Climate Dynamics (2000) 16:587–601

M. Stendel · J. R. Christy · L. Bengtsson

Assessing levels of uncertainty in recent temperature time series

Received: 5 May 1999 / Accepted: 15 December 1999

Abstract We examine to what degree we can expect to obtain accurate temperature trends for the last two decades near the surface and in the lower troposphere. We compare temperatures obtained from surface observations and radiosondes as well as satellite-based measurements from the Microwave Soundings Units (MSU), which have been adjusted for orbital decay and non-linear instrument-body effects, and reanalyses from the European Centre for Medium-Range Weather Forecasts (ERA) and the National Centre for Environmental Prediction (NCEP). In regions with abundant conventional data coverage, where the MSU has no major influence on the reanalysis, temperature anomalies obtained from microwave sounders, radiosondes and from both reanalyses agree reasonably. Where coverage is insufficient, in particular over the tropical oceans, large differences are found between the MSU and either reanalysis. These differences apparently relate to changes in the satellite data availability and to differing satellite retrieval methodologies, to which both reanalyses are quite sensitive over the oceans. For NCEP, this results from the use of raw radiances directly incorporated into the analysis, which make the reanalysis sensitive to changes in the underlying algorithms, e.g. those introduced in August 1992. For ERA, the bias-correction of the one-dimensional variational analysis may introduce an error when the satellite relative to which the correction is calculated is biased itself or when radiances change on a time scale longer than a couple of months, e.g. due to orbit decay. ERA inhomogeneities are apparent in April 1985, October/November 1986 and April 1989. These dates can be identified with the replacements of satellites. It is possible that a negative bias in the sea surface temperatures (SSTs) used in the reanalyses may have been introduced over the period of the satellite record. This could have resulted from a decrease in the number of ship measurements, a concomitant increase in the importance of satellite-derived SSTs, and a likely cold bias in the latter. Alternately, a warm bias in SSTs could have been caused by an increase in the percentage of buoy measurements (relative to deeper ship intake measurements) in the tropical Pacific. No indications for uncorrected inhomogeneities of land surface temperatures could be found. Near-surface temperatures have biases in the boundary layer in both reanalyses, presumably due to the incorrect treatment of snow cover. The increase of near-surface compared to lower tropospheric temperatures in the last two decades may be due to a combination of several factors, including high-latitude near-surface winter warming due to an enhanced NAO and upper-tropospheric cooling due to stratospheric ozone decrease.

1 Introduction

The question as to whether a climate warming, due to the increase of greenhouse gases in the atmosphere, is detectable or not has recently caused some controversy. For the most recent two decades, which is the only period for which reliable three-dimensional global observations are available, a lack of agreement is found between different temperature datasets. Surface temperatures measured by synoptic weather stations (Jones 1994; Karl et al. 1994) show a warming trend of +0.15 K/decade for the period 1979–1997 (Hurrell and Trenberth 1996). Other data sets such as tropospheric temperatures obtained from microwave sounding data (Christy 1995) and radiosonde observations (Angell 1988 and updates; Parker et al. 1997) do not show such a

M. Stendel (✉)
Danish Meteorological Institute, Lyngbyvej 100,
2100 Copenhagen, Denmark
E-mail: mas@dmi.dk

J. R. Christy
University of Alabama in Huntsville, Huntsville, Alabama, USA

L. Bengtsson
Max-Planck-Institut für Meteorologie, Hamburg, Germany

warming, but rather a slight cooling of about 0.05 K/decade for the same period. In fact, one would rather expect the opposite, since climate modelling experiments with an increased amount of greenhouse gases show a somewhat larger warming in the lower and middle troposphere than at the surface, which is due to an enhanced warming with height, mainly caused by the lapse-rate feedback (e.g. Santer et al. 1996).

This discrepancy between surface observations and upper-air measurements has caused some confusion, even leading to the assertion that model simulations are unreliable due to the apparent failure to reproduce the global temperature trends since 1979 (e.g. Singer 1996). While such a two-decade record certainly is too short to confidently show whether a global warming has taken place or not, due to internal low frequency fluctuations in the climate system lasting for decades and longer (Manabe and Stouffer 1996), it is nevertheless of great importance to understand why the trends differ and to suggest an explanation for such a difference.

The present study is one in a series of three which analyze in depth the reasons for these differences. Bengtsson et al. (1999) discuss the reliability of model-predicted trends by means of simulations with a coupled general circulation model (Roeckner et al. 1999) taking into account the effect of greenhouse gases, tropospheric and stratospheric ozone, sulfate aerosol and the eruption of Mt. Pinatubo. The reliability of the microwave temperature soundings is discussed in detail by Christy et al. (2000b) as well as by a recently published paper by Santer et al. (1999). This study addresses the reliability of the observational records, based on monthly mean temperatures. We will consider land surface and ocean observations, upper air measurements from radiosondes and satellites as well as reanalyses from NCEP (Kalnay et al. 1996) and ECMWF (Gibson et al. 1997). In Sect. 2 the data used in this study are briefly described. Section 3 shows comparisons for several regions of the Earth. An assessment of the accuracy of lower tropospheric temperature observations and a discussion of trend differences between surface and tropospheric temperature are given in Sect. 4. Section 5 gives some concluding remarks.

2 Data

2.1 Land and sea surface data

The surface data used here are a combination of near-surface air temperature anomalies over land (Jones 1994) and in situ SST anomalies (MOHSST6) over the oceans. They were used in the IPCC (1996 Intergovernmental Panel on Climate Change) assessments (Nicholls et al. 1996, updated) and have been described in detail e.g. by Parker et al. (1994, 1995b).

There are a number of well-known problems with the surface temperature data set. The record is not global, so that most regions of the Earth will be undersampled compared to the midlatitudes of the Northern Hemisphere (NH). While probably no bias is introduced by these data coverage effects, Nicholls et al. (1996) estimate a trend uncertainty of ±0.15 K for the twentieth century (mainly from the early parts of the record), whereas the more recent investigation by Jones et al. (1997a) suggests that this uncertainty range is probably too large. It should also be noted that it is unclear in the IPCC literature what confidence interval is represented by this uncertainty range. Inhomogeneities due to urbanization effects, station relocation etc. have been corrected. Over the oceans, night-time marine air temperatures (NMATs) and SSTs are used. Folland and Parker (1995) and others have demonstrated the general consistency between these datasets, although there are indications for differences in the tropics (Christy et al. 2000a).

2.2 Radiosonde data

At the Hadley Centre, radiosonde temperatures from over 400 sites, which provide the most homogeneous records and largest spatial distribution, have been gridded into boxes of 5° latitude by 10° longitude. The data set used here (HadRT2.0, Parker et al. 1997, updated) is completely independent of the MSU. Temperatures are produced for every mandatory level, and a simulated MSU lower tropospheric value is calculated based on the static weighting function.

Two other radiosonde data sets are available, from the Russian Research Institute for Hydro-Meteorological Information – World Data Center (RIHMI, Sterin et al. 1997) and from Angell (1988 with updates).

The RIHMI analysis is based on the National Climatic Data Center's (NCDC, Asheville, N.C.) Comprehensive Aerological Reference Data Set (CARDS, see e.g. Wallis 1998) for the period 1958 to 1990 and on data collected via the global telecommunication system (GTS) at RIHMI from 1991, consisting of data from all available radiosondes (more than 800) since 1958. Values are objectively analyzed to a full global grid at the mandatory tropospheric levels. Though the analysis method allows for proper spatial weighting of geographical regions when computing the global mean, it also introduces errors of unknown magnitude in data-sparse regions. There are also homogeneity problems: the CARDS dataset at present is not homogeneized, and missing and inhomogeneous GTS data as well as the merging procedure of CARDS and GTS data may introduce further errors.

The Angell dataset uses 63 sondes widely distributed around the world and with long periods of record. The station values are binned in latitude bands from which a global mean is produced. Trenberth and Olson (1991) showed that the sparse network used by Angell leads to an overestimation of the amplitude of low- frequency temperature fluctuations outside the tropics.

Recently, Gaffen et al. (2000) have demonstrated that corrections for radiosonde data inhomogeneities can have a non-negligible effect on estimated temperature trends, in particular in the stratosphere. Corrections for instrumentation changes, observation time, location, etc. are not explicitly incorporated in the data sets discussed in this section.

2.3 MSU version "d" data

Satellite upper-air temperatures have been retrieved operationally since 1979 by microwave sounders on board NOAA polar-orbiting satellites which measure the vertically integrated thermal emission of molecular oxygen in different spectral intervals near 60 GHz. Lower stratospheric temperatures are obtained by MSU channel 4 (57.95 GHz) and tropospheric temperatures by channel 2 (53.74 GHz). Since part of the radiation that is measured in channel 2 comes from the stratosphere, a synthetic channel temperature, T_{2LT} (where LT stands for lower troposphere), is constructed from the outer eight view angles of channel 2 (Spencer and Christy 1992). The weighting function of MSU2LT peaks slightly lower in the troposphere than channel 2 (near 740 hPa compared to 595 hPa for channel 2), but it is more exposed to surface effects.

The MSU was not originally designed as an instrument for climate monitoring. Therefore the key question is the inter-satellite stability and the merging of the data. Orbit decay of the satellites in periods of high solar activity causes a small artificial cooling in T_{2LT} (Wentz and Schabel 1998), and the non-linear response of the instrument-body temperature leads to a small artificial warming in T_{2LT} and T_2 (Christy et al. 1998). These effects have been adjusted for in the MSU "d" data that is used here. A detailed description of the merging procedure and the correction of drift errors and biases is given in Christy et al. (2000b).

For global averages over the period 1979 to 1998, a trend[1] of +0.06 K/decade for T_{2LT} and of +0.05 K/decade for HadRT2.0 is found (Christy et al. 2000b). Since such a comparison does not take into account the limited spatial coverage of the radiosonde data set, Christy et al. (2000b) have also estimated trends for collocated MSU and radiosonde data over the western part of the NH. For 97 US-controlled stations (in the conterminous USA, Canada, Alaska, Iceland, Mexico, the Caribbean and Bermuda) and for subsets of these stations, they find trend agreement within ±0.02 K/decade and correlations of annual as well as monthly anomalies to the MSU of above 0.96 in middle and high latitudes. In the subtropics, the correlations are slightly lower.

The close agreement between lower tropospheric temperature trends derived from MSUs and radiosondes may reflect a real cooling during the last two decades. However, other authors have pointed out uncertainties that may impede the interpretation. These may be related to inhomogeneities in either the MSU (Hurrell and Trenberth 1996, 1997, 1998; Jones et al. 1997b) or radiosonde data (Gaffen 1994; Gaffen et al. 2000), different quantities actually measured (Hansen et al. 1995), sampling and data coverage problems (e.g. Santer et al. 1999) and the choice of method for trend calculation (Santer et al. 2000).

2.4 Reanalyses

Reanalyses use numerical forecast models with a "frozen" observational data assimilation system. Since the model output is influenced by the data assimilation strategy and the numerical model used, reanalyses do not give a direct observation of the climatic state of the atmosphere. However, they should yield data that are uninfluenced by the numerous changes in model physics by which operational analyses are typically affected (Bengtsson and Shukla 1988).

The reanalyses used in this study cover the period 1979 to present for NCEP and 1979 to 1993 for ERA. Monthly mean temperature anomalies were interpolated to a common 2.5° by 2.5° grid. For both reanalyses, data are available on 17 pressure levels. Further aspects of the different layout of the reanalyses are discussed by Stendel and Arpe (1997). As mentioned by Santer et al. (1999), it is important to note that in addition to differences in resolution, data assimilation and model physics between the reanalyses, the lower boundary conditions are slightly different. In the ECMWF reanalysis, the following were used: 1° monthly SST analyses from the UK Meteorological Office (GISST1.1; Parker et al. 1995a) for December 1978 and January 1980 to October 1981, GISST2.2 (Rayner et al. 1996) analyses for 1979 and 1° weekly optimum- interpolated SST analyses of Reynolds and Smith (1994) for November 1981 to the end of the period. NCEP uses GISST2.2 data to October 1981 and Reynolds and Smith (1994) SSTs thereafter. The treatment of sea ice is identical for both reanalyses and follows the post-analysis procedure of Nomura (1995).

3 Results

In order to simplify the comparison of analyzed fields from the NCEP and ECMWF reanalyses and microwave data we have calculated the equivalent microwave temperature, $T2_{LT}$, of the reanalyzed tropospheric temperature, representing the temperature of the lower troposphere (Stendel and Bengtsson 1997). Errors introduced by this approach compared to solving the full radiative transfer equation are generally small compared to other uncertainties in the data sets (Santer et al. 1999). For 1979–1993, the period common to all data sets used here, global monthly T_{2LT} anomalies show a relatively high correlation between MSU temperatures and the reanalyses from both NCEP and ERA (see Tables 1 and 2). These correlations are, however, dominated by similarities in seasonal and interannual variability, so that the three data sets behave differently on longer time scales: compared to the MSU, each of the reanalyses shows a number of sudden (but generally not simultaneous) jumps (Fig. 1). We do not know which of the data sets (if any) is correct. There is, however, evidence that inhomogeneities can be identified in both reanalyses (discussed later), but this does not imply that there are no problems with the MSU.

Figure 1 gives some indication that the jumps in the difference time series of MSU and reanalyses may be related to changes in the satellite availability. To estimate the effect of the incorporation of satellites, lower tropospheric temperature anomalies are investigated for two regions in detail. Both reanalyses exclude satellite data over land below 100 hPa. Santer et al. (1999) point out that even when no satellite data are used, the incorporation of such data elsewhere is not confined to the region of incorporation. This effect is largest in the lower stratosphere. Since the 2_{LT} retrieval has no stratospheric contribution, we can conclude that the MSU has no major influence on the reanalysis in the data-rich regions over the Northern Hemisphere continents (Europe and Asia, 40°N to 70°N, 10°W to 135°E and North America, 30°N to 70°N, 125°W to 75°W). On the other hand, almost no conventional observations exist over the oceans, in particular the tropical Pacific (20°S to 20°N, 160°E to 80°W), so that radiances measured from satellites have to be included in the data assimilation.

However, this raises a number of difficulties. In the NCEP reanalysis the original operational TOVS[2] retrievals from NESDIS, taking into account all radiances (i.e. from HIRS, SSU and MSU), are used (Kalnay et al. 1996). Raw NESDIS temperature retrievals are used over all oceans (Basist and Chelliah 1997). No explicit intersatellite calibration in the NESDIS retrievals is done in NCEP, in contrast to MSU and ERA. Since it is not clear a priori whether or not a given change in e.g. the attenuation coefficients for water vapour will have a significant effect on the data assimilation, this procedure may cause problems, considering the numerous changes

---

[1] It is important to recall that trend differences of a few hundredth of Kelvin per decade must be interpreted with caution due to the large variability of tropospheric temperatures. To consider an extreme case, the global T_{2LT} trend for 1979–1997 was −0.01 K/decade. The inclusion of 1998 (dominated by the strong 1997/98 El Niño) changed this trend to +0.06 K/ decade.

[2] TOVS stands for TIROS-N Operational Vertical Sounder, NESDIS means National Environmental Satellite Data and Information Service, HIRS is the High Resolution Infrared Sounder and SSU the Stratospheric Sounding Unit.

Table 1 Correlations between lower tropospheric temperature data sets

	HadRT2.0	MSU "d"	NCEP	ERA
Global				
HadRT2.0	1.00	0.80	0.82	0.68
MSU "d"		1.00	0.93	0.83
NCEP			1.00	0.85
ERA	–			1.00
North America (30°N–70°N, 125°W–75°W) and Eurasia (40°N–70°N, 10°W–135°E)				
MSU "d"	–	1.00	0.97	0.97
NCEP	–		1.00	0.98
ERA	–			1.00
Tropical Pacific (20°S–20°N, 160°E–80°W)				
MSU "d"	–	1.00	0.97	0.83
NCEP	–		1.00	0.87
ERA	–			1.00

Correlations for the global temperatures and for temperatures over the data-rich regions of the Northern Hemisphere (North America and Eurasia) and the tropical Pacific using monthly mean data. Correlations to HadRT2.0 are given only for global mean temperatures

Table 2 RMS differences between lower tropospheric temperature data sets

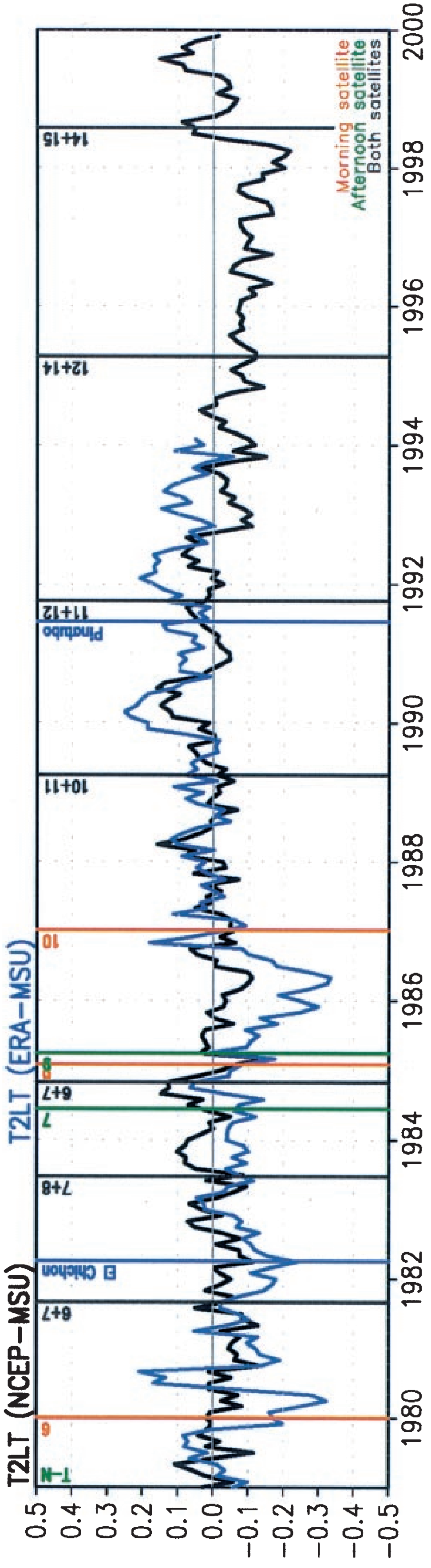
	HadRT2.0	MSU "d"	NCEP	ERA
Global				
HadRT2.0	0.000	0.206	0.215	0.219
MSU "d"		0.000	0.075	0.123
NCEP			0.000	0.114
ERA				0.000
North America (30°N–70°N, 125°W–75°W) and Eurasia (40°N–70°N, 10°W–135°E)				
MSU "d"	–	0.000	0.163	0.146
NCEP	–		0.000	0.120
ERA	–			0.000
Tropical Pacific (20°S–20°N, 160°E–80°W)				
MSU "d"	–	0.000	0.099	0.248
NCEP	–		0.000	0.225
ERA	–			0.000

RMS differences for the global temperatures and for temperatures over the data-rich regions of the Northern Hemisphere (North America and Eurasia) and for the tropical Pacific using monthly mean data. RMS differences to HadRT2.0 are given only for global mean temperatures

in NESDIS reported e.g. by Andersson et al. (1991) and Kelly et al. (1991).

ERA does not directly incorporate raw TOVS radiances into the analysis (Gibson et al. 1997). Instead they are "cloud-cleared", i.e. MSU radiances are used for clear retrievals only. Radiances from cloudy retrievals may have a cold bias due to scattering effects from precipitation (Prabhakara et al. 1996) and have consequently been removed in ERA. However, the effect appears to be negligible (Christy 1995; Spencer et al. 1996). The radiances are then converted to temperatures by means of a "one dimensional variational analysis" (1D-Var) scheme (Eyre et al. 1993), where the variational part of the analysis consists of finding the atmospheric temperature and humidity profile that fits the measured radiances best. 1D-Var temperatures are used below 100 hPa over the oceans only. Between 30°S and 30°N, only cloud-cleared radiances are included (Gibson et al. 1997). The so obtained profiles then have to be bias-corrected (Eyre 1992). This is done by using corrections calculated from the previous months' biases with respect to a selection of radiosonde stations that is regarded as reliable. The bias correction does not explicitly take into account radiance changes over a period longer than a couple of months. When a new satellite is introduced, an erroneous bias correction for the old satellite may result in an artificial bias for the new one. Furthermore, the inclusion of the HIRS water vapour channels, which can have significant impact on the temperature and water vapour structure in the tropics, bears the possibility of introducing substantial errors (Uppala 1997).

Over the NH continents, a reasonable agreement of MSU and reanalysis temperatures is found, with correlations of 0.97 for both NCEP and ERA (Fig. 2 and Table 1). We note, however, that the surface warms relative to the lowest tropospheric pressure levels in both reanalyses (surface minus 850 hPa residual trend: +0.13 K/decade in ERA for 1979 to 1993, +0.10 (+0.19) K/decade in NCEP for 1979 to 1993 (1998)). How can this be explained? Such systematic differences in adjacent layers could e.g. be caused by small variations in the static stability of the lower troposphere, which in particular during winter can be quite substantial. One



◀ **Fig. 1** Time series of residual lower tropospheric equivalent MSU T2LT anomalies (K) for NCEP-MSU (*black curve*) and ERA-MSU (*blue curve*). Also given (as *vertical lines*) are the replacement dates of polar-orbiting satellites used for the reanalyses. *Orange lines* mean that only one "morning" satellite is available from the given data, *green lines* denote "afternoon" satellites, and *black lines* give periods with both a "morning" and an "afternoon" satellite. The *blue vertical lines* denote the dates of the eruption of the volcanoes El Chichón and Mt. Pinatubo. All anomalies with respect to the period 1979–1993

may speculate that for example more marked inversions during the early part of the record (leading to extra cold surface temperatures) than during the latter part could generate such a systematic difference.

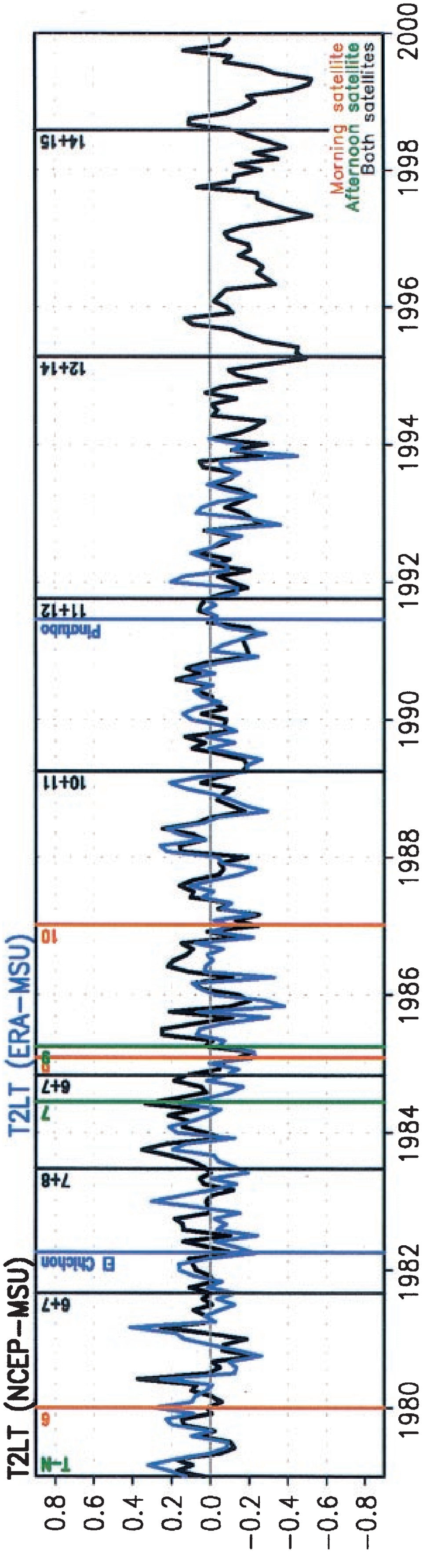
One way we can check this is to consider the near-surface (2 m) temperature from the reanalyses. NCEP reanalyses are not suitable for this type of comparison, since they were erroneously forced with the snow cover of 1973 throughout the period 1974 to 1994 (Kalnay personal communication).[3] For 1979–1993, ERA T_{2m} over North America and Eurasia warms by 0.14 K/decade less than observations (Jones 1994). Although there is no common period of correct snow cover treatment in NCEP and ERA, the T_{2m} anomalies from both reanalyses behave similarly. This might be due to erroneous observed data; however, as discussed in Sect. 4.2, an improper treatment of the planetary boundary layer, in particular snow cover, is more likely.

For the tropical Pacific (Fig. 3 and Table 1), several jumps are visible in the time series of the difference between ERA and MSU. These can partly be traced back to transitions from one polar orbiting satellite to another. For example, there is a cold bias in ERA starting in September 1981, which coincides with the introduction of NOAA-7 as an "afternoon" satellite in addition to the "morning" satellite NOAA-6 ("morning" and "afternoon" refer to the local northbound equator-crossing time of the satellite). This cold bias stays present just up to El Chichón's eruption in March 1982. Several problems were experienced with the NOAA satellites in 1984 and 1985 (Christy et al. 1998). The repeated changes from and to NOAA-6 show up in the residuals, and the introduction of NOAA-9 in April 1985 coincides with a clearly visible cooling in ERA as large as 0.5 K. This bias decreases with time and is near zero when NOAA-10 finally replaced NOAA-6 in January 1987. Until April 1989, this satellite was the only active one, and no further problems appeared, but with the introduction of NOAA-11 as "afternoon" satellite a positive bias in ERA temperatures is introduced which remains visible most of the time, even after the replacement of NOAA-10 with NOAA-12 in October 1991.

Further jumps are visible in ERA in early 1980 and in 1986. These can be traced back to problems with MSU

---

[3] Another error that affects the boundary layer, though not important for the period considered here, is the erroneous rejection of sea level pressure values below 1000 hPa over parts of the NH prior to 1968 (Kistler, personal communication).



◀

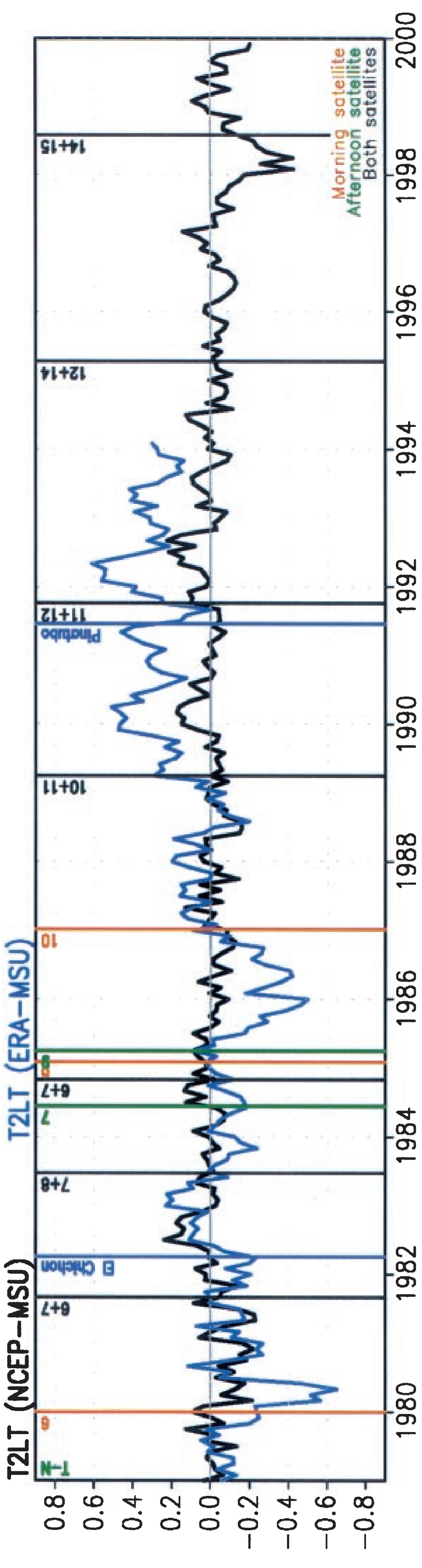
Fig. 2 As for Fig. 1, but for the data-rich regions of the Northern Hemisphere, North America (30°N–70°N, 125°W–75°W) and Eurasia (40°N–70°N, 10°W–135°E)

channel 3 in the profile reconstruction. This channel is a key parameter in the retrieval schemes, and without accurate information for the upper troposphere, the entire vertical temperature profile can be erroneously skewed. MSU channel 3 experienced erratic behaviour and sudden shifts on NOAA-6, NOAA-7 and NOAA-9 that are not apparent in any other channel. While no clear jumps can be identified in ERA relative to NCEP for the NH continents, a trend is clearly visible in regions where ERA has to rely on TOVS data. This is evidently a factor in the significantly cold temperatures in ERA in 1980 and 1986 (Fig. 3).

What does the vertical distribution of the temperature trends look like? Figure 4 shows normalized global pressure-level and corresponding equivalent microwave temperature differences of ERA and NCEP. Quite large differences between the reanalyses are visible in the stratosphere (e.g. at 10 hPa). These relate to an unrealistic representation of the quasi-biennal oscillation (QBO) in both reanalyses (compare Pawson and Fiorino 1998; Santer et al. 1999).

Tropospheric differences between ERA and NCEP are particularly large from 1992 on, which suggests a connection to a change in the cloudy algorithm over the tropical oceans (Reale et al. 1994) in August 1992. Since such retrievals are directly assimilated, they are likely to result in spurious trends in the NCEP reanalysis (see Santer et al. 1999, Plate 1).

A jump in the difference between ERA and NCEP is visible in the lower troposphere in October/November 1986 (near the surface, no jump, but rather a smooth transition and a residual annual cycle are apparent, with largest differences of ERA and NCEP in winter of the NH extratropics). Both reanalyses had problems during this period. While for NCAR a change in the rawinsonde correction scheme (Basist and Chelliah 1997) may have had an effect, for ERA this can be traced back to the assimilation of satellite data in data-sparse regions. ERA's 1D-Var analyzes departures from the background (i.e. the first guess), so that errors in the first guess eventually become biases in the 1D-Var and therefore in the final analysis. The October/November 1986 jump seems to be due to two events: the transition to only one satellite in 1985 with subsequent warming of the first guess and erratic behaviour of NOAA-9 and MSU3 in October 1986, so that an inappropriate bias correction keeps the warming of the first guess (Kållberg personal communication). An erroneous rejection of all tropical 1D-Var retrievals in February 1987 (Fiorino personal communication) caused further errors. Consequently, we find a jump in the ERA global mean temperature in late 1986, and the second half of ERA (i.e. 1987–1993) is warmer and wetter in the lower tropical troposphere over the oceans than the first half. However,



◀

Fig. 3 As for Fig. 1, but for the tropical Pacific (20°S–20°N, 150°E–90°W)

as Fig. 5 demonstrates, this trend is considerably weaker or even absent at isolated island stations such as Diego Garcia in the tropical Indian Ocean, Guam, Kwajalein, Ponape and Tahiti in the tropical and subtropical Pacific and St. Helena in the tropical Atlantic.

The comparison of temperature time series between the NH continents (where MSU and reanalysis data are most independent because the radiosondes provide the upper atmospheric information for the reanalyses) and the tropical Pacific highlights the fact that the results of such reanalysis exercises may only be expected to be as good as the input databases. This corroborates recent results by Santer et al. (1999), who found that NCEP's lower and middle tropospheric temperature differ markedly from radiosonde data, caused by a likely warm bias in NESDIS temperature retrievals which could only be corrected from 1979 on so that the pre-satellite-era NCEP data are too cold.

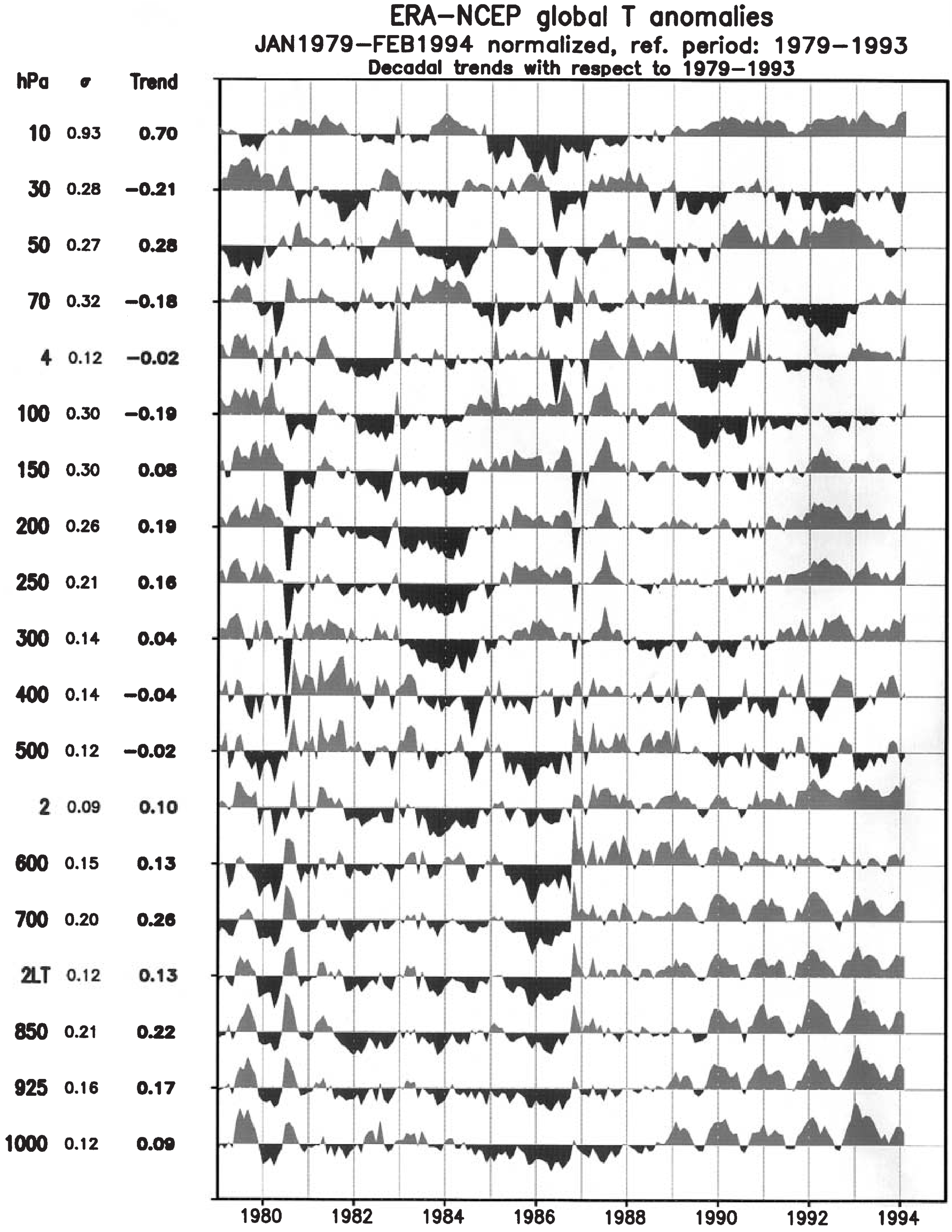
4 Discussion

Two questions arise from the results presented so far: (1) to what degree of certainty are we able to determine recent trends in the troposphere? and (2) how can we explain the difference in trends between tropospheric and near-surface temperatures? These issues will be discussed in the following subsections.

4.1 Accuracy of tropospheric temperature measurements

The first question reveals the need to examine the accuracy of tropospheric temperatures derived from satellites and reanalyses. The reliability depends crucially on the geographic region and on the datasets involved. Four dimensional data assimilation as it is used for the reanalyses from ECMWF and NCEP is a complex, nonlinear method to incorporate data of all types, with very different error characteristics, into a forecast model. In regions where no conventional observations are available, the data assimilation scheme has to rely on satellite radiances. These are incorporated into the analysis in a way that differs substantially between NCEP and ERA. Furthermore, neither of the models is perfect. The potential unreliability of reanalyzed temperatures can be estimated for the period 1979 to 1993, which is common to both reanalyses. For this period, ERA gives a lower tropospheric trend of +0.24 K/decade for the tropical Pacific, whereas for NCEP the trend is –0.10 K/decade and for the MSU –0.14 K/decade (Table 3). These values may be directly compared to the results of Santer et al. (1999, Table 5).

How do we know which of the data sets is correct or what amount of error we assign to each? The warming


ERA–NCEP global T anomalies
JAN1979–FEB1994 normalized, ref. period: 1979–1993
Decadal trends with respect to 1979–1993
hPa σ Trend
10 0.93 0.70
30 0.28 −0.21
50 0.27 0.28
70 0.32 −0.18
4 0.12 −0.02
100 0.30 −0.19
150 0.30 0.08
200 0.26 0.19
250 0.21 0.16
300 0.14 0.04
400 0.14 −0.04
500 0.12 −0.02
2 0.09 0.10
600 0.15 0.13
700 0.20 0.26
2LT 0.12 0.13
850 0.21 0.22
925 0.16 0.17
1000 0.12 0.09
1980
1982
1984
1986
1988
1990
1992
1994

◀

Fig. 4 Time series of normalized global temperature differences (K) of ERA and NCEP with respect to the period 1979–1993 at pressure levels from 1000 hPa to 10 hPa. The standard deviations (K) of the monthly differences are given on the *left*, and the *tick marks* are every 4σ. Also given are the differences of the equivalent microwave temperatures near the pressure levels where the respective weighting function has its maximum. Decadal trends for the period 1979–1993

trend in the ERA data can to a large extent be attributed to the treatment of satellite observations, in particular of NOAA-11, which leads to a positive temperature bias of about 0.3 K over the tropical oceans with respect to both NCEP and MSU after 1989 (Fig. 3). On the other hand, the decrease in NCEP temperature does not agree with Santer et al.'s (1996) estimate of 0.08–0.30 K/decade for the expected 2_{LT} trend. Trenberth and Hurrell (1997) point out a number of potential consistency problems with the MSU data, mainly related to the circumstance that eight different satellites, with slightly different sampling characteristics, have been used during the period. However, the MSU is calibrated globally (Christy et al. 2000b), so one might expect better agreement of MSU and reanalyses over the tropical Pacific, taking into account the good agreement over the Northern Hemisphere continents.

There is still some possibility that the NCEP reanalyses and the microwave data do indeed have a similar systematic error. This could be the case for example if the MSU data had a major influence on the reanalysis. Mo et al. (1995) showed that NESDIS retrievals had a warm bias in the tropical troposphere which affected the NCEP reanalysis. However (although there are differences between both reanalyses in the forecast models and data assimilation schemes as well as the parametrization of sub-grid scale processes), for regions with ample data coverage from radiosondes, decadal trends in both reanalyses agree reasonably: for the same 15-year period (1979–1993), the difference over North America and Eurasia is only 0.02 K/decade (NCEP: +0.14 K/decade, ERA: +0.16 K/decade; the MSU trend is +0.25 K/decade, see Table 3), in general agreement with Santer et al. (1999, Plate 10). We note that this is not valid for the NH as a whole (compare Santer et al. 1999, Table 5); though we can conclude that the similarity of trends in NCEP and MSU cannot be attributed to an interdependence of the two data sets, this does not imply that both are "correct".

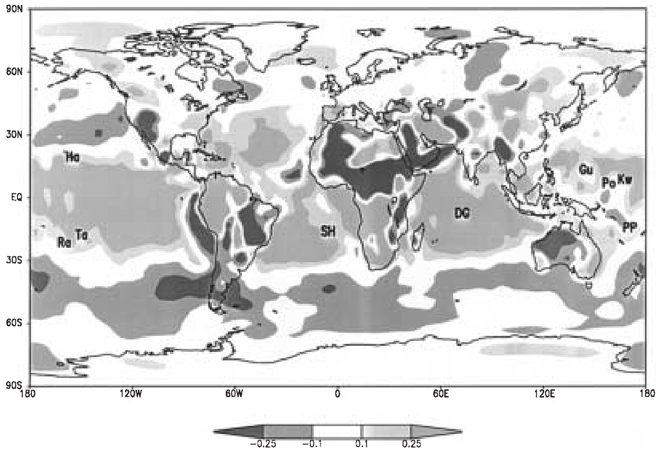


Fig. 5 Specific humidity at 925 hPa from ERA, difference (g/kg) 1987–1993 minus 1979–1985. Also given are the positions of some tropical radiosonde stations (*Ha*, Hawaii; *Ra*, Rarotonga; *Ta*, Tahiti; *SH*, St. Helena; *DG*, Diego Garcia; *Gu*, Guam; *Po*, Ponape; *Kw*, Kwajalein; *PP*, Pago-Pago)

Table 3 Linear temperature trends (K/decade) for 1979–1993

	Global	North America and Eurasia	Tropical Pacific
HadRT2.0	0.06	–	–
MSU "d"	−0.06	0.25**	−0.15
NCEP	−0.04	0.14*	−0.10
ERA	0.09	0.16**	0.24
IPCC T_{2m}	0.15	0.36**	0.12
NCEP T_{2m}	−0.05	0.20**	–
NCEP SST	–	–	0.02
ERA T_{2m}	0.01	0.22**	–
ERA SST	–	–	0.07

Temperature trend estimates for the globe, for the data-rich regions of the Northern Hemisphere, North America (30°N–70°N, 125°W–75°W) and Eurasia (40°N–70°N, 10°W–135°E), and for the tropical Pacific (20°S–20°N, 160°E–80°W) using monthly mean data. Trends for the 2 m temperature and the SST in the NCEP and ERA reanalyses are also given. A trend estimate from HadRT2.0 is given only for the global mean temperature

* Significant at the 10% level

** Significant at the 5% level

With respect to the reanalysis data a temperature bias can be caused by systematic errors in the forecasting model used in the data assimilation. These can affect the temperature in data sparse regions, an error which can eventually be eliminated in a later period when new observations become available. This is a serious problem which influences studies over longer periods during which major changes in the observing system have taken place (see also Nicholls et al. 1996). It would therefore be of considerable interest to quantify the importance of data coverage inhomogeneities in the reanalyses by creating a subset which is in this respect as homogeneous as possible.

4.2 Differences between tropospheric and near-surface temperatures

Possible mechanisms that might shed some light on the question of why near-surface trends exceed those in the free troposphere include biases in either the SST, the land surface station data or the MSU measurements, changes in data coverage and/or measurement techniques, or real lower tropospheric lapse rate changes during the last two decades. These possibilities will be discussed in the following subsections. The surface "global" means are based on only about 70% coverage of the globe, and considerable spatial and temporal coverage differences exist between the data sets. It is therefore possible that the trends differ because different regions of the Earth have been sampled. This issue has been discussed in detail by Santer et al. (1999) and will not be covered here. Another possible explanation is the uncertainty in the trends. The records are noisy, so fitting a linear trend is not exact. The statistical significance has been tested by means of a standard *t* test against the null hypothesis that the trend is zero. Autocorrelation of the data has been accounted for, as in Santer et al. (1999). Due to the high autocorrelation of the time series, the effective sample size is reduced considerably, and none of the trends for the globe or the tropical Pacific differ significantly from zero. For North America and Eurasia, we find trends slightly different from zero at the 5% significance level (except NCEP T_{2LT}, which is significant at the 10% level). Linear trends near the surface and in the lower troposphere do not differ significantly due to the large standard error caused by the autocorrelation of the time series. Nevertheless, we can gain some insight into the physical mechanisms behind the differences between lower tropospheric and near-surface temperature.

4.2.1 SST properties

Figure 6 gives a time series of the difference of global SST anomalies from the Reynolds and Smith SST (used in the reanalyses from October 1981 on) and in situ data (MOHSST). The difference trend is near −0.15 K/decade. Figure 7 shows the geographical distribution of the difference between the Reynolds and Smith (1994) SST used for ERA and the in situ SST over the oceans. Over land, the difference between the reanalyzed and the Jones (1994) 2 m temperature is given. It is evident that the Reynolds and Smith (1994) SST and the reanalysis data cools with respect to the in situ data.

The reanalyses require global SST (and sea ice) datasets. However, in situ SSTs incorporate only observations from ships and buoys, which means that for large regions of the Earth, particularly in the Southern Hemisphere, no data are available. Such gaps are filled in the Reynolds and Smith (1994) SST with satellite observations, where Laplace's equation is solved and the satellite data determine the second derivative of the field to be interpolated, with the in-situ data acting as boundary conditions, thus largely compensating for the satellite's typically cold bias (Reynolds 1988; Reynolds et al. 1989). As demonstrated by Hurrell and Trenberth (1999), the Reynolds and Smith (1994) SST analysis has a cold bias. This can be traced back to the fact that in the 1980s, when in situ ship observations from the Comprehensive Ocean-Atmosphere Data Set (COADS) were used, roughly 30 000 observations per week were available for the bias correction, whereas in the 1990s only ship data from the GTS, covering about 15 000

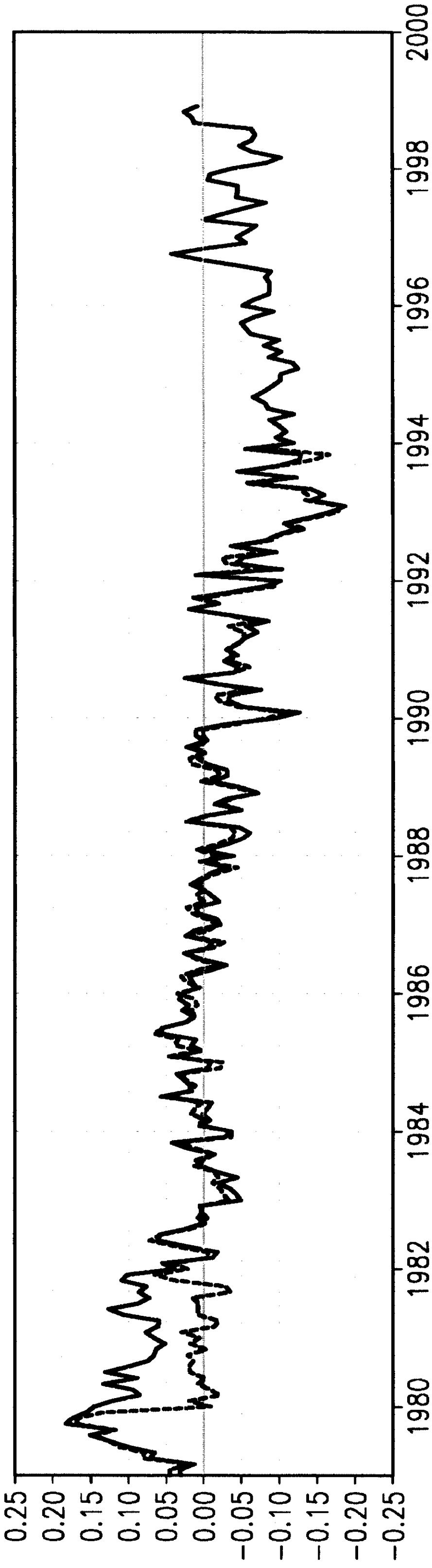


Fig. 6 Time series of the difference (K) between the SST used for NCEP and from IPCC (*solid line*) and between the SST used for ERA and from IPCC (*dashed line*). Details about the SST used for the reanalyses are given in Sect. 2

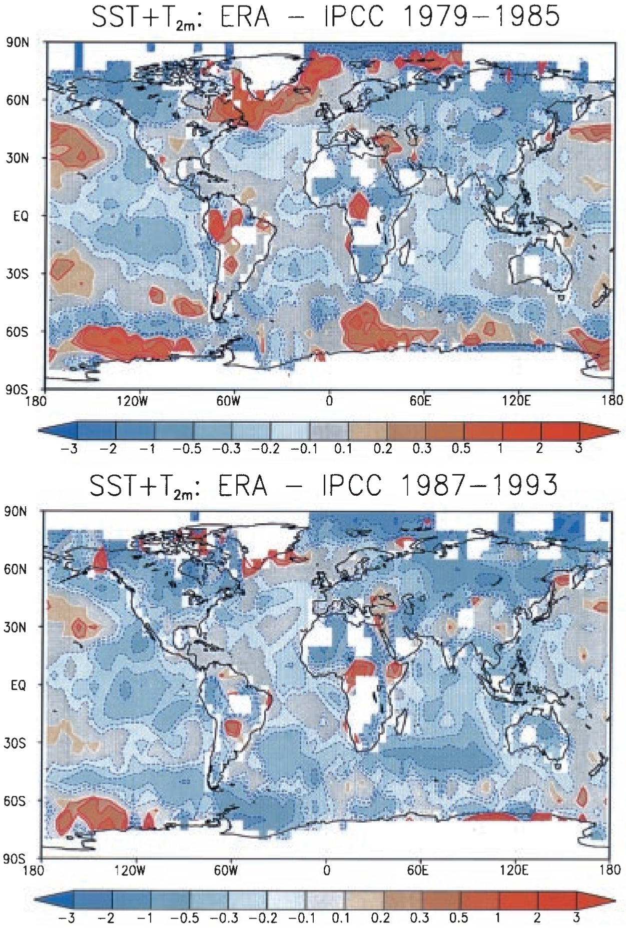


Fig. 7 Difference (K) of the 2 m temperature from ERA and from IPCC over land and of the SST used for ERA and from IPCC over the oceans for the periods 1979–1985 (*upper panel*) and 1987–1993 (*lower panel*). IPCC is the combined Jones et al. (1994)/MOHSST6 data set used in the IPCC second assessment report. Details about the SST used for the reanalyses are given in Sect. 2

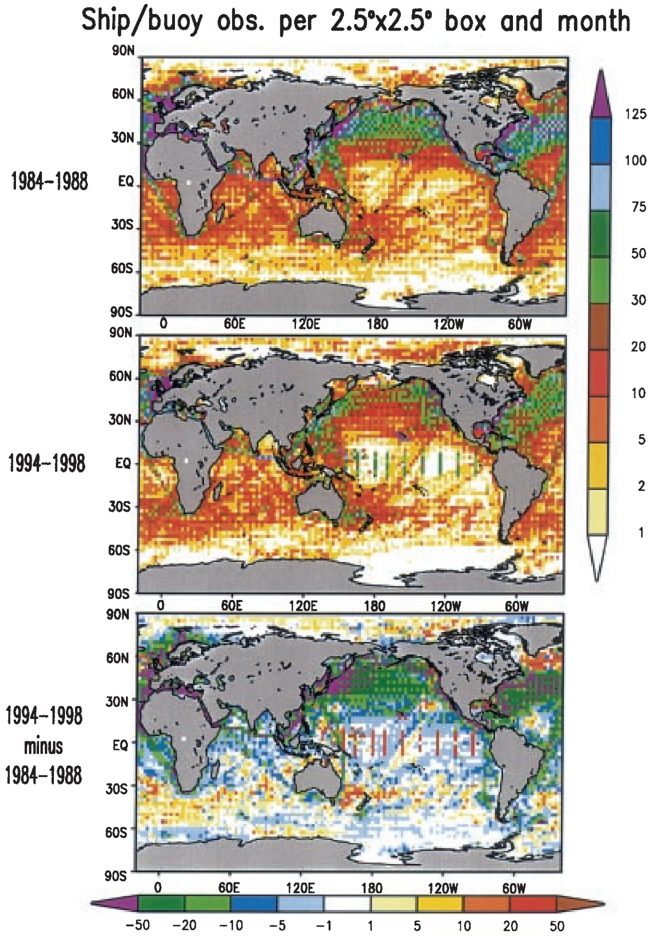


Fig. 8 Number of ship and buoy observations per 2.5° by 2.5° grid box and month. *Upper panel* average for 1984–1988; *middle panel* average for 1994–1998; *lower panel* 1994–1998 minus 1984–1988

Table 4 Data coverage from ships and buoys in the tropical Pacific 1951–1995

Month	Number of ship observations		Number of buoy observations	
	Tropical Pacific	Niño 3.4	Tropical Pacific	Niño 3.4
January 1951	2117	166	0	0
January 1961	5589	140	0	0
January 1971	7838	308	0	0
January 1981	7363	221	221	0
January 1991	12476	307	14772	2292
January 1995	9813	145	7960	443

The tropical Pacific covers the area 20°S to 20°N, 160°E to 80°W. The Niño 3.4 region extends from 5°S to 5°N and from 170°W to 120°W

observations per week, could be utilized. The bias correction is underestimated in situations with sparse in situ data, which may lead to an estimated artificial cooling trend of 0.1 ± 0.1 K/decade (Reynolds 1998) in the Reynolds and Smith (1994) SST compared to the in situ SST. Another reason for a cold bias in the Reynolds and Smith (1994) SST is that this data set uses a temperature of −1.8 °C at the ice edge, whereas in GISST (as well as MOHSST) higher temperatures in marginal ice zones are allowed (Rayner et al. 1996). Furthermore, in situ measurements of the SST observe a different quantity from what is seen from the satellite, which measures the ocean skin temperature, a quantity dependent on conditions (e.g. air-sea heat fluxes and sea state) other than "bulk" SST values as reported by in-situ measurements.

Since the mid-1980s, ship measurements have been supplemented by reports from moored and drifting buoys, in particular in the tropical Pacific (Fig. 8 and Table 4). Although buoy temperatures are observed at a depth of nominally 1 m (whereas ships measure temperatures at depths as much as 10–15 m), both are treated equally in the analysis. Webster et al. (1996) have shown that tropical Pacific SSTs decrease with depth, therefore a warm bias may have been introduced into the in situ data due to the change in the mix of observational platforms. Our preliminary analyses indicate that for the tropical Pacific buoys tend to be warmer than collocated ships by up to 0.13 K for a four-year period (1992–1995). However, for large areal averages there seems to be very little impact (Parker personal communication), so clearly more investigations are necessary, including other factors such as the uncertainty of anomalies produced from buoys which now occupy areas that have had extremely few observations in the past and thus unknown climatologies.

4.2.2 Land surface temperatures

Another possible explanation might be a systematic error in the Jones (1994) data, probably due to a warming bias caused by the ongoing urbanization (for the estimation of sampling errors see also Jones et al. 1997a). To shed some light on this issue, we have compared the reanalyzed temperature anomalies to station data from the North Atlantic Climatological Dataset (NACD, Frich et al. 1996). Temperature time series from stations covering the northern North Atlantic region (Greenland to Finland and Svalbard to Belgium), generally covering the period 1890–1990, have been carefully checked for homogeneity. Out of the 89 available stations, five large cities (Copenhagen, Helsinki, Reykjavik, Oslo, Bergen) were rejected due to environmental changes (urbanization), 42 were found to be inhomogeneous, and four were rejected for other reasons. The time series of the remaining 38 stations are regarded as homogeneous. Comparisons to the reanalyzed and Jones (1994) temperature anomalies (not shown) gave no indication of a warm bias in the Jones dataset, which confirms recent findings by Peterson et al. (1999).

4.2.3 MSU problems

There are periods when problems with the MSU become apparent. Since the MSU was not originally designed as a climate monitoring instrument, special care has to be taken when merging the brightness temperatures of the nine satellites that form the MSU time series up to now. The procedure is described in detail in Christy et al. (2000b) and includes the elimination of the orbit drift of two afternoon satellites, NOAA-7 and NOAA-11. The artificial warming in the drifting afternoon satellites is due to (a) the sampling of Earth emissions at later points in the diurnal cycle and (b) the changing shadowing effect on the instrument itself, which alters the temperature of the instrument. This second effect relates to the non-linear response of the instrument since the response is a function of the difference between the temperature of the scene and the body temperature of the radiometer. As the temperature of the radiometer changes due to the shadowing effect, the calculated Earth temperatures must be adjusted accordingly. In general, these effects, though amounting to only a few hundredths of a degree in any particular year, were not exactly quantified when the instruments were calibrated before launch, but may be potentially important.

In addition to the MSU channel 3 problems in 1980 and 1986 already discussed, Fig. 1 indicates that the MSU is progressively and almost consistently warmer than NCEP between 1991/92 and 1998. Part of this difference may be due to NCEP's snow-cover problems. The redistribution of infrared emitters after the eruption of Mt. Pinatubo in 1991, which may have led to the assignment of the emission to too cold levels, may also explain part of the difference in the early part of the period. Otherwise, the MSU on NOAA-12 (which Mo, 1995, provided new coefficients for) may still need further, slight adjustments, i.e. NOAA-12 may have a slight warming trend remaining in its data.

4.2.4 Radiosonde instrumental changes

Another possible explanation may be related to spurious trends in the radiosonde data due to e.g. instrumental changes. This has been discussed in detail by Gaffen (1994) and Gaffen et al. (2000). In the mid-1990s, US controlled stations began replacing VIZ by Vaisala radiosondes. Figure 9 shows as a preliminary result that this replacement for two stations in the tropical Pacific in December 1995 results in a ~0.5 K cooling of the radiosonde compared to the collocated MSU temperature. This may be responsible for a spurious cooling in the NCEP time series.

4.2.5 Changes in the stability of the lower troposphere

There is evidence for changes in the vertical stability of the lower troposphere on decadal time scales. As

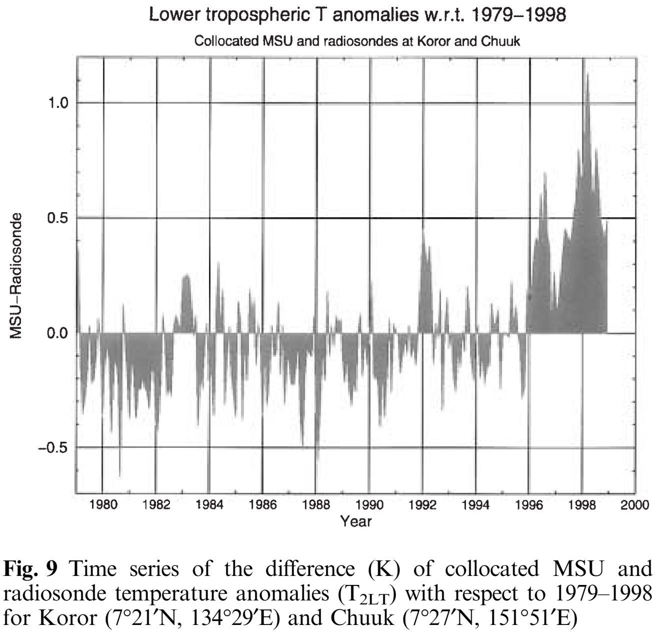


Fig. 9 Time series of the difference (K) of collocated MSU and radiosonde temperature anomalies (T_{2LT}) with respect to 1979–1998 for Koror (7°21′N, 134°29′E) and Chuuk (7°27′N, 151°51′E)

demonstrated by Hurrell (1995, 1996), the North Atlantic oscillation (NAO), which modulates wintertime temperatures over large parts of Eurasia, has shown an upward trend (i.e. toward high index values) in the last three decades. A similar trend was recently also found for the Arctic Oscillation (Thompson and Wallace 1998). Such conditions are related to a more pronounced westerly flow at temperate and high latitudes, leading to markedly warmer conditions in the boundary layer (e.g. due to the destruction of wintertime inversions). Near-surface warming exceeding that in the lower troposphere has been reported for the NH, in particular Russia and Canada (Ross et al. 1996) and for the tropical oceanic regions (Christy et al. 2000a). For the polar regions, Fig. 10 corroborates this finding for the NCEP reanalysis, showing quite large changes of lower tropospheric temperature gradients in particular over the Arctic, where the near-surface temperature has risen considerably with respect to the 700 hPa temperature (trend for 1949–1999: +0.29 K/decade). Although the standard error, accounting for autocorrelation of the time series, is quite large (0.3 K/decade), this indicates that the temperature profile of the lowest 2–3 km of the atmosphere has almost reversed since the 1950s. A similar trend, although smaller in magnitude, is also found in austral winter over Antarctica. For other seasons, the decadal changes are much smaller.

A warming in surface observations compared to both reanalyses is also found over tropical land (Fig. 7). A possible explanation for this might be land-use changes, as recently suggested by Pielke et al. (1999). Apart from the decadal trend in the temperature difference between observed and reanalyzed near-surface temperature anomalies already mentioned, an annual cycle with a winter and spring maximum over the NH land masses, in particular Canada and Russia (not shown), is apparent from Figs. 2 and 4. This annual cycle clearly demonstrates that both reanalyses are flawed with respect to the parametrization of the boundary layer. For NCEP, this may be caused by the improper treatment of snow-cover mentioned already.

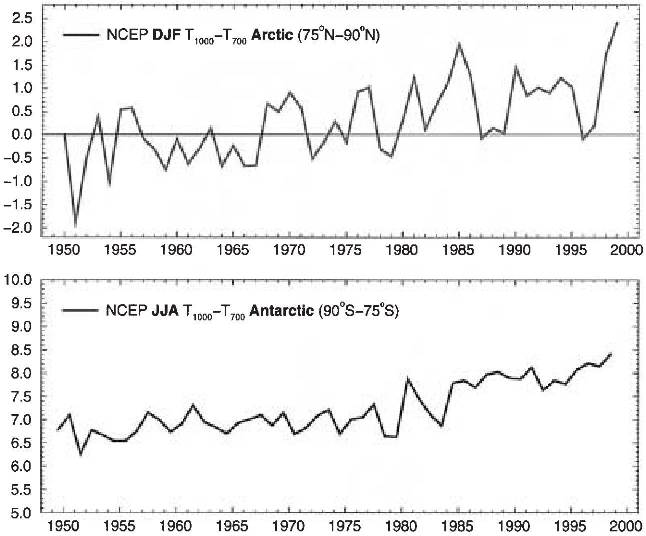


Fig. 10 Time series of lower tropospheric (1000 hPa minus 700 hPa) temperature gradients (K) from the NCEP reanalysis for (*upper panel*) the Arctic (75°N–90°N) in boreal winter (DJF 1950–1999; the year refers to the January of the respective winter) and (*lower panel*) Antarctica (90°S–75°S) in austral winter (JJA 1949–1998)

The vertical distribution of reduced upper- and middle-tropospheric warming due to the decrease of stratospheric ozone may also be a possible cause for differences in temperature trends further below, as first demonstrated by Santer et al. (1996) and Tett et al. (1996). Recently, Bengtsson et al. (1999) demonstrated in an ensemble simulation that the observed warming of the surface relative to the lower troposphere may partly be explained by the combined forcing by well-mixed greenhouse gases, tropospheric ozone and sulfate aerosol as well as stratospheric ozone depletion and sulfate aerosol from the eruption of Mt. Pinatubo.

The period 1979–1998 witnessed two of the three largest volcanic eruptions and two of the three strongest ENSOs of the twentieth century. The lower troposphere responds to such forcing with greater amplitude and later timing of anomalies than the surface (Christy and McNider 1994; Jones 1994). This may also contribute to the difference between near-surface and lower tropospheric trends.

One may furthermore speculate that effects due to changed cloud cover and subsequent changes in the hydrological cycle may also contribute to the difference between near-surface and tropospheric temperatures. However, much more research needs to be done before quantitative comments can be made.

5 Conclusions

The major findings of this investigation can be summarized as follows.

1. The estimation of recent temperature trends from surface measurements, radiosondes, satellite observations or reanalyses is subject to a considerable degree of uncertainty. In the data-rich regions of the NH, temperature anomalies obtained from MSUs, radiosondes and from both ECMWF and NCEP reanalyses agree reasonably. For regions without conventional observations from radiosondes, large differences are found between the MSU and either reanalysis. Since the MSU is calibrated globally, this implies that in data-sparse regions (and thus the globe) tropospheric temperature trends cannot reliably be estimated from reanalyses, corroborating earlier findings by Santer et al. (1999).

2. Both reanalyses are sensitive to changes in satellite data availability. There are a number of specific problems with the treatment of satellite data in both of the reanalyses. NCEP employs NESDIS retrievals based on raw radiances using algorithms that have undergone alterations over time. Thus, intersatellite biases are not

specifically calculated, and the product is very sensitive to changes in the algorithms (e.g. August 1992, Fig. 3). ERA uses only cloud cleared radiances and performs a "one-dimensional variational analysis". The results of such an analysis have to be bias-corrected, which introduces an error when the satellite relative to which the correction is calculated is biased itself or when radiances change on a time scale longer than a couple of months, e.g. due to orbit decay. ERA inhomogeneities are visible in April 1985, October/November 1986 and April 1989.

3. After accounting for the uncertainties with temperature trends calculated from reanalyses, residual negative trends remain in the reanalyses compared to both SST and land surface temperature. While trends in the in situ SST may be introduced by inhomogeneities in the observing system (degradation of the number and, probably, also quality, of ship measurements during the last decade, replacement of ship observations by buoy measurements), there are no indications of inhomogeneities in land surface temperature trends. Near-surface temperatures in both reanalyses have biases, particularly in winter, indicating problems with the parametrization of the boundary layer, presumably due to the incorrect treatment of snow-cover. There is evidence for a slight increase of the tropospheric lapse rate over the last two decades, which may be caused by a combination of several factors, including recent circulation changes, stratospheric ozone decrease and volcanic effects.

Acknowledgements The authors would like to thank Klaus Arpe, Chris Folland, Per Kållberg, David Parker, Erich Roeckner, Ben Santer and Sakari Uppala for helpful discussions and clarifying comments. Karl Taylor and an anonymous reviewer helped to improve the manuscript. Technical assistance was given by the German Climate Computing Centre. The study was supported under the EU contract ENV4-CT95-0099.

References

Andersson E, Hollingsworth A, Kelly G, Lönnberg P, Pailleux J, Zhang Z (1991) Global observing system experiments on operational statistical retrievals of satellite sounding data. Mon Weather Rev 119: 1851–1864

Angell JK (1988) Variations and trends in tropospheric and stratospheric global temperatures, 1958–87. J Clim 1: 1296–1313

Basist AN, Chelliah M (1997) Comparison of tropospheric temperatures derived from the NCEP/NCAR reanalysis, NCEP operational analysis and the microwave sounding unit. Bull Am Meteorol Soc 78: 1431–1447

Bengtsson L, Shukla J (1988) Integration of space and in situ observations to study global climate change. Bull Am Meteorol Soc 69: 1130–1143

Bengtsson L, Roeckner E, Stendel M (1999) Why is the global warming proceeding much slower than expected? J Geophys Res 104: 3865–3876

Christy JR (1995) Temperature above the surface layer. Clim Change 31: 455–474

Christy JR, McNider NT (1994) Satellite greenhouse signal. Nature 367: 325

Christy JR, Spencer RW (1995) Assessment of precision in temperatures from the microwave sounding units. Clim Change 30: 97–102

Christy JR, Spencer RW, Lobl ES (1998) Analysis of the merging procedure for the MSU daily temperature time series. J Clim 11: 2016–2041

Christy JR, Parker DE, Brown SJ, Macadam I, Stendel M (2000a) Differential trends in tropical sea surface and atmospheric temperatures since 1979. Geophys Res Lett (submitted)

Christy JR, Spencer RW, Braswell WD (2000b) MSU tropospheric temperatures: dataset construction and radiosonde comparisons. J Atmos Ocean Technol (in press)

Eyre JR (1992) A bias correction scheme for simulated TOVS brightness temperatures. ECMWF Techn Mem 186

Eyre JR, Kelly GA, McNally AP, Andersson E, Persson A (1993) Assimilation of TOVS radiance information through one-dimensional variational analysis. Q J R Meteorol Soc 119: 1427–1463

Folland CK, Parker DE (1995) Correction of instrumental biases in historical sea surface temperature data. Q J R Meteorol Soc 121: 319–367

Frich P, Alexandersson H, Ashcroft J, Dahlström B, Demarée GR, Drebs A, van Engelen AFV, Hanssen-Bauer I, Heino R, Jónsson T, Jonasson K, Keegan L, Nordli PØ, Schmith T, Steffensen P, Tuomenvirta H, Tveito OE (1996) North Atlantic climatological dataset (NACD version 1) - Final report. Danish Meteorological Institute, Scientific Report 96–1, 47 pp + app

Gaffen DJ (1994) Temporal inhomogeneities in radiosonde temperature records. J Geophys Res 99: 3667–3676

Gaffen DJ, Sargent MA, Habermann RE, Lanzante JR (2000) Sensitivity of tropospheric and stratospheric temperature trends to radiosonde data quality. J Clim (accepted)

Gibson JK, Kållberg P, Uppala S, Hernandez A, Nomura A, Serrano E (1997) ECMWF reanalysis project report series, 1. ERA description. European Centre for Medium-Range Weather Forecasts, Reading/UK, 72 pp

Hansen J, Wilson H, Sato M, Ruedy R, Shah K, Hansen E (1995) Satellite and surface temperature data at odds? Clim Change 30: 103–117

Hurrell JW (1995) Decadal trends in the North Atlantic Oscillation: regional temperatures and precipitation. Science 269: 676–679

Hurrell JW (1996) Influence of variations in extratropical wintertime teleconnections on Northern Hemisphere temperature. Geophys Res Lett 23: 665–668

Hurrell JW, Trenberth KE (1996) Satellite versus surface estimates of air temperature. J Clim 9: 2222–2232

Hurrell JW, Trenberth KE (1997) Spurious trends in satellite MSU temperatures from merging different satellite records. Nature 386: 164–167

Hurrell JW, Trenberth KE (1998) Difficulties in obtaining reliable temperature trends: reconciling the surface and satellite microwave sounding unit records. J Clim 11: 945–967

Hurrell JW, Trenberth KE (1999) Global sea surface temperature analyses: multiple problems and their implications for climate analysis, modeling and reanalysis. Bull Am Meteorol Soc 80: 2661–2678

Jones PD (1994) Recent warming in global temperature series. Geophys Res Lett 21: 1149–1152

Jones PD, Osborn TJ, Briffa KR (1997a) Estimating sampling errors in large-scale temperature averages. J Clim 10: 2548–2568

Jones PD, Osborn TJ, Wigley TML, Kelly PM, Santer BD (1997b) Comparisons between the microwave sounding unit temperature record and the surface temperature record from 1979 to 1996: real differences or potential discontinuities? J Geophys Res 102: 30135–30145

Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen J, Zhu Y, Leetmaa A, Reynolds R (1996) The NCEP/NCAR 40-year reanalysis project. Bull Am Meteorol Soc 77: 437–471

Karl TR, Knight RW, Christy JR (1994) Global and hemispheric temperature trends: uncertainties related to inadequate spatial sampling. J Clim 7: 1144–1163

Kelly G, Andersson E, Hollingsworth A, Lönnberg P, Pailleux J, Zhang Z (1991) Quality control of operational physical retrievals of satellite sounding data. Mon Weather Rev 119: 1866–1880

Manabe S, Stouffer R (1996) Low-frequency variability of surface air temperature in a 1000-year integration of a coupled atmosphere-ocean-land surface model. J Clim 9: 376–393

Mo T (1995) A study of the Microwave Sounding Unit on the NOAA-12 satellite. IEEE Trans. Geosci Remote Sens 33: 1141–1152

Mo K, Wang XL, Kistler R, Kanamitsu M, Kalnay E (1995) Impact of satellite data on the CDAS-reanalysis system. Mon Weather Rev 123: 124–139

Nicholls N, Gruza GV, Jouzel J, Karl TR, Ogallo LA, Parker DE (1996) Observed climate variability and change. In: Houghton JT, Meira Filho LG, Callander BA, Harris N, Kattenberg A, Maskell K (eds) Climate change 1995: the science of climate change Cambridge University press, Cambridge, UK, pp 133–192

Nomura A (1995) Global sea-ice concentration data set for use with the ECMWF re-analysis system. ECMWF Tech Rep 76

Parker DE, Jones PD, Folland CK, Bevan A (1994) Interdecadal changes of surface temperature since the late 19th century. J Geophys Res 99: 14 373–14 399

Parker DE, Folland CK, Bevan A, Ward MN, Jackson M, Maskell K (1995a) Marine surface data for analysis of climatic fluctuations on interannual to century time scales. In: Martinson DG, Bryan K, Ghil M, Hall MM, Karl TR, Sarachik ES, Sorooshian S, Talley LD (eds) Natural climate variability on decade-to-century time scales. National Academy Press, Washington D.C., pp 241–250 and color figures on pp 222–228

Parker DE, Folland CK, Jackson M (1995b) Marine surface temperature: observed variations and data requirements, Clim Change 31: 559–600

Parker DE, Gordon M, Cullum PN, Sexton DMH, Folland CK, Rayner N (1997) A new global gridded radiosonde temperature data base and recent temperature trends. Geophys Res Lett 24: 1499–1502

Pawson S, Fiorino M (1998) A comparison of reanalyses in the tropical stratosphere, 3, The quasi-biennal oscillation. Clim Dyn 14: 645–658

Peterson TC, Gallo KP, Livermore J, Owen TW, Huang A, McKittrick DA (1999) Global rural temperature trends. Geophys Res Lett 26: 329–332

Pielke RA Sr, Walko RL, Steyart LT, Vidale PL, Liston GE, Lyons WA, Chase TN (1999) The influence of anthropogenic landscape changes in weather in South Florida. Mon Weather Rev 127: 1663–1673

Prabhakara C, Cadeddu M, Yoo JM, Arking A, Dalu G (1996) Examination of global atmospheric temperature monitoring with satellite microwave measurements, 3. Cloud and rain contamination. Clim Change 33: 491–496

Rayner NA, Horton EB, Parker DE, Folland CK, Hackett RB (1996) Version 2.2 of the global sea-ice and sea surface temperature data set, 1903–1994. Hadley Centre Climate Research Technical Note CRTN-74

Reale A, Chalfant M, Wagoner R, Gardner T, Casey L (1994) TOVS operational sounding upgrades: 1990–1992. NOAA Tech Rep NESDIS 76, US Department of Commerce, Washington, DC, 67 pp

Reynolds RW (1988) A real-time global sea surface temperature analysis. J Clim 1: 75–86

Reynolds RW (1998) SST comparison. Available via Internet at http://nic.fb4.noaa.gov:8000/research/cmb/reynolds/uk_oi.html

Reynolds RW, Folland CK, Parker DE (1989) Biases in satellite derived sea-surface-temperatures. Nature 341: 728–731

Reynolds RW, Smith TM (1994) Improved global sea surface temperature analyses using optimum interpolation. J Clim 7: 929–948

Roeckner E, Bengtsson L, Feichter J, Lelieveld J, Rodhe H (1999) Transient climate change simulations with a coupled atmosphere-ocean GCM including the tropospheric sulfur cycle. J Clim 12: 3004–3032

Ross RJ, Otterman J, Starr DO'C, Elliott WP, Angell JK, Susskind J (1996) Regional trends of surface and tropospheric temperature and evening-morning temperature difference in northern latitudes: 1973–93. Geophys Res Lett 23: 3179–3182

Santer BD, Taylor KE, Wigley TML, Johns TC, Jones PD, Karoly DJ, Mitchell JFB, Oort AH, Penner JE, Ramaswamy V, Schwarzkopf MD, Stouffer RJ, Tett S (1996) A search for human influences on the thermal structure of the atmosphere. Nature 382: 39–46

Santer BD, Hnilo JJ, Wigley TML, Boyle JS, Doutriaux C, Fiorino M, Parker DE, Taylor KE (1999) Uncertainties in observational estimates of temperature change in the free atmosphere. J Geophys Res 104: 6305–6333

Santer BD, Wigley TML, Boyle JS, Gaffen D, Hnilo JJ, Nychka D, Parker DE, Taylor KE (2000) Statistical significance of trend differences in layer-averaged temperature time series. J Geophys Res (accepted)

Singer SF (1996) Climate change and consensus. Science 271: 581–582

Spencer RW, Christy JR (1992) Precision and radiosonde validation of satellite gridpoint temperature anomalies, II: a tropospheric retrieval and trends during 1979–90. J Clim 5: 858–866

Spencer RW, Christy JR, Grody NC (1996) Analysis of 'Examination of global atmospheric temperature monitoring with satellite microwave measurements'. Clim Change 33: 477–489

Stendel M, Arpe K (1997) ECMWF reanalysis project report series, 6. Evaluation of the hydrological cycle in reanalysis and observations. European Centre for Medium-Range Weather Forecasts, Reading/UK, 62 pp

Stendel M, Bengtsson L (1997) Toward monitoring the tropospheric temperature by means of a general circulation model. J Geophys Res 102: 29 779–29 788

Sterin AM, Orshekhovskaya VA, Mishina NM (1997) Comparison of upper-air temperature variations in the past and current decade, derived from the global radiosonde database and from the microwave sounding unit. Proc 22nd Ann. Climate Diagnostics and Prediction Workshop, Berkeley, CA, USA, 214–218. US Department of Commerce, Sills Bldg., 5285 Port Royal Road, Springfield, VA 22161

Tett SFB, Mitchell JFB, Parker DE, Allen MR (1996) Human influence on the atmospheric vertical temperature structure. Science 274: 1170–1173

Thompson DWJ, Wallace JM (1998) The Arctic Oscillation signature in the wintertime geopotential height and temperature fields. Geophys Res Lett 25: 1297–1300

Trenberth KE, Olson JG (1991) Representativeness of a 63-station network for depicting climate changes. In: Schlesinger ME (ed) Greenhouse-gas-induced climatic change: a critical appraisal of simulations and observations. Elsevier, Amsterdam, pp 249–259

Trenberth KE, Hurrell JW (1997) Reply to "How accurate are satellite 'thermometers'?" Nature 389: 342–343

Uppala S (1997) ECMWF reanalysis project report series, 3. Observing system performance in ERA. European Centre for Medium-Range Weather Forecasts, Reading/UK, 261 pp

Wallis TWR (1998) A subset of core stations from the Comprehensive Aerological Reference Dataset (CARDS). J Clim 11: 272–282

Webster PJ, Clayson CA, Curry JA (1996) Clouds, radiation and the diurnal cycle of sea surface temperatures in the tropical western Pacific. J Clim 9: 1712–1730

Wentz FJ, Schabel M (1998) On the discrepancy between observed in situ surface warming and the cooling trend in MSU tropospheric temperatures. Nature 394: 661–664