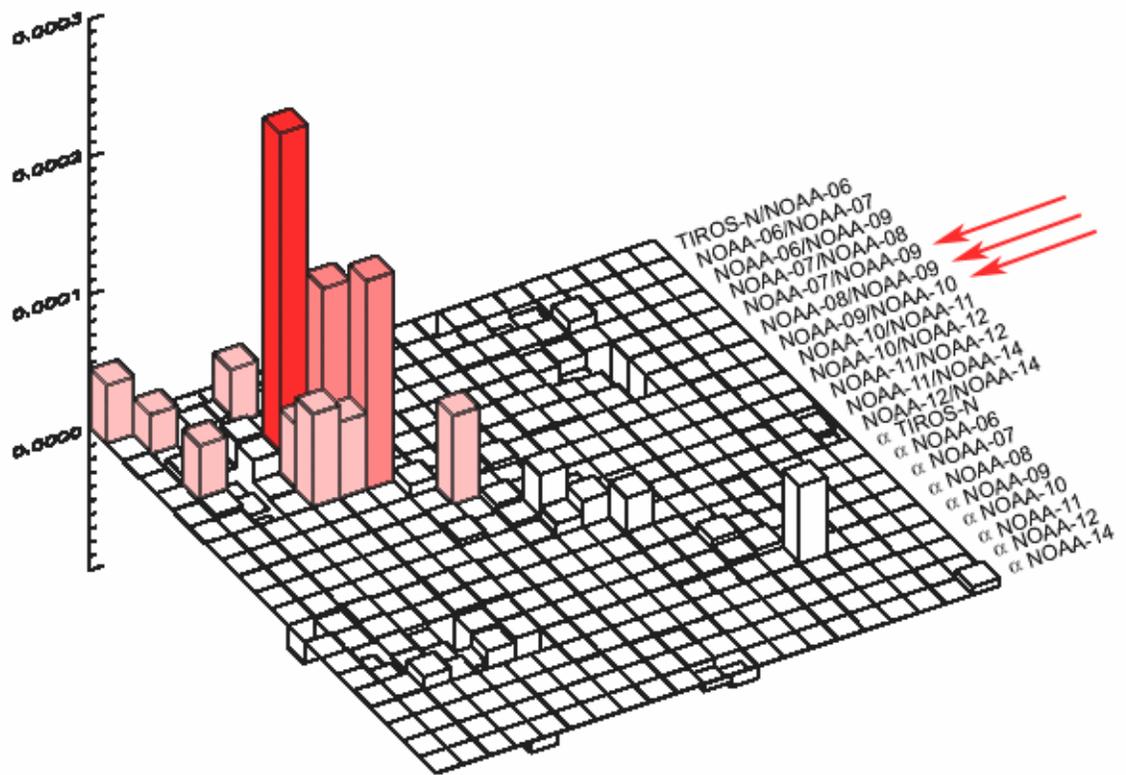
## MSU Channel 2 Time Series Comparison RSS vs. UAH



**Figure 23:** Comparison of global middle troposphere time series from MSU Channel 2 as determined by RSS and UAH from 1979 to October of 2003. The upper curves are based on an “Ocean Only” merge to characterize hot target calibration coefficients and co-orbiting satellite offsets. The lower curves are based on a “Land and Ocean” merge. The black curves show the difference between RSS and UAH for the merge in question. Taken from Mears et. al. (2003b).



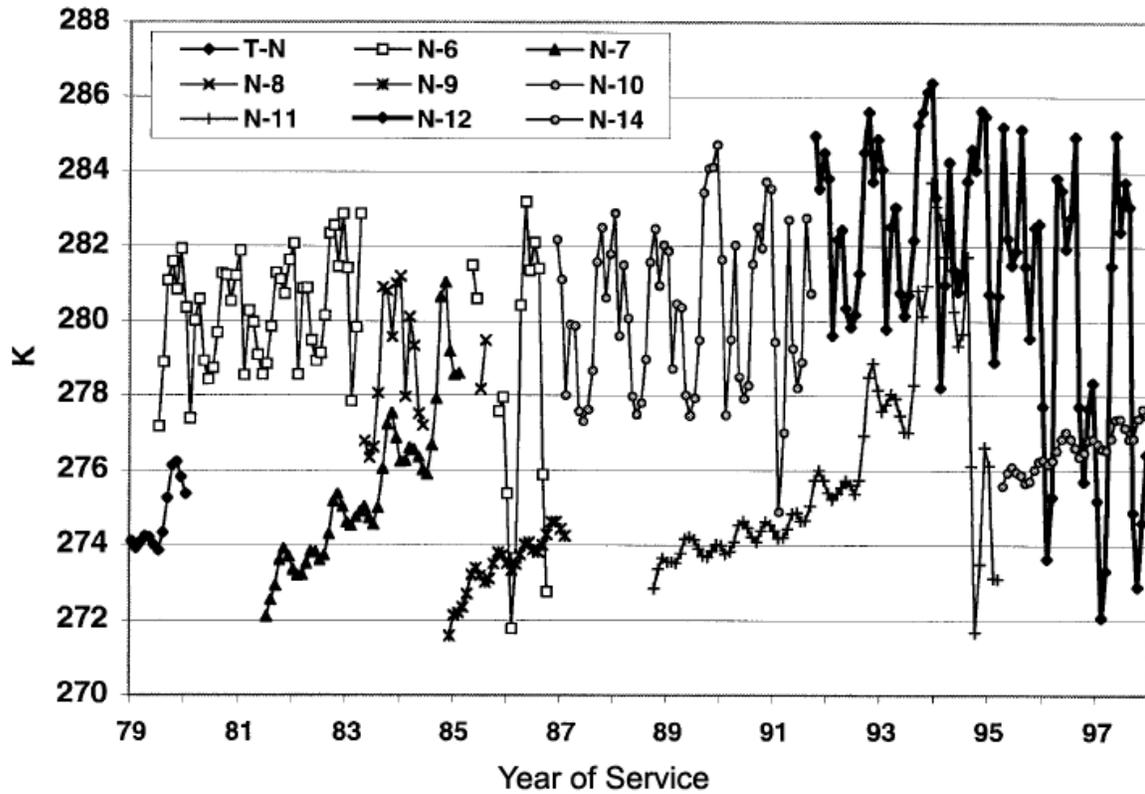
**Figure 24:** Comparison of Hot Target Calibration Factors ( $\alpha_i$ ) as determined by RSS and UAH merge calculations. The constant value of 0.03 used by Prabhakara et. al (2000) to collectively characterize all POES satellite hot target variations and all sources of linear and non-linear error (except diurnal drift and orbital decay) is shown for comparison. Adapted from Mears et. al. (2003b).



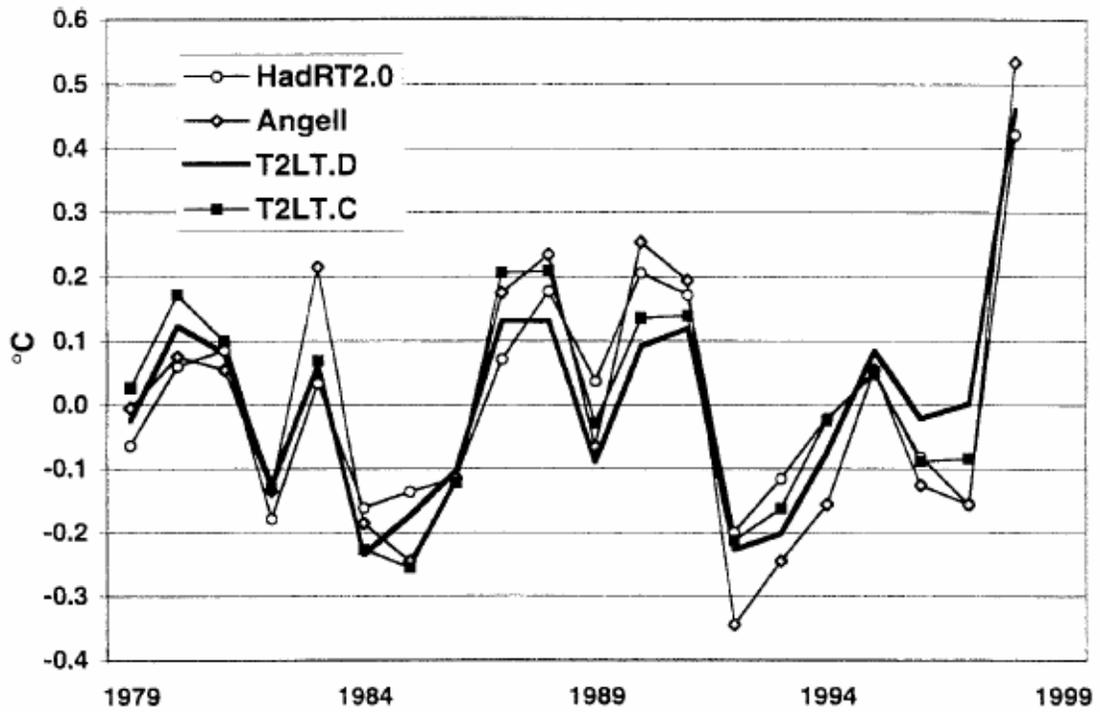
**Figure 25:** Results of a Monte-Carlo analysis of the covariance matrix of standard deviations of the residuals from the RSS Ocean-Only merge calculation. The covariance matrix was derived by superposing statistical noise on the Ocean-Only merge to create a set of 30,000 “noisy” merge calculations and groups of offsets prior to Monte-Carlo analysis. Results have been adjusted to account for the reduction in degrees of freedom among the residuals caused by a significant lag-1 autocorrelation (0.40). Taken from Mears et. al. (2003).

	Analysis Method	Trend (K/Decade)	RMS of Resid.(K)
Ocean-only	Diurnally corrected	0.098 +/- 0.028	0.033
	No diurnal correction	0.091 +/- 0.030	0.034
	Spencer and Christy, Vers. 5.0	-0.011	
Land-only	Diurnally corrected	0.087 +/- 0.046	0.064
	No diurnal correction	0.023 +/- 0.046	0.077
	Spencer and Christy, Vers. 5.0	0.050	
Global	Diurnally corrected	0.097 +/- 0.028	0.031
	No diurnal correction	0.067 +/- 0.028	0.040
	Spencer and Christy, Vers. 5.0	0.009	
	Prabhakara et al	0.13	

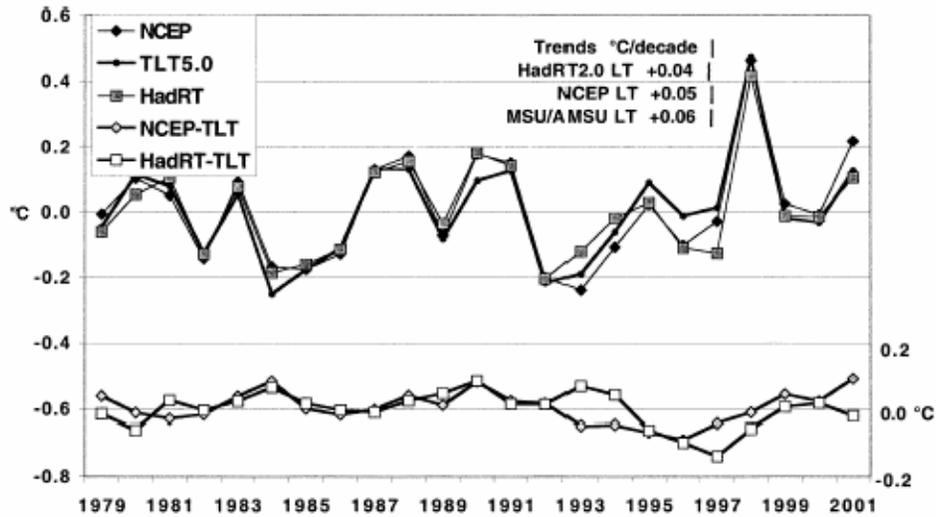
**Figure 26:** 1979-2001 global Channel 2 temperature trends in deg. K/decade for RSS Ver. 1.0 and UAH Ver. 5.0 as derived with and without diurnal corrections for Ocean-Only, Land-Only, and Global merge calculations. The corresponding global trend of Prabhakara et. al. (2000) is shown for comparison. From Mears et. al., 2003.



**Figure 27:** Monthly average MSU Channel 2 Hot Calibration Target temperature as measured by the two platinum resistance thermocouples (PRTs) monitoring it for each of the individual MSU's during the satellite era. This represents the temperature of the instrument itself. Variations in this temperature can be used to estimate the magnitude of IBE impacts on MSU2 trends. From Christy et al., 2000.



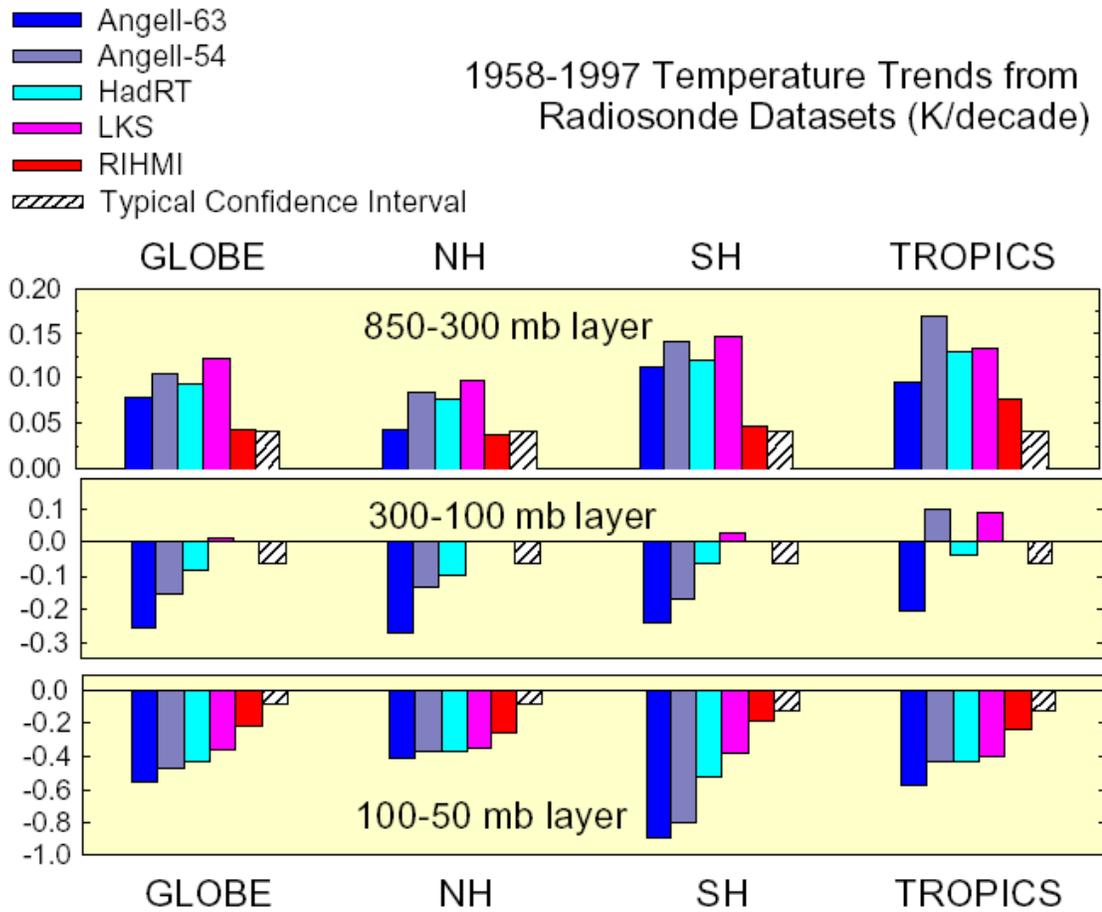
**Figure 28:** Time series of annual global lower troposphere temperature anomalies for 2 UAH MSU2LT products and two tropospheric radiosonde products weighted with a 2LT weighting function. The Angell curve represents the Angell 63 network (Angell, 1988). From Christy et al., 2000.



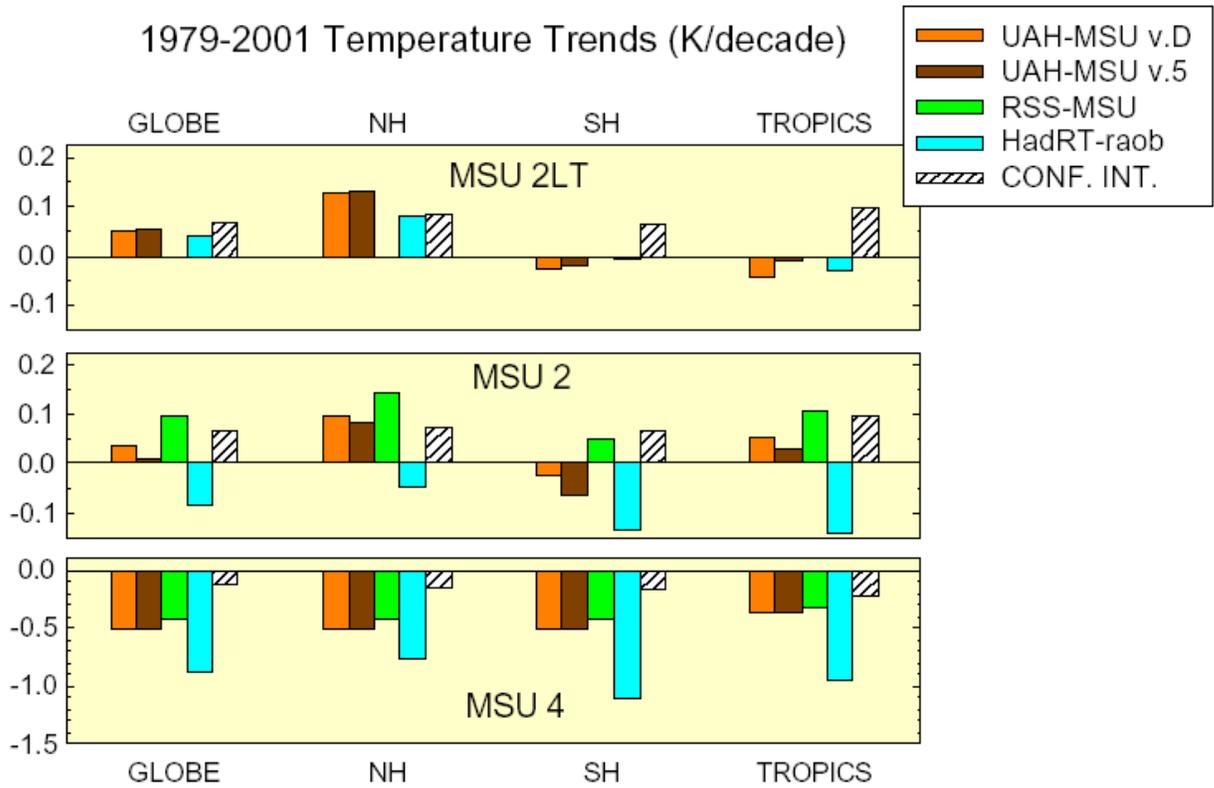
**Figure 29:** Time series of annual global lower troposphere temperature anomalies for 2 UAH MSU2LT products and two tropospheric radiosonde products weighted with a 2LT weighting function. NCEP is a reanalysis product that uses a global circulation model, in situ data, and satellite data other than MSU, but with NOAA sounding profiles (Kalnay, 1996). HadRT data are from the HadRT2.1 product which has been corrected for anomalous temperature discontinuities using MSU data. HadRT2.1 is not, strictly speaking, independent of UAH Ver. D, and NCEP is not entirely independent of radiosonde data. From Christy et. al., 2003.

	TLT °C decade <sup>-1</sup>	TMT °C decade <sup>-1</sup>	TLS °C decade <sup>-1</sup>
Global 95% CI trend			
Minqin	±0.031	±0.015	±0.033
U.S. VIZ sondes, composite $\sigma_{\Delta ANN}$	±0.016	±0.039	±0.113
U.S. VIZ sondes, rms $\Delta b$	±0.043	±0.039	
HadRT	±0.075		±0.127
NCEP	±0.067		?
Rms consensus	±0.051	±0.033	±0.100
95% CI monthly global anomalies			
	°C	°C	°C
Minqin	±0.24	±0.12	±0.25
U.S. VIZ radiosondes	±0.20	±0.15	±0.35
NCEP	±0.16		±0.41
Rms consensus	±0.20	±0.14	±0.35
95% CI annual global anomalies			
	°C	°C	°C
Minqin	±0.08	±0.04	±0.09
U.S. VIZ radiosondes	±0.04	±0.10	±0.27
HadRT, NCEP	±0.20, ±0.18		±0.38, ?
Rms consensus	±0.14	±0.08	±0.27

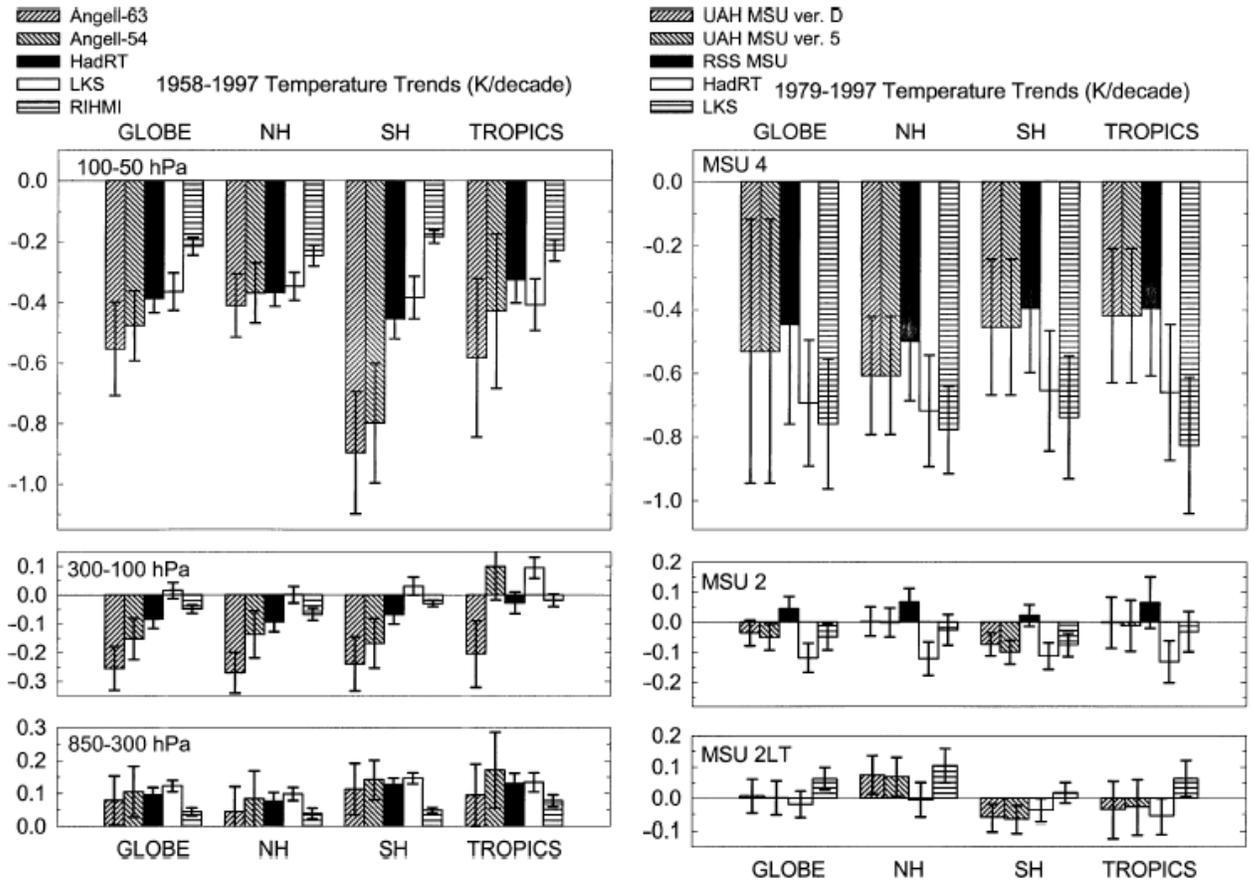
**Figure 30:** Summary of 95 percent confidence interval estimates for calculations of global troposphere temperature statistics for UAH Ver. 5.0 based on UAH analysis of the Minqin radiosonde station in China, UAH selected U.S. radiosonde stations, the NCEP reanalysis product, and HadRT2.1. TLT corresponds to the lower troposphere, TMT the middle troposphere, and TLS the lower stratosphere. From Christy et. al., 2003.



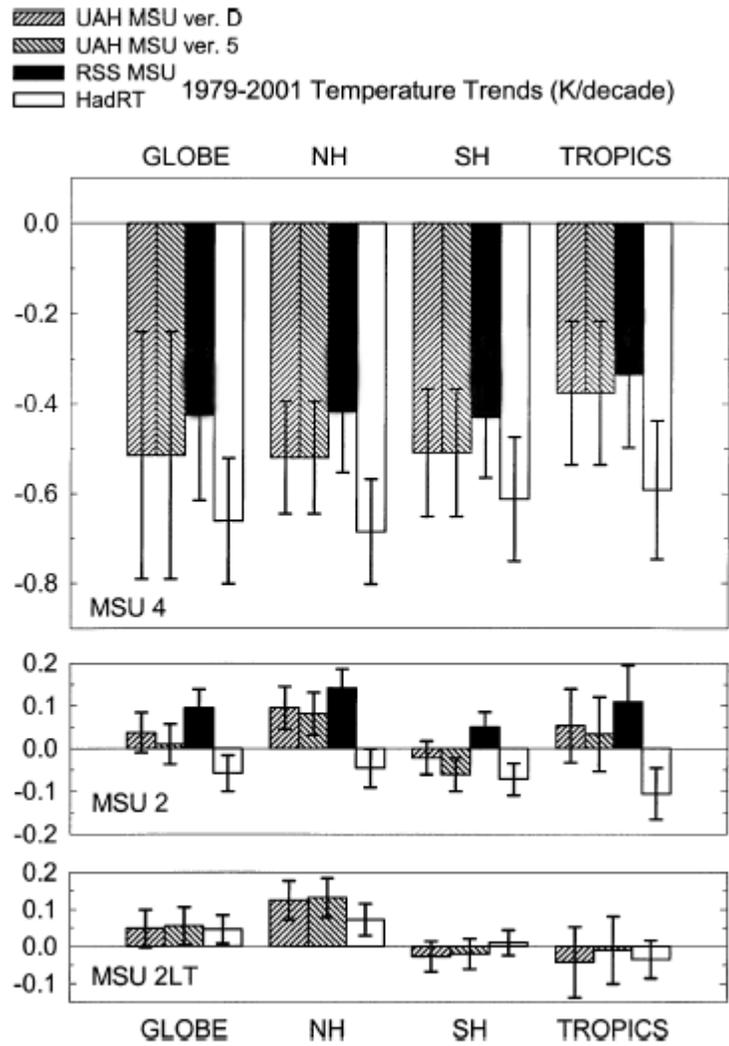
**Figure 31:** Trends in global temperature for 1958-1997 for troposphere (top), tropopause (middle), and lower stratosphere (bottom), in four regions, from 5 radiosonde datasets. The confidence intervals shown are typical values of the  $\pm 2$  sigma uncertainty estimates. Imagining placing the midpoint of these confidence intervals at the value of each trend, and determining if there is overlap, will give a sense of whether there are statistically significant differences within groups of trend estimates. From Seidel et. al., 2003.



**Figure 32:** Trends in global temperature for 1970-2001 for the lower troposphere (MSU2LT), middle troposphere (MSU2), and the lower stratosphere (MSU4), in four regions, from 3 MSU datasets and one radiosonde dataset. The confidence intervals shown are typical values of the  $\pm 2$  sigma uncertainty estimates. Imagining placing the midpoint of these confidence intervals at the value of each trend, and determining if there is overlap, will give a sense of whether there are statistically significant differences within groups of trend estimates. From Seidel et. al., 2003.

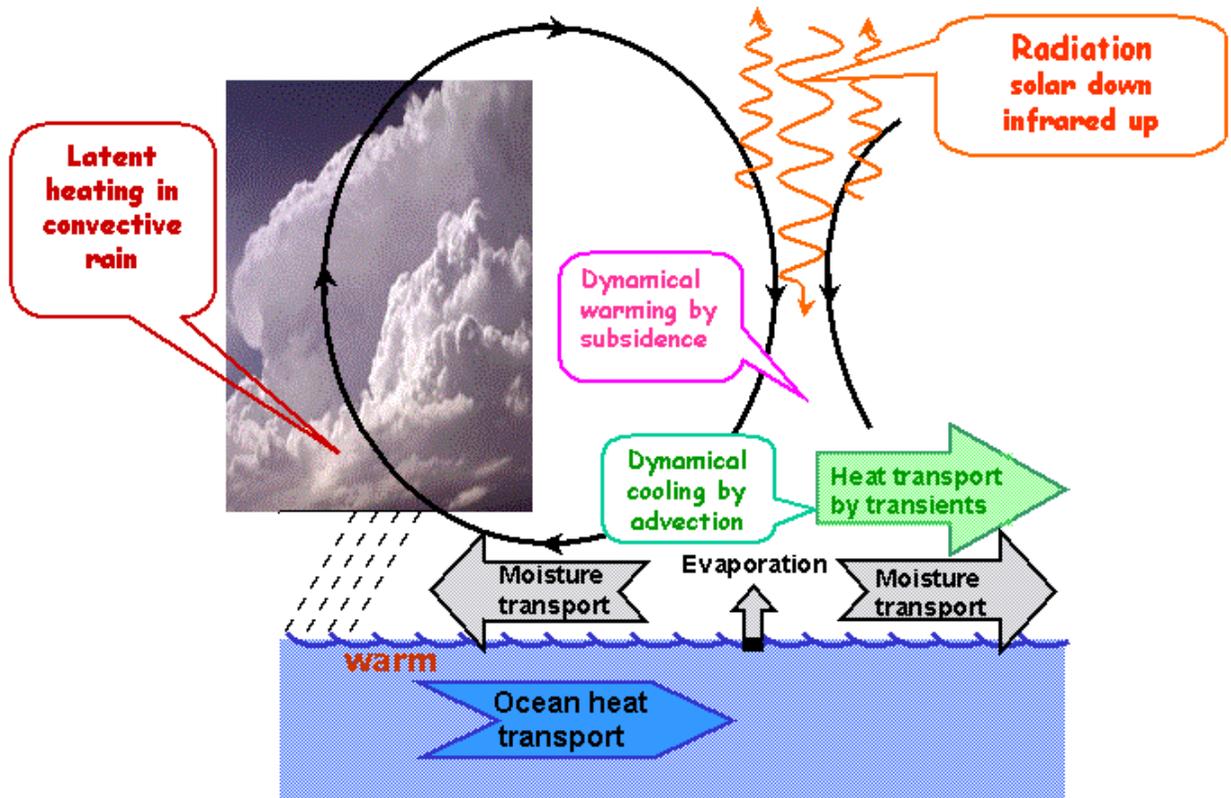


**Figure 33:** Trends (deg. K/decade) in global temperature for 1958–97 for three atmospheric layers (top) 100–50 hPa (top), 300–100 hPa (middle), and 850–300 hPa (bottom), in four regions, from radiosonde datasets (left side), and for 1979–97 for three layers (top) MSU4, (middle) MSU2, (bottom) MSU2LT, in four regions, from MSU/AMSU and radiosonde datasets. Confidence intervals shown are +/- one Standard Error estimates. HadRT data are for the HadRT2.1 release. From Seidel et al. (2004).

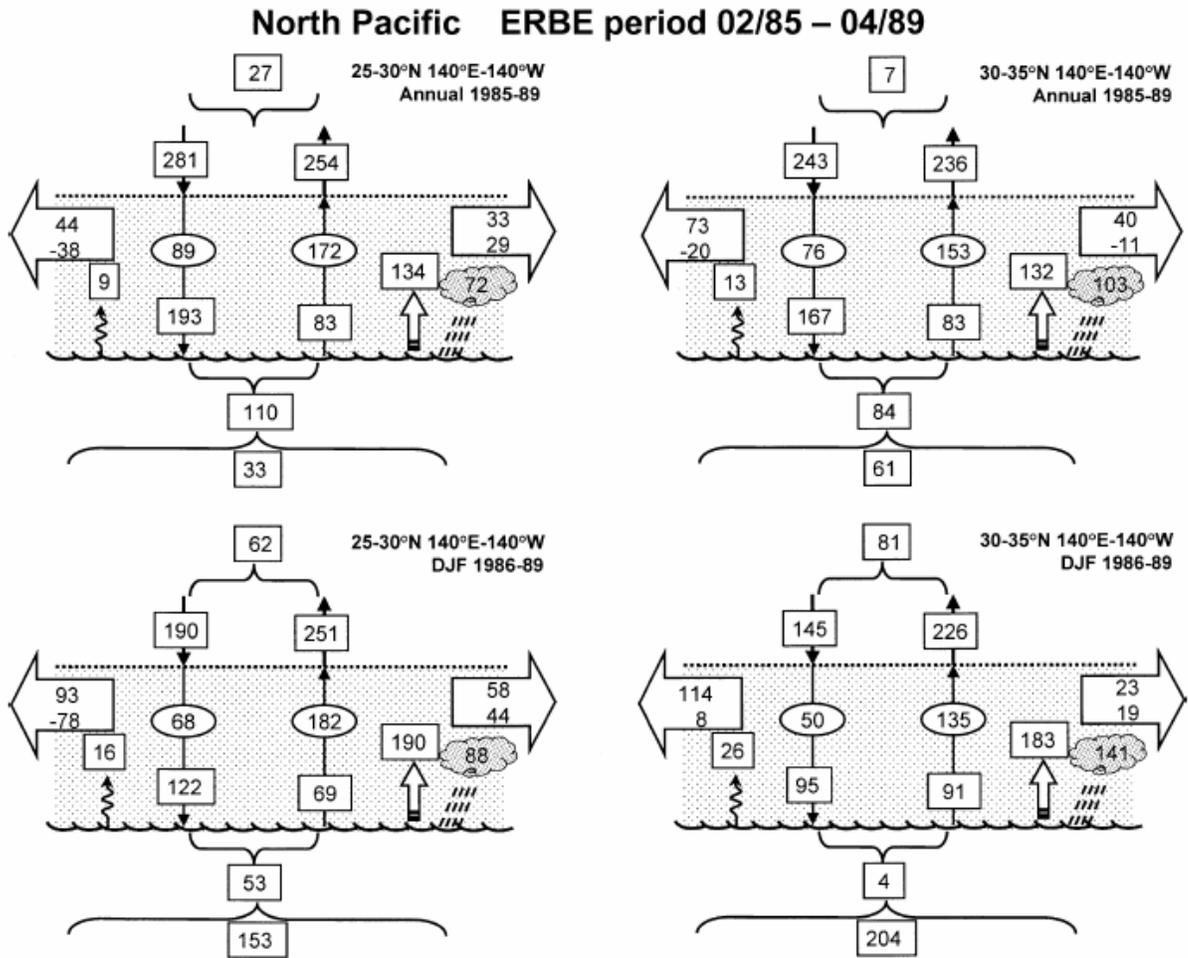


**Figure 34:** Temperature trends for 1979–2001 for three vertical layers MSU4 (top), MSU2 (middle), and MSU2LT (bottom), in four regions, from MSU/AMSU and radiosonde datasets. Confidence intervals shown are +/- one Standard Error estimates. HadRT data is for the HadRT2.1 Version. From Seidel et al. (2004).

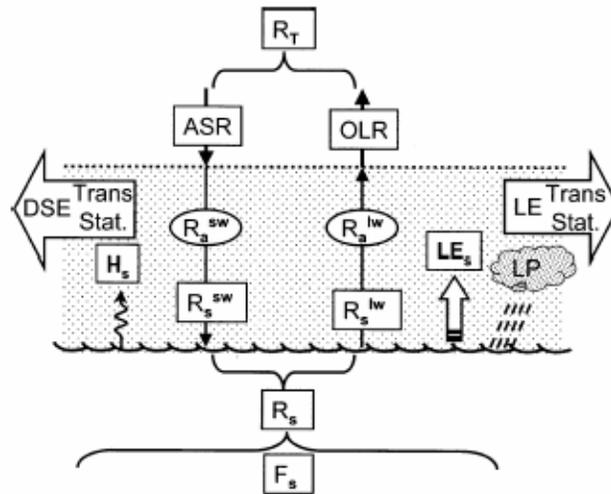
## Hadley circulation and heat budget in subtropics



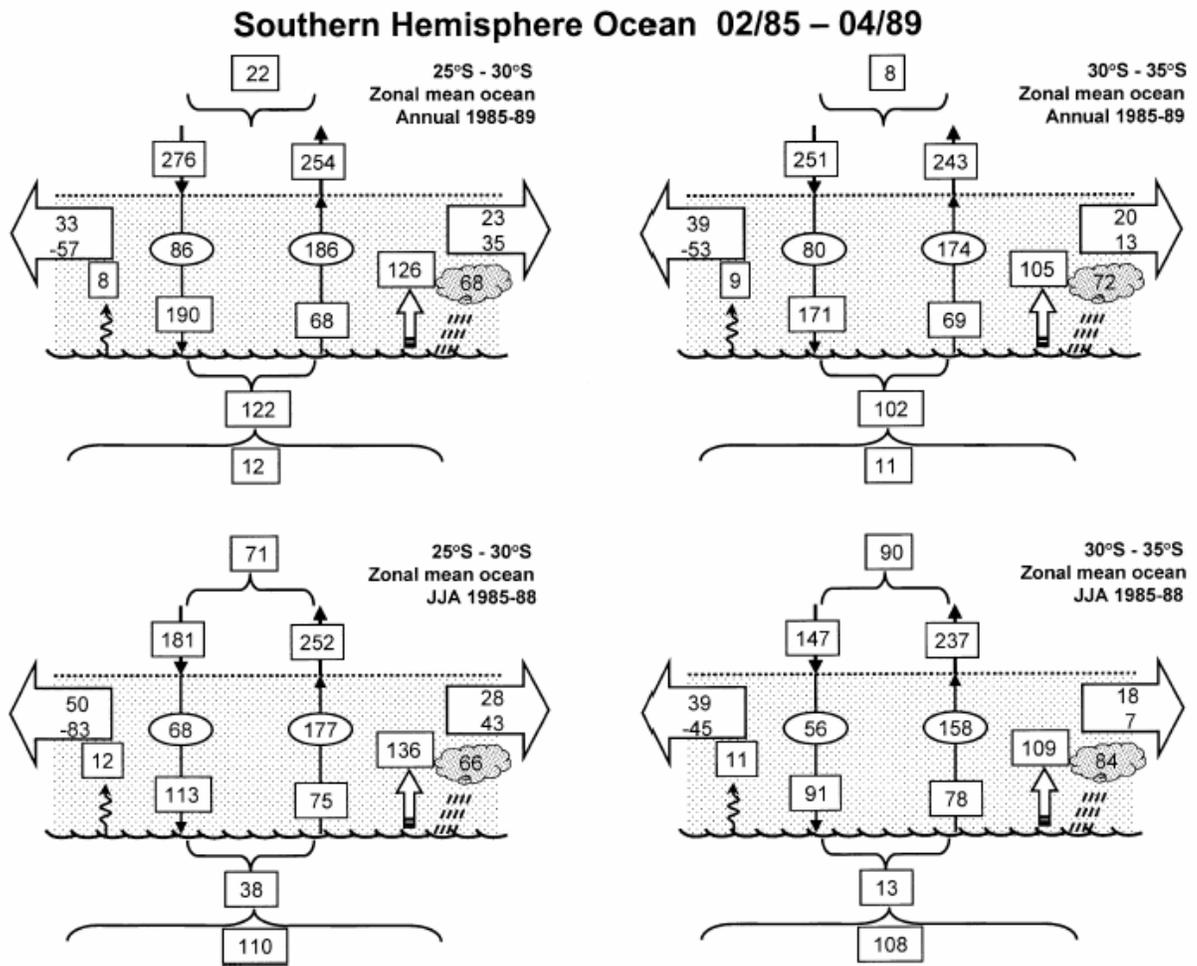
**Figure 35:** Schematic of the Hadley Cell circulation in the Subtropics, and the total heat budget associated with it. The various key processes and heat transports associated with a potential decoupling of the surface and troposphere are shown. The view is zonal (along lines of latitude) with the equatorial direction to the left and the poleward direction to the right. From Trenberth and Stepaniak, 2003b.



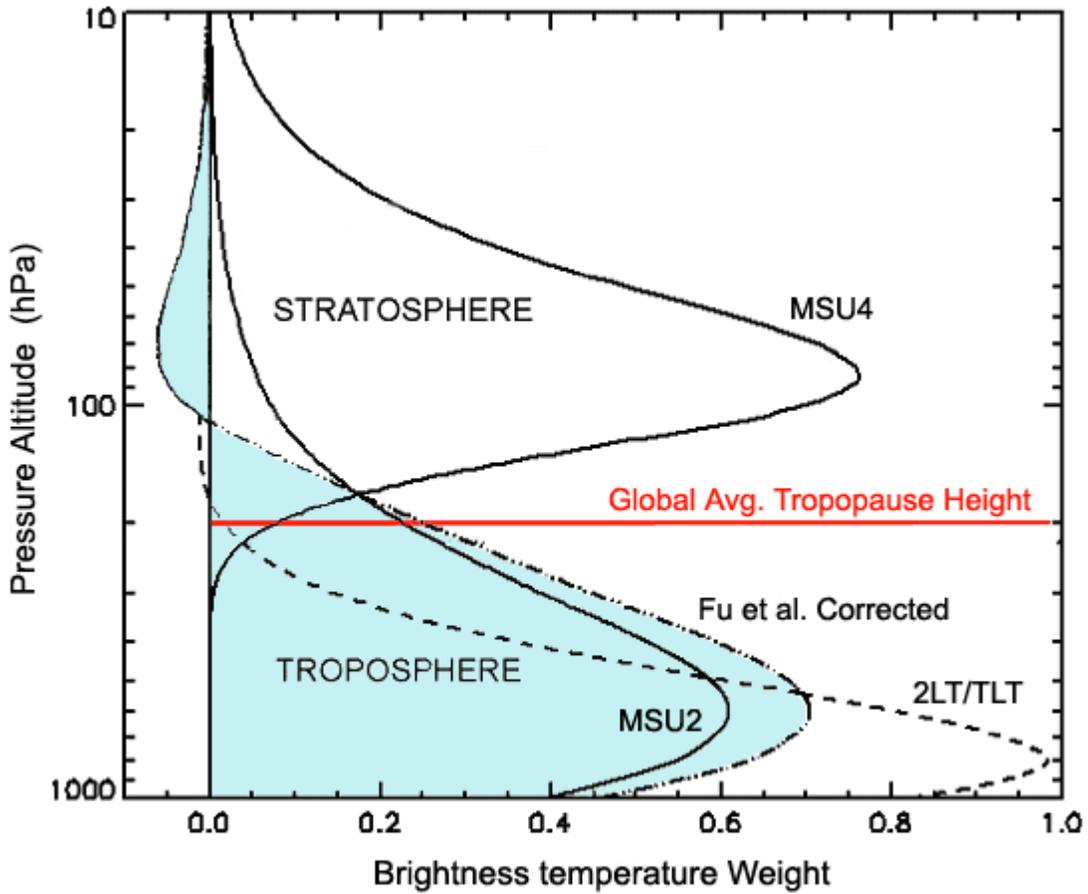
**Figure 36a:** North Pacific energy flow diagrams showing average annual magnitudes of incoming, outgoing, and poleward transports as described by Trenberth and Stepaniak (2003). The poleward components of this budget (horizontal arrows and energy flows) are thought to be responsible for the decoupling of the troposphere and surface temperature trends in the tropics and subtropics. Energy flows are derived from Earth Radiation Budget Experiment (ERBE) data, Southampton Oceanographic Centre (SOC) heat budget atlas data for the marine surface (Josey et al. 1998, 1999), and the Climate Prediction Center (CPC) Merged Analysis of Precipitation (CMAP) precipitation estimates from Xie and Arkin (1997). Sectors shown are for 140°E–140°W for (left) 25°–30°N and (right) 30°–35°N, (top) annual and (bottom) DJF for the ERBE period Feb 1985–Apr 1989 in  $W/m^2$ . From Trenberth and Stepaniak, 2003b.



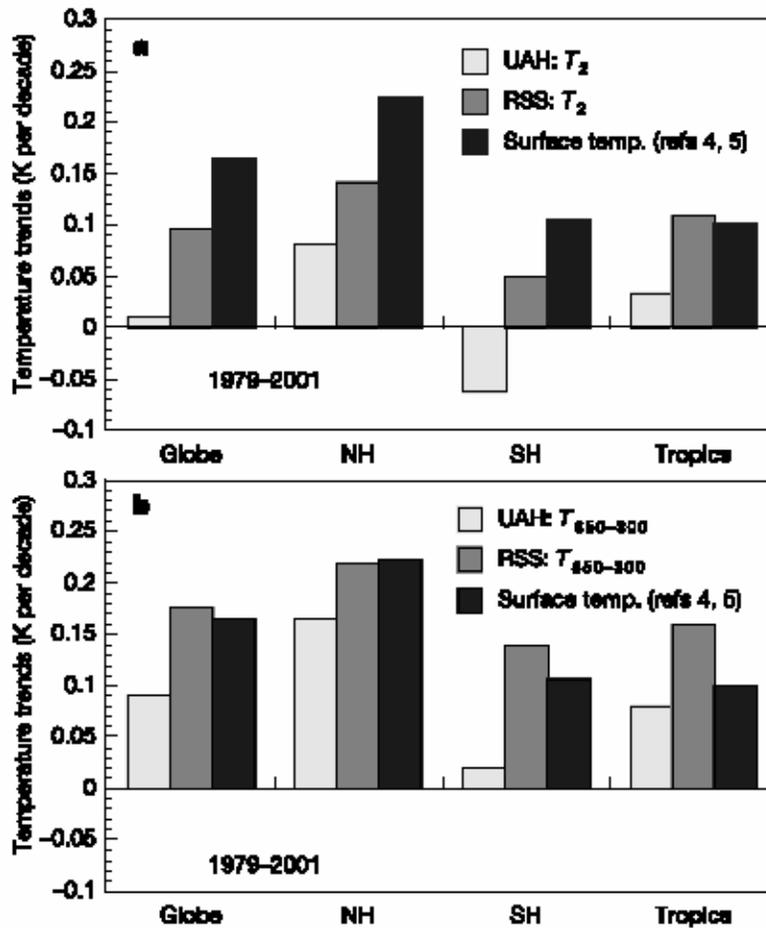
**Figure 36b:** Schematic layout of the energy flows given in Figure 34a and 35 as they are defined in Trenberth and Stepaniak (2003). Energy flows are shown in zones through the TOA (dotted line, subscript “T”), surface (scalloped line, subscript “s”), and atmosphere (stippled, subscript “a”). Arrows or brackets indicate direction of flows. Here  $R$  indicates radiation flows with amounts in square boxes and amount deposited in ovals for shortwave (superscript “sw”) or longwave (superscript “lw”). The large horizontal arrows indicate divergence of energy by the atmosphere for the (left) DSE and (right) LE broken down into transient and quasistationary components. All units are in terms of energy, hence latent heating from evaporation is given here by  $LE_s$  (to distinguish it from latent energy LE) and precipitation by  $LP$ , while the sensible heat is  $H_s$  and the net surface flux is  $F_s$ . Taken from Trenberth and Stepaniak, 2003b.



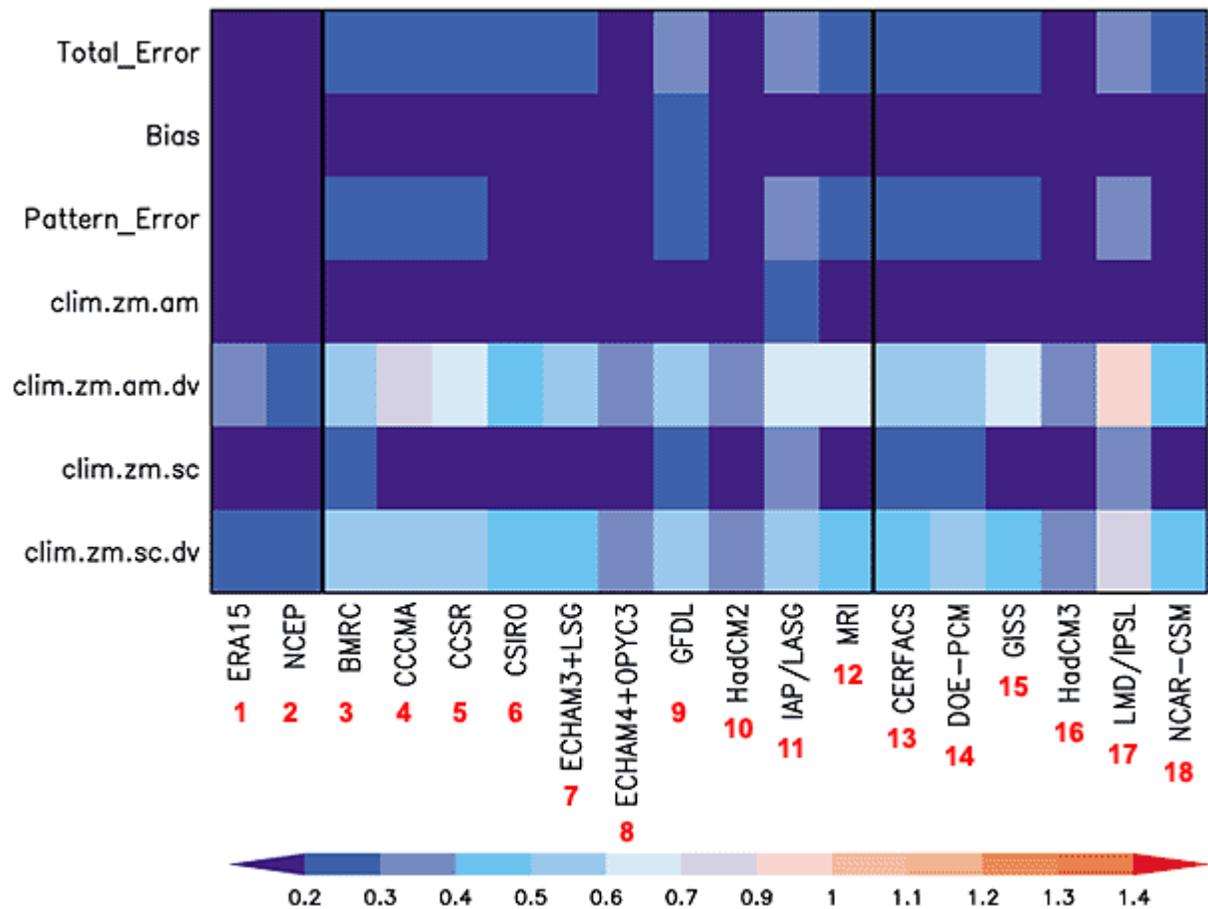
**Figure 37:** South Pacific energy flow diagrams showing zonal mean annual magnitudes of incoming, outgoing, and poleward transports as described by Trenberth and Stepaniak (2003). The poleward components of this budget (horizontal arrows and energy flows) are thought to be responsible for the decoupling of the troposphere and surface temperature trends in the tropics and subtropics. Energy flows are derived from Earth Radiation Budget Experiment (ERBE) data, Southampton Oceanographic Centre (SOC) heat budget atlas data for the marine surface (Josey et al. 1998, 1999), and the Climate Prediction Center (CPC) Merged Analysis of Precipitation (CMAP) precipitation estimates from Xie and Arkin (1997). Sectors shown are for (left) 258–308S and (right) 308–358S (top), annual and (bottom) JJA for the ERBE period Feb. 1985–Apr 1989 in  $W/m^2$ . From Trenberth and Stepaniak, 2003b.



**Figure 38:** The corrected MSU Channel 2 weighting function derived by Fu et al. (2004) compared with the uncorrected MSU2, MSU4, and 2LT/TLT channels (Christy et al., 2003; Mears et al., 2003). Whereas the actual Channel 2, 4, and TLT functions are everywhere positive, as required for real weighting, the Fu et al. function goes negative above 100 hPa to remove stratospheric effects from the uncorrected MSU2 channel. Global average tropopause height is shown for comparison.

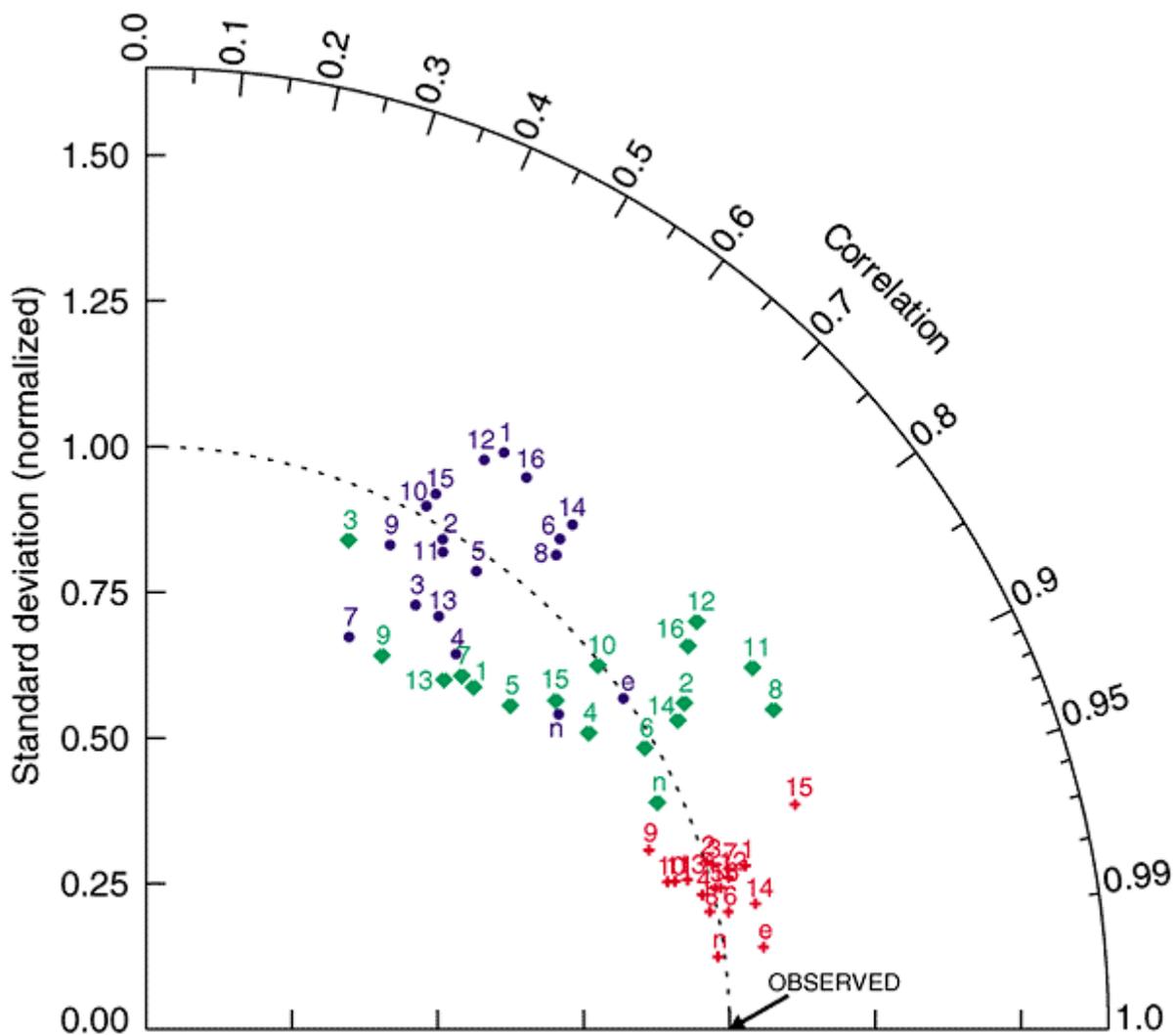


**Figure 39:** Trends in monthly mean troposphere temperature anomalies for MSU channel 2 without correction for stratospheric influence (top), and for the MSU-derived 850–300-hPa layer with correction (bottom). Trends are given for the globe, Northern Hemisphere (NH), Southern Hemisphere (SH) and tropics (308 N–308 S). Uncorrected UAH values are from Version 5.0 (Christy et. al., 2003) and uncorrected RSS values are from Version 1.0 (Mears et. al., 2003). Surface temperature trends for the same regions are also shown for comparison. From Fu et. al., 2004.



### Surface Air Temperature compared against Jones et al. (1999)

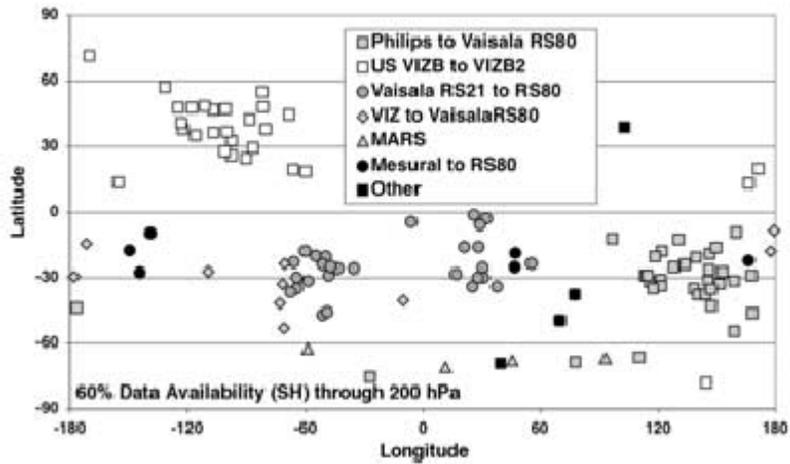
**Figure 40:** Components of space-time errors of surface air temperature (climatological annual cycle) simulated by Coupled Model Intercomparison Phase 2 CMIP2 model control runs. Shown are the total errors, the global and annual mean error (“bias”), the total rms (“pattern”) error, and the following components of the climatological rms error: zonal and annual mean (“clim.zm.am”); annual mean deviations from the zonal mean (“clim.zm.am.dv”), seasonal cycle of the zonal mean (“clim.zm.sc”); and seasonal cycle of deviations from the zonal mean (“clim.zm.sc.dv”). For each component, errors are normalised by the component’s observed standard deviation. The two left-most columns represent alternate observationally based data sets, ECMWF and NCAR/NCEP reanalyses, compared with the baseline observations (Jones et al., 1999). Remaining columns give model results: the ten models to the left of the second thick vertical line are flux adjusted and the six models to the right are not. From Covey et al. (2000) and the IPCC (2001).



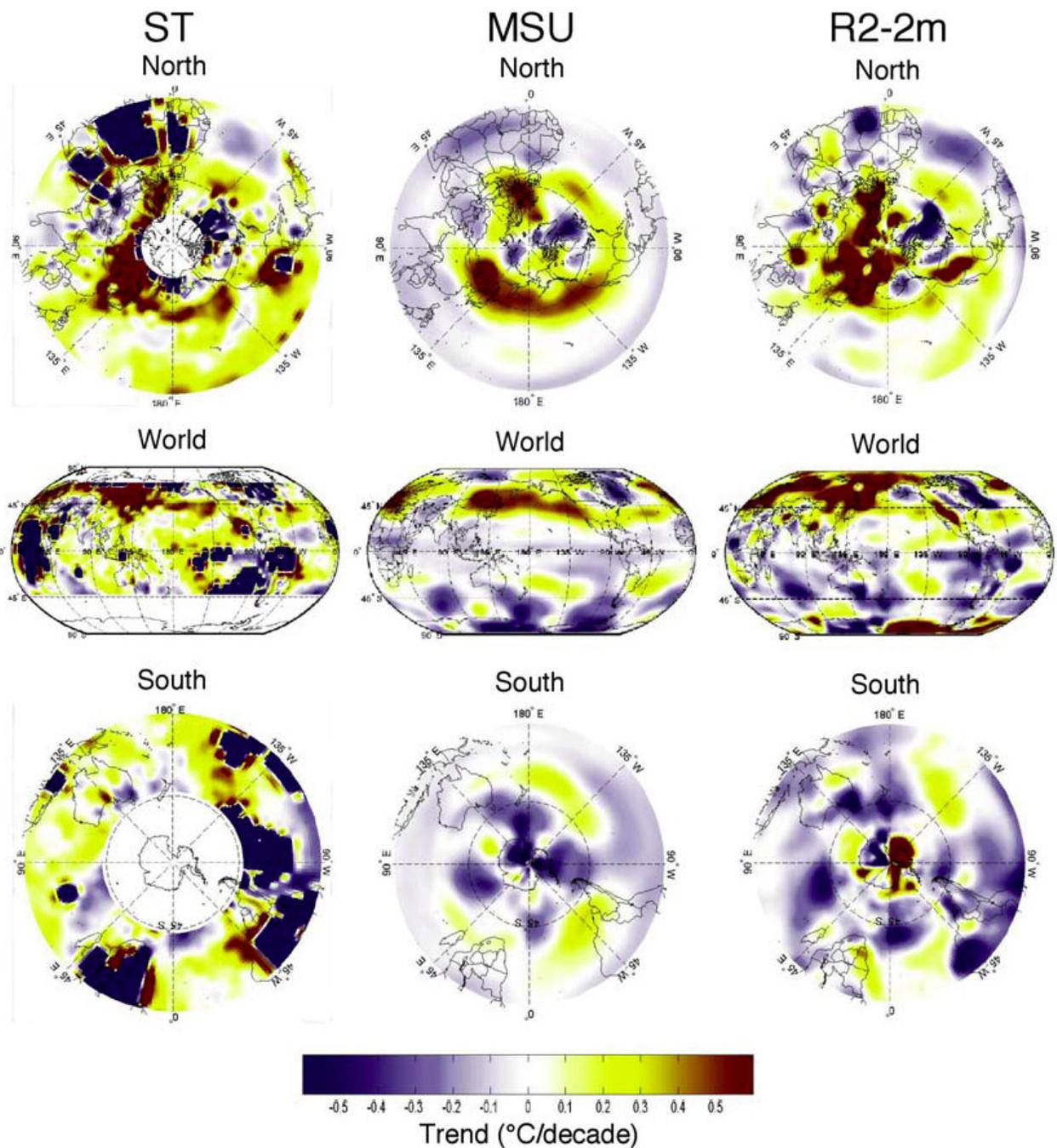
- Purple:** Precipitation compared against Xie and Arkin (1996).
- Green:** Sea Level Pressure compared against ERA-15.
- Red:** Surface Air Temperature compared against Jones et al. (1999).

**Figure 41:** Second-order statistics of surface air temperature, sea level pressure and precipitation simulated by the Coupled Model Intercomparison Phase 2 CMIP2 model control runs (Meehl et al. 2000). The radial co-ordinate gives the magnitude of total standard deviation, normalized by the observed value, and the angular co-ordinate gives the correlation with observations. It follows that the distance between the OBSERVED point and any model's point is proportional to the rms model error. Numbers indicate models counting from left to right in Figure 38. Letters indicate alternate observationally based data sets compared with the baseline observations: e = 15-year ECMWF reanalysis ("ERA"); n = NCAR/NCEP reanalysis. From Covey et al. (2000) and the IPCC (2001).

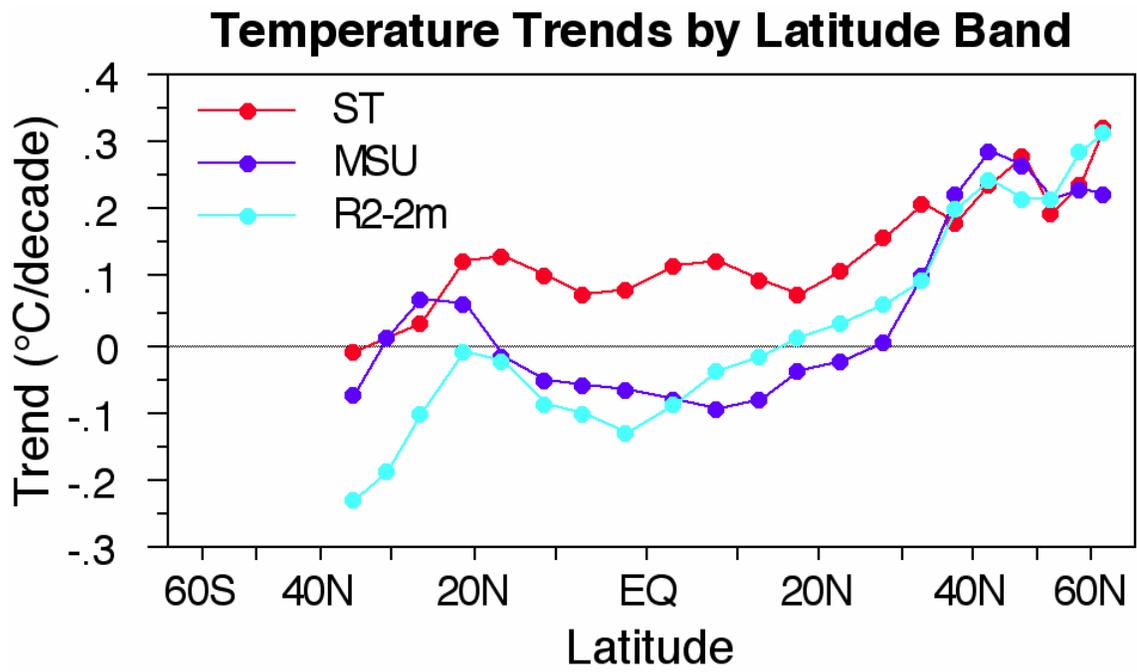
### Radiosonde Stations of Christy & Norris, 2004



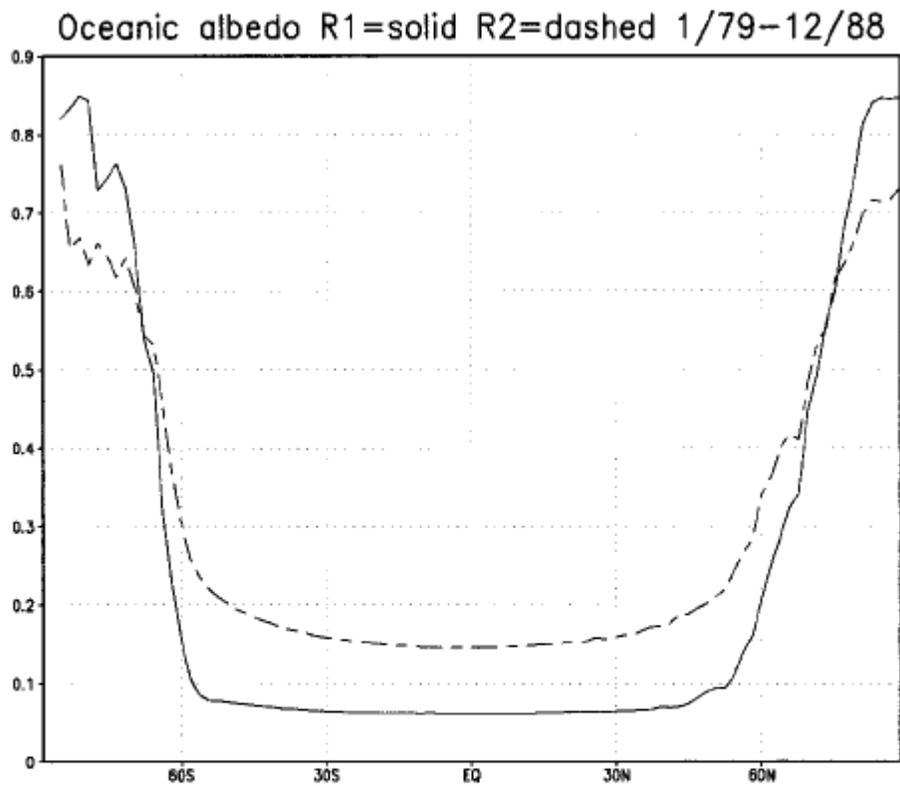
**Figure 42:** The 89 station radiosonde network used by Christy and Norris for their Year 2004 MSUTLT-radiosonde intercomparison study. Stations are shown by their global latitude and longitude location, though continental land masses are not shown for comparison. Stations are delineated by instrument type and recorded instrument changes experienced during the period of record (Jan. 1979 – July 2001). From Christy and Norris, 2004.



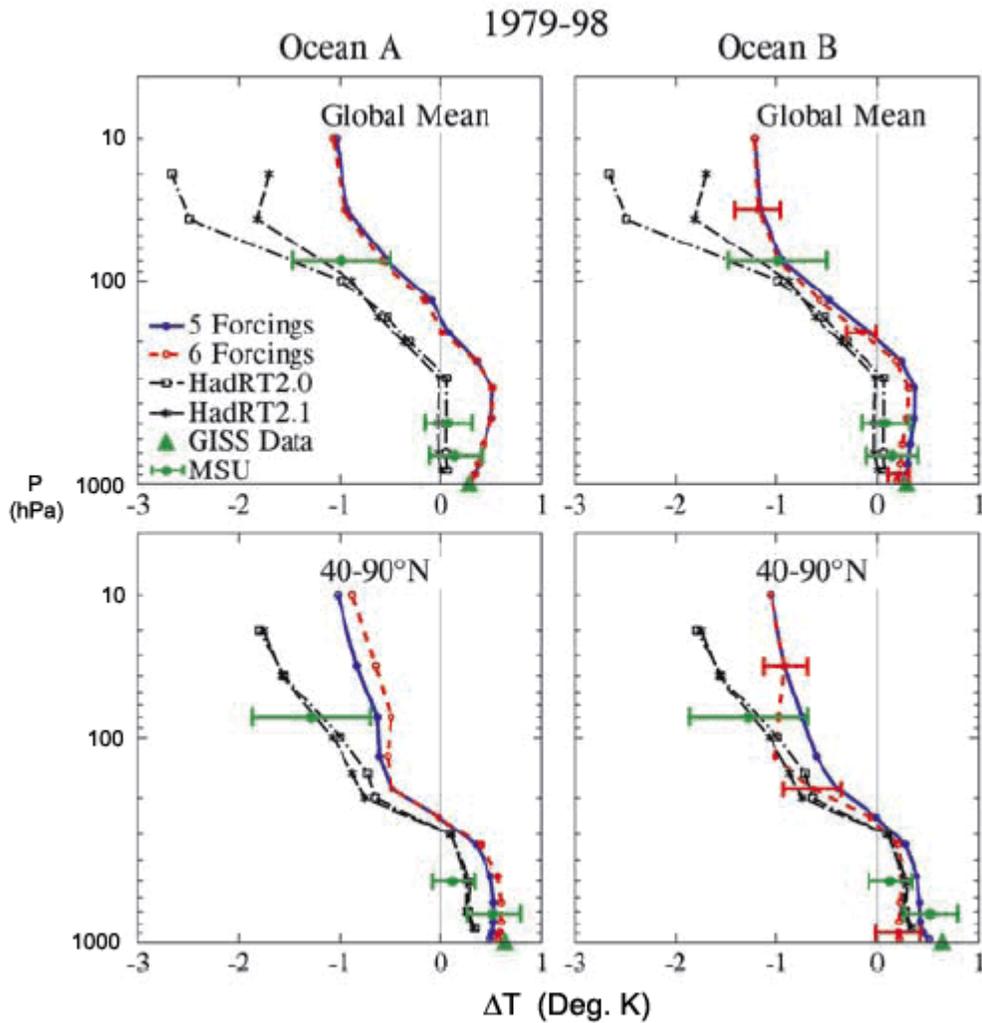
**Figure 43:** Trend-line maps of Surface Temperature, UAH Ver. D MSU 2LT, and R2-2m for 1979-1996 viewed from North Pole, Full World, and South Pole Projections as reported in Douglass et al. (2004). Note that apart from polar regions (which are shown as colorless circles) cells where Surface Temperature data are missing are made dark blue – and are therefore indistinguishable from their cells that show strong regional cooling. Taken from Douglass et al. (2004).



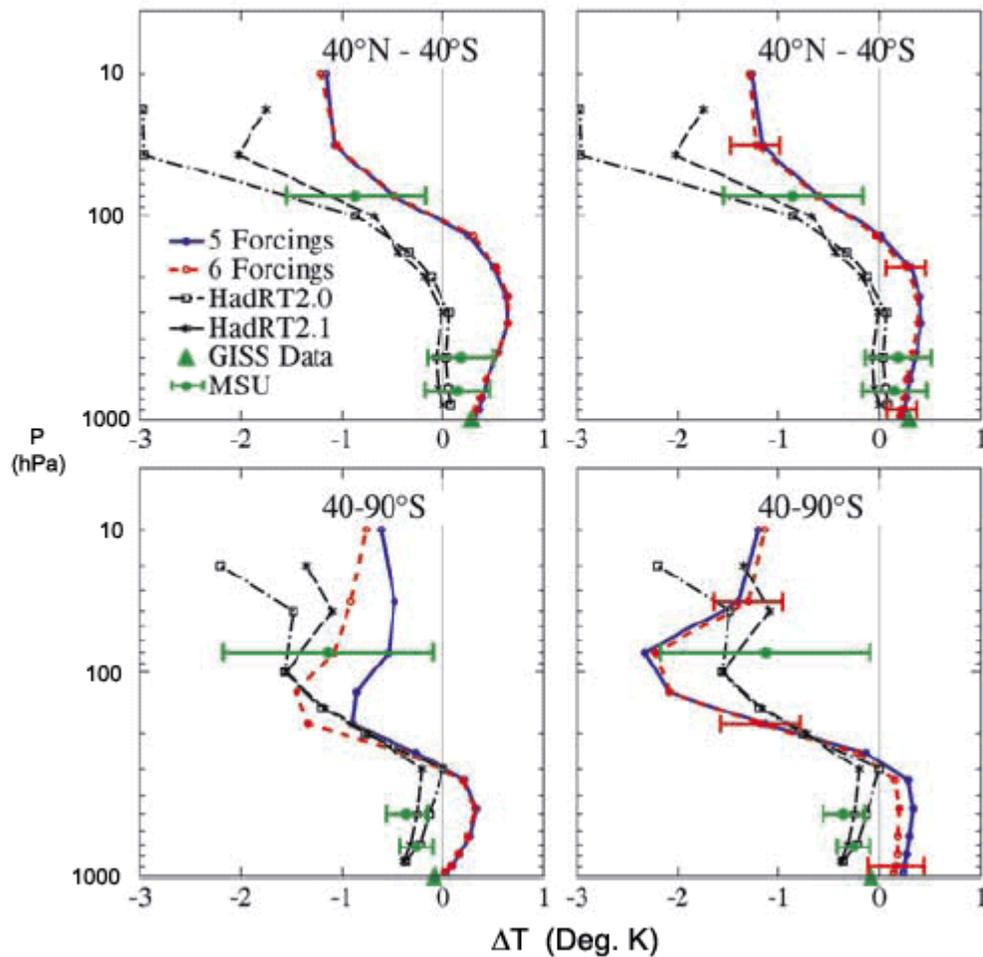
**Figure 44:** Zonally averaged temperature trends for the period 1979-1996 from the Surface Record, MSU2LT, and the NCEP/NCAR 2-Meter Reanalysis as determined by Douglass et al. (2004) and plotted as a function of latitude. Taken from Douglass et al. (2004).



**Figure 45:** Comparison of 10-yr mean (1979–88) zonally averaged albedo over ocean regions in the original NCEP/NCAR R-1 Reanalysis (dashed - Kalnay et al., 1996) and the R2-2m update (solid – Kanamitsu et al., 2002) shown as fractions of 1.0. Albedos increase significantly beyond 60 deg. N. or S. latitude toward either pole. Taken from Kanamitsu et al. (2002).



**Figure 46A:** Change of annual-mean temperature profile in the GISS SI2000 AOGCM for the globe and Northern Hemisphere over the period 1979–1998 based on linear trends. Model results are for oceans A (left) and B (right), with five and six forcings as applied by Hansen et al. (2002). Surface observations are the land-ocean data of Hansen et al. (1999), with SSTs of Reynolds and Smith (1994) for ocean areas. The bars on the MSU satellite data (Christy et al., 2000) are twice the standard statistical error adjusted for autocorrelation (Santer et al., 2000b). Radiosonde profiles become unreliable above about the 100-hPa level. Twice the ensemble standard deviation is shown at three pressure levels for ocean B with six forcings. Taken from Hansen et al. (2002).



**Figure 46B:** Change of annual-mean temperature profile in the GISS SI2000 AOGCM for the Tropics/Extra-tropics and Southern Hemisphere over the period 1979–1998 based on linear trends. Model results are for oceans A (left) and B (right), with five and six forcings as applied by Hansen et al. (2002). Surface observations are the land-ocean data of Hansen et al. (1999), with SSTs of Reynolds and Smith (1994) for ocean areas. The bars on the MSU satellite data (Christy et al., 2000) are twice the standard statistical error adjusted for autocorrelation (Santer et al., 2000b). Radiosonde profiles become unreliable above about the 100-hPa level. Twice the ensemble standard deviation is shown at three pressure levels for ocean B with six forcings. Taken from Hansen et al. (2002).



**Figure 47:** The network of surface weather stations used by McKittrick and Michaels (2004) in their study of correlations of 1979-2000 surface temperature trends to parameterized climate, economic, and social factors. The stations were selected from GISS surface records (Hansen et al., 1999) and records from the Climate Research Unit, University of East Anglia. Taken from Michaels et al. (2004).

**Surface Temperature Trends from McKittrick & Michaels  
with and without Economic and Social Effects**

<b>Adjustment</b>	<b>Adjusted Sample Average Trend (deg. K/decade)</b>	<b>Adjusted Sample Standard Deviation</b>
Original Sample Average	0.270	0.237
Remove Economic Effects	0.110	0.125
Remove Social Effects	0.063	0.136
Remove Soviet Effects	0.011	0.053

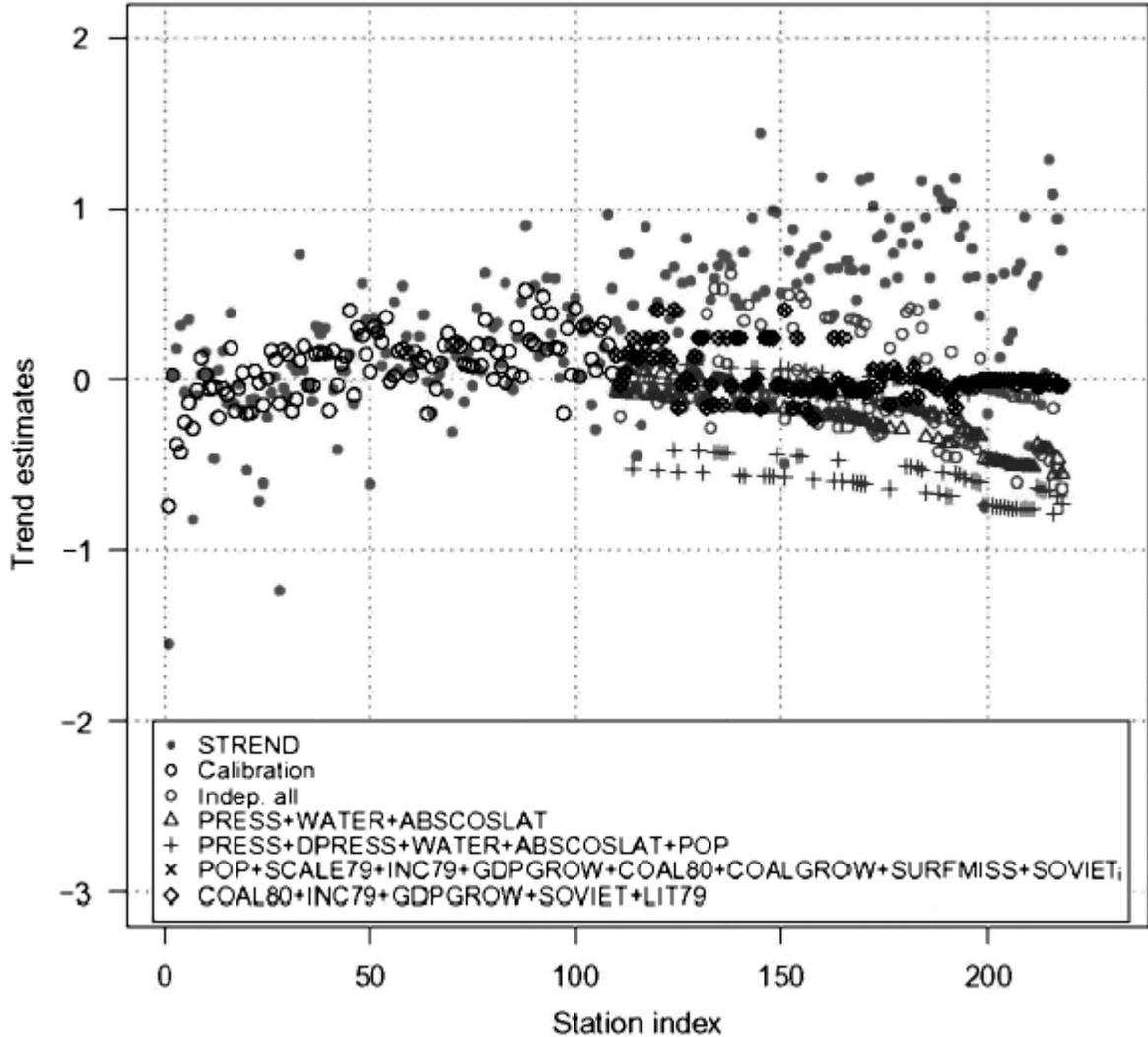
**Figure 48:** Global, land based average surface temperature trends with and without economic and social influences, and their associated standard deviations, as reported by McKittrick and Michaels (2004). Taken from Michaels et al. (2004).

**Surface Temperature Trends from McKitrick & Michaels  
with and without Economic and Social Effects**

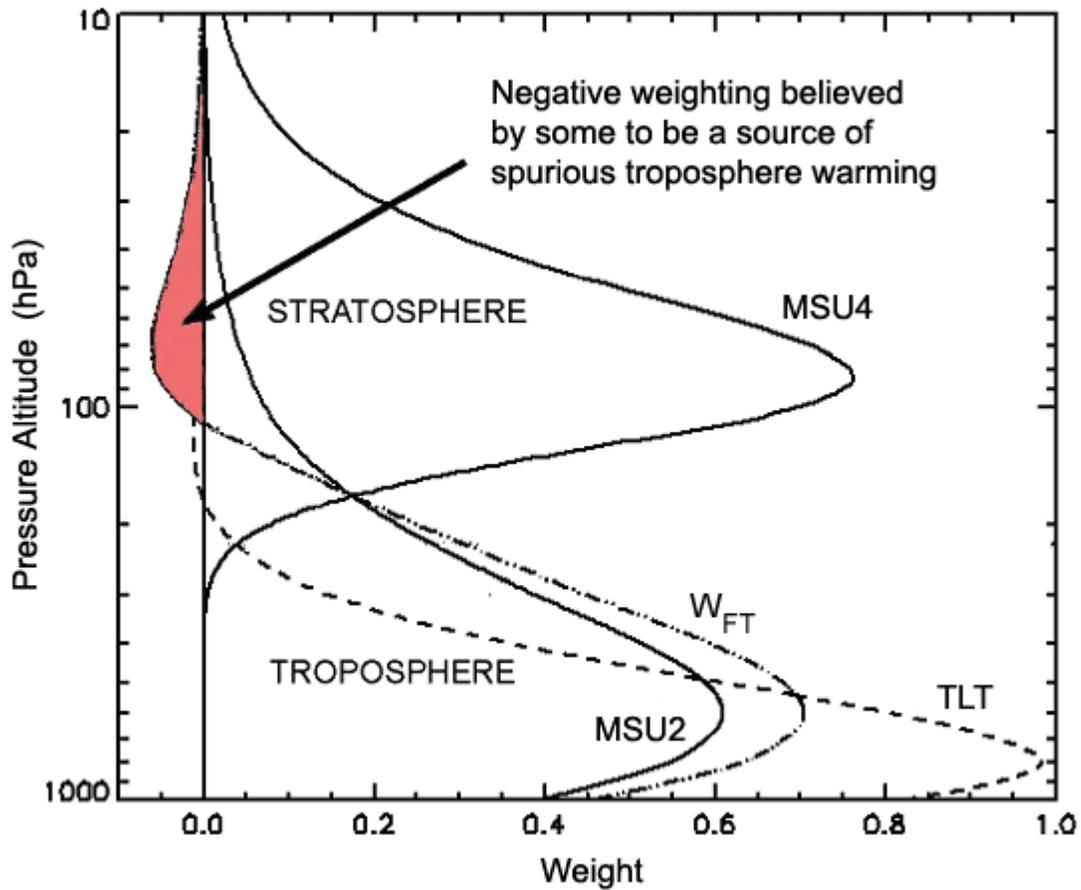
**(After corrections for erroneous latitude inputs)**

<b>Adjustment</b>	<b>Adjusted Sample Average Trend (deg. K/decade)</b>	<b>Adjusted Sample Standard Deviation</b>
Original Sample Average	0.270	0.237
Remove Economic Effects	0.182	0.135
Remove Social Effects	0.134	0.165
Remove Soviet Effects	0.111	0.133
Set Income to Sample Average	0.226	0.133

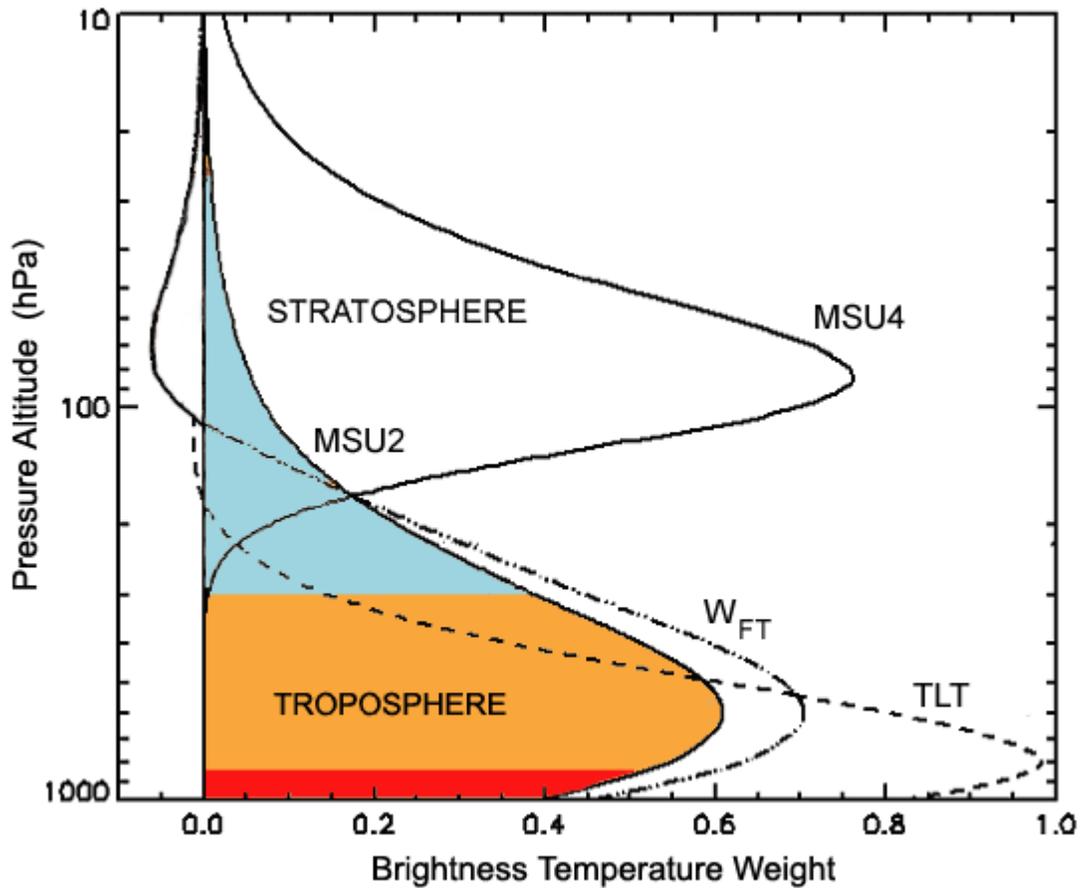
**Figure 49:** Global, land based average surface temperature trends with and without economic and social influences, and their associated standard deviations, as reported by McKitrick and Michaels after correction of erroneous latitude inputs to their original SHAZAM regression run. Taken from McKitrick and Michaels (2004b).



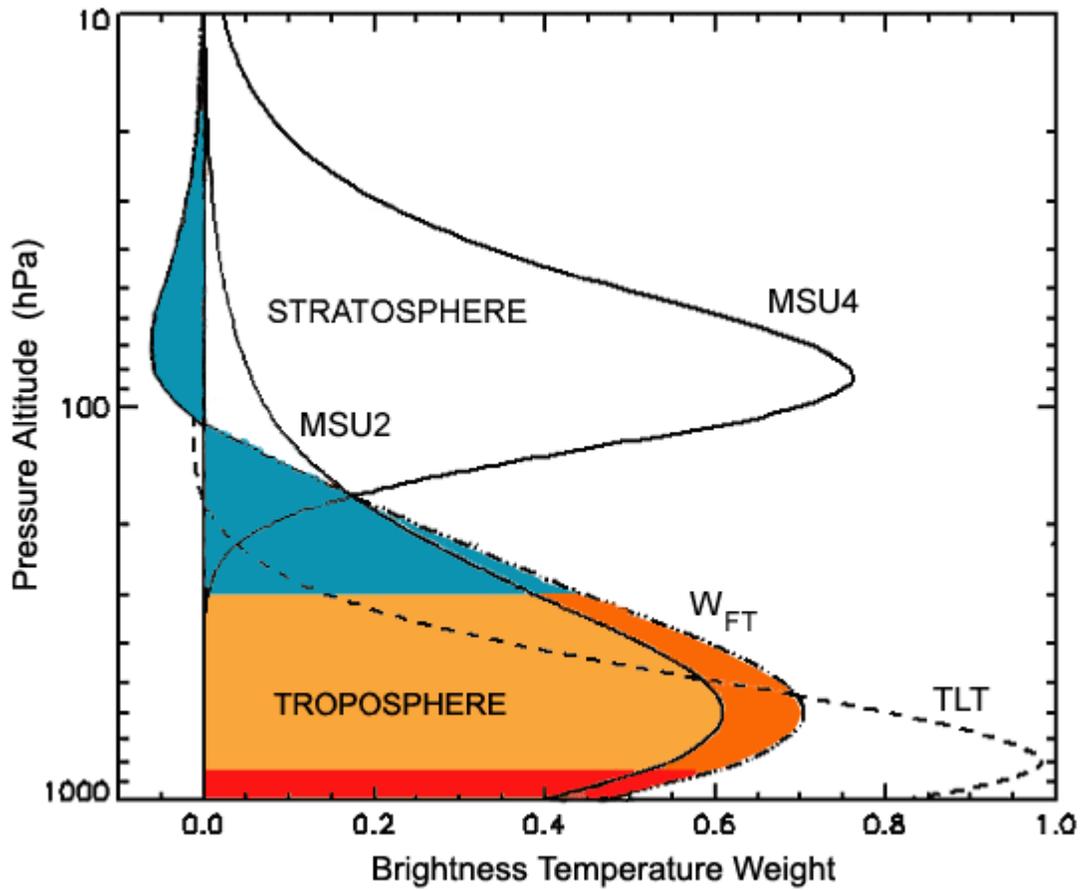
**Figure 50:** Results of regression analyses with 5 different models designed to reproduce the modeled results of McKittrick and Michaels (2004) to test their derived correlations of surface temperature trends with economic, social, and climatic variables. One model was run using all of McKittrick and Michaels' data and the remaining 4 were run using various subsets of their dependent variables. Each model run shown used data from stations within the latitude range 75.5° S to 35.2°N for calibration and stations in the latitude range 35.3° to 80.0° N and corresponding depending variables for prediction and evaluation. The trend estimates shown are in deg. K/decade. Taken from Benestad (2004).



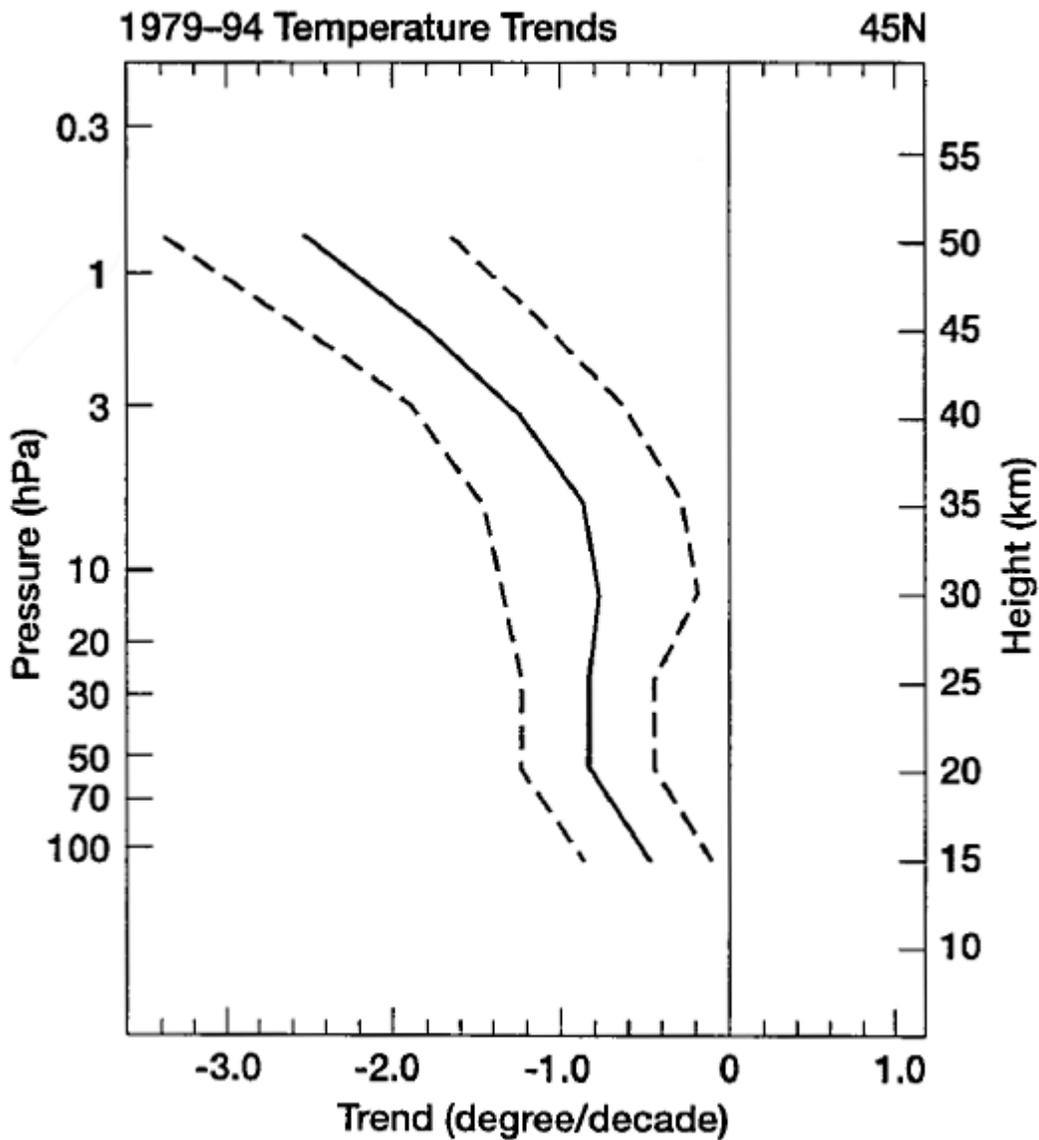
**Figure 47:** The Fu et al. corrected weighting function  $W_{FT}$  for the free troposphere (850-300 hPa layer) compared with uncorrected weighting functions. TLT, MSU2, and MSU4 weighting functions are from UAH Version 5.0 (Christy et al., 2003).  $W_{FT}$  is from Fu et al. (2004). Some critics of the Fu et al. methods have claimed that the area shown in red aliases a spurious warming into the free troposphere trend. Figure taken from Spencer (2004).



**Figure 48:** Figure 47 modified to reflect the layers being detected and trended by MSU2. The region shown in orange is the free troposphere (850-300 hPa layer), the light blue region reflects the tropopause and lower stratosphere, and the red region reflects the surface affected layer. MSU2 measures the entire shaded region, but the layers shown in orange and light blue are known to have differing trends during the satellite era. Adapted from Spencer (2004).



**Figure 49:** Figure 47 modified to reflect the layers being detected and trended by the effective weighting function of Fu et al. (2004). The region shown in dark blue averages to zero above 300 hPa. The combined area shaded in light orange, dark orange, and dark blue averages to the actual temperature and trend of the free troposphere (850-300 hPa layer), shown here in light orange. The red region reflects the surface affected layer. Adapted from Spencer (2004).

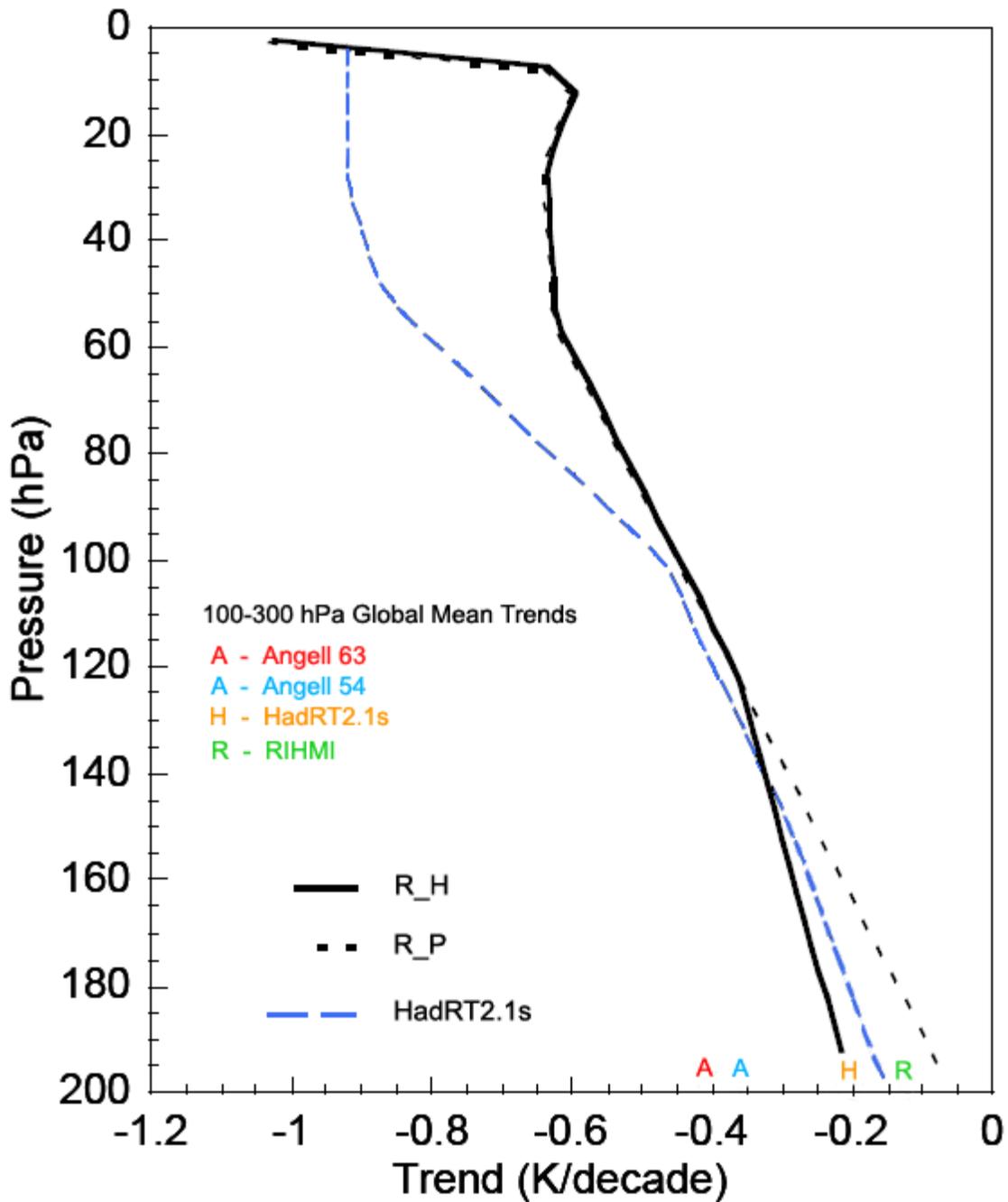


**Figure 50:** Multi-dataset vertical stratospheric temperature trend profile for the period 1979-1994 at 45 deg. N. Latitude. This profile was derived by Ramaswamy et al. (2001) using a mix of radiosonde, satellite, and other analyzed datasets, and was used by Fu and Johanson (2004) for the derivation of their updated stratosphere-corrected MSU2 weighting function. Taken from Ramaswamy et al. (2001).

Zonal Mean Stratospheric Temperature Time Series used by the SPARC-STTA Project

<i>Data Set</i>	<i>Reference</i>	<i>Period</i>	<i>Location</i>	<i>Monthly</i>	<i>Levels, hPa</i>
Angell	<i>Angell [1988]</i>	1958–1994	eight bands	3-monthly	100–50
Oort	<i>Oort and Liu [1993]</i>	1958–1989	four bands	3-monthly	50, 30, 20, 10
Russia	<i>Koshelkov and Zakharov [1998]</i>	1959–1994	85°S–85°N	monthly	100, 50
		1961–1994	70°N, 80°N	monthly	100
			70°N, 80°N	monthly	50
UK RAOB (or RAOB)	<i>Parker and Cox [1995]</i>	1961–1994	87.5°S–87.5°N	monthly	100, 50, 30, 20
Berlin	<i>Labitzke and van Loon [1994]</i>	1965–1994	10°–90°N	monthly	100, 50, 30
Lidar	<i>Hauchecorne et al. [1991]</i>	1979–1994	44°N, 6°E	monthly	10, 5, 2, 1, 0.4
MSU	<i>Spencer and Christy [1993]</i>	1979–1994	85°S–85°N	monthly	90
Nash	<i>Nash and Forrester [1986]</i>	1979–1994	75°S–75°N	monthly	50, 20, 15, 6, 5, 2, 1.5, 0.5
CPC	<i>Gelman et al. [1994]</i>	1979–1994	85°S–85°N	monthly	70, 50, 30, 10, 5, 2, 1
		1964–1978	20°N–85°N	monthly	50, 30, 10
Reanal	<i>Kalnay et al. [1996]</i>	1979–1994	85°S–85°N	monthly	100, 70, 50, 30, 10
GSFC	<i>Schubert et al. [1993]</i>	1979–1994	90°S–90°N	monthly	100, 70, 50, 30, 20
UKMO/SSUANAL	<i>Bailey et al. [1993]</i>	1979–1994	90°S–90°N	monthly	50, 20, 10, 5, 2, 1

**Figure 51:** Zonal mean stratospheric temperature time series datasets used by the SPARC-STTA project for their analysis of stratospheric trend profiles. The 1979-1994 datasets shown were used to derive the vertical stratospheric temperature trend profile shown in Figure 54 (Ramaswamy et al., 2001) and used by Fu and Johanson (2004) for the derivation of their updated stratosphere-corrected MSU2 weighting function. The sources shown for the 1979-1994 datasets are reproduced in the references of this paper also. Taken from Ramaswamy et al. (2001).

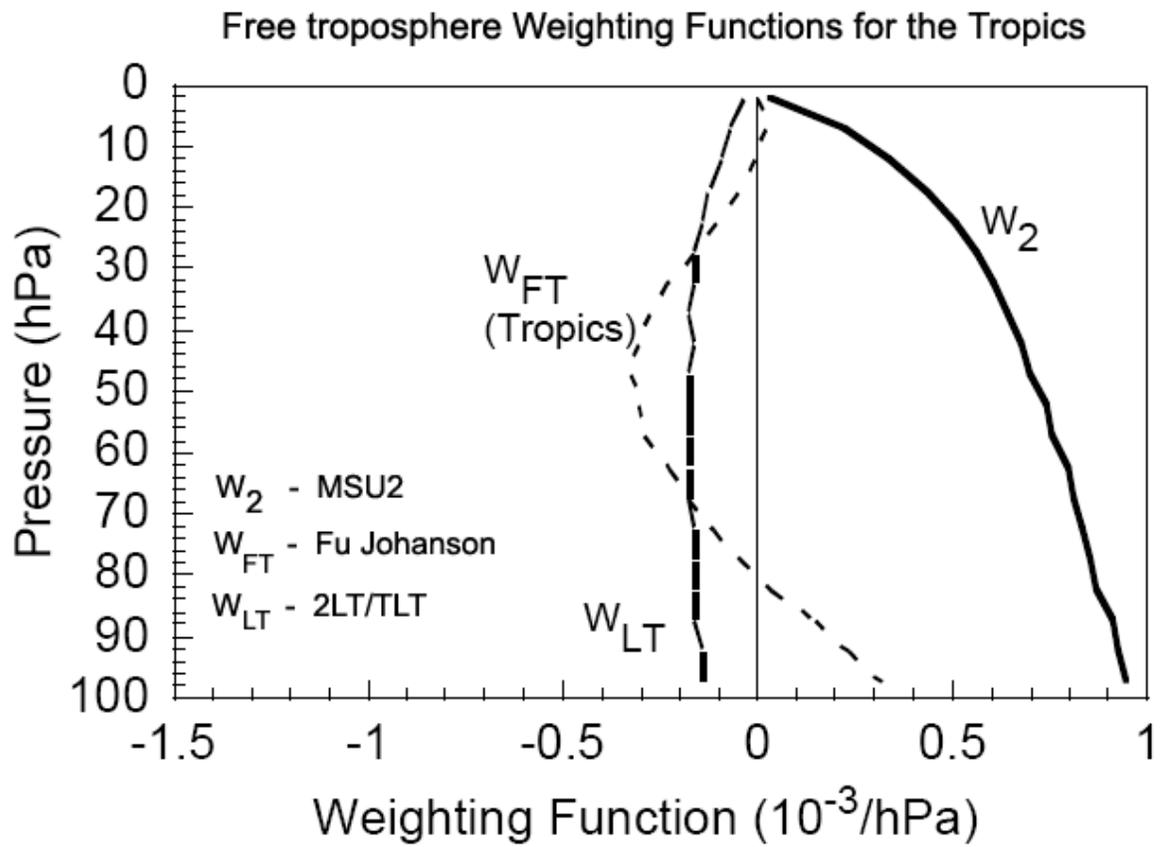


**Figure 52:** Vertical stratospheric temperature trend profiles above 200 hPa for the period 1979-2001 as derived by Fu and Johanson (2004). The solid and short-dashed curves are based on a multiple dataset profile derived by Ramaswamy et al. (2001) rescaled to the global MSU4 trend of UAH Version 5.0 (Christy et al., 2003) using linear extrapolations with respect to height ( $R_H$ ) and pressure ( $R_P$ ). The purple dashed line ( $R_{Had}$ ) is the corresponding profile from the HadRT2.1s radiosonde product (Parker et al., 1997). Also shown for comparison are global mean trends for the same period from 4 other radiosonde products as described in Seidel et al. (2004). Taken from Fu and Johanson (2004).

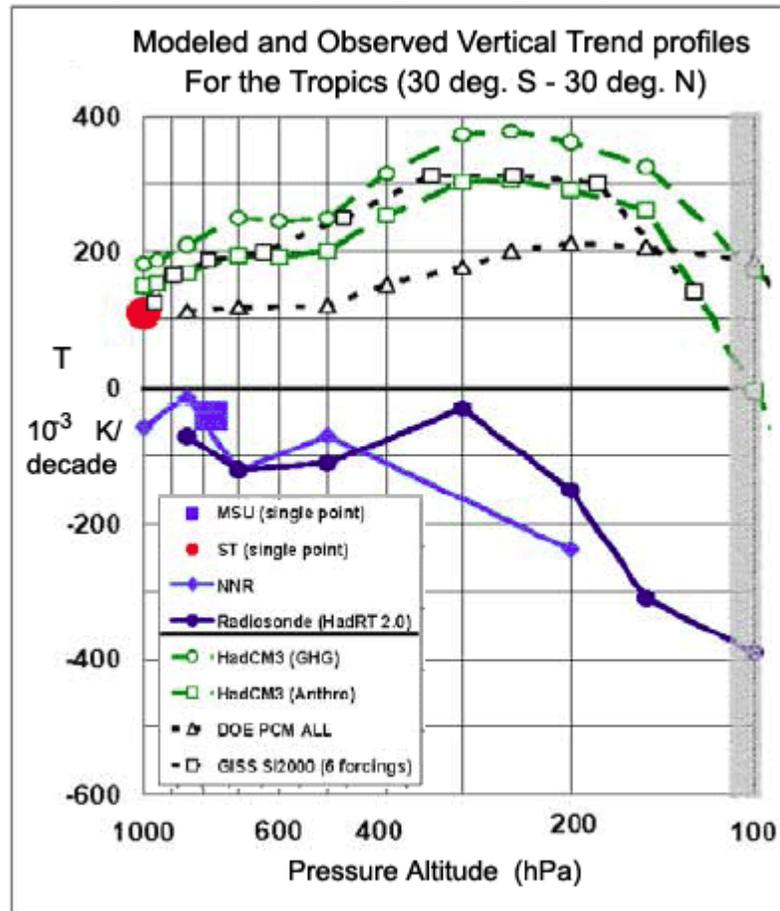
Tropical 1978-2002 Deep-Layer Temperature Trends from Tett and Thorne (2004)

Data source	$T_4$	$T_2$	$T_{fws}$	$T_{2LT}$	$T_{850-300}$	$T_{1,000-100}$	Surface
MSU (Remote Sensing Systems )	-0.35	0.12	0.18				0.13
MSU (University of Alabama, Huntsville )	-0.39	0.05	0.10	0.00			0.13
Radiosonde (HadRT2.1s )	-0.60	-0.08	-0.02	0.03	0.00 (0.09)	-0.03 (0.07)	0.13
ERA40 reanalysis (1 Dec 1978-1 Dec 2001)	-0.20	0.06	0.09	0.02	0.03 (0.07)	0.10 (0.02)	0.10
HadAM3 Model average	-0.29	0.19	0.25	0.20	0.22 (0.03)	0.23 (0.02)	0.14
Smallest trend	-0.31	0.17	0.22	0.18	0.21	0.21	0.12
Largest trend	-0.26	0.22	0.29	0.23	0.25	0.27	0.15
HadCM3 Model average	-0.45	0.16	0.23	0.21	0.21 (0.03)	0.22 (0.01)	0.16
Smallest trend	-0.42	0.07	0.13	0.12	0.12	0.13	0.10
Largest trend	-0.51	0.24	0.32	0.28	0.30	0.30	0.21

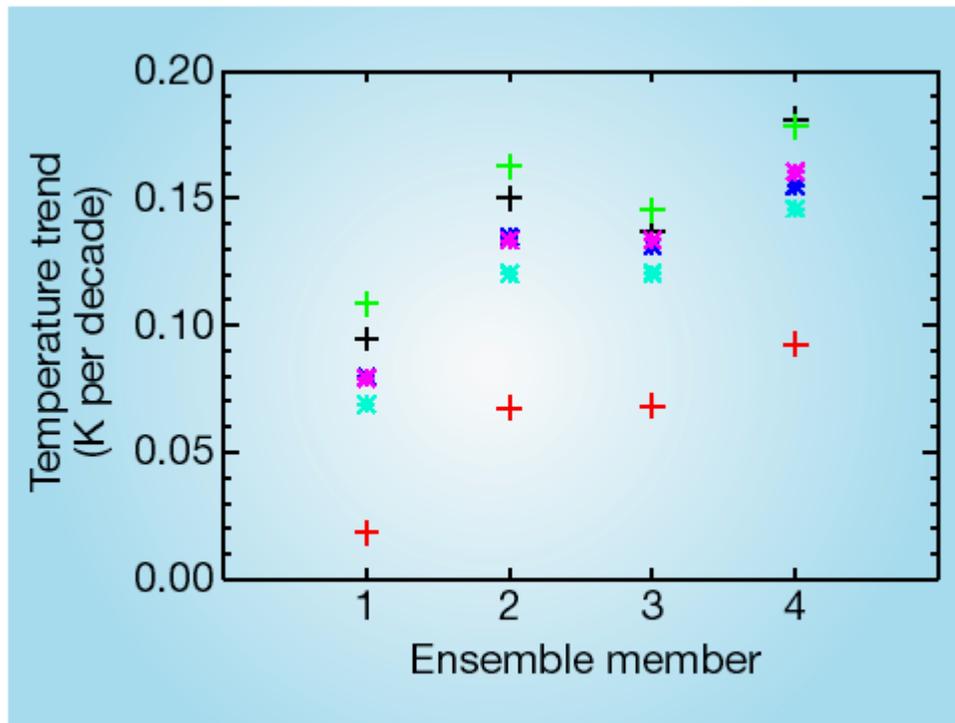
**Figure 53:** Tropical (30° S to 30° N Latitude) temperature trends (deg. K/decade) for the period 1978-2002 as derived by Tett and Thorne (2004) using the Fu et al. (2004) method and data from radiosonde, reanalysis, and model run products. For the non-satellite data sets, static weighting functions were used to estimate MSU2 and MSU4 equivalents.  $T_{fws}$  is the free troposphere trend they derived for each data set using the Fu et al. published coefficients applied to the  $T_2$  and  $T_4$  data. All datasets were zonally averaged, then cosine-weighted and least-square estimates of the linear trends were computed from annual-means. Indian data were removed from the HadRT2.1s analysis. Also shown are the logarithms of the pressure-weighted 850–300 hPa temperatures and the pressure-weighted 1,000–100-hPa temperatures. The RMS of the annual-mean differences between those trends and  $T_{fws}$  is shown in brackets. Surface trends are from data averaged over land and ocean. For ERA-40, 2-meter temperatures were used over land and sea surface temperatures over the oceans. Surface temperatures from HadCRUT2v were used for RSS, UAH and HadRT2.1s. For the two model ensembles, the average, largest and smallest trends are shown. The difference between largest and smallest gives an indication of uncertainty in the ensemble average. The coupled (HadCM3) and atmosphere-only (HadAM3) simulations differ in their forcings, with the main differences being a correction of an error in ozone loss and changes to the sulphur cycle in the HadAM3 simulations. The HadAM3 (HadCM3) ensemble consists of six (four) simulations. Taken from Tett and Thorne (2004).



**Figure 54:** The free troposphere weighting function of Fu and Johanson (2004) for the tropics (30 deg. S to 30 deg. N Latitude) compared to the corresponding weightings for MSU2 and 2LT/TLT. Taken from Fu and Johanson (2004).



**Figure 55:** Modeled and observed vertical trend profiles for the tropics (30 deg. S to 30 deg. N Latitude) as reported by Douglass et al. (2004b). HadCM3 trends are for 1975-1995, DOE PCM trends are for 1979-1999, and GISS SI2000 trends are for 1979-1998. The MSU trend (single point) gives TLT data from UAH Version D (Christy et al., 2000) truncated to 1996. The surface trend (single point) is from Jones et al. (1999). The NNR profile is for the NCEP/NCAR 2-Meter Reanalysis (Kisteler et al., 2001). MSU, surface, and NNR trends are for the period 1979-1996. Taken from Douglass et al. (2004b).



**Figure 56:** Simulated trends in global-mean free-tropospheric temperature as derived by Gillett et al. (2004) using the Fu et al. method applied to 1958-1999 results from the DOE PCM coupled AOGCM. Black crosses, are 850-300 hPa layer trends in each of four realizations of an experiment with anthropogenic and natural forcing. Asterisks indicate free-tropospheric temperature trends reconstructed from synthetic MSU2 and MSU4 trends using the method of Fu et al. These are calculated using three different sets of regression coefficients, which are derived from radiosonde observations by Fu et al. (pink asterisks), estimated from the PCM experiments (dark blue asterisks), and obtained directly from the MSU2 and MSU4 weighting functions (light blue asterisks). Red crosses, simulated trends in MSU2; green crosses, simulated trends in TLT. The simulated trend in MSU4 is  $-0.36 \pm 0.03$  deg. K per decade. The model's surface warming over 1890-1999 ( $0.62$  °C) is consistent with that observed. Taken from Gillett et al. (2004).