GCOS Reference Upper-Air Network

GRUAN Report 3

Review of Operational Requirements for Temperature Sonde Measurements

DAVID BUTTERFIELD AND TOM GARDINER

Publisher

GRUAN Lead Centre

Number & Version

GRUAN-RP-3
Rev. 1.0 (2014-05-05)

Document Info

Title:	Review of Operational Requirements for Temperature Sonde Measurements
Topic:	Temperature Measurements
Authors:	David Butterfield and Tom Gardiner
Publisher:	GRUAN Lead Centre, DWD
Document type:	Report
Document number:	GRUAN-RP-3
Page count:	34
Revision / date:	1.0.0 / 2014-05-05

Abstract

This document draws together the results of a literature review on the operational requirements for temperature measurements made from radiosondes to underpin a global reference network of such measurements. The objective of this review was to bring together the information in the peer-reviewed literature to provide guidance to the GRUAN community on the requirements for sonde temperature measurements, covering aspects such as measurement scheduling, measurement uncertainty, change management and network design.

The report is focussed on covering issues relating to climate/trend measurement requirements and does not specifically cover issues relating to satellite calibration / validation or Numerical Weather Prediction (NWP).

Revision History

Version	Author / Editor	Description
0.7 (2013-12-12)	D. Butterfield	First document
0.8 (2013-12-30)	T. Gardiner	Reformatted as GRUAN Technical Note – first draft
0.9 (2014-02-11)	M. Sommer [Ed.]	Renamed as GRUAN Report 3 – prepared for review process
1.0-pre (2014-04-30)	D. Butterfield	Redrafted in response to review comments
1.0 (2014-05-05)	M. Sommer [Ed.]	Reformatted for publishing → first published version

Table of Contents

1 Scope

This document draws together the results of a literature review on the operational requirements for temperature measurements made from radiosondes to underpin a global reference network of such measurements. The review was led by David Butterfield as part of the activities of the GRUAN Scheduling Task Team.

The objective of this review was to bring together the information in the peer-reviewed literature to provide guidance to the GRUAN community on the requirements for sonde temperature measurements, covering aspects such as measurement scheduling, measurement uncertainty, change management and network design.

The report focusses on covering issues relating to climate/trend measurement requirements and does not specifically cover the requirements relating to satellite calibration / validation or Numerical Weather Prediction (NWP). However, where the papers reviewed include aspects on these issues these are noted in the bibliography. The operational measurement requirements surrounding satellite cal/val and NWP are extensive, but it is hoped to extend this review to cover these issues in the future.

The main conclusions of the review are given in sections 2 to 5, and a bibliography containing short summaries of the papers used to inform these requirements and recommendations is given in the Annex.

2 Scheduling Issues

Sampling twice daily, at 0000 and 1200 UTC, ensures that monthly statistics will be statistically significantly different from those based on four observations per day in only ~5% of the cases [A.4]. However, sampling once daily introduces biases in monthly mean temperatures. Large errors result from changing from 0000 to 1200 UTC observations (or vice versa). Twice-daily sampling must be done at least once every two days to ensure that monthly means are accurate to within 2 K. Sampling every two days, or every three days (but not every seven days), yields monthly means and standard deviations that are not significantly different from the true values at least 99.5% of the time.

The effects of changing observing schedule [A.4] can be seen in the table below:

Schedule 1 to Schedule 2	Trend error rate*	Trend period
0000 and 1200 daily to 0000 and 1200 every other or every 3rd day	5%	20 – 50 year
0000 and 1200 daily to 0000 and 1200 weekly	5% 13%	20 year 50 year
0000 and 1200 daily to 0000 or 1200 daily	4% 17%	20 year trend 50 year trend
0000 to 1200 daily or vice versa	14% 17%	20 year trend 50 year trend

Note:

* *The percentage of trends in time series with an observing schedule change introduced at the midpoint that are statistically significantly different from trends in unaltered time series.*

The smaller error rate for the shorter time periods is because the uncertainties are larger for trends with shorter time periods, and therefore differences in trends are less likely to be significant.

The conclusions are:

- Maintaining a constant time of observation is more important than maintaining daily observations for avoiding errors in temperature trend estimates.
- Measurements should be made at least twice daily at 0000 and 1200 UTC to try and avoid bias in monthly means.

3 Temperature Measurement Uncertainty

In atmospheric profile measurements using sondes the uncertainty needs to be determined for each data point (at each altitude) individually. All sources of uncertainty should be combined to produce an uncertainty budget [A.7].

The guidance document for the GCOS Upper Air Reference Network [A.7] (GRUAN) outlines requirements for high quality reference sites that would be used to underpin the Network. These sites would be few in number (100 or less across the globe) but be able to provide measurements of the highest quality and can also be used as inter comparison sites for less accurate measurement technologies. GRUAN recommends overall measurement uncertainties, expressed with a level of confidence of 95%, as:

- 0.1 K with a vertical resolution of 100 m in the troposphere
- 0.2 K with a vertical resolution of 500 m in the stratosphere

The following assumptions about individual uncertainty contributions to the overall measurement uncertainty have been made:

1. Temperature sensor calibration uncertainty <0.1 K, and calibrations are traceable to national standards.
2. Calibration uncertainty is considered to be an altitude-independent absolute systematic contribution to the uncertainty profile. Temperature transducers are normally calibrated on the ground across the full temperature range and this calibration is independent of the altitude that the actual temperature measurement was made at.
3. The contribution of the uncertainty in frequency measurement of the resonant circuit within the temperature transducer to the total uncertainty of the temperature sensor is negligible.
4. The effect of radiative balloon heating or adiabatic balloon cooling on the temperature data is considered to be negligible, provided the rope between balloon and radiosonde is at least 40m.
5. The time-lag of modern temperature sensors is generally less than 1 second over the entire temperature range. The temperature during the ascent generally varies by less than a tenth of a degree in this time frame. Therefore, the bias caused by the time lag of the temperature sensor during the ascent can be neglected.
6. When the radiosonde emerges into dryer air above a cloud, evaporation of the condensed water cools the sensor and creates a cool bias in this region. This effect can lead to deviations up to 1K above a cloud and will need to be flagged in the data accordingly with an increased uncertainty during this period.
7. Radiation correction is normally the largest source of uncertainty and must be clearly documented in the QA procedures.

The following recommendations are made:

- All raw data should be recorded before calibration corrections are applied. All calibration coefficients and corrections (i.e. radiation effect) must be recorded so that data can be reprocessed at a later date.
- In order to provide traceable data it is necessary to use common and well documented processing algorithms so that common and well defined (with respect to the applied corrections and filtering) temperature profiles are produced. This is necessary for the processing of higher (data quality) level data.

Uncertainty requirements for temperature measurements set in December 2012 by the World Climate Research Programme[A.14] (WCRP) are given in the following table:

Requirement	Standard Uncertainty
Threshold	2.0 K
Breakthrough	0.5 K
Goal	0.2 K

Note that these uncertainties are expert opinion and not from peer reviewed literature.

Definitions:

- The “threshold” is the minimum requirement to be met to ensure that data are useful.
- The “goal” is an ideal requirement above which further improvements are not necessary.
- The “breakthrough” is an intermediate level between “threshold” and “goal” which, if achieved, would result in a significant improvement for the targeted application. The breakthrough level may be considered as an optimum, from a cost-benefit point of view, when planning or designing observing systems.

The proposed GRUAN limits are the same or stricter than the WCRP requirements.

4 Effects of Procedural and Instrumental Changes

The greatest challenge for detecting temperature trends in observational data is presented not by measurement precision or sampling frequency but by temporal inhomogeneities in time series. These can be introduced by changes in instrumentation, observing practices, or data-processing methods, and result in time-varying biases that can masquerade as, or mask, true trends.

The table below shows the effect of an unknown, but constant, single random timed intervention on a time series[A.4]:

Stability, K	Error rate*
0.10	1% for 20 – 50 year period
0.25	2 – 5% for 20 – 50 year period
0.50	3 – 12% for 20 – 50 year period
2.00	15 – 35% for 20 – 50 year period

Note:

* *The percentage of trends that are statistically significantly different from the actual reanalysis trends.*

Long-term data stability is the primary determiner of trend estimate accuracy. By maintaining temperature measurement stability to within 0.1 K, for periods of 20 to 50 years, errors in trend estimates can be avoided in at least 99% of cases [A.4]. The worst case scenario [A.5], where an intervention occurs in the middle of a measuring period and the magnitude of intervention is unknown but constant, results in the time required to detect a trend increasing by a factor of 1.59.

Reducing the precision (increasing the random error) of temperature data has minor effects on monthly means and standard deviations [A.4] and is not an important factor in determining multi-decadal trends.

The main instrumental recommendations are:

- To ensure long-term data stability all procedural and instrumentation changes must be fully documented and parallel intercomparisons performed to allow for the accurate assessment of such a change.
- All raw data should be recorded before calibration corrections are applied. All calibration coefficients and corrections (i.e. radiation effect) must also be recorded so that data can be reprocessed at a later date.

5 Network Requirements

McCarthy[A.9] proposes the maximum spacing of radiosonde sites to keep the uncertainty of global mean temperature trends <0.05 K·decade^{-1} in the troposphere and <0.1 K·decade^{-1} in the stratosphere of:

- 30° longitude and 15° latitude, north of 30° N
- 20° longitude and 10° latitude, south of 30° N

To implement the above a network of 248 stations would be required (49 stations at or north of 30° N and 199 stations south of 30° N).

Analysis of GRUAN station location data (2008 [A.9]) shows that there is a lack of coverage over India, central Africa and northern hemisphere high latitudes and potential over sampling in parts of Europe and East Asia.

In contrast Free & Seidel [A.1] suggest a relatively small [A.1] (<100 stations) well designed network may be almost as good as larger ones for monitoring long-term temperature trends due to unrepresentative distribution of additional sites.

A high quality reference series is important for trend recovery, and this may be a problem when a biased sparse network is used [A.14]. If traceable data was available from a high quality GRUAN or similar-sized network, then it is very likely that this could adequately constrain the uncertainties in the trends for the rest of the global network. It is therefore recommended that the GRUAN network is maintained to a high standard, including adherence to the GCOS monitoring principles to provide a better understanding of future trends in the free atmosphere. Additional RS92 data from GUAN sites could be processed in the same manner as the GRUAN data to provide additional global representation.

Annex A Bibliography and Notes

A wide range of papers were reviewed, but only those of direct relevance to the scope are discussed below. The following papers were included for their contributions on temperature trends, relationships between radiosondes, satellites and modelled measurements, measurement uncertainty and operational requirements. As this document's main focus is on the operational requirements for current and future GRUAN measurements, papers on measurement uncertainty and temperature artefacts from radiosondes no longer in regular use were not included. The key points of relevance from the selected papers are summarised below, followed by a summary of the trend results in the different papers.

A.1 Causes of differing temperature trends in radiosonde upper air data sets, M Free & D Seidel, Journal of Geophysical Research, Vol 110, D07101, doi:10.1029/2004JD005481, 2005.

Link → http://dx.doi.org/10.1029/2004JD005481

A.1.1 Main conclusion

Homogeneity adjustments, sampling differences, and differences in input data make roughly comparable contributions to total differences between trends in the LKS and HadRT temperature data set for 1979–1997. It follows that to narrow uncertainties in radiosonde temperatures, we must consider all three sources of disagreement.

Several radiosonde temperature data sets have been created in recent years, including the Angell, Lanzante-Klein-Seidel (LKS) and Hadley Centre (HadRT) data sets. These temperature records are important for climate change detection and attribution studies, however the data sets do not give the same trends for large-scale means on multi-decade timescales. They also vary from the measured trends using data from layer mean and microwave sounding unit (MSU) satellite equivalent temperatures. Differences between temperature trends in the tropics derived from the HadRT and LKS radiosonde data sets for a mid-tropospheric layer corresponding to that measured by the MSU channel 2 satellite product were 0.1 K·decade^{-1} (-0.132 for HadRT versus -0.032 for LKS) for 1979–1997. Global radiosonde temperature trends differed from actual MSU satellite trends by 0.16–0.31 K·decade^{-1} in the stratosphere, or 26–51% of the mean trend.

The paper assesses the relative contribution of special sampling differences, temporal sampling differences, differences in the original data, adjustments for inhomogeneity and other differences in processing. It then moves to examine the differences in trends in the HadRT and LKS trends. The paper focuses on the HadRT and LKS data sets.

The two data differ in the observation times used at individual stations, as well as in the methods used for quality control and homogeneity adjustment. The HadRT data has been adjusted for changes in instruments and procedures after 1979. HadRT has been adjusted only above 200 mbar. The LKS data set uses a different method and removes inhomogeneities at all levels in the atmosphere. In addition, HadRT data sets are gridded products, while LKS consists of individual station data.

A.1.2 Sampling Error

Spatial Sampling Error Estimated From Globally Complete Data Sets

The global data sets for MSU and National Centres for Environmental Prediction were subsampled for the 444 stations used in the HadRT data set. The standard deviation in trend for the global data set was in the order of one half to two thirds to the actual radiosonde data. From this difference it is assumed that the uncertainty estimates from the subsampling method are likely to be smaller than the actual sampling uncertainties.

The differences between full and subsampled global mean trends for 1979–1997 range from less than 0.002 K·decade^{-1} for the GUAN at 50 mbar to 0.071 K·decade^{-1} for the same network at 200 mbar (or 24% of the trend from the complete reanalysis data set). At 850, 500, and 50 mbar, most differences are less than 0.05 K·decade^{-1}. For comparison, standard error of the trends in the reanalysis is ~0.07 K·decade^{-1} for 850 mbar and 500 mbar, ~0.2 K·decade^{-1} at 200 mbar, and ~0.5 K·decade^{-1} at 50 mbar. There are no clear overall best or worst networks and no apparent relationships between size of network and size of sampling error. Trends for 1960–1997 were also examined and similar overall results were found. When errors are plotted as a function of network size, there is again no apparent relationship between network size and spatial sampling error. Although errors for the LKS network are smaller overall than for the Angell networks, large outliers from the GUAN and HadRT networks suggest no reliable improvement with increasing network size.

To explore this result 13 hypothetical networks were constructed with approximately evenly spaced locations with network sizes from 48 to 1200 sites. The spatial sampling trend uncertainties in these networks decline as network size increases from 48 to around 400 locations and are then roughly similar for larger networks. The uncertainty for networks larger than 360 stations is less than 0.02 K·decade^{-1}, in contrast to the much larger errors for the HadRT Network. The paper proposes that the advantage of larger network size is overcome by uncertainties due to unrepresentative distribution of additional sites. **The results suggest that a relatively small (fewer than 100 stations) well designed network may be almost as good as larger ones for monitoring long-term temperature trends.**

Temporal Sampling Error From Globally Complete Data Sets

Temporal sampling differences resulting from missing monthly means at radiosonde sites can be simulated in complete data sets by removing the corresponding data and then reanalysing.

The difference between trends in data subsampled in time only and trends in data sampled in both space and time gives an estimate of the effect of temporal sampling error in the LKS and HadRT networks The uncertainties for 1979–1997 trends are mostly less than 0.02 K·decade^{-1} except in the stratosphere and for reanalysis at 200 mbar. In some cases, particularly in the stratosphere, temporal sampling errors are larger than spatial sampling errors.

Total sampling uncertainties (differences between full grid trend and trend from data subsampled in both time and space) from reanalysis tests range from less than 0.001 K·decade^{-1} for HadRT at 850 mbar in the tropics to 0.098 K·decade^{-1} for LKS at 850 mbar in the SH for 1979–1997. In the SH, LKS errors are noticeably larger than those for HadRT, but in the global mean the two networks have similar overall performance. For MSU data the largest uncertainty is for HadRT in the SH stratosphere. MSU total sampling error estimates elsewhere are much smaller than those from reanalysis and show little difference between HadRT and LKS.

Estimates of Sampling Error Using Radiosonde Data

The HadRT radiosonde data set was subsampled in space and time for 71 stations common to both the HadRT and LKS networks, out of the 444 available and the uncertainty in trends compared. While the overall magnitude of sampling effects is similar in the radiosonde and reanalysis experiments, the differences at individual levels and regions show little relation. For example, at 850 mbar in the SH, where the reanalysis shows a difference of 0.140 K·decade^{-1} between trends in the 444 and 71 station networks, the actual radiosonde data show a difference of only -0.003 K·decade^{-1}.

The same test was performed using the UAH MSU data. As with the reanalysis experiment the actual sampling effect and that estimated from UAH MSU data often differ in size and show different vertical and regional patterns. Thus the sampling uncertainty estimates from reanalysis and MSU data are not a fool proof guide to the effect of network selection in actual radiosonde data. Some differences between estimated and actual errors are expected given the differences between the quantities measured by radiosondes and those measured by satellites and the reanalysis. Nevertheless, the results suggest the possibility that the large-scale geographic patterns of trends in the reanalysis and MSU may not be sufficiently similar to those in radiosonde upper air data to permit confident assessment of sampling uncertainty. Alternatively, the effects of small-scale sampling uncertainty and random instrumental uncertainties present in the radiosonde data sets but not in the globally complete data sets may be so large as to overwhelm effects of large-scale sampling errors in the radiosonde data.

A.1.3 Comparison of Trends in Radiosonde and MSU Data Sets Using Similar Sampling

MSU versus LKS and HadRT

The results show that inadequate spatial coverage and missing months of data are not major causes of differences between MSU and radiosonde trends. More likely sources of discrepancies are time-varying biases in one or more data sets, differences in processing and adjustments, and the inherent difference between the point measurement of a radiosonde and the horizontally and vertically averaged measurement of the satellite instrument.

HadRT versus LKS

71 locations common to the LKS and HadRT networks (16 of these locations included more than one station in the relevant HadRT grid box) were compared after removing data for missing months. From this comparison it was found that stations in Calcutta and Bombay had significantly large differences in trends between the two networks. These stations were deleted from further trend analysis. Of the 59 identical stations present, 7 have trends that differ by more than 0.2 K·decade^{-1} and 8 more differ by more than 0.1 K·decade^{-1}.

A.1.4 Comparison of Sampling, Homogeneity Adjustment, and Source Data Effects on Trends in LKS and HadRT Data Sets

LKS and HadRT Networks use different methods to adjust for differences in temporal homogeneity in their data sets. Comparing the unadjusted and adjusted data provide a measure of the impact of the homogeneity adjustments. The effects of the LKS and HadRT adjustments are opposite in sign in the troposphere (LKS adjustments increase the warming, and HadRT adjustments increase the cooling trend in their respective data sets), while both reduce the cooling trend in the stratosphere in the global, hemispheric, and tropical means.

The effects of adjustments in the troposphere in the global mean (1979–1997) is roughly similar in size to the effects of sampling differences and differences in input data and is of the same sign. In the tropics the pattern is similar, but the effects are larger, with almost 0.2 K·decade^{-1} total difference in trend at 300 mbar. If the tropics are limited to 20° S–20° N the total difference between data sets is larger, and sampling effects predominate, while source data effects become minimal. In the stratosphere, large effects of opposite sign result in smaller net differences between trends in the two data sets. The relative roles of the three factors differ in the SH and NH and when we use a smaller subset of locations with identical stations.

A.2 An Update of Observed Stratospheric Temperature Trends, W Randal et al, Journal of Geophysical Research, Vol 114, D02107, doi:10.1029/2008JD010421, 2009.

Link → http://dx.doi.org/10.1029/2008JD010421

A major difficulty in developing understanding of stratospheric temperature trends are uncertainties regarding the homogeneity of observational data. The available monitoring systems have been designed primarily to provide information for weather forecasting or shorter-term research foci, rather than for the detection of long-term trends, and hence continuity of record has not been a priority.

The focus of this study is both to update the observed trends in radiosonde, lidar and satellite data to recent periods, and to subject the data sets to renewed critical scrutiny.

Datasets

- Satellites
  - Microwave sounding observations (satellite) – Microwave sounding unit (MSU) channel 4 (MSU4) and the Advanced Microwave Sounding Unit (AMSU) channel 9 (AMSU9)
  - Stratospheric Sounding Unit (SSU) - satellite
- Radiosondes

○ RATPAC	1958 onwards	Seasonal and large area means
○ HadAT2	1958 onwards	Monthly gridded
○ RATPAC-lite	1979 onwards	Monthly gridded
○ IUK	1958 onwards	Monthly gridded
○ RAOBCORE1.4	1958 onwards	Monthly gridded
○ RICH	1958 onwards	Monthly gridded

- Lidar 3 stations – Southern France (OHP), Germany (Hohenpeissenberg) and California (Table Mountain)
- Analyses and Reanalyses – disregarded due to discontinuities due to changes in satellite instrumentation.

Trend Calculation

Estimates of trends are derived using a multivariate linear regression analysis. The statistical model includes terms to account for linear trends, together with variability associated with the 11-year solar cycle, plus two orthogonal time series to model the quasi-biennial oscillation (QBO).

To remove the effect of warming due to major volcanic eruptions (1963, 1982 and 1991) on temperature trends, data was omitted for 2 years following each eruption.

Uncertainty estimates for the trends are calculated using a bootstrap resampling technique, which includes the effects of serial autocorrelation. This represents the statistical uncertainty associated with overall atmospheric variability, but does not include systematic measurements uncertainties.

A.2.1 Temperature Changes in the Lower Stratosphere

1979 to 2007

The satellite data show cooling trends of -0.2 to -0.4 K·decade^{-1} over mid-latitudes of both hemispheres, with slightly smaller values in the tropics and NH high latitudes (although polar trends are strongly seasonally dependent). The cooling trends are somewhat larger in the UAH version of MSU4, and these differences are largest over NH mid-latitudes. Overall there is reasonable agreement between the satellite and radiosonde-based results, with the exception that the IUK data exhibit larger cooling trends, especially in the tropics. The good agreement between satellite and radiosonde data encourages exploration of further details of the variability and trends in radiosonde data.

Trends in the lower stratosphere (70–30 hPa) exhibit an overall 'flat' latitudinal structure, with relatively constant trends of order -0.5 K·decade^{-1} over ~60° N–60° S. The vertical profile of temperature trends for 1979–2007 derived from the radiosonde data in the Tropics (30° N–60° S) and the NH and SH extra-tropics (30°–90° N and S) show a reasonably coherent pattern of trends, with stratospheric cooling of -0.5 K·decade^{-1} (or larger) seen over all latitudes above 100 hPa. The scatter of stratospheric trends among the data sets is largest in the tropics and SH extra-tropics, reflecting the sparse station network and the influence of different homogeneity adjustments.

There is substantial variability in the Tropics regarding the detailed vertical structure of trends between ~300 and 70 hPa, and the height of the changeover from tropospheric warming to stratospheric cooling.

Polar stratospheric temperature trends are highly seasonally dependent. Antarctic temperature trends show strongest cooling during austral spring (September–November) and summer (December–February), with trends of order -1.0 to -1.5 K·decade^{-1} over 200–50 hPa. These strong cooling trends are associated with development of the Antarctic ozone hole during the 1980s. Relatively weaker and less statistically significant stratospheric trends are observed for other seasons. In the Arctic there are significant trends in the lower stratosphere during spring and summer, but not during autumn or winter. Highly significant warming trends are observed throughout the Arctic troposphere for all seasons, with largest values near the surface.

1957 to 2005

For 50 hPa temperature data there is good agreement in the variability among the data; the spread of the data is somewhat larger in the SH, particularly for the pre-1970 time period, and this can probably be attributed to the poorer spatial and temporal sampling of the radiosonde network in the SH. In the NH, the RATPAC data show distinctly smaller long-term changes than the other data for the early part of the record.

There is large variability in vertical temperature trend results among the different data sets for the 1958–1978 time period, and the uncertainty is especially large in the SH extra-tropics. Quite large cooling trends (near -1.0 K·decade^{-1}) are found in the NH extra-tropical stratosphere in all data sets except RATPAC, which shows trends of order -0.3 K·decade-1. The sparse observational database and known instrumental uncertainties for this period, together with the large trend uncertainties im-

plied by the spread of results suggest an overall poor knowledge of trends for the period 1958–1978. Therefore, the accuracy of trends for 1958 – 2005 are highly suspect given data uncertainties.

A.2.2 Temperature Changes in the Middle and Upper Stratosphere

From 1979 to 2005 (SSU Data)

Volcanic effects in 1982 and 1991 and long-term cooling can be seen in MSU4 and SSU data. The volcanic effects are largest in the lower altitude data, while the long-term cooling is highest at higher altitudes. Mean temperatures following each volcanic warming episode are lower than before the eruption.

Overall the near-global SSU time series show relatively constant temperatures throughout the stratosphere for the most recent decade 1995–2005.

Vertical profiles of linear trends show the largest trends are observed in the upper stratosphere (SSU channels 27 and 36x (2 hPa)), with values of -1.0 to -1.3 K·decade^{-1}; somewhat lower-magnitude trends, -0.5 K·decade^{-1} are found in the middle and lower stratosphere. Overall there is very good agreement between radiosonde and MSU4 trends in the lower stratosphere. The RATPAC-lite data set shows relatively constant trends of -0.5 K·decade^{-1} over 70–20 hPa, whereas the other radiosonde data show cooling that increases with altitude.

The latitudinal structure of temperature trends in the middle and upper stratosphere (SSU channels 25, 26, and 27) show a relatively flat latitudinal structure for channels 25 (20 hPa) and 26 (10 hPa), but with weaker cooling over high latitudes in the winter hemisphere. Channel 27 (2 hPa) shows more latitudinal structure, with enhanced cooling at polar latitudes and relative minima near 40° N and S (which are most pronounced during winter).

Comparisons between SSU and Lidars

There are only a few long-term records available for lidar stations. This paper only uses measurements from 3 stations, all located in the northern Hemisphere (OHP, Hohenpeissenberg and Table Mountain). A key limitation in these comparisons is the very different sampling between the lidars and the SSU data. Lidar data is very localised while SSU data is zonal means. The lidar data suffers from limited temporal sampling while the SSU data are true monthly means. For the analyses the daily lidar observations have been binned into monthly samples, deseasonalised, and then formed into three-monthly seasonal means.

Time series of lidar and SSU channel 27 temperature anomalies for the three lidar stations reveal significantly more variability from lidar than SSU, but this is expected considering the very different sampling. There is significant correlation (~0.5–0.6) between the lidar and SSU seasonal temperature anomalies at each location, and also some reasonable agreement for the low-frequency inter-annual variations.

Trends from the OHP record for 1979–2005 show statistically significant cooling of ~1.5 K·decade^{-1} over altitudes ~35–55 km, and these trends are larger than the corresponding SSU trends for this period. Comparing the vertical profile trends over 1988–2005 from the 3 stations gives substantially different results, although the statistical uncertainty is relatively large (especially for Hohenpeissenberg, where there are typically less than 10 lidar observations per month). It is unclear if these differences among the lidar station trends are associated with temporal sampling uncertainties at the individual stations, or to spatial differences of the actual trends (or both); however, the large difference between lidar trends at the nearby stations OHP and Hohenpeissenberg suggest that temporal sampling is a key issue. Given the large statistical uncertainties for this shorter record, together with the space-time sampling differences between the lidar and SSU data sets, it is difficult to constrain

uncertainties in either data set by these comparisons.

A.2.3 Solar Cycles in Temperature

The 11-year solar cycle is an important component of low-frequency variability in the stratosphere. Quantifying solar effects in the stratosphere is important for understanding forcing mechanisms and coupling with the troposphere and for comparison with model simulations. The available satellite data now span more than two full solar cycles, and the most recent cycle is not influenced by any large volcanic signal.

The MSU4 satellite results show a statistically significant positive solar signal of approximately 0.4 K in the tropics (30° N–S), with insignificant signals in extra-tropics. There is a relatively large projection over the Antarctic that is not statistically significant given the high natural inter-annual variability in this region. The solar signal derived from all of the radiosonde data sets show reasonable agreement with the satellite data in the tropics.

For the period 1979–2007 statistically significant positive values near ~0.5 K are observed in the lower stratosphere for both sets of satellite data and for the different radiosonde data sets. A consistent stratospheric signal is also observed in the radiosonde data for the longer record 1958–2005. The radiosonde-derived solar signal is not statistically significant in the tropical troposphere for 1979–2007, but the longer record 1958–2005 shows a significant signal of approximately 0.2 K.

A.2.4 Remaining Uncertainties

To improve the analysis of stratospheric temperature trends, the reliability of both the satellite and radiosonde data sets needs to be improved. Some specific requirements are:

1. The details of the processing of the SSU data need to be clarified in the peer-reviewed literature, and the raw radiances used to derive the SSU data need to be made publicly available in order to produce alternative, independent SSU climate data products.
2. MSU, SSU and AMSU data need to be combined to provide stratospheric climate information. The SSU operational data record ended in 2005, and the continuation of time series in the middle and upper stratosphere will be based on AMSU data (linked to SSU measurements using the 1998–2005 year overlap period).
3. The differences between the trends derived from the RSS and UAH MSU channel 4 data products should be further explored and reconciled.
4. There remain regions of large variability and uncertainty among the different radiosonde-based data sets, especially regarding trends in upper levels (e.g., 30 hPa) and for the pre-satellite time period.
5. Future climate observing system should include reference observations of upper air temperatures, such as those proposed by the Global Climate Observing Systems Program and appears to be nearing implementation as the Global Climate Observing System Reference Upper Air Network (GRUAN).

A.3 Comparison of Radiosonde and GCM Vertical Temperature Trend Profiles: Effects of Dataset Choice and Data Homogenization, J Lanzante & M Free, Journal of Climate, Vol 21, 5417-5435, 15th October 2008.

Link → http://dx.doi.org/10.1175/2008JCLI2287.1

Homogenization is a crucial step in the production of datasets intended for use in assessing long-term climate change because of the potential corrupting effect of changes in instruments and recording practices that have occurred over time. Homogenization is an attempt to eliminate the non-climatic (i.e., artificial) component of change from the data. Because of the complexities and ambiguities involved in homogenization, different analysts can use different approaches, each of which seems scientifically defensible, yet create datasets whose trends are substantially different. The paper tries to assess how homogenization influences the correspondence between observations and climate models.

Datasets

- Radiosondes
  - Radiosonde Atmospheric Temperature Products for Assessing Climate (RATPAC-B)
  - Hadley Centre Atmospheric Temperatures, version 2 (HadAT2), complemented by Hadley Centre Climatic Research Unit (CRU) surface temperature dataset (HadCRUT2).
- General Circulation Models (GCM), 1960-1999
  - World Climate Research Programme's (WCRP's) Coupled Model Intercomparison Project, Phase 3 (CMIP3) multi-model dataset.

In all comparisons, the GCM spatial (horizontal and vertical) and temporal sampling has been degraded to match that of the observations.

A.3.1 Temperature changes from a time series perspective

In the troposphere, adjusted versions of RATPAC and HadAT2 agree well at the global scale and their short-term variability is dominated by El Niño Southern Oscillations (ENSO). The observations almost always fall within the GCM ensemble spread. All datasets show a warming of the troposphere over time. For longer time scales the observations track the GCM mean reasonably well, with both showing monotonic warming interrupted by major volcanic eruptions, as reported in other papers.

Stratospheric series show gross agreement between models and observations in that both indicate long-term cooling in the latter half of the record and dramatic short-term warming for a couple of years following each major volcanic eruption. During the satellite era (1979–1999) the radiosonde cooling trend is greater than that in the GCMs. After the 1991 eruption radiosonde data is significantly lower than the GCM prediction, with the observed datasets lying near the GCM ensemble minimum.

A.3.2 Vertical profiles of temperature trends

Troposphere

In the troposphere temperature trend profiles from the models, and to a somewhat lesser extent those from the observations, exhibit some distinctive differences between latitude zones. During the longer era, trends in the Northern Hemisphere are largely uniform and positive in the main body of the troposphere, with a sharp decrease in the near-tropopause region leading to negative trends in the stratosphere. Overall, observation–model agreement is poorer during the satellite era, both in terms of the overall magnitude of tropospheric warming as well as the shapes of the profiles. Nevertheless, results from both eras are in general agreement that:

1. Adjustment makes trends more positive

2. Adjustment enhances the agreement between models and observations.

Stratosphere

Long-term temperature change in the stratosphere is characterized by cooling at almost all levels for almost all cases, with trend magnitudes several times larger than in the troposphere. Comparison of models and observations in the stratosphere yields broad conclusions similar to that for the troposphere. The modelled minus observed trend difference is almost always positive, and this difference is larger during the satellite era than the radiosonde era. The magnitude of this difference is roughly comparable to the model trends. The effect of homogeneity adjustment is usually to decrease the estimated observed stratospheric cooling and hence the model–observation difference (trends are negative due to cooling). **The effect of adjustment is largest in the tropics, especially for the satellite era, consistent with the notion that tropical radiosonde temperature data are particularly subject to time-varying measurement biases.**

A.3.3 Quantitative assessment of temperature trend profiles

Bivariate plots -Troposphere

For the radiosonde era (1960–1999) both the RATPAC and HadAT2 show reasonably good agreement with the models, with the models showing slightly more warming than the observations. Homogeneity adjustments on both datasets improve the observation – model agreement and increase the correlation (shape agreement).

For the satellite era (1979–1999) positive relative biases dominate with better agreement for RATPAC than HadAT2 and better agreement after adjustment. In addition, agreement in the tropics is poorer than in the northern and southern hemispheres. However, in contrast to the radiosonde era, the agreement is poorer and the intermodal spread is much greater.

Bivariate plots -Stratosphere

Stratospheric bivariate plots for RATPAC and HadAT have a distinctly different character than those for the troposphere. Except for a few instances, adjustment has very little effect on shape agreement. This may be due in part to the fact that the stratospheric trend profiles have less complexity than those for the troposphere. The dominant effect of adjustment is typically to reduce the positive relative bias, the magnitude of which can be several times larger than that in the troposphere. However, even after adjustment there is often a big relative bias, especially during the satellite era where observed cooling can be approximately twice as large as in the models.

Observation – model agreement is worst in the southern hemisphere. **Sampling variability may play a role in the unusual behaviour due as the number of stations is far fewer than for the other zones. This deficiency is magnified by the fact that there are fewer observations at higher altitudes because of premature bursting of radiosonde balloons that is more likely at the low temperatures and pressures of the stratosphere. Another factor that may play a role in the odd behaviour is the large radiative forcing resulting from stratospheric ozone depletion. Uncertainties either in this forcing or in the feedbacks involved may act to magnify the differences between different model simulations.**

A.4 Measurement Requirements for Climate Monitoring of Upper-Air Temperature Derived from Reanalysis Data, D Seidel & M Free, Journal of Climate, Vol 19, 854 – 871, 1st March 2006.

Link → http://dx.doi.org/10.1175/JCLI3666.1

In specifying requirements for upper-air temperature observations for climate monitoring, several issues must be addressed. These include the spatial and temporal resolution of the observations, and their accuracy, precision, and long-term stability. To address them requires an understanding of the expected future variations in temperature, the types of climate statistics that will be required from the observations, and the way in which individual observations will be assembled to develop those statistics. The reanalysis of the climate of the past half century is used as a model of the spatial and temporal variations in temperature that might be expected over the next half century, from the surface to 30 hPa.

Datasets

- Radiosondes
  - Reanalysis sampled at the locations of 15 GCOS stations, avoiding polar regions + mirror image w.r.t. Prime Meridian (longitudes swapped E for W).
- Simulated measurement protocols
  - Precision of temperature measurement (sensitivity to random uncertainty)
  - Sampling of the diurnal sample
  - Number of sample days per month
  - Long term measurement stability
  - Change f sampling time

The table below give results from simulations on the effects of variable precision of temperature measurements, variable sampling of the diurnal cycle and, variable sampling of the month. The right hand column gives recommendations for minimum measurement precision and launch scheduling.

Simulation	Parameter changed	Effect of change on mean	Effect of change on standard deviation	Recommendation
Temperature measurement precision	Measurement precision (K): 0.01 0.10 0.50 1.00	 0.01 K 0.01 K 0.06 K 0.11 K	 <10% <10% <10% <30%	Minimum measurement precision 0.50 K
Diurnal cycle Base measurement (0000, 0900, 1200, 1800)	Reduce to 2 (0000, 1200) Reduce to 1 (0000)	5.5% 20.2%	3.9% 8.7%	Twice daily sampling (0000 and 1200)

Simulation	Parameter changed	Effect of change on mean	Effect of change on standard deviation	Recommendation
Sampling of the month Base measurement daily (0000, 1200)	Every other day Every 3rd day Every 7th day	Max 1.8 K (<=0.09 K, 50% of the time) Max 3.4 K (<=0.16 K, 50% of the time) Max 8.6 K (<=0.39 K, 50% of the time)	Max 30% (10%, 50%) Max 50% (10%, 50%) Max 100% (50%, 50%)	Every 2 or 3 days with consistent time coverage.

Trends

For 20-yr data segments, less than 10% of the trends are significant. With longer segments, the frequency of significant trends increases, but even for 50-yr segments it is nowhere larger than 40% and is less than 20% in more than half the cases. This result emphasizes the fact that, even with optimal “observations”, with perfect precision, full temporal sampling, and no artificial discontinuities, statistically significant temperature trends in the reanalysis are not frequent, especially for short data records.

1) Effects of measurement precision and temporal sampling

Simulation	Parameter changed	Effect of change on Trend	Recommendation
Temperature measurement precision	Measurement precision: 0.01 K 0.10 K 0.50 K 1.00 K	<1% all cases	Precision <=0.5 K, reduces effect considerably
Diurnal cycle Base measurement (0000, 0900, 1200, 1800)	Reduce to 2 (0000, 1200) Reduce to 1 (0000)	11% 17%	Twice daily sampling (0000 and 1200)
Sampling of the month Base measurement daily (0000, 1200)	Every other day Every 3rd day Every 7th day	12% for all periods 13% for all periods 10% for 20-yr period, 27% for 50-yr period	Every 2 or 3 days with consistent time coverage.

Combining the effects of measurement precision and sub-monthly sampling shows that measurement precision has a minor effect on error rates, which are dominated by period length (higher error rates for longer periods) and are higher for less frequent sampling.

2) Effects of temporal inhomogeneities

The greatest challenge for detecting temperature trends in observational data is presented not by measurement precision or sampling frequency but by temporal inhomogeneities in time series. These can be introduced by changes in instrumentation, observing practices, or data-processing methods, and result in time-varying biases that can masquerade as, or mask, true trends.

These effects on long-term data stability can be simulated by randomly introducing step changes, or

interventions, in the time series. Trends were calculated using twice daily sampling, every day at full precision.

Single random timed intervention per time series:

Stability, K	Error rate*
0.1	1% for 20 – 50 year period
0.25	2 – 5% for 20 - 50 year period
0.5	3 – 12% for 20 - 50 year period
2.0	15 – 35% for 20 - 50 year period

Note:

* *The percentage of trends that are statistically significantly different from the actual reanalysis trends*

Almost identical data is produced when using 0.5K precision data with sampling every other or every third day.

Conclusion: Long-term data stability is the key factor in determining the accuracy of trend estimates.

3) Multiple random timed interventions per time series

With a single intervention, the 25-yr trend error rate increases from 0.4% to 20% as the maximum intervention size increases from 0.1 to 2.0 K, and from 1% to 35% for 50-yr trends. Introducing a second intervention increases the error rate by up to several percent but has a much smaller impact than the first intervention. Third, fourth, and fifth interventions have negligible effects. This is because the interventions are uncorrelated, to simulate the effects of changes in instruments whose accuracies (bias errors) are unrelated from one type to the next. In this scenario, the first intervention introduces an artificial trend, but subsequent interventions can either aggravate or meliorate the trend error depending on whether the second intervention is of the same or opposite sign as the first.

Almost identical data is produced when using 0.5 K precision data with sampling every other or every third day.

4) Effects of changes in observing schedule

Two time series, each based on a different observing schedule are joined at the midpoint of each time period and the effect examined.

Schedule 1 to Schedule 2	Trend error rate*	Trend period
0000 and 1200 daily to 0000 and 1200 every other or every 3rd day	5%	20 – 50 year
0000 and 1200 daily to 0000 and 1200 weekly	5% 13%	20 year 50 year
0000 and 1200 daily to 0000 or 1200 daily	4% 17%	20 year trend 50 year trend
0000 to 1200 daily or vice versa	14% 17%	20 year trend 50 year trend

Note:

* *The percentage of trends in time series with an observing schedule change introduced at the midpoint that are statistically significantly different from trends in unaltered time series.*

Conclusion: Maintaining a constant time of observation is more important than maintaining daily observations for avoiding errors in temperature trend estimates.

A.5 Factors affecting the detection of trends: Statistical considerations and applications to environmental data, Weatherhead et al, Journal of Geophysical Research, Vol 113, No D14, 17149-17161, July 1998.

Link → http://dx.doi.org/10.1029/98JD00995

Environmental data is normally auto correlated, for instance, higher than normal temperature on one day is often associated with higher than normal temperature on the next day. This positive autocorrelation tends to confound with a linear trend and therefore increases the length of time required to detect a given trend.

Measurement = constant + slope + noise + seasonality

In this paper the seasonality term is ignored.

Noise is random, has a mean of zero and a variance. The variance term contains the normally positive autocorrelation term.

Trend detection – effects of autocorrelation and variability

For a month to month variability of 10% and moderate autocorrelation of 0.5, it would take 23.4 years to detect a 5% per decade trend. If autocorrelation is set at 0 then the period drops to 16.3 years, therefore autocorrelation has a significant effect.

In comparison, using the same parameters for variability and autocorrelation a trend of 1% per decade would take 68.7 years! The long times required for trend detection come from the random variability in the data sets.

The precision of a long term linear trend estimate is highly dependable on the variance and autocorrelation of the noise. Estimates of these can allow the estimation of the number of years necessary to detect a trend of given magnitude of trend, or the magnitude of trend that can be detected by a fixed number of years of data.

Trend evaluation with interventions

Measurement = constant + slope + noise + seasonality + mean level shift due to intervention

The uncertainty of trend estimate is greatest when the intervention occurs half way through the collection period; the stand deviation of the slope is twice that obtained when there is no intervention. If the intervention occurs 25% or 75% the way through the collection period, the standard deviation of the slope increases by a factor ~1.5.

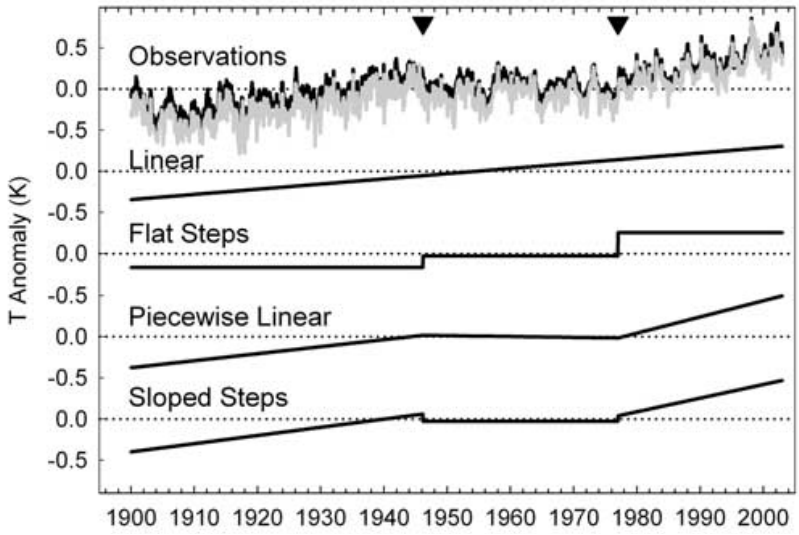
If the relative change caused by the intervention is known then this can greatly reduce the standard deviation of the trend. If our confidence in our knowledge of this change in level is: 0.1, 0.5 or 0.9 then the increase in standard deviation is increased by a factor of 1.8, 1.6 or 1.2, respectively, compared to the standard deviation where no intervention is present.

Using this information on increased standard deviation of the trend, the number of years required to detect a trend can be modelled. The worst case scenario, where an intervention occurs in the middle of a measuring period and the magnitude of intervention is unknown but constant, results in the time required to detect a trend increasing by a factor of 1.59.

A.6 An assessment of three alternatives to linear trends for characterizing global atmospheric temperature changes, D Seidel and J Lanzante, Journal of. Geophysical. Research, Vol 109, D14108, doi:10.1029/2003JD004414.

Link → http://dx.doi.org/10.1029/2003JD004414

Data from surface temperature, radiosondes and satellites were analysed using linear, flat steps, piecewise linear and sloped step statistical models as shown in the following figure:



Datasets

- Surface temperature
  - University of East Anglia/Climatic Research Unit
  - National Oceanic and Atmospheric Administration (NOAA)/National Climatic Data Centre's Global Historical Climatology Network
- Radiosonde
  - Hadley Centre for Climate Prediction and Research HadRT2.1
- Satellite
  - Microwave Sounding Unit (MSU)

The results of these different statistical models are given in the table on actual trends in Annex B .

A.7 Reference Quality Upper-Air Measurements: guidance for developing GRUAN data products, F Immler et al, Atmospheric Measurement Techniques, Vol 3, 1217–1231, 2010.

Link → http://dx.doi.org/10.5194/amt-3-1217-2010

General

In atmospheric profile measurements the uncertainty needs to be determined for each data point (at each altitude) individually. All sources of uncertainty should be summarized to an uncertainty

budget.

If measurements are uncorrelated the uncertainty of a mean decreases by 1/√N, while if the measurements are correlated (systematic effects) then the uncertainty does not reduce with more measurements. An example of this consistent uncertainty is if smoothing is applied to a vertical profile where the uncertainties are caused essentially by the same systematic effect and are therefore highly correlated.

Metrological traceability is the property of a measurement result whereby the result can be related to a reference through a documented, unbroken chain of calibrations, each of which contributes to the measurement uncertainty.

Establishing operational upper-air reference observations

The establishment of upper-air reference observations on an operational basis consists of definition, execution and evaluation phases. These phases are broken down into the following tasks:

- Defining requirements
- Reviewing existing instruments and choosing candidate(s)
- Identifying and quantifying sources of uncertainty
- Defining and validating a GRUAN data product
- Implementing a GRUAN data product
- Data archiving and processing issues

Example: Determining uncertainty in radiosonde temperature profiles

How a reference quality measurement, as described above can be achieved for radiosonde temperature measurements. The Vaisala RS92 and Graw DFM-06 radiosondes will be used as the basis of the example.

The requirements for GRUAN measurements of temperature have been specified in the GCOS 2007 Report: “GCOS Reference Upper-Air Network (GRUAN): Justification, requirements, siting and instrumentation options” (Link → http://www.wmo.int/pages/prog/gcos/Publications/gcos-112.pdf). These requirements are summarised below.

An uncertainty of:

- 0.1 K at a vertical resolution of 100 m in the troposphere
- 0.2 K at a vertical resolution of 500 m in the stratosphere

Currently there is no commercial instrument that meets these requirements, but this is the ultimate goal of GRUAN.

Review of existing instruments

Instrument review is an on-going process within the initial phase of GRUAN. It is not expected that all sites use identical instrumentation. Establishing the uncertainty budgets of these instruments is an important step in ensuring the comparability of the measurements from different sites and identifying the technology that is best suited to fulfil the long-term goals of the network.

Establishing the uncertainty budget

a) Uncertainty arising from of the indication of the measuring system

The capacitive sensors change the frequency of a resonant circuit depending on the sensor temperature. This frequency is of the order of 10 kHz and is measured and transmitted with a resolution of 0.01 Hz. The dependency of the frequency on temperature is roughly 0.5 Hz/K. The accuracy of the indication is therefore about 0.02 K and much lower than the stated uncertainty of the sensor of 0.15 K. It can be assumed that the contribution of the frequency measurement to the total uncertainty of the temperature sensor is negligible.

b) Calibration

Sensors should be supplied with a calibration certificate with a proven traceability to National Standards and therefore to SI units. The uncertainty of these calibrations should be well below 0.1K throughout the entire temperature range. The calibration uncertainty is considered to be an altitude-independent absolute systematic contribution to the uncertainty profile. Altitude-dependent uncertainties are characterized separately.

Some radiosondes are recalibrated before launch by a ground check. The reference sensors of the ground check station should be regularly calibrated by a certified agency to ensure traceability to SI. In this case the reference sensor could be considered a “GRUAN site working standard”

c) Radiation correction

The largest part of the overall uncertainty arises from the radiation that is absorbed or emitted by the sensor, in particular during day-time measurements. Radiation can affect the measurement in different ways:

- Incoming radiation heats the sensor directly
- Indirect radiative heating: Incoming radiation heats the sensor framework, the mount that surrounds the radiosonde or any other part of the sounding equipment (incl. the balloon). This heat can then reach the sensor by conduction or via air passing over this part, warming up and then passing over the temperature sensor.
- The sensor emits (long-wave) radiation and is thereby cooled. This effect plays a significant role for sensors with white coatings, but is considered negligible for metallic coatings.

Generally, a radiation correction is applied to the temperature by the software in the receiving station. This correction should be documented in the accessible literature and depends on pressure, ventilation (ascent rate), and the incoming solar radiation.

d) Other sources of uncertainty

The effect of radiative balloon heating or adiabatic balloon cooling on the temperature data is considered to be negligible, provided the rope between balloon and radiosonde is at least 40 m.

When the radiosonde emerges into dryer air above a cloud, evaporation of the condensed water cools the sensor and creates a cool bias in this region (wet bulb effect). This effect can lead to deviations up to 1 K above a cloud and the data need to be flagged appropriately, e.g., by assigning a correspondingly increased uncertainty to data in such regions.

Time-lag bias during the ascent is of the order of less than a second over the entire temperature range. The temperature during the ascent varies generally by less than a tenth of a degree in this time frame (along an adiabatic profile at a typical balloon ascent speed of 5 m/s the temperature gradient is 0.04 K/s). Therefore, it may be assumed that the bias caused by the time lag of the temperature sensor can be neglected.

Validating the temperature measurements

From parallel measurements on a single balloon launch (2008, Lindenberg) it can be concluded that the estimated uncertainties are consistent with measurements from other instruments in the troposphere and into the lower stratosphere, where there is no wet bulb effect. In the stratosphere some instruments show significant differences to each other. This is most probably due to larger (direct or indirect) effects of solar radiation on these other sensors. It should be noted, that this was not a proper validation experiment since there was no reference instrument available.

Data archiving issues

All raw data should be recorded before calibration corrections are applied. All calibration coefficients and corrections (i.e. radiation effect) must also be recorded so that data can be reprocessed at a later date.

In order to provide homogeneous data it is necessary to install a common 'GRUAN' script on all participating stations that produces a common and well defined (with respect to the applied corrections and filtering) temperature profile necessary for the processing of higher level GRUAN data.

A.8 Uncertainties in climate trends – Lessons from upper air temperature records, P Thorne et al, American Meteorological Society, 1437 – 1442, October 2005.

Link → http://dx.doi.org/10.1175/BAMS-86-10-1437

Historical datasets on atmospheric temperature have been compromised due to the lack of a robust reference network. If such a network was implemented in the future it would significantly reduce uncertainty in future climate modelling activities.

A reference network does not necessarily require extensive global coverage, but should act as anchor points for global networks, which may have operational weather requirements as their primary output. Although such sites will be more expensive than standard sites they are relatively less expensive than satellite based measurements.

There would also be additional benefits for instrument development, radiative transfer code development and model development.

Without such a reference network being established it is likely that in 20 years' time the climate community will still be struggling with uncertainty and unable to ascertain true climatic variations.

A.9 Spatial sampling requirements for monitoring upper-air climate change with radiosondes, MP McCarthy, International Journal of Climatology, 985-993, Vol 28, Aug 2008.

Link → http://dx.doi.org/10.1002/joc.1611

The original 64 GUAN measuring stations were selected based upon their contribution to a spatially homogenous network. In 2008 the network was made up of 164 stations. An assessment of the spatial sampling requirements for such a network has been conducted.

Datasets

- Radiosondes

  - Hadley Centre for Climate Prediction and Research HadAT
  - Integrated Global Radiosonde Archive (IGRA)
- Satellites
  - Microwave Sounding Unit (MSU)

Correlations between measuring stations that fell within 5° latitude or longitude were calculated from the radiosonde data and a similar method using 5° grid boxes for the MSU data. The analysis suggests that there are acceptable levels of correlation within the troposphere if measurements are made within 10°–15° latitude by 24°–36° longitude, with less longitude resolution required near the equator. This would equate to a requirement for between 100 and 240 globally distributed stations.

This distribution assumption was tested at a measurement height of 500 hPa, using a stratified grid method and a randomised grid method, both with 90 nodes. The site numbers in both methods were then reduced to the closest actual radiosonde sites, giving 76 stations for the stratified grid method and 75 sites for the randomised grid method. Analysis showed that for near global (80° N–80° S) mean temperature there appears to be little additional benefit in reducing station separations less than 30° longitude. Extra stations do not reduce sampling uncertainties but are likely to be useful in assessing other uncertainties. The data shows that there is a poorly sampled zonal band in the southern hemisphere mid-latitudes where radiosonde distribution is much sparser.

The stratified and random grid methods can be used to assess the likely magnitude of trend uncertainties due to spatial sampling. The number of stations required delivering a temperature trend uncertainty of 0.05 K/decade and 0.1 K/decade and a humidity trend uncertainty of 0.33 %/decade and 0.65 %/decade were calculated. These requirements are based upon quarter and half of expected trends of 0.2 K/decade and 1.3 %/decade for air temperature and humidity at 500 hPa. From this, radiosondes distributed approximately 30° longitude and 15° latitude should keep sampling errors below 0.05 K/decade for global and northern hemisphere mean trends, while for the southern hemisphere, tropics and zonal bands (30° N–60° N and 10° S–10° N) measurements need to be made approximately every 20° longitude and 10° latitude. These requirements would be less stringent if radiosonde stations were evenly spaced without gaps.

From the trend uncertainty calculations the current GUAN network is inadequate for measuring global mean trends in humidity.

Analysis of GUAN station data shows that there is a lack of coverage over India, central Africa and northern hemisphere high latitudes. There is also potential over sampling over parts of Europe and East Asia.

A.10 Assessing bias and uncertainty in the HadAT adjusted radiosonde climate record, MP McCarthy, Journal of Climate, Vol 21, 817-832, Feb 2008.

Link → http://dx.doi.org/10.1175/2007JCLI1733.1

The paper assess the performance of an automated method for detecting break points in radiosonde data, compared to previously used manual methods.

Datasets

- Radiosondes
  - Hadley Centre for Climate Prediction and Research HadAT

- Integrated Global Radiosonde Archive (IGRA)

- Simulated data
  - Third Hadley Centre Atmospheric Model (HadAM3)

Break points can be successfully identified and adjusted for by the automated method where they are at least 0.4 K in magnitude, but are poorly resolved if smaller than this.

A number of factors discussed in this work are general to all related attempts to homogenise radiosonde temperature records:

- Biases exist in day relative to night radiosonde data.
- Quality of the background or near neighbour reference is a major influence on trend recovery.
- The time interval used for estimating breakpoint adjustment is an important consideration for homogenisation methods.
- Methodological choices can result in significant parametric uncertainties in radiosonde trend estimates, and methods should be objectively tested in their ability to recover climate signals from data with trend biases.

A.11 Impact of missing sounding reports on mandatory levels and tropopause statistics: a case study, JC Antuna et al, Annales Geophysicae, Vol 24, 2445-2449, Issue 10, 2006.

Link → http://dx.doi.org/10.5194/angeo-24-2445-2006

The paper describes the effect of missing sounding data reports on mean temperature and pressure values for one station (WMO station number 78335, Cuba) for a time lag of 8 years. It was found that around 40% of the data was missing from the IGRA archive. The effect of this missing data on the mean values of altitude and temperature was analysed compared to the complete dataset.

The test of the statistical significance of the differences between means showed no significant differences at all levels, both for temperature and altitude. The significance of the test ranged between 99.51% and 99.66% for altitude and between 99.50% and 99.57% for temperature. Likely explanations for the fact that there is no statistical difference between the means in both datasets is probably related to the normal distribution of the variables, and that the main cause of missing values was transmission difficulties, which has a random origin.

A.12 A quantification of uncertainties in historical tropical tropospheric temperature trends from radiosondes, P Thorne et al, Journal of Geophysical Research-Atmospheres, Vol 116, Article Number: D12116 DOI: 10.1029/2010JD015487, JUN 29 2011.

Link → http://dx.doi.org/10.1029/2010JD015487

Datasets

- Radiosondes
  - Hadley Centre for Climate Prediction and Research HadAT

The main conclusion of the paper

A comprehensive analysis of the uncertainties in historical radiosonde records has yielded trend uncertainties of the same order of magnitude as the trends themselves. It is highly unlikely that these uncertainties can be reduced using the near-neighbour HadAT approach.

Over the full period of the radiosonde record, the estimates produced are in statistical agreement with model expectations all the way up to the tropical tropopause.

Over the shorter satellite era, a discrepancy remains, particularly in the upper troposphere. Potential explanations range from likely residual observation errors (either at the surface or once airborne) or statistical end point effects, to more far reaching reasons involving physical processes or forcings missing from some (known to be the case) or all climate models.

The present analysis cannot provide definite conclusions in this regard. However, the high degree of agreement over the 45 year radiosonde record provides a strong degree of confidence in overall climate model behaviour in the tropics on the longest time scales.

A.13 Critically Reassessing Tropospheric Temperature Trends from Radiosondes Using Realistic Validation Experiments, H. Titchner et al, Journal of Climate, Vol 22, 465 – 485, Feb 2009.

Link → http://dx.doi.org/10.1175/2008JCLI2419.1

Datasets

- Radiosondes
  - Radiosonde Observation Correction Using Reanalyses (RAOBCORE)
  - GUAN
- Simulations
  - HadAM3 forced with observed sea surface temperature and sea ice distributions from the Hadley Centre Sea Ice and Sea Surface Temperature (HadISST) dataset and prescribed anthropogenic and natural external forcings.

The aim of the paper is to reassess the uncertainty in the manually homogenized Met Office Hadley Centre radiosonde temperature dataset (HadAT). Microwave Sounding Unit (MSU) datasets all yield trends that are more or less consistent with model predictions. So do some more recent radiosonde temperature datasets and temperatures inferred from radiosonde winds. However, other recently produced radiosonde datasets have all reported less warming than expected since 1979.

To test the homogenisation system artificial breakpoint profiles were applied to the daytime and night time control datasets from the climate model. Four error models were tested based on different assumptions regarding the size and distribution of the break points. After examining the full radiosonde dataset, analyses were done on data from the GRUAN stations.

The main findings from the GUAN analyses were:

1) It is likely that the GUAN night time coverage is too sparse to create a sufficiently homogeneous neighbour composite series when only GUAN stations are used.
2) A high quality reference series is important for trend recovery, and this may be a problem when a biased sparse network is used. If we could gain a high quality GUAN or similar-sized network, then it is very likely we would be able to adequately constrain the uncertainties in the trends for

the rest of the global network.

3) Much would be gained from a coordinated comparison of independent radiosonde homogenization methods, particularly if realistic validation experiments are performed.

Recommendations

Results were much more encouraging when a GUAN network consisting of perfect station records was used as a reference series to homogenize the rest of the available stations. We therefore recommend that the GUAN network is maintained to a high standard, including adherence to the GCOS monitoring principles (GCOS 2004) so that we can have a better understanding of future trends in the free atmosphere.

A.14 Observing Systems Capability Analysis and Review Tool (OSCAR) – World Meteorological Organisation

Link → http://www.wmo-sat.info/oscar/

This database is the official repository of requirements for observation of physical variables in support of WMO Programmes and Co-sponsored Programmes. These requirements are maintained by the focal points designated for each application area.

Using the Rolling Review of Requirements (RRR) process defined by the Manual on the Global Observing System (WMO-No. 544) (Part II, Requirements for observational data), user requirements for observations are compared with the capabilities of present and planned observing systems. User requirements are collated in a comprehensive, systematic and quantitative way in the WMO Observing Requirements database, which attempts to capture observational requirements to meet the needs of all WMO programmes.

The comparison of user requirements with observing system capabilities for a given application area is called a Critical Review. The output of this is reviewed by experts in the relevant application and used to prepare a Statement of Guidance (SOG), the main aim of which is to draw attention to the most important gaps between user requirements and observing system capabilities, in the context of the application. A wide range of applications within WMO programmes have already been addressed.

The following uncertainties for atmospheric temperature observations used for climate modelling research are expert opinion and not from peer reviewed literature. These values were set in December 2012 by the World Climate Research programme (WCRP):

Requirement	Standard Uncertainty
Threshold	2.0 K
Breakthrough	0.5 K
Goal	0.2 K

Note: These uncertainties are expert opinion and not from peer reviewed literature.

Definitions:

- The “threshold” is the minimum requirement to be met to ensure that data are useful.
- The “goal” is an ideal requirement above which further improvements are not necessary.

- The “breakthrough” is an intermediate level between “threshold” and “goal” which, if achieved, would result in a significant improvement for the targeted application. The breakthrough level may be considered as an optimum, from a cost-benefit point of view, when planning or designing observing systems.

Annex B Summary of Trends and Differences in Trends due to Measurement Method

The following 2 tables summarise the findings in the above papers with regards to measured temperature trends and differences in temperature trends due to measurement methods. Although these tables are outside of the scope of this document they provide useful background information.

B.1 Measured Temperature Trends

Source	Trend K·decade^{-1}	Period	Location	Measurement	Comments
2	+0.2	1958 – 2005	LS	Sonde	Solar cycle
2	+0.5	1979 – 2005	LS	Sonde + MSU4 + SSU	Solar cycle
2	-0.2 to -0.4	1979 – 2007	LS	MSU4	
2	-0.5	1979 – 2007	LS	Sonde +MSU4	
2	-1.0	1957 – 2005	LS	Sonde	Large uncertainties in data 1958 – 1978 compromise the results.
2	-1.0 to -1.5	1979 – 2007	LS Antarctic	Sonde	Ozone hole during 1980s
2	-0.5	1979 – 2005	M	SSU	
2	-1.0 to -1.3	1979 – 2005	M – US	SSU	
2	-1.5	1979 – 2005	US	Lidar	
6	0.87 0.87 0.66	1900 – 2002	Surface	Sloped step Piecewise linear Linear	
6	0.32 0.52	1958 – 2001	T	Sloped steps Linear	Radiosonde
6	0.13	1979 – 2001	T	Linear Change dominated by inter-annual changes, no real slope.	Satellites Radiosonde results over the same period give 0.14
6	-1.82 -1.82 -1.90	1958 – 2001	S	Sloped step Linear Linear	Radiosonde Volcanic periods removed, linear give the best fit
6	-0.88 -1.13 -0.83 -0.99	1979 – 2001	S	Sloped step Linear Flat step Linear	Satellite Volcanic eruptions account for 94% of the cooling Volcanic activity removed

Key: T Troposphere
MT Mid Troposphere
S Stratosphere
LS Lower Stratosphere
US Upper Stratosphere

B.2 Differences in Temperature Trends due to Measurement Method

Source	Difference in trend K·decade^{-1}	Period	Location	Measurement	Comments
1	-0.1	1979 – 1997	MT	Sonde – MSU2	
1	+0.16 to -0.31	1979 – 1997	LS	Sonde – MSU2	
1	0.071	1979 – 1997	200 mbar	Sonde – MSU2	Full and subsampled mean global trend
1	<0.05	1979 - 1997	50, 500 – 850 mbar	Sonde – MSU2	Full and subsampled mean global trend
1	0.02	1979 – 1997	MT	Sonde – MSU2	Due to temporal effects
1	0.2 0.1 <0.1	1979 – 1997	MT	Sonde	LKS vs HadRT 7 stations 8 stations 44 stations
3	Observed > modelled	1960 – 1999	T	Sonde x 2 – Modelled x 6	Homogeneity adjustments (sonde) improve agreement and correlation
3	Observed >> modelled	1960 – 1999	S	Sonde x 2 – Modelled x 6	Similar to troposphere, but sampling ozone depletion may cause problems in southern hemisphere
4	0 to 60%	1955 – 2005	T to S	Sonde	Analysis of the effects of precision, sampling time and frequency and measurement stability
5				Effect of autocorrelation, variability and measurement intervention on the detection of trends	

Key: T Troposphere
MT Mid Troposphere
S Stratosphere
LS Lower Stratosphere