P4.5 ADJUSTING ARCHIVED RADIOSONDE DATA USING COMPLETE VALIDATED AND INFERRED RADIOSONDE METADATA TO COMPUTE UNBIASED ATMOSPHERIC TEMPERATURE AND MOISTURE TRENDS

Steven R. Schroeder *

Texas A&M University, College Station, Texas

ABSTRACT

Archived radiosonde data is extensive enough that global and regional trends of atmospheric temperature and moisture (total precipitable water) can potentially be calculated back to about 1958. However, all trends computed from radiosonde data are uncertain because instrument changes cause artificial discontinuities and because station histories are incomplete and inaccurate.

The first goal of the ongoing Validated Atmospheric Profiles for Operations and Research (VAPOR) project is to develop complete historical station and instrument metadata. Three major steps are involved: (1) Combine and add all available metadata updates to the most extensive historical radiosonde metadata sources. (2) At stations which are well-documented, search for consistent characteristics of specific instrument types, and discontinuities which indicate instrument changes, in time series of variables which are very sensitive to different instrument types. (3) At stations and in time periods where metadata is absent or incorrect, infer the use of specific instrument types by similar characteristics, and the dates of changes by discontinuities, in time series of the same variables.

The second goal is to develop and apply adjustments for each distinct instrument type to allow computation of unbiased atmospheric temperature and moisture trends. While the metadata is not yet complete, it is most complete and best checked in Japan, China, the Russian Federation, India, and Australia in the period starting 1973. This reports on initial efforts to develop adjustments to compensate for instrument changes and their biases using the detailed Japanese instrument history as a test case. In general, because technological improvements have caused radiosondes to gradually become more sensitive and better protected from radiation errors, the instrument-caused trend is generally an erroneous cooling and drying, with the largest errors (and improvements) at the highest levels.

The proposed method of data adjustment is called equiprobability transformation. Basically, readings from all instruments are adjusted so the probability distribution of each variable of climate interest (temperature and dew point depression) in the same climate environment (stratified by pressure interval and sun angle, and by temperature interval in the case of dew point depression) matches the probability distribution of the same variable using a target “reference” instrument type. The best comparisons of instruments are obtained from data either before and after a change to (or from) a reference instrument at the same station, or from simultaneous use of the two instrument types at different closely-located stations.

At this time, only preliminary tests of temperature differences between instrument types at mandatory pressure levels have been performed for Japan. Differences between consecutive instruments are unexpectedly large, and are similar for day and night observations (although the night differences are slightly larger). This indicates that, if the adjustment applied to readings from one instrument type to make it comparable to the readings from another instrument type is simply the difference before and after the change (the adjustment is the average of the second instrument minus the first instrument), natural interannual variations occurring around the transition are projected into the other periods being adjusted. In addition, a portion of the long-term actual trend is removed.

One adjustment procedure being investigated is to define the adjustment as the difference between instruments in day minus night temperatures. This means that an instrument usually will have no adjustment applied to night temperatures unless data evidence supports a change in the night radiative and other characteristics of that instrument type.

1. INTRODUCTION

One of the ongoing controversies relating to the possibility of anthropogenic climate change caused by the buildup of greenhouse gases is that observed temperature trends above the surface tend to differ from trends expected using model and theoretical studies. Radiosonde-based trends for the last few decades usually show less tropospheric warming than at the surface, while model runs show faster warming in the upper troposphere than at the surface (Lanzante 2007). Models predict stratospheric cooling due to ozone depletion, but the observed cooling in radiosonde data is larger than expected. The radiosonde trends are affected by frequent instrument changes. Generally, instruments have improved over time, with newer models having faster-responding sensors and better protection from radiation. The ongoing improvements have been hypothesized to add an erroneous cooling and drying trend to the actual climate trend.

To determine the actual trend, the nature and timing of instrument changes at each station needs to be defined. When instrument changes are known, it is potentially feasible to determine the differences between

---

***** *Corresponding author address:* Steven R. Schroeder, Department of Atmospheric Sciences, Texas A&M University, 3150 TAMU, College Station, TX 77843-3150; e-mail: steves@ariel.met.tamu.edu.

the instruments and develop adjustments to compensate for the artificial differences. The most extensive global radiosonde metadata history, listing station locations, instruments, and the dates of changes, is part of the Integrated Global Radiosonde Archive (IGRA) project (IGRA 2006; Durre et al. 2006; also see Gaffen 1996), but is incomplete and contains some inconsistent or erroneous information. A project to develop complete metadata is underway, and instrument histories are nearly complete back to 1973 or earlier in Japan, Australia, the Russian Federation, China, and India. This project uses two approaches to complete metadata, including seeking additional published and unpublished metadata, and examining the data for discontinuities in variables that tend to be very sensitive to different instrument types. That project is mentioned only briefly here in Section 1.2, but is described in detail in Schroeder (2007).

1.1. Indirect methods to identify and correct biases

While determining unbiased trends in data is a two-step process, including constructing complete metadata and developing data adjustments, many researchers address the second step and attempt to develop trends while the first step is still quite incomplete. Without complete metadata, indirect methods such as the following are used to attempt to remove biases from climate trends:

(1) An automated discontinuity detection and removal method (Gaffen et al. 2000). This was found to remove essentially all trends, regardless of the tuning of the level of sensitivity.

(2) Removing identified discontinuities only if they approximately coincide with known transitions (Gaffen et al. 2000, Lanzante et al. 2003). Trends from this approach appear mostly reasonable, but there is no assurance that the adjustments are correct for the “right” reason because they are subjective.

(3) Using only stations which appear homogeneous over a long period (Ross and Elliott 1999). The authors found only 7 stations out of 188 that appeared homogeneous from 1948 to 1995. Actually, no station is homogeneous for such a long period, although some Chinese stations still (as of 2007) have used GZZ-2 radiosondes since 1964.

(4) Deleting data around each suspected discontinuity and computing the trend using “first differences,” or the difference from one year to the next in each month in the remaining short segments (Free et al. 2004). This method does not produce a time series for any individual station. The first differences are area-averaged, and are then summed over time (the first difference in the first year is zero) to produce a time series which retains the trend of the accepted data.

(5) Comparing radiosonde minus satellite temperature retrieval time series (Randel and Wu 2006, Christy et al. 2007). Various approaches are used to make comparisons and detect and compensate for discontinuities. Of course, the satellite record is also not homogeneous, so discontinuities when satellites change need to be accounted for.

(6) Comparing day and night temperatures (Sherwood et al. 2005). At 50 and 300 hPa, the authors find daytime cooling relative to night temperatures, with the largest trends in longitudes where these times are close to local noon and midnight. The trend is consistent with decreases in uncorrected radiative heating, and allows adjustment of the temperature trend in a group of stations to a night-only equivalent. This adjustment does not correct errors other than radiative heating. Christy et al. (2007) use this approach by separately comparing day and night radiosondes with satellite retrievals.

(7) Comparing each station time series to a constructed “neighbor” time series (Thorne et al. 2005), which assumes that discontinuities in the neighbor series are averaged out with respect to any discontinuities at the target station. A discontinuity in the station minus neighbor time series is likely to be caused by an instrument or processing change at the target station. This approach is assumed to detect undocumented instrument changes, so 70 percent of the change points are not associated with documented changes.

(8) Using radiative theory applied to instrument configurations to construct adjustments. Luers and Eskridge (1998) developed theoretical “temperature correction models” of the radiation and lag errors of major radiosonde types. Durre et al. (2002) found that applying the corrections where transitions appear well-documented often made discontinuities worse. It is possible that the main reason for the apparent failure of these adjustments is that the archived data already was adjusted up to 2 times, first by the originating station or country when the observation was transmitted for operational forecasting, and possibly a second time by the forecasting center (NCEP in this case) which archived the data. The remaining error would be whatever error was undercorrected (or overcorrected) in these adjustments, and might not have characteristics that resemble the modeled radiation and lag errors.

(9) Make no adjustments. Trends from unadjusted data sets (Angell 2003, Sterin 1999) are now considered to primarily reflect errors resulting from not considering instrument and processing changes, not actual trends.

(9) Applying a statistical test to radiosonde minus reanalysis temperatures at each station to identify and adjust for discontinuities (Haimberger 2007). As with Thorne et al. (2005), undocumented change points are detected and adjusted.

(10) Simultaneously estimating trends, change points, and natural variability using a procedure called “iterative universal Kriging” or IUK (Sherwood 2007). In theoretical tests, no variation of the sensitivity parameter used in this method resulted in a hit rate above 45 percent (a “hit” is detection of an instrument-caused discontinuity, and almost all variations had a false detection rate above 50 percent (a “false detection” is considering a natural variation to be caused by an instrument change), but certain sensitivity settings still tended to produce nearly-correct trends.

Since adjustments are mostly subjective, a method is considered "successful" if the trend is close to the expected magnitude or at least is close to results from another approach. However, these trends are still questioned, and it is not known whether the errors are undercorrected or overcorrected, because the adjustment methods tend to produce the “expected” results. As stated in Sherwood (2007), most methods without external data inputs tend to remove some of the real trend, while if a method compares radiosonde data to a reference data series, the adjusted radiosonde trend tends to approach the trend of the reference series.

If complete metadata makes the nature and timing of all transitions known, the methods above can be applied much more confidently because false detections would be nearly eliminated. The main remaining error could be considering a natural variation at the time of an instrument change to be caused by the instrument change, but most such “false detections” could be identified if the same instrument change occurs at different times at other stations, because an instrument-caused discontinuity should be similar at each station making the same instrument change. In the long run, adjustments will be validated by similar trends produced using different methods and data sources.

1.2. Summary of approach to produce complete metadata

The fact that all upper air temperature data sets require elaborate adjustment schemes to achieve an uncertain level of credibility indicates that future progress to develop accepted atmospheric climate trends will be very limited until historical metadata is much more complete and accurate.

It is a slow process to obtain metadata from the usual sources such as agencies making radiosonde observations, so it is an attractive possibility to develop complete metadata using the archived data itself. Hypothetically, instrument changes can be identified by data discontinuities, but this approach has not been very successful so far because in variables of research interest, primarily temperatures at specific levels, the discontinuities caused by instrument changes are often not obvious enough to be confidently identified even though they are clearly large enough to greatly contaminate all derived trends.

This research more successfully completes metadata using the archived data itself by intentionally searching for especially sensitive variables that amplify differences between instruments by several times, even though such variables have little or no research interest. Using station time series, consistent characteristics can be attributed to a specific instrument at well-documented stations, and similar signals at a station or in a period without metadata allow inference of the use of the same instrument type. Using the same methods to examine time series of the same variables at all stations, with or without documentation, the available metadata is validated based on consistency with the data and missing metadata is constructed.

Variables computed from the soundings which are the most sensitive to different instrument types include the lowest relative humidity reported above the surface, the lowest temperature or pressure with a reported dew point, day minus night differences in such variables, and even the number of temperature or dew point levels reported per sounding. Signatures derived from these variables have considerable commonality at all stations using the same instrument, with smooth variations among levels and seasons at a station and in differing environments among stations.

While some metadata must be available as a starting point to identify instrument types, with even partial metadata it should be possible to attribute a particular set of characteristics to an instrument type. Characteristics of each instrument type are not absolutely unique, but when similar characteristics are found at a station where the metadata is missing or questionable, the number of candidate instrument types which might have been used is greatly narrowed down.

Because the most extensive historical metadata compilation, the IGRA metadata file, has only had a few updates since the middle 1990s, a recent metadata source which was not used in the initial IGRA metadata file should be discussed. This is the 5-digit instrument code reported as part of the 31313 group in transmitted soundings. The 31313 group has been used to report the instrument type ("31313 code") and launch time in some United States soundings since 2 March 1989 and since early 1992 in many countries, and now appears in about 85% of all soundings. The "31313 code" is a 1-digit solar and radiation correction code (WMO-No. 306, Vol. I Part B, BUFR [Binary Universal Form for the Representation (of meteorological data)] Table 0 02 013, or Code Table 3849), a 2-digit radiosonde and ground equipment code (BUFR Table 0 02 011 or Code Table 3685), and a 2-digit tracking (wind finding method, such as radar or GPS) or sounding status code (BUFR Table 0 02 014 or Code Table 3872).

Ideally, 31313 codes provide an exact instrument history, but the code has limitations. The solar and radiation correction code is generic (such as "country solar correction") so the 5 different Vaisala RS80 corrections are not distinguished. The radiosonde code does not specify varieties such as Vaisala RS80 A-Humicap and H-Humicap, which had differing dry biases from packaging contamination (Wang et al. 2002). All 2-digit radiosonde codes are now assigned, but obsolete codes can be reassigned, as described in a recent version of Code Table 3685 at http://www.wmo.ch/pages/prog/www/WMOCodes/Operational/CommonTables/BufrCommon-11-2007.pdf. At least one informal code reassignment has already occurred. French stations reported code 34 (Czechoslovakia Vinohrady) starting November 2004 to refer to Vaisala RS92 with a STAR ground station (confirmed in the 2006 WMO upper air catalog). Usually, stations report code 90 ("unknown instrument") if no code is assigned. The wind finding code is also generic and sometimes a code value such as "systems operating normally" is reported instead. Finally, tables do not give any references or further details about code values, so they are sometimes used incorrectly. For example, some countries report the code for “no wind finding method,” but interpret this to mean that the radiosonde does not determine the wind, while the wind is actually obtained by radar tracking of the balloon.

It should also be mentioned that data used in this project is from National Center for Atmospheric Research (NCAR) Data Set 353.4 (DS353.4), which contains observations back to 1973 transmitted over the Global Telecommunications System (GTS) and processed by the National Centers for Environmental Prediction (NCEP, formerly National Meteorological Center (NMC) before 1994). IGRA combines this data set with 10 other data sets and has many observations back to 1963 or earlier, but DS353.4 still accounts for over half of IGRA. Other researchers may use other radiosonde data sources, but IGRA is the most popular because it is arranged as time series of individual stations. However, the disadvantages of IGRA are that newly-established stations, including stations with new ID numbers which replace nearby closed stations, are added infrequently, and IGRA does not contain data for ships, including fixed ships, so the amount of oceanic data is significantly reduced.

Since a forecasting center often applies adjustments to radiosonde data, it is possible that an archived sounding has actually been adjusted twice, first by the station and then by the forecasting center. If a data archive is obtained from a source other than NCAR or IGRA, the data values for the same observation may differ, and instrument adjustments would then differ from the adjustments derived in this project. However, the dates and nature of instrument changes derived from the data should not differ. One of the reasons why the variables that are most sensitive to instrument changes are mostly moisture-related is that processing centers usually do not apply adjustments to reported dew points.

2. SUMMARY OF JAPAN RADIOSONDE HISTORY

While the methods to develop instrument metadata for each station from archived observations, including validating available information and constructing metadata where information is missing or erroneous, is not discussed in detail here, the Japanese instrument history needs to be discussed in some detail because this information is relevant for guidance in developing data adjustments.

The Japan instrument history since 1957 is relatively uncomplicated because Japan has changed radiosondes infrequently, older instrument types have distinct signatures, and stations have routinely reported 31313 codes since November 1995. Based on a Meisei radiosonde history covering 1938-1979 (Ishikawa 1981), histories before 1957 are probably complex with nearly annual model changes.

Gaffen (1996) states that stations change from Meisei S-50M-L to Meisei RSII-56 in 1957 and to Meisei RSII-80 in 1981, and lists the instrument used at most stations in 1993 or 1994. However, 3 Vaisala models (RS 53, RS 56, and RS 2 56) listed in 1960 never existed (Ken Goss, personal communication, 2003). The Antarctic station Syowa (89532) is documented in issues of *Japanese Antarctic Research Expedition Data Report* (e. g., JMA 1994), and 22 Japanese weather ships are partly documented in some journal papers and field program reports.

Meisei RSII-56, with a bimetal thermometer in a duct and a hair hygrometer (Ninomiya 1975, p. II), is used starting 1973. In the archived data, dew points are reported at levels warmer than -30° C. Humidity is usually moist, but can be 1% in rare cases.

The next instrument is Meisei RSII-80, with an external thermistor and a carbon hygristor (JMA 1994). In Gaffen (1996), the change from RSII-56 to RSII-80 occurs in March 1981 at 18 civilian stations but is not reported at 9 military stations. In the archived data, the lowest temperature with a reported dew point changes from about -30° to -40° at civilian stations in March 1981 and at military stations in April 1982. Time series of more conventional variables show less distinct or no discontinuities at these times, but the total precipitable water becomes drier. However, 0000 UTC (daytime in Japan) minus 1200 UTC (night) temperatures at 100 hPa do show a discontinuity, with RSII-56 averaging about 0.8° C warmer at 100 hPa in daytime than at night, and RSII-80 averaging only about 0.3° warmer. Discontinuities are even larger and more consistent at higher altitudes.

The dates of instrument changes from RSII-56 to RSII-80 are confirmed exactly from time series of individual observations. The clearest signal of the use of Meisei RSII-80 is a change from about -30° C to -40° in the coldest temperature with a reported dew point. At most stations, RSII-56 is not used again after the first RSII-80 sounding. However, one of the military stations (47580) has dew points reported to temperatures alternating around -40° and -30° for about 3 weeks in April 1982, indicating use of both RSII-56 and RSII-80 in that period before a complete change to RSII-80 (Russian stations, especially, show similar alternations on a more or less irregular schedule, often for several years, indicating that it is very common for stations to use multiple radiosonde types for prolonged periods).

While reporting the dew point until the temperature decreases to -30° or to -40° is not an inherent characteristic of a specific radiosonde (but some early Japanese radiosondes had a mercury thermometer wired to break a contact and terminate the humidity signal transmission at -30°), the observed behavior is almost definitely a signal of the instrument change. If the change in reporting had simply been an administrative decision unrelated to an instrument change, it could have been implemented on or about the same day at all stations, no drying would have been noticed at that time, and no station would have alternated back and forth between reporting policies.

Since Japanese stations cover a wide range of climate environments, the particular average values of variables vary with station location and season. However, all stations changing from RSII-56 to RSII-80 (89532, the Japanese Antarctic site at Syowa, used special radiosondes with carbon hygristors starting 1966, so it did not make this psecific transition) show drying coinciding with the instrument change.

The transition to RSII-80 illustrates several points: (1) Sensitive variables can make an instrument transition quite obvious. (2) Signals of an instrument type are consistent over a wide range of climate conditions. (3)

An instrument transition changes all variables simultaneously. (4) In a less sensitive variable, the discontinuity exists but may be indistinct or not detectable. (5) The same signals are seen whether metadata has or has not been found. (6) It is common for a station to use more than one instrument type at a time.

The next major Japanese model, Meisei RSII-91, has Vaisala RS-80 sensors (Yagi et al. 1996). Gaffen (1996) reports RSII-91 in use at 16 stations in 1993 or 1994 but does not list transition dates. When Japanese stations report 31313 codes starting 8 November 1995, these 16 stations use RSII-91 (code 74702), 1 station uses Vaisala RS80 Omega (code 46105), and 5 stations use RSII-80 (code 72202 or 02202). Transition dates from RSII-80 before 8 November 1995 (16 transitions to Meisei RSII-91 and 1 transition to Vaisala RS80) need to be determined.

In monthly averages, even of sensitive variables, transitions to RSII-91 are somewhat ambiguous, but individual soundings show a distinct change. RSII-91 is drier than RSII-80, with lower humidity, larger dew point depressions, and colder dew points. While both RSII-80 and RSII-91 report dew points to temperatures around -40°, with RSII-91 the coldest temperature with a reported dew point is almost always exactly -39.9 or -40.1° C, while with RSII-80 the coldest temperature with a reported dew point may vary by several degrees. This minor change in reporting is enough to identify the instrument transition with a maximum error of a day or so. The instrument-related explanation for the change is that RSII-80 uses a baroswitch, so it alternates between transmitting a temperature and humidity signal, while the electronic processing of the RSII-91 transmits a complete measurement each second. Stations started using RSII-91 between 21 February 1992 and 27 February 1997, except for 2 stations (47881 and 47981) that still use RSII-80 in 2007.

However, 7 Japanese stations started reporting dew points almost always to a temperature exactly -39.9 or -40.1° between 1986 and 1991 without drying until later (between 1 October 1992 and 25 February 1996). The beginning of drying is interpreted as the beginning of the use of RSII-91. It is possible that these stations used an undocumented radiosonde with an RSII-91 capacitive pressure cell and an RSII-80 thermistor and hygristor.

While the transitions from Meisei RSII-80 to Vaisala RS80 at 2 Japanese military stations are not documented, those transitions are identified by slightly different dew point reporting. Military stations report dew points to temperatures slightly below -40° starting July 1982, and Vaisala reports dew points to a temperature at or above -40.1°. Also, in the data Vaisala RS80 is slightly drier than Meisei RSII-80 and RSII-91.

Four Meisei RSII-91 varieties are documented. Kitaoka (1997) mentions a "Type 93" uncoated and "Type 94" antiradiation coated thermistor, and Ishihara (2004) describes a new humidity sensor starting about July 1999 which corrects a dry bias but causes a moist bias in cold conditions, and a humidity correction starting February 2003.

Evidence of the thermistor change is ambiguous. Northern Japan stations show steplike warming in day minus night temperatures at 20 to 30 hPa if RSII-91 starts before mid-1993, but such a change is less evident at the southernmost Japanese stations. Also, the day minus night difference decreases somewhat gradually through 1995, but an increase in the difference is not evident at some stations where RSII-91 starts in 1994 or 1995. It is possible that several experimental coatings were tried and instroduced gradually.

The other transitions between RSII-91 varieties can be identifed within about 2 days even though the changes are extremely small. All stations using Meisei RSII-91 show the same signals. Before the new humidity sensor is introduced, there are quite a few reports of 1 percent relative humidity. With the new humidity sensor there are few or no humidity reports under 3 percent, and reports of 1 percent humidities resume after the correction. The new humidity sensor begins at civilian stations in late June or early July 1999, and at military stations in December 1999 or January 2000. The bias correction starts 10 December 2002 at station 47971 and between 28 January and 2 February 2003 at all other stations using RSII-91. Since the humidity sensor change and the later algorithm change are barely noticeable in sensitive variables, if Ishihara (2004) had not been found, these changes would not have been noticed.

The recent Japan history is diverse and is documented by 31313 codes. Most stations in 2007 use Meisei RSII-91, 2 stations each use Meisei RSII-80 and Meisei RS-01G, and single stations use Vaisala RS92 Autosonde and Sippican Mark IIA. In sensitive variables, these instruments have only small differences, mainly in distributions of lowest reported humidity. Meisei RS-01G is described at http://www.meisei.co.jp/english/product/p0111.htm. The Syowa (89532) history since 1973 includes special Antarctic models, with a carbon hygristor used with all models starting 11 February 1966, so differences between models through Meisei RSII-80 are small. Some Japanese ships used special radiosonde models such as ES61A Echosonde and SCM (Ninomiya 1975, p. II), but according to 31313 codes since about 1997, most Japanese ships use Vaisala RS80. Overall, the instrument history for Japan starting 1973, at least in GTS data, is very close to exact and complete.

3. PROPOSED INSTRUMENT ADJUSTMENT PROCEDURE

Most researchers make adjustments for each discontinuity at each station, working backward from the latest data. Some difficulties with this approach are (1) The current instrument is not adjusted even if it is known to be biased, (2) A new breakpoint requires all earlier adjustments to be recomputed, (3) It is difficult to adjust for closely-spaced transitions or use of multiple radiosonde types at the same station in a period, (4) An excessive number of adjustments is needed if actual complex station histories are accounted for, (5) As the number of adjustments rises, the statistical uncertainty of the reconstructed trend increases substantially, and (6) As the number of adjustments rises, more and more of the real trend tends to be removed.

The approach proposed here differs fundamentally by developing adjustments for each distinct instrument type instead of for each station. The technique is "equiprobability transformation" (Eskridge et al. 1995), which replaces data values to transform the cumulative probability distribution observed by each instrument to the same percentiles of the distribution observed by a chosen "reference" instrument, which makes the readings statistically equivalent to the reference instrument. For example, if the observed temperature is the 14th percentile, the 14th percentile observed by the reference instrument is substituted for the temperature observed by this instrument (actually, an amount is added to or subtracted from the observed temperature to transform it to the specified percentile). Variations in environment, including differences between stations, are accounted for by stratifying probability distributions by pressure layer, sun angle, and (for dew point adjustments) temperature interval.

Proposed steps are summarized as follows, with some issues involved in each step described afterward:

(1) Select a "reference" instrument, which is the average of certain widely-used models. Each instrument type is to be adjusted to be statistically equivalent to the reference instrument.

(2) Using completed metadata, for each instrument type, determine a short "chain" of transitions to the reference instrument. For each transition, make a list of stations and time periods using each of the two instrument types.

(3) Develop and apply temperature adjustments first. Determine the cumulative distribution of temperatures at each station and for the group of stations for each instrument type in each list from step 2, stratified by pressure interval and sun angle category. Each adjustment is the difference between the cumulative distributions.

(4) Develop and apply dew point adjustments after temperatures are adjusted. Determine the cumulative distribution of dew point depressions at each station and for the group of stations for each instrument type using the same lists from step 2, stratified by pressure and temperature interval and category of sun angle. Develop dew point adjustments in the same way as the temperature adjustments are developed.

In step 1, the main issue is choosing instruments to be included in the reference, since no instrument is error-free. Should only recent models be reference instruments, or should some older widely-used models be included? Also, the probability distribution for the combined reference instrument is an average of the distributions for each model included, so each instrument in the reference needs to be adjusted to the combined reference.

In step 2, a major issue is to decide how many separate instrument type adjustments should be used. With over 1000 radiosonde types, applying 1000 adjustment schemes would probably remove almost all of the natural trend. Most instrument families will be initially treated as homogeneous. For example, Vaisala RS80 varieties may fall into as few as 2 distinct groups, those with A-Humicap and H-Humicap humidity sensors. The combined instrument is basically the average of the underlying types. If inhomogeneous instruments are combined, the adjustment will not satisfactorily remove biases, and this step will need to be repeated with a revised grouping of minor instrument types.

Step 2 needs to identify a "chain" of instruments to compare, such as "Type A" to "Type B" to a reference. Each "chain" of comparisons should be short to minimize the uncertainty of the final adjustment. For a pair of instrument types, suitable comparisons include a change from one type to the other (in either order) at the same station, frequent alternations if the instrument of each sounding is identified, and simultaneous use at nearby stations. Formal intercomparisons conducted by the WMO, with multiple instruments attached to the same balloon, would be ideal but are not suitable for global data adjustments because the number of soundings is too small. All comparison types can be used to compute the probability distribution for each radiosonde, but each instrument type should have a similar amount of data in each comparison, preferably in an integer number of years to cover the annual cycle, and as many stations as possible using these instruments should be included to provide adequate data to define the probability distributions.

In step 3, it is possible that temperature adjustments for many instrument types should be zero because operational radiation corrections are statistically adequate. For a pair of instruments, if differences in distributions are small and unsystematic, the instruments are probably not detectably different. In this step, the reported dew point depressions can be kept the same, which changes the dew point by the same amount as the temperature, but any bias caused by this change is incorporated into the cumulative probability distribution of dew point depressions in step 4.

In step 4, many issues are the same as in step 3. Probably all instrument pairs will require dew point adjustments (the differences will not be deemed negligible), but for some instruments differences as sun angle varies will be insignificant. In this step, an additional issue is how to develop cumulative probability distributions when moisture data is not always reported. Special procedures are needed if the dew point is not reported when the temperature or humidity is below a certain value, or dew points are randomly missing or end at random levels, or if dew point "censoring" or statistical humidity reporting is used.

The initial effort to develop adjustments will not emphasize estimating statistical uncertainty of adjusted data, because most adjustments are based on such a large number of soundings that the statistical uncertainty is small. However, as in most homogenization efforts, the structural uncertainty can be large. Structural uncertainty arises from issues such as whether the correct discontinuities are identified, whether adjustments remove some of the natural trend, and whether the available stations adequately sample the global average. Because of the large number of validated soundings (about 15 million since 1973), experiments can be performed to estimate structural uncertainty by using only some of the stations in each "chain" of transitions, or using alternate "chains" of transitions to a reference, or including different instruments in the reference.

It will take more than one iteration of these steps to develop reliable adjustments. Time series of sensitive variables and variables of climate interest should be derived from adjusted data. An incorrectly inferred instrument type or period of use should cause an obvious discontinuity for the duration of the erroneous instrument. The metadata needs to be modified, and the adjustments need to be recomputed if this station was used to develop the adjustments. In this way, the adjustment process feeds back into and validates the metadata development process.

Similarly, with complete metadata produced in this project, other researchers can also apply their data adjustment procedures to their data. Where this metadata is correct, their methods should be more successful, and where a discontinuity persists, further testing is needed to determine if their data (such as satellite retrievals) or incorrect radiosonde metadata is the cause of the error.

4. PRELIMINARY EXAMINATION OF JAPANESE RADIOSONDE ADJUSTMENT STEPS

As a first test of the proposed adjustment procedure above, the Japanese radiosonde history was selected because the history is relatively uncomplicated and well-validated. So, the issues mentioned above or other issues arising from performing the proposed steps will not be substantially complicated by issues arising from doubts about actual instrument histories.

Since reported dew point data cannot be properly adjusted before temperature data is satisfactorily homogenized, the discussion in this paper focuses on temperature comparisons. The primary issue in each comparison is whether the observed differences between radiosonde models adequately define the adjustments which should be applied.

Because the most common Japanese station instrument history starting 1973 has been successive use of Meisei RSII-56, RSII-80, and RSII-91, it is logical to select Meisei RSII-91 as the reference instrument in this preliminary test. For a global adjustment, the most logical reference instrument will probably be an average of widely-used relatively recent Vaisala and Sippican (or VIZ) models, but this preliminary test chooses Meisei RSII-91 as the reference because it is still widely used in Japan as of 2007. Most of each station history will be adjusted forward to be equivalent to RSII-91, but stations which have transitioned from RSII-91 to other instruements will be adjusted backward to be equivalent to RSII-91.

As discussed in Section 3, Meisei RSII-91has had at least 4 varieties, with small differences detected in sensitive variables, so even RSII-91 is not homogeneous. The decision about whether all RSII-91 varieties should be combined as a single instrument, or whether only certain RSII-91 varieties should be considered in the reference, or all RSII-91 varieties should be considered in the reference but should be separately adjusted, should be made after examining the differences between the varieties in the data.

So, the first preparatory step was to compare all relevant instrument pairs in terms of differences before and after each transition. With the transition dates known (except for the Type 93 and 94 thermistors, for which the transitions are uncertain), lists were made with each station and time period. To even out seasonal cycles, where possible the instruments were compared for 6 years at each station, 3 years before and 3 years after the transition date. Because formal intercomparisons are not available, at each station the comparisons should be evenly matched in terms of as close to the same sampling of common conditions before and after the transition. So if 3 years are not available before and after the transition, the comparison should cover an integer number of years with each instrument type.

The first major transition, from Meisei RSII-56 to RSII-80, occurred in early March 1981 at all Japanese civilian stations and in April 1982 at 2 military stations, as mentioned above, so the RSII-56 period of comparison was about March 1978 to February 1981 or April 1979 to March 1982 (with dates varying slightly by station), and the RSII-80 period was about March 1981 to February 1984 or April 1982 to March 1985. At each level except the lowest and highest levels, the statistics below are based on approximately 40,000 observations on each side of the comparison.

Average differences (RSII-80 minus RSII-56) including all stations were negative (cooling) in the troposphere and positive (warming) in the stratosphere. This is not the expected difference. If RSII-80 had smaller radiation errors than RSII-56, then RSII-80 should be cooler than RSII-80, probably at all altitudes and especially in the stratosphere. In daytime observations, the difference is lower (less positive or more negative) than in night observations. Specific average differences (°C, RSII-80 minus RSII-56) are as follows:

Pressure level (hPA)	All	Night	Day	Day minus Night
1000	-0.33	-0.31	-0.35	-0.04
500	-0.62	-0.41	-0.81	-0.40
250	-0.93	-0.67	-1.18	-0.51
100	+0.32	+0.62	+0.04	-0.58
70	+1.08	+1.50	+0.68	-0.82
30	+0.39	+1.10	-0.22	-1.32

In the table above, the average of all observations is not necessarily the same as the average of the day and night numbers due to varying relative proportions of day and night observations, and because observations with low sun angles (-4 to 4°) are not counted as either day or night observations. The interpretation of the last column is given later.

The main explanation for this unexpected difference is the eruption of El Chichon on 28 March 1982. The stratospheric warming is included in the RSII-80 period, so RSII-80 is warmer than RSII-56 in the stratosphere. This is supported by warming seen at night as well as in the daytime. It is possible that the night difference is the true (natural) difference between the two periods. The average difference between instruments is lower (less positive or more negative) in the daytime. Possibly a better idea of the instrumental difference is the "difference of the differences" in the last column. This is specifically (the RSII-80 minus RSII-56 average temperature at a specific level using daytime observations), minus (the RSII-80 minus RSII-56 average using night observations), which is column 4 minus column 3 in the table.

To make the RSII-56 observations statistically equivalent to the RSII-80 observations in this sense, the number in the last column is added to each temperature at the specified level with a daytime observation (similar adjustments can be computed at other mandatory levels, and adjustments can be interpolated between mandatory levels). The adjustments are negative, meaning that daytime RSII-56 temperatures are lowered, and implying that RSII-56 was subject to larger daytime radiative heating errors than RSII-80.

Note that after making the proposed adjustments, RSII-80 is even more warmed relative to RSII-56, and the apparent discontinuity is made larger. However, the proposed adjustments would be applied to RSII-56 even before the comparison period of March 1978 to February 1981, and the RSII-80 data continues until the instrument is replaced by RSII-91. Considering the entire RSII-56 and RSII-80 period, the warmth concentrated in 1982 and 1983 would be seen as an outlier, but the entire period seems more homogeneous after the adjustment.

There are two additional factors that mean that this proposed adjustment is preliminary and needs to be studied further:

First, while instrument differences were developed by sun angle interval as well as by pressure level, the discussion above develops differences only in terms of day and night observations. Further checking will be needed to determine if the differences by sun angle seen in the data are legitimate. Generally, a radiative heating error is expected to increase with sun angle, but in the data the pattern is more irregular. Part of the reason for this is that, with observations almost always taken at fixed times (usually specified as 0000 and 1200 UTC), the sun angles do not vary randomly so a given sun angle represents specific times of the year. In the comparison period, natural deviations in months when observation times are in one sun angle range (such as 45 to 54°) may differ considerably from the natural deviations in months when the sun angle at the observation time is in a different range (such as 65 to 74°). When different stations are grouped together, the same seasons will represent different sun angle ranges if the stations cover a considerable latitude range. So, a table of average differences of one instrument type minus another instrument type, stratified by pressure level and sun angle, may not show a smooth pattern in the portion of the table which shows variations according to sun angle. For all observations, for all night observations, or for all day observations, the patterns of differences according to pressure level are much more consistent.

Second, the above adjustment is zero for night observations, and it is not always true that the instrument error at night is zero. Based on theoretical calculations (Luers and Eskridge 1998), some radiosonde models including MeiseiRSII-80 (Computed errors are not shown for RSII-56) may have had radiative errors about as large (too cold) at high altitudes at night as in the daytime. Confirming such an error will probably require extensive comparison with other data sources such as satellite retrievals, and it will be difficult to make an adequate comparison if the instrument change also occurs near the time of a satellite change.

Similar instrument differences were computed for the RSII-80 to RSII-91 transition, the introduction of the new humidity sensor on Meisei RSII-91, the introduction of the humidity correction algorithm on RSII-91, and the transition of Meisei RS-01G "from" RSII-91. The last transition is computed in the opposite sequence (instrument group 1 is RS-01G, the most recent instrument, and instrument group 2 is RSII-91, the older instrument). Adjustments computed in the usual way would be applied to the RS-01G observations to adjust RS-01G readings to be equivalent to RSII-91.

The RSII-80 to RSII-91 transition, like the RSII-56 to RSII-80 transition, was affected by a volcano, which in this case was the eruption of Pinatubo on 15 June 1991. While some stations introduced RSII-91 as early as 1992, most stations introduced RSII-91 in 1994 or 1995. This means that the stratospheric warming from Pinatubo was mostly in the RSII-80 comparison period. Specific average differences (°C, RSII-91 minus RSII-80) are as follows:

Pressure level (hPA)	All	Night	Day	Day minus Night
1000	-0.25	-0.05	-0.42	-0.37
500	-0.62	-0.54	-0.70	-0.26
250	-0.18	-0.08	-0.27	-0.19
100	-0.61	-0.52	-0.70	-0.18
50	-1.20	-1.09	-1.28	-0.19
20	-0.74	-0.69	-0.70	-0.01

Computing the day minus night "difference of differences" as done previously, RSII-91 is assumed to be slightly colder than RSII-80. However, the difference between RSII-91 and RSII-80 is smaller than the difference between RSII-80 and RSII-56, and diminishes with height. The amount in the last column would be added to each daytime RSII-80 temperature and also would be added to each previously-adjusted RSII-56 temperature. As with the previous adjustment, this assumes that the night error of all radiosonde types involved is zero. Also, in this case, the RSII-91 comparison period includes both Type 93 and Type 94 thermistors.

For the two instrument changes within RSII-91 with similar statistics computed, the instrument difference involves only the humidity sensor or humidity computations, so the temperature differences probably should be zero. The difference computations (not shown) are not all small, but the "difference of differences" values are small and unsystematic through 100 hPa. There is enough data to compute differences up to 10 hPa, and the "difference of differences" is about

+0.5° for each transition at 10 hPa. This means that the RSII-91 temperatures at 10 hPA before 1999 (with the previous humidity sensor, but with no documented temperature sensor changes) would be increased 1.0° relative to the most recent values, while presumably no corrections should be made. Note that the comparison period includes the 1997-1998 El Niño and the 1999-2000 La Niña, so some natural variations are large in these periods.

Some general lessons from the sample comparisons above, and other comparisons which are not discussed because the instrument changes involved only a few stations, are as follows:

(1) Even when a large number of observations at a substantial number of stations over a large geographical area are used to make a comparison, it is simplistic to assume that the proper adjustment to make two instrument types comparable is equal to the difference between instruments (specifically, the average of the second instrument type minus the first type, with this quantity added to readings from the first instrument type to make it statistically equivalent to the second type).

(2) When the comparisons are stratified by intervals of sun angle, because the observations are generally not taken at random times of the day, the differences between instruments in a sun angle interval are based on a restricted subset of stations (according to latitude) and seasons. Therefore, it may be difficult to determine if an instrument error increases with sun angle. However, the vertical structure of instrument differences at various levels for all night observations or all observations should show a coherent structure, both at an individual station and averaged over the entire group of stations used to compare two instrument types.

(3) Even if a known instrument change should not affect a specific data element (such as temperature at various levels), the difference between instrument types based on a substantial period before and after the change probably will still be nonnegligible due to real climate variations which differ between the comparison periods.

(4) If the time periods are not carefully matched on each side of the comparison, the apparent differences between instruments are exaggerated. This was found in comparing Vaisala RS80 at 2 stations with Meisei RSII-91 for a year around the transition. While Meisei RSII-91 was used beyond the selected 1-year period, Vaisala RS80 was used only for 6 months (October through March) at each station, so the differences between instruments were large. The comparison should include only the same time of the year for each instrument, and even then, there may be a noticeable difference between years.

As an additional comparison, at the 21 stations with a sufficient number of day and night observations in all or nearly all years between 1973 and February 2007 to compute linear regression trends, the trends at 20 and 500 hPa were computed for all observations and for day and night observations separately. Here, an observation with a sun angle of 15 to 90° is considered a daytime observation and a night observation has a sun angle of -90 to -5°, so observations with sun angles between -4 and 14° are excluded. Also, all Japanese ship reports are combined as a single "station," and its trend is less reliable because Japanese ships reported from different locations in different years. At the Japanese Antarctic station (Syowa, 89532), the daily observations are counted as "day" and "night" only close to the equinoxes, and the number of comparisons is too small for the trends of day minus night temperatures to be reliable.

First, in the day minus night temperatures at 20 hPa, all stations except 47681 and 47991 show distinct discontinuities at the time of the change from Meisei RSII-56 to RSII-80 (a reduction in the day minus night difference, indicating a fairly large radiative error with RSII-56). At station 47681, few soundings reach 20 hPa so the data for comparison is sparse, and at station 47991, there is a weak discontinuity, evident in annual instead of monthly averages.

This discontinuity should explain some of the differing trends in temperatures in the day compared to at night. Including all observations, the regression slopes show cooling at higher altitudes and warming near the surface, with the change from a positive slope near the surface and a negative slope at higher altitudes occurring around 250 to 350 hPa in most of the Japanese mainland and around 180 to 200 hPa in southern Japan and at island stations. For daytime observations, the corresponding change is at about 300 hPa in northern Japan to 600 to 750 hPa at southern island stations. For night observations, this change occurs between 150 and 200 hPa at nearly all stations.

At 20 hPa, the computed regression slope using daytime observations is about -0.07 to -0.10° per year (multiply by about 33 to estimate the total linear trend since the data checked ends in February 2007), -0.01 to -0.04° per year using night observations, and -0.05 to -0.08° per year using all observations at each station except 2 of the southern island stations. The northern stations tend to show stronger trends. At the southern island stations, the trends were somewhat less, about -0.02 to -0.04° per year using daytime observations, -0.01° per year using night observations, and -0.02 to -0.03° per year using all observations, and their day minus night discontinuities at the transition from RSII-56 to RSII-80 are somewhat weak. Overall, the patterns of averages, trends, and other statistics have fairly smooth spatial variations, which is expected because almost all stations have similar data quality.

At 500 hPa, the computed regression slopes using daytime observations show weak warming in northern Japan (about 0.01° per year) and weak cooling in southern Japan (about -0.01° per year), trends using night observations around +0.03 to 0.04° per year in northern Japan to +0.01° per year in southern Japan, and trends using all observations from about +0.02 to 0.03° per year in northern Japan to very weak warming or cooling in southern Japan.

While the trends computed using night observations are probably more nearly correct, this hypothesis cannot be completely proved using only radiosonde data. If some radiosonde model has a large radiative cooling error at night, then a persistent warming discontinuity, increasing with height, should be seen when that model is replaced by an instrument with a less severe error. However, no consistent discontinuities are seen in night temperature data, even at high levels, when instruments change. The largest signals at those levels are temporary warming after volcanic eruptions and some other transient anomalies for which the causes have not been analyzed.

5. DATA AVAILABILITY

While global metadata is still being prepared, incomplete versions of various files are being made available periodically at the Texas A&M University Atmospheric Sciences FTP site by anonymous ftp at ftp.met.tamu.edu. The files are in directory /data/ftp/pub/schroeder. For files with different versions, the latest version contains the largest number, such as rg5.f. For some files, the version number is the date in the form YYMMDD, such as RaobMetadata.080101.

Acknowledgements

This project has been partially supported by NOAA Office of Global Programs grant NA04OAR310125 and by contracts with the NOAA Applied Research Center for Data Set Development through the National Climatic Data Center.

References

Angell, J. K., 2003: Effect of exclusion of anomalous tropical stations on temperature trends from a 63-station radiosonde network, and comparison with other analyses. *J. Climate*, **16**, 2288-2295.

Christy, J. R., W. B. Norris, R. W. Spencer, and J. J. Hnilo, 2007: Tropospheric temperature change since 1979 from tropical radiosonde and satellite measurements. *J. Geophys. Research*, **112**, doi:10.1029/2005JD006881.

Durre, I., T. C. Peterson, and R. S. Vose, 2002: Evaluation of the effect of the Luers-Eskridge radiation adjustments on radiosonde temperature homogeneity. *J. Climate*, **15**, 1335-1347.

Durre, I., R. S. Vose, and D. B. Wuertz, 2006: Overview of the Integrated Global Radiosonde Archive. *J. Climate*, **19**, 53-68.

Eskridge, R. E., O. A. Alduchov, I. V. Chernykh, Z. Panmao, A. C. Polansky, and S. R. Doty, 1995: A Comprehensive Aerological Reference Data Set (CARDS): Rough and systematic errors. *Bull. Amer. Meteor. Soc.*, **76**, 1759-1775.

Free, M., J. K. Angell, I. Durre, J. Lanzante, T. C. Peterson, and D. J. Seidel, 2004: Using first differences to reduce inhomogeneity in radiosonde temperature datasets. *J. Climate*, **17**, 4171-4179.

Gaffen, D. J., 1996: *A Digitized Metadata Set of Global Upper-Air Station Histories*. NOAA Technical Memorandum ERL ARL-211, Silver Spring, MD, 38 pp. (available through http://www.ncdc.noaa.gov/oa/cab/igra/papers/gaffen1996.pdf and http://www1.ncdc.noaa.gov/pub/data/stnhistory)

Gaffen, D. J., M. A. Sargent, R. E. Habermann, and J. R. Lanzante, 2000: Sensitivity of tropospheric and stratospheric temperature trends to radiosonde data quality. *J. Climate*, **13**, 1776-1796.

Haimberger, L., 2007: Homogenization of radiosonde temperature time series using innovation statistics. *J. Climate*, **20**, 1377-1403.

IGRA, 2006: Integrated Global Radiosonde Archive Metadata, http://www.ncdc.noaa.gov/oa/cab/igra/index.php?name=metadata.

Ishihara, M., 2004: Presentation on Recent National Tests/Comparisons. Recent Tests and Comparisons of Radiosonde Operated by Japan Meteorological Agency. WMO, Geneva. [Available online at http://www.wmo.int/pages/prog/www/IMOP/meetings/Upper-Air/Systems-Intercomp/Doc3-2(3).pdf]

Ishikawa, K., 1981: *Radiosonde History of Meisei Electric Company* (in Japanese). Tokyo, Meisei Electric Co., 17 pp. [Available online at http://www.meisei.co.jp/museum/rsonde/index.html]

JMA, 1994: *Antarctic Meteorological Data Obtained by the Japanese Antarctic Research Expedition, Special Volume No. V, Summary of Surface and Aerological Observations at Syowa Station (1957-1993)*. Tokyo, Japan Meteorological Agency, 341 pp.

Kitaoka, T., 1997: Radiational errors of temperature sensors of radiosondes. *Proceedings of the First SPARC General Assembly*, Melbourne, Australia, 229-232

Lanzante, J. R., 2007: Diagnosis of radiosonde vertical trend profiles: Comparing the influence of data homogenization versus model forcings. *J. Climate*, **20**, 5356-5364.

Lanzante, J. R., S. A. Klein, and D. J. Seidel, 2003: Temporal homogenization of monthly radiosonde temperature data. Part I: Methodology. *J. Climate*, **16**, 224-240.

Luers, J. K., and R. E. Eskridge, 1998: Use of radiosonde temperature data in climate studies. *J. Climate*, **11**, 1002-1019.

Ninomiya, K., 1975: *AMTEX ’75 Data Report, Vol. 2, Aerological Data*. Tokyo, Management Committee for AMTEX, 510 pp.

Randel, W. J., and F. Wu, 2006: Biases in stratospheric and tropospheric temperature trends derived from historical radiosonde data. *J. Climate*, **19**, 2094-2104.

Ross, R. J., and W. P. Elliott, 1999: A search for tropospheric water vapor trends: 1948-1995. *Preprints, Tenth Symposium on Global Change Studies*, Dallas, TX, 126-128.

Schroeder, S. R., 2007: Using sensitive variables to validate and complete global historical radiosonde metadata - Toward computing atmospheric climate trends adjusted for instrument changes. *14th Symposium on Meteorological Observations and Instrumentation*, San Antonio, TX, Poster JP1.2, http://ams.confex.com/ams/pdfpapers/119792.pdf.

Sherwood, S. C., 2007: Simultaneous detection of climate change and observing biases in a network with incomplete sampling. *J. Climate*, **20**, 4047-4062.

Sherwood, S. C., J. R. Lanzante, and C. L. Meyer, 2005: Radiosonde daytime biases and late-20th century warming. *Science*, **309**, 1556-1559.

Sterin, A. M., 1999: An analysis of linear trends in the free atmosphere temperature series for 1958 - 1997 (in Russian). *Meteorol. Gidrol.*, **5**, 52-68.

Thorne, P. W., D. E. Parker, S. F. B. Tett, P. D. Jones, M. McCarthy, H. Coleman, and P. Brohan, 2005: Revisiting radiosonde upper air temperatures from 1958 to 2002. *J. Geophysical Research*, **110**, D18105, doi:10.1029/2004JD005753.

Wang, J.H., H. L. Cole, D. J. Carlson, E. R. Miller, K. Beierle, A. Paukkunen, and T. K. Laine, 2002: Corrections of humidity measurement errors from the Vaisala RS80 radiosonde - Application to TOGA COARE data. *J. Atmos. Oceanic Tech.*, **19**, 981-1002.

Yagi, S., A. Mita, and N. Inoue, 1996: *WMO International Radiosonde Comparison (Phase IV, Tsukuba, Japan, 15 February - 12 March 1993), Final Report* (WMO/TD-No. 742, Instruments and Observing Methods Report No. 59). Geneva, Switzerland, WMO.