JP1.2 USING SENSITIVE VARIABLES TO VALIDATE AND COMPLETE GLOBAL HISTORICAL RADIOSONDE METADATA - TOWARD COMPUTING ATMOSPHERIC CLIMATE TRENDS ADJUSTED FOR INSTRUMENT CHANGES

Steven R. Schroeder *

Texas A&M University, College Station, Texas

ABSTRACT

All long-term temperature and moisture trends derived from radiosonde data are questioned because of instrument changes, which superimpose artificial cooling and drying (on the average) onto the actual trend as sensors have become more responsive and better protected from radiative errors. Proper adjustments for instrument biases require complete metadata, listing the location and instrument history for each station including dates of changes, but the best available metadata compilations have many gaps and uncertainties.

Two methods to fill in missing metadata have limited success. First, ongoing efforts to obtain historical metadata by personal contacts have provided only a small amount of additional information so far, and it is known that much information is no longer available or was never preserved. Second, it is logical to attempt to identify transitions using the data itself by seeking discontinuities in station time series of temperature and moisture variables of research interest, but even though instrument-related discontinuities are definitely large enough to seriously contaminate trends, they are often not large enough to be distinguished from natural variations.

This research uses variations of both methods, both to validate existing metadata and fill in missing metadata. First, an extensive search of peer-reviewed and nontraditional sources has uncovered a large amount of unexploited metadata. Second, and most importantly, an extension of the second method is to systematically examine time series of especially sensitive variables to identify instrument characteristics and changes at each station. These variables have little or no climatic interest but amplify instrument differences to show very consistent signatures of instrument types at documented stations. Similar signals at stations or in time periods where metadata is missing or inaccurate indicate the use of the same instruments. The combination of expanded metadata and systematic examination of sensitive variables leads to very reliable inferred metadata, including the timing of transitions.

While examination of all data archived from the Global Telecommunication System back to 1973 is not complete, this approach is validated by producing complete metadata in regions which are transitioning to new instrument types, including the Russian Federation, India, Japan, China, and their Antarctic and ship stations. After this approach is used to develop complete global metadata, researchers will be able to use knowledge of the nature and timing of instrument changes to more confidently adjust for the effects of instruments on trends of atmospheric temperatures and other variables of research interest.

1. INTRODUCTION

Archived radiosonde observations are potentially of great value for detecting climate responses to the ongoing greenhouse gas buildup, because most of the industrial-era increase in greenhouse gases has occurred since near-global radiosonde data became available around 1958. Preliminary indications of the nature of hypothesized feedbacks, which may either amplify or partly offset future greenhouse gas warming and other climate changes, should already be detectable in large-scale averages. Unbiased radiosonde time series also provide better inputs to reanalyses and are needed to validate satellite retrievals and model simulations.

However, as temperature and humidity sensors have become more responsive and better protected from errors such as radiative heating, artificial trends are superimposed on the true natural trends. Instrument changes cause steplike discontinuities at individual stations. In global averages, the instrument-induced artificial trend is mostly in the direction of cooling and drying, making all computed long-term trends questionable. Factors which complicate the identification of instrument-related discontinuities are that many stations use multiple instrument types for long periods, interim fixes to a known problem may cause a series of small (mostly undetectable) discontinuities, operational data adjustments may overcorrect or undercorrect a bias, and some biases are a sampling problem (for example, where balloons do not reach high levels in cold cases).

To quantify and adjust for instrument-caused changes, complete metadata is needed, listing the history of instruments and procedures at each station including the timing of changes. The Integrated Global Radiosonde Archive (IGRA) (Durre et al. 2006) contains the most extensive available metadata. Over 97 percent of the IGRA metadata records are from Gaffen (1996), which states that the metadata is quite incomplete, and sometimes questionable or inconsistent, even though it comes from documented sources. In addition, most listed events have limited use for deriving adjustments because they are “static events” (“snapshots”) reporting the use of a procedure or instrument at a stated time, but not the starting or ending date.

This research has uncovered a large amount of metadata documentation, and IGRA personnel continue

---

***** *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.

to request historical information from national weather agencies. However, many gaps and inconsistencies remain, and historical metadata will never be completed from documented sources and national inputs. In some cases, information has been lost due to wars or governmental changes, but often the information now considered important was not kept, whether in manuscript or computer-readable form, due to document storage limitations.

1.1. Indirect methods to identify and correct biases

Due to the incomplete state of available metadata, researchers are forced to use indirect methods to attempt to remove biases from climate trends. The adjusted trends are still quite uncertain due to lack of knowledge of all instruments used and when transitions occurred. The main criterion for evaluation of the "success" of a method is whether the subjective adjustments involved lead to a trend that is close to "expected" or at least close to results obtained using another approach. Some methods explored to quantify and adjust temperature errors are as follows:

(1) An automated method to detect and remove discontinuities (Gaffen et al. 2000). This was found to remove essentially all trends, both natural and instrument-caused, regardless of the tuning of the level of sensitivity.

(2) Removal of identified discontinuities only if they approximately coincide with known transitions (Gaffen et al. 2000, Lanzante et al. 2003). The 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) Attempting to use 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 late 2006) use GZZ-2 radiosondes, similar to the Russian A22 series, which was introduced in the USSR in 1957.

(4) Computing "first differences," or the difference from one year to the next in each month, with data deleted around each suspected discontinuity (Free et al. 2004). The remaining first differences are area-averaged, either regionally or globally. The area averages are 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 segments. Again, the decision to delete data segments is subjective, and this method does not produce a time series for any individual station.

(5) Comparison of radiosonde minus satellite time series, where the radiosonde temperatures are averaged in an altitude band corresponding to the satellite weighting function (Randel and Wu 2006). The authors divided station time series into "high bias" and "low bias" groups (where the radiosonde minus satellite time series showed a large or small trend, respectively) and found that the "low bias" group showed a more moderate overall stratospheric cooling trend than the "high bias" group. However, they found many discontinuities that are not coincident with documented or suspected instrument changes.

(6) Comparison of day and night temperatures (Sherwood et al. 2005). A common way to deal with radiative errors in climate statistics is to only include night soundings, but some stations only make daytime soundings, and in summer at high latitudes all soundings can be in the daytime. The authors computed 0000 minus 1200 UTC temperature trends at 50 and 300 hPa, and found 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. The authors state that this adjustment does not correct errors other than radiative heating.

(7) Comparison of each station time series to a constructed "neighbor" time series (Thorne et al. 2005). For any "target" station, the neighbor series is the average of a large number of stations, weighted by the expected correlation (from a reanalysis) of each station location with the "target" location. The assumption is 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. In this research, it was a human decision to choose over 3500 breakpoints at 676 stations. Only 30 percent of the breakpoints were associated with documented changes at the stations. Each adjustment was the temperature change needed to homogenize the difference series. This procedure, even with a large number of adjustments, appears to retain the natural trend.

(8) Computing adjustments based on radiative theory applied to instrument configurations. Luers and Eskridge (1997, 1998) developed theoretical "temperature correction models" of the radiation and lag errors of major radiosonde types. Durre et al. (2002) applied the corrections to station time series where transitions appear well-documented, and found that discontinuities were often larger after the adjustments. While radiation and lag errors are not the only data differences between radiosonde models, the authors acknowledge that some transition dates are not accurate and some stations may have gradually transitioned to a new radiosonde.

(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.

1.2. A new 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. Almost all papers about upper air trends derived from radiosondes state that obtaining more metadata is a critical need. For example, in the Climate Change Science Program report *Temperature Trends in the Lower Atmosphere: Steps for Understanding and Reconciling Differences* (Karl et al. 2006), the first three recommendations on page 14 and the first two recommendations in Chapter 6 emphasize the necessity of obtaining more metadata, as well as previously-unexploited data, to create climate-quality data sets of temperature and additional variables such as atmospheric water vapor content.

This research explores the possibility that metadata can be completed by using the archived data because it is unlikely that historical metadata will ever be completed by seeking information from each country and station. The basic hypothesis is that each instrument type has a consistent signature in especially sensitive variables which amplify differences between instruments by several times, compared to the usual variables of research interest. Using station time series, 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. Since all metadata is subject to errors, even metadata from published sources or personal contacts must be checked. 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.

This project originated in an attempt to develop a long-term satellite-based precipitable water data set (Schroeder and McGuirk 1998a). Because those water vapor retrievals are based on statistical regression calibrated using radiosondes, they reflect errors of radiosonde humidity sensors (Ross and Gaffen 1998) as well as temperature errors. The conclusion that most of the observed tropical drying from 1979 to 1995 was real, without specific knowledge of instrument change dates, was reinforced by searching for discontinuities and noting that 71 moistening and 86 drying discontinuities were seen from 1973 to 1995 at 122 stations with relatively complete time series (Schroeder and McGuirk 1998b). With a steady transition to drier instrument types, there should have been few moistening discontinuities. However, the stated conclusion that much of the observed drying trend was real because the drying and moistening errors nearly cancel out is not very convincing in the absence of metadata.

At the time the research was performed, the metadata of Gaffen (1996) was not widely known. The availability of this extensive metadata collection, combined with the detailed search of precipitable water time series for discontinuities, suggests that a deliberate search for data discontinuities could pinpoint all significant transitions. Specifically, signatures of distinct instrument types are sought, and discontinuities indicate transitions.

A reported sounding contains a large amount of interrelated information, all of which is greatly influenced by instrument characteristics and processing procedures. Variables computed from the soundings which are the most sensitive to different instrument types have little or no meteorological interest, in part because some of these variables are at the extremes of sensitivity of even modern instruments, so they amplify the differences between instruments. These variables 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. As discussed below, 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. With the expanded metadata found in this research, most instrument types and many station transitions are supported by more than one reference, in addition to consistent data signals.

Section 2 summarizes available sources of archived data, and section 3 discusses metadata and its completeness and accuracy. Section 4 discusses how instrument signatures are obtained at stations with reported metadata. Section 5 describes how these signatures are used to construct missing metadata, and section 6 describes a few cases where reported instrument types are validated and instruments are inferred if metadata is not available. Section 7 discusses how the complete instrument metadata is expected to be used to develop instrument adjustments to compensate for biases. Section 8 describes some limitations that may reduce the accuracy of inferred instruments or the accuracy of computed adjustments. While adjustments have not yet been derived in detail, section 9 shows a time series from 1973 to September 2006 of global monthly area-averaged precipitable water using the unadjusted data. The adjusted data and all other variables are expected to be gridded and averaged using the same procedures. Finally, section 10 describes how to obtain interim versions of metadata and some of the data.

2. ARCHIVED RADIOSONDE DATA AND ITS QUALITY

Most 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. IGRA contains only time series of prespecified stations, so most stations with short records are omitted. The disadvantage of that limitation is that newly-established stations, including stations with new ID numbers which replace nearby closed stations, need to be added manually. Also, IGRA does not contain data for ships, including fixed ships, so the amount of oceanic data is significantly reduced.

Little worthwhile climate research can be performed without access to archived individual observations. Historical soundings are available for analysis because of long-term data archiving efforts. For example, all accumulated United States Weather Bureau pilot-balloon observations were punched onto cards starting in 1938 (Dashiell 1938), followed quickly by radiosonde data. However, archived soundings are not exactly the same as the original observations because of limitations and losses in the steps of observing, processing, reporting, and archiving.

First, the instrument does not correctly register atmospheric conditions because of errors such as sensor lags, radiative heating in sunlight or cooling at night, heat from the battery and transmitter, latent heating or cooling from cloud or precipitation particles and evaporation, heating or cooling of the balloon if the instrument rises through its wake, sensor inability to register extremes, and the time interval between transmissions of each meteorological parameter. Inherent instrument limitations cause most of the biases that need to be corrected.

Second, processing on the ground may correct some radiosonde errors (primarily lag and radiation errors), but due to limited time to perform computations needed to prepare an observation, the atmospheric profile is simplified relative to the profile recorded by the instrument. In general, increasing computation capability and communication capacity has allowed the vertical resolution of reported soundings to increase gradually. The increased vertical resolution over time can cause a sampling bias. For example, computing mandatory levels first and then selecting significant levels according to their deviations from a straight line connecting mandatory levels led to a disproportionate number of tropopause reports at exactly 100, 150, 200, and 300 hPa (Endlich 1954).

Third, coding and communication of the observation causes some further losses. For example, since 1968 temperature has been reported to the even tenth of a degree if 0° C or above, or to the odd tenth if below 0°, and the dew point depression is reported to the nearest 0.1° up to 5° C or to the nearest whole degree from 6 to 49° C (WBAN, 1968). From 1949 through 1967, the temperature and dew point were both reported with a precision of about 0.3° C (Ratner 1964, pp. 40-41). Before 1949 the temperature was reported to the nearest whole degree and moisture data was reported in the form of relative and specific humidity (USWB, 1941). Other major reporting changes have occurred which are omitted here. As long as units of measure are correctly interpreted, these changes should not cause biases at the reported levels, but they can cause false trends in some variables such as inversion or tropopause heights as vertical profiles have become less smoothed because of higher reporting precision.

Finally, not all observational data has been archived in digital form. Most early digital archiving used 80-character punched cards. The NCDC collection peaked at around 600 million cards or microfilmed cards in the early 1970s (personal communication), equivalent to a capacity of nearly 50 gigabytes, now the size of a small personal computer hard drive. Observations sometimes were truncated to minimize the number of cards, but although most weather reporting codes since the 1940s were designed to fit card formats, choices were made about what was archived. DS353.4 uses a format which includes all upper air data elements, but some items (such as the 31313 instrument code) may not have been added to the structure until after some stations reported those data elements for a while. DS353.4 omits some data elements in certain periods for unknown reasons, such as significant wind levels in most of the world from August 1989 to August 1991 (In Kalnay et al. (1996, pp. 441-442) the European Centre for Medium-Range Weather Forecasts (ECMWF) archive has missing significant wind levels in different parts of the world in that period). Also, data values are archived after quality control, and the procedures change occasionally. For example, from 1 April 1997 to 18 February 1999, many dew point depressions which reported in whole degrees are changed by 0.1°, such as to 8.9 or 19.1°. Fortunately, DS353.4 appears to only minimally change or omit temperature and dew point data, so it is quite satisfactory for use in detecting most instrument-related changes.

3. AVAILABLE METADATA AND ITS QUALITY

The location and elevation of an operational land station is not included with observations but must be obtained from a station catalog. Similarly, observations did not include instrument type information until the 31313 group was established in the late 1980s. Since the late 1990s, about 80 percent of all soundings contain a 31313 group, and the WMO uses that information as the primary source for updates to the *Catalogue of Radiosondes and Upper-Air Wind-Finding Systems* (WMO, 2006a and earlier dates).

The most extensive global radiosonde station and instrument history is a composite of three sources. Gaffen (1993) obtained responses about radiosonde histories from 50 countries. Gaffen (1996) combines this information with additional published documentation sources and some personal communications. IGRA (2006) extracts the metadata for IGRA stations and adds recently-obtained updates. However, only 3 percent of the IGRA metadata records are not in Gaffen (1996), which illustrates how slowly additional historical metadata is obtained by contact with individual countries. One reason for the slow rate of updates to the IGRA metadata is that IGRA does not yet consider the instrument type reports from the 31313 group.

A literature search starting with the sources referenced in Gaffen (1993, 1996) revealed that there is a large amount of station and instrument metadata which can be added to the metadata. Sources include scientific journals (both United States and international, many not in English), reports of field programs, some books, military manuals and reports, conference preprints, about

100 actual radiosondes in collections (at TAMU, the NCDC Weather Museum, and the Smithsonian National Museum of American History), manufacturer web sites, exhibits at American Meteorological Society annual meetings, WMO documents, other web searches, and even advertisements in journals. As examples of additional information, field program reports often contain detailed descriptions and operating instructions for the radiosondes used in those programs, and foreign journals were especially helpful in identifying radiosondes manufactured or used in India, Japan, East and West Germany, Finland, Italy, Canada and France. Also, on-line WMO documents used in working groups help clarify the recent India and China radiosonde changes.

A major source of radiosonde metadata is the *WMO Catalogue of Radiosondes and Upper-Air Wind Systems in Use by Members*, which is updated about every 2 years (WMO 2006a). New stations (both surface and upper air) are usually also listed in WMO Publication 9A, *Observing Stations* (WMO 2006b), which is updated nearly every week. Instrument types in the latest WMO radiosonde catalog are mostly based on 31313 instrument codes reported in observations, with some personal communications to clarify apparent erroneous codes (For example, France reported using an early 1980s instrument from Czechoslovakia starting in late 2004, while it actually introduced the Vaisala RS92 at that time). Other catalogs back to 1942 have been found, and it appears that upper air stations started using the 5-digit WMO station numbering system at some time between 1949 and 1953. The disadvantage of these catalogs is that they contain “snapshots” and the time of a location or elevation change is not stated.

The National Weather Service (NWS) upper air web site (http://www.ua.nws.noaa.gov) contains dates of instrument changes and station moves back to the middle 1990s, Elliott et al. (2002) contains similar information on events from 1988 to about 2000, and Schwartz and Govett (1992) lists station locations and some instrument changes from the beginning of most station records to about 1990. The first two sources include operational NWS stations only, and Schwartz and Govett (1992) adds military, Canadian, Mexican, Caribbean, and some supplemental NWS stations. So, the North American station history is quite accurately documented, although there are a few inconsistencies and errors. One error is that some supplementary stations, operated briefly in field programs, were assigned the same 5-digit station numbers as nearby operational stations by Schwartz and Govett (1992), and those errors propagated into Gaffen (1996) and the IGRA (2006) metadata files. The IGRA file shows erratic locations, names, and moves (for example, at Pittsburgh, PA), when actually several stations operating simultaneously are listed as a single station.

DS353.4 attaches a latitude, longitude, and elevation (but not a station name) to each observation from an NCEP catalog. Location metadata for a station changes with the first observation after the change is entered into the NCEP catalog, but updates were very infrequent until 1996. Many elevations were erroneous by 30 to over 100 meters from 20 January 1976 to 4 February 1980, and from 4 June 1986 to 30 June 1989. Surface and upper air observing sites with the same WMO station number may be different, and it appears that the 20 January 1976 and 4 June 1986 updates erroneously listed the surface observing site. Also, if a new land station opens, observations do not appear in DS353.4 until the station ID is added to the NCEP metadata, so some data gaps in DS353.4 when one station is replaced by another may not have actually occurred. Overall, the metadata in DS353.4 from the NCEP catalog is more extensive than any other source which has been located, and includes recent military stations with nonstandard station station numbers (starting with 69 or 99) in the Balkans and Middle East, and the locations appear to be of fair quality even if most elevations are inaccurate, as discussed in section 4a.

The 31313 group was established in the late 1980s to report instrument and launch time information with each observation. It was reported by some United States stations irregularly before 1990, and in quite a few countries starting in early 1992, with a gradual increase in the number of stations and accuracy of 31313 reports so roughly 80 percent of the soundings in 2006 contain instrument information. The instrument section of the 31313 group is 5 digits, with the first digit containing a code for the solar or radiation correction, the next 2 digits identifying the radiosonde type and sometimes the ground processing unit, and the last 2 digits identifying the wind finding method and some other conditions of the sounding.

If every observation reports a 31313 group, the instrument transition history for a station is potentially exact. However, codes omit some important radiosonde differences. First, the solar and radiation correction is vague so, for example, different Vaisala RS80 corrections (V82, V86, and V93) are not distinguished. Second, the A-Humicap and H-Humicap variations of Vaisala RS80, which have various levels of erroneous drying due to contamination of the sensor from packaging as well as progressive attempts to fix the problem (Wang et al. 2002, section 4a) are not distinguished. Third, almost all 2-digit instrument codes are assigned, so future models will either need to reuse obsolete codes or will be coded as “unknown instruments.” Fourth, the wind finding method is mixed with other codes such as “systems operating normally, ” so sometimes the wind finding method is not reported. Fifth, parameters as defined in the 31313 code do not have a unique correspondence to announced radiosonde models, so several Vaisala RS80 models may use the same 31313 code, and the same Vaisala RS80 model may use several 31313 codes. Finally, the published tables do not give any references to further details about the instruments or methods. However, as discussed in section 4b, recently the 31313 codes appear to be well over 90 percent correct, and many erroneous values are easily corrected. Because the 31313 codes are not detailed enough, the need for a station and instrument history will still persist in the future.

When development of station metadata was begun, the lack of a suitable list of radiosonde types was noticed. Gaffen (1993) has a list of radiosonde types organized by manufacturer, and a list of VIZ sondes, but many radiosondes were not included, the lists were not organized for convenient insertion of new radiosonde models, and the lists were not systematically documented. A major task in this research was to develop a list of radiosondes and other upper air weather observing equipment (with references), and a code scheme that allows for the systematic insertion of newly-identified models. This task has grown beyond what was originally expected, and currently 2023 codes are assigned. Even with over 2000 codes, this list of instruments is still quite incomplete, some sondes listed are incorrect, some were designed or proposed but not produced, and some instruments may be on the list more than once with different names.

Many sonde types which were considered questionable by Gaffen (1993) have been verified by additional references. For example, in Gaffen (1993) a MARS radiosonde was introduced in Czechoslovakia in 1969, and a widely-used MARS radiosonde series was introduced in the USSR in 1983. These are determined to be different instruments using Air Weather Service Master Station Catalogs from 1977 to 1984, and there are probably at least three Czech MARS varieties. The instrument used in Prague in 1977 was called “MARS 1K” and the instrument used in 1984 (based on an undated list of names from about 1981) was called “ZAP MARS 4WF BERLIN (1680)”, made by Vinohrady. Based on Gaffen (1993), the third Czech MARS model was introduced in 1986 and still used in 1990, made by Metra Praha, with an improved pressure capsule and a USSR transmitter and MMT-1 thermistor (also used in Russian RKZ, MARS, and MRZ radiosondes according to Zaitseva (1993)). So, to add to the confusion, the late 1980s Czech MARS radiosondes probably resembled the Russian MARS radiosondes, while earlier Czech MARS radiosondes most likely resembled the Russian RKZ radiosondes.

4. VALIDATION OF AVAILABLE METADATA

As additional metadata sources are added, the number of inconsistencies increases. Therefore, all metadata must be evaluated for consistency as much as feasible.

Fortunately, any relatively complete radiosonde observation contains a large amount of information which should be internally consistent. The first data check is to examine each observation for errors, to attempt to correct certain errors, and otherwise to reject erroneous data elements or entire soundings. While many tests are applied to each sounding in this project, most of which are similar to Comprehensive Quality Control (Collins 2001a, b), and a few errors are corrected (such as an incorrect temperature sign, or a height error of 500 or 1000 meters), the tests which are most relevant for validation of metadata are discussed below. In most cases, even if a sounding is rejected due to violation of some criterion, calculation of indicators and statistics continues for that sounding to see if the error can potentially be corrected (such as by changing the surface elevation), although the statistics for that sounding are not included in the monthly and annual statistics for that station.

4.1. Validation of locations and elevations

Section 3 mentions that location and elevation data is obtained from station catalogs, but few sources specify the dates of station relocations. For forecasting applications, a small error has little effect, partly because the radiosonde drifts away from the launch location (but the latest mesoscale models are starting to account for the radiosonde drift, and also the flight time to reach each pressure level). Even some climate applications are little affected, such as the temperature trend at a specific level, and globally the location and elevation errors do not cause systematic trends of any variable. These are reasons why location metadata has not been systematically maintained in the past.

However, precipitable water (PW) climatology is affected by inaccurate station elevations. PW is computed by integrating specific humidity over the full atmospheric pressure range, so an incorrect surface elevation causes no error, but if the surface pressure level is omitted from an observation, PW is underestimated. Also, a station PW time series contains a discontinuity if the elevation changes, which cannot be compensated for properly if an elevation is inaccurate. Similarly, when gridding PW to determine spatial averages, each computed PW value is adjusted to the average surface elevation in the grid box.

Because a long-term goal of this project is to determine unbiased moisture trends including PW, it is desirable to identify station locations and elevations as accurately as possible. Based on the number of discrepancies between metadata sources, location and elevation metadata is probably less accurate than instrument metadata, partly because much instrument metadata is simply omitted.

Fortunately, the surface elevation used as the basis for height computations can be reconstructed more accurately than any other data element if the surface pressure is reported. To prepare the original sounding, heights at specified pressure levels are computed by accumulating the thickness of each pressure layer from the surface elevation (if heights are obtained by radar, the pressure thicknesses of height layers are computed instead). Here the calculation proceeds downward from the first reported above-surface height to compute the surface elevation. If the computed and reported surface elevations differ by more than about 30 meters, the sounding is rejected because the surface level is missing, the surface pressure is wrong, or the observation may actually be from another station.

Alternatively, the surface elevation in the metadata could be wrong. Time series of the computed surface elevation are constructed for each station. Alduchov and Eskridge (2002) recommend use of this method to check all elevations in the CARDS (now IGRA) metadata.

In this project, station processing starts with collection of location and elevation metadata, and computation of surface elevations to document the elevation history and attempt to reconcile elevations with reported locations.

The timing of elevation changes is detected first by scanning annual averages (because some stations have large annual cycles in the computed surface elevation), then monthly averages, then time series of individual observations.

The main findings of this step are as follows:

(1) On a global basis, there is little systematic error in the computed surface elevation. The average computed minus reported surface elevation from 1973 through September 2006 is -0.09 meter.

(2) Because pressures are reported to the nearest whole millibar in the reporting code, it is normal to have a variation of about 6 to 8 meters in the computed elevation from observation to observation, equal to the thickness of a 1-hPa layer near sea level, depending on the temperature. Variations are larger at a high-elevation station because a 1-hPa layer is thicker, and the height at the 500-hPa level and above is reported to the nearest 10 meters. In any case, if the reported surface pressures are correct the computed elevation variations should be relatively evenly distributed within the observed range, so the monthly (or at least annual) averages are consistent.

(3) If the computed surface elevation differs greatly from the catalog elevation, and the computed elevation is consistent from observation to observation as above, then the average elevation is likely to be the actual radiosonde launch elevation. Other patterns, such as larger variability or frequent occurrence of surface pressure exactly equal to a mandatory pressure, may indicate a missing surface pressure or an incorrect station mixed in. Even after the metadata in this project is revised to contain hydrostatic elevations, individual observations will be rejected if their computed elevations deviate by more than about 30 meters from the metadata elevation because such observations almost always have a missing or wrong surface pressure, are for another station, or have erroneous reported heights.

(4) This procedure can usually detect elevation changes of 5 meters or more to the exact observation, and changes as small as 1 meter within a few weeks, unless the “annual cycle” of the computed elevation is large.

(5) There are many consistent elevation changes of only 1 or 2 meters. It is likely that many of these are not station moves but result from a resurvey of the station elevation or a minor change in computations, such as a change in procedures associated with a new radiosonde model. Some small discontinuities may also be seen as changes in the observed minus first guess heights in operational forecast center evaluations.

(6) Dates of reported and computed elevation changes almost never match because the catalogs are updated irregularly and do not indicate the date of a change. The only exception is that the location and elevation history for most stations operated by the United States is quite well documented, as mentioned in Section 3. So, hydrostatically-derived elevation change dates are generally more accurate than change dates in most documentation unless the elevation change is only 1 or 2 meters.

(7) While it was hoped that horizontal station moves could be identified by corresponding changes in the computed surface elevation, discrepancies in the metadata are so numerous that moves often cannot be reconciled exactly without more information. In some cases, a computed elevation change is in the opposite direction of an elevation change reported at approximately the same time. It is probable that most moves reported about 20 January 1976 and 4 June 1986 in the NMC metadata are spurious and are actually the surface observing sites. For example,station 22820 (Petrozavodsk, Russia) moved from 111 to 40 meters on 20 January 1976 (with a station move), to 109 meters on 5 February 1980, to 40 meters on 4 June 1986, and back to 109 meters on 29 August 1989. Air Weather Service Master Station Catalogs from 1977 to 1984 show the surface observing site at 40 meters and the upper air site at 109 meters, both with the same latitude and longitude. In this case, the hydrostatic elevation is 111 meters, decreasing to 109 meters on 16 March 1987, which might not have indicated an actual move, and Gaffen (1996) and IGRA (2006) have a constant 109-meter elevation in the entire period. Some apparent discrepancies persist for more than 20 years in different metadata sources. For example, at station 61901 (St. Helena Island) from 5 August 1977 (when the station opens) to 4 February 1980, NMC metadata reports a 630-meter elevation but the hydrostatic elevation is 436 meters, Gaffen (1996) reports this elevation change in September 1993, and IGRA (2006) still has the 630-meter elevation.

(8) Some strange elevation changes are apparently real. For example, at station 21504 (Ostrov Preobrazheniya, Russia), the hydrostatic elevation was 35 meters, but starting 1 December 1990 the hydrostatic elevation was 60 meters in the 0000 UTC observations and 35 meters in the 1200 UTC observations, and starting 1 April 1991 the hydrostatic observation was 60 meters in all observations. Other changes were consistent with using the A22 instrument at the old site, changing to MRZ-3 at the new site, and a 4-month period in between when both sites operated. Also, some fixed ship locations show changes between two different elevation ranges every few weeks because two ships alternated, sometimes using different instruments, at the location. Such changes can result in very complex station histories, but if different locations indicate different instruments at such a station, the instrument history can be precisely documented.

(9) Even if the station location cannot be directly computed from observations, various techniques can be applied to catch most errors. Gross error checks for values outside a normal climate range for a reported location can detect substantially mislocated observations, such as an Arctic observation in the tropics. A change to a new location and back to the former location in metadata is likely to be spurious (as mentioned above), or it could be a temporary move due to damage or construction at the previous site (which should be confirmed by the data if there is an elevation change). For ships, plotting the path as a time series shows many location errors of a multiple of 10 degrees. Sometimes, the wrong hemisphere is stated. Almost any move of about 30 km or more should be checked, especially if the station name is unchanged, and in general, the city in the

station name should be close to the station location. Finally, the reported station elevation should not be substantially too low or too high for the region. An illustration of an error of the last type is station 99877, which started reporting in June 2004 with an NCEP location of 34.27° N, 67.27° E, 679 meters. However, the computed surface elevation is consistently 303 m, the stated location is a very rural area of Afghanistan with an elevation between 2000 and 3000 meters, and the summer surface temperatures are much warmer than expected. Since station 38927 (Termiz, Uzbekistan) is 3° north of the stated location, with an elevation of 312 m at the end of its earlier data series, and climatological values of station 99877 are consistent with the data at Termiz, it appears that this station is Termiz, although the location of 99877 and 38927 may not be exactly the same due to the small difference in the computed elevation.

4.2. Developing signatures to validate instrument types

The previously-available metadata, and the additional metadata found in the literature search, identify the instruments used at many stations in many periods. In addition, the 31313 codes make it theoretically possible to provide exact instrument metadata (subject to the limitations of the 31313 codes) for most stations worldwide since the late 1990s. Using well-documented stations, the main tasks are to determine apparent signatures of documented instrument types, and to assess the degree to which these signatures are consistent between seasons, stations in the same country, and stations in different countries using the same instrument.

To perform the first task, selection of the appropriate variables that seem to be most sensitive to instrument types has been an ongoing process. For the first test of the concept of identifying instrument types from the data (in 1999), annual and monthly averages of various variables were computed (with a limited capability to examine time series of individual observations), and instruments were best distinguished by the average dew point depression at 500 hPa and the average total precipitable water. These or similar variables have been used by other researchers to demonstrate the problem of uncorrected instrument biases, such as examples of time series of precipitable water anomalies shown in Ross and Elliott (1996, pp. 39-40).

The initial investigations in 1999 were quite broad-brush, with instruments divided into only 22 different models for separate corrections. The current examination of characteristics still starts with annual and monthly averages of variables, but time series of variables computed from each observation need to be examined to properly distinguish the far larger number of instrument types which have been used and to identify the observation when the actual transition to a new model is most likely to have occurred.

Since the initial investigations, the preferred variables that best distinguish instrument types have changed, although most variables are still moisture-related. Some of the main variables are the pressure at the top of the sounding, the lowest pressure with a dew point reported, the tropopause pressure, the number of temperature levels, the number of dew point levels, the coldest temperature of the sounding, the tropopause temperature, the 100-hPa temperature, the coldest temperature with a reported dew point, the coldest reported dew point, the largest dew point depression, the surface relative humidity, the lowest and highest relative humidity reported above the surface, the average dew point depression in 200-hPa thick layers centered on 700, 500, and 300 hPa, the number of wind by pressure levels, the lowest pressure level with reported wind data, and the highest height with reported wind by height data. In addition to these and other variables for each sounding, monthly and annual averages are computed for all soundings, for soundings near 0000 and 1200 UTC only (because off-hour soundings often do not reflect a random sample of conditions), for soundings near 0000 UTC separately, near 1200 UTC separately, and for the 0000 minus 1200 UTC difference. Solar angles along with the last set of statistics can help determine if instruments have a noticeable day minus night bias.

For a while, the lowest relative humidity of each sounding was calculated, but recently it was decided to calculate the surface relative humidity, because it is usually measured by a permanent surface-based instrument instead of the radiosonde, as well as the lowest and highest relative humidity above the surface. The highest relative humidity may be a distinguishing variable because some instruments rarely report humidity close to 100 percent even in clouds.

For a particular instrument type, only a few variables tend to be the most consistent indicators of that model, and the other variables simply support those most sensitive variables.

A common policy regarding the reporting of dry conditions is to set a lower limit on the reportable relative humidity, so the lowest relative humidity reported above the surface is the most frequent distinguishing variable. The most frequent lower limit is 10 percent, but for many instruments the limit appears to be 1, 2, 3, or 5 percent. An apparent lower limit is supported by a disproportionate number of reports of that lowest relative humidity and no lower relative humidities, except possibly for a few lower values that are probably typographical errors (At some stations, it is moderately common to have a spurious 49° C dew point depression at the highest level of the sounding). Alternatively, the relative humidity may be reported down to some apparent lower limit, and it is reported as "missing" when the humidity is lower. The number of cases of humidity at the lower limit may not be disproportionately high, but no or very few cases with lower humidities are found. At stations with such a policy, usually the dew point is reported on the average to a much lower altitude in the local dry season than in the wet season. The most frequent lower limits in that case appear to be 10, 15, 20, 25, and 30 percent. A third practice is to substitute a "statistical" relative humidity when "motorboating" occurs. The signal from many radiosondes, played through a speaker, sounded like an outboard motor when the relative humidity was low enough that the sensor was not sensitive, and when that occurred a nearly-constant relative humidity (varying somewhat with temperature and pressure) was substituted. This could be detected as a disproportionate number of relative humidities that are close to, but not exactly, the same value, such as a large number of relative humidities between 11 and 15 percent.

Finally, the National Weather Service (NWS) and most other stations operated by the United States followed a practice called dew point "censoring" from 1 April 1973 to 30 September 1993 (As of late 2006, two military stations, Diego Garcia (61902) and Cape Canaveral (74794) still follow this practice). "Censoring" refers to reporting an artificial 30° C dew point depression when the relative humidity is below 20 percent, so this is detected by 3 tests: At least one 30° dew point depression, no dew point depressions above 30° C, and no other relative humidities below 19 percent (this was used instead of 20 percent to allow for small differences in computing the relative humidity). At stations which have not practiced dew point "censoring," is unlikely to see more than one or two observations per year that meet these criteria by chance (So far, no other variations of dew point "censoring" have been found, such as routinely reporting a 40° C dew point depression when the relative humidity is below 15 percent). In this research, each instrument model with dew point "censoring" is treated as a separate model from the same instrument without "censoring," but it is well-documented that (at least at NWS stations) the beginning or ending of "censoring" was not an instrument change. Also, it should be mentioned that soundings are too dry if the 30° C dew point depression is taken as a legitimate value. In the context of the VIZ B (model 1499-520) instrument used in 1993, comparisons of probability distributions during and just after "censoring" showed that the instrument would have reported an average relative humidity of 17 percent at cold temperatures, decreasing gradually at temperatures above about 0° C to around 12 percent at 30° C or warmer.

Sometimes the driest reportable condition is expressed as a maximum dew point depression. In that case, there would be a disproportionate number of dew point depressions having a particular value and few or no larger dew point depressions, so the lowest relative humidity is more variable than the lowest dew point depression.

Another common reporting limit, sometimes in conjunction with a limit discussed above, is to not report the dew point when the temperature is below a threshold, or when the pressure decreases below a certain value, or both. The most common limit is -40° C, and variations include allowing the lowest temperature to be slightly colder than -40°, reporting or not reporting a significant level slightly warmer than -40°, and either resuming or not resuming reporting of the dew point if the temperature rises above -40° in the middle stratosphere. For example, the standard procedure in 1957 at stations operated by the United States was to select a temperature level between -37° C and -40° as the top level with a dew point, except that if the temperature rose to -35° C or warmer at a higher level, then dew point reporting would resume, with the lowest level reported having a temperature between -37° and -40° (USWB, 1957). In Japan (see section 6), it appears that more recent radiosondes are distinguished by the coldest temperature with a reported dew point more and more precisely at -40° C (actually -40.1° due to reporting code limitations).

The discussion above distinguishes instruments based on reporting practices, and it does not matter whether a reporting limit is based on true limitations of the humidity sensor or is simply a policy. A reporting practice change is most likely at the time of an instrument change, but a policy can change with no instrument change (such as the beginning and ending of "censoring" mentioned above). Recent Vaisala ground systems allow the operator to specify the limits of humidity reporting (so a change in reporting does not imply an instrument change), but the two stations practicing dew point "censoring" in 2006 use the early 1980s Meteorological Sounding System (Bellue et at. 2005), and apparently it is impractical to change the software to eliminate "censoring." Intentional hardware limits are illustrated by hypsometer radiosondes controlled by a baroswitch and early Japanese radiosondes. In a hypsometer radiosonde, at low pressures there are usually no baroswitch contacts connected to the humidity sensor, but in their place the contacts are connected to the hypsometer. Certain Japanese radiosondes used a mercury thermometer to cut off the humidity signal at -30° C (DuBois et al. 2002, p. 60).

If a reporting practice is consistent, it may be possible to identify the instrument type for each sounding even when instrument types alternate frequently. For example, in the USSR, the A22, RKZ, and MRZ series reported dew points to, or at least close to, the top of each sounding, while the MARS series reported dew points at temperatures above -40°, including in some cases in the stratosphere. According to Zaitseva (1993), the A22 series used a bimetal thermometer in a duct, and the RKZ, MARS, and MRZ series use the same MMT-1 rod thermistor, but a goldbeaters skin hygrometer which appears unchanged is used in all models. While sensor mounting differs, the differences in data characteristics between models are likely to be small, except for the possibility that the A22 series may read warmer than the other models if the bimetal thermometor was not adequately radiation-corrected as compared to the response of the thermistor (Examination of data shows daytime temperatures averaging less than 0.5° C warmer than night temperatures at 100 hPa). Fortunately, the USSR station history in Gaffen (1996) is quite extensive, except for a few stations, and many stations used the RKZ, MARS, and MRZ series in sequence.

Even though change points differ between stations, the most frequent observed pattern is a gradual or sudden increase in the number of soundings with dew points reported to a temperature above -40° at a time close to the reported introduction of MARS, a period with all soundings reporting dew points to a temperature above -40°, and a gradual or sudden increase in the number of soundings with dew points reported to or near the top of the sounding at a time close to the reported introduction of MRZ. Based on the data, it is more difficult to distinguish between the A22 and RKZ series. Some stations are reported to transition from A22 to MRZ, and again the transition time is difficult to determine precisely (see section 6) although MRZ appears to be slightly drier than A22 radiosondes. The main points of this discussion are that in this case, reporting practices apparently make it easy to distinguish radiosonde models which have minimal differences in sensors, and it is quite common for a station to use more than one radiosonde type for a long period. Also in this case, if an instrument is misidentified, there is little impact on corrections or trends because there is little difference between these instruments. However, even in this case, further tests should reinforce the accuracy of the identification of instruments.

Surprisingly, if the primary difference between instruments is a change in characteristics, the attribution of each sounding to a particular instrument type is somewhat less definite. For example, Vaisala soundings made with a capacitive humidity sensor tend to be drier than soundings made with a different sensor, but a “dry” radiosonde launched into a thunderstorm environment may be more moist than a “moist” radiosonde type launched on a dry day. In this case, sensitive variables usually show clear differences between instrument types in time averages, and the change to a new instrument may be identified to within a few observations rather than to the exact sounding.

Random variations or defects in instruments may in some cases make it difficult to identify instrument types, at least for individual soundings. For example, a defective humidity sensor may not produce a usable signal in part or all of a flight. The presence of a few (or even many) defective humidity sensors can be inferred from time series where the lowest pressure with a reported relative humidity in each sounding varies irregularly, rather than being (for example) almost always at the last level with temperature above -40°. In an extreme case, many Russian stations had a high proportion of soundings with no moisture data above the surface around 1991. It is likely that many radiosondes were flown without hygrometers due to limited supplies. If humidity values are reported as missing below a certain value such as 15 percent, the top pressure level with a reported relative humidity is also quite irregular, but usually the top level with a reported dew point is fairly dry, dew point reporting may resume if the radiosonde rises into a less dry layer, and the lowest pressure with a reported dew point tends to have a distinct annual cycle.

In the discussion above, it does not matter whether the limiting values are even remotely close to correct. All that matters for identifying an instrument is a minimal amount of metadata at one or more stations, and a characteristic pattern of data reporting. It is a separate task to quantify the errors of each instrument type.

As an example of possibly incorrect relative humidities, the lowest reported humidity is not necessarily more moist with “moist” humidity sensors (hair, goldbeaters skin, or lithium chloride hygristors) than with “dry” sensors (carbon hygristors or capacitive polymers), although averages of moisture variables are almost always wetter with “moist” than “dry” humidity sensors. Early kite, balloon, and airplane meteorographs used hair hygrometers, and sometimes relative humidities of 1 percent (Gregg 1918) and 0 percent (Blake 1933) were claimed. Japanese radiosondes with hair hygrometers also occasionally reported 1 percent relative humidity (e. g., Ninomiya 1975). The Russian A22 series, introduced in the USSR in 1957, and the very similar Chinese GZZ-2 radiosonde, still used in 2006, both of which which have a goldbeaters skin hygrometer, occasionally show relative humidity reports of 5 percent (with rare drier cases that are likely to be typographical errors).

It is probable that these humidity sensors were not calibrated at such low relative humidities, and excessively low readings could have resulted from nonlinear responses or variations in calibration. Humidities above 100 percent are also sometimes found, but are arbitrarily set to 100 percent, since reporting codes have no provision for a negative dew point depression.

While the carbon hygristor used in VIZ radiosondes is now considered capable of registering dry conditions, in the United States the data processing practices tended to prevent reporting of dry conditions (Brousaides, 1975, Wade 1991) until the late 1990s. For example, the humidity evaluator used before computers were available (USWB No. 500, VIZ part number 1063-65B, which is a circular slide rule to convert recorder chart divisions to relative humidity based on empirically-derived factors) shows no humidity values below 10 percent and humidity intervals are irregularly-spaced. According to Wade (1991), computerized radiosonde data processing was simply designed to emulate the slide rule output. Just before “censoring” started in 1973, soundings made with VIZ radiosondes generally reported relative humidities down to 10 percent, and after “censoring” ended in 1993 in the NWS, the distribution of humidity values was not much different. However, many radiosonde types were tested or used operationally at Berlin - Templehof Airport (station 10384) and Bendix AN/AMT-12 radiosondes with carbon hygristors made by Aerological Research Inc. were found to be responsive to low relative humidity, so Maedlow and Pantzke (1966) drew extra lines on the USWB-500 circular slide rule to register relative humidity as low as 2 percent at temperatures below -50° C. This is an example of possible differing data characteristics when the same instrument is used in more than one country.

Since the signature of an instrument type is derived by examining well-documented stations first, the signature should be quite consistent between levels, seasons, and stations. However, some complicating factors are as follows (more are mentioned in section 8).

Different agencies or countries using the same instrument may have different data processing and reporting practices, which may obscure some instrument differences, either intentionally or unintentionally. Alternatively, the same practice applied to different instruments may hinder the ability to distinguish the instruments. For example, “censoring” continues in 2006 at two stations (as discussed above), but those stations do not continue to use the same instrument models as in 1993 or earlier. They are reported to be using recent

Sippican (the company which was formerly called VIZ) radiosondes, which at other stations report relative humidity down to 1 percent. At these stations, the change to a drier model may only be distinguished by an increase in the proportion of levels per sounding which have an arbitrary 30° C dew point depression. In some cases, the same instrument with a different practice (such as "censoring") should be treated as a different instrument.

If instruments are distinguished mainly by data characteristics, some climatic environments may partly obscure the characteristics. For example, Vaisala RS80 radiosondes usually report some quite dry levels in almost every sounding, but in the monsoon season in Bangladesh these instruments produce quite wet readings because the dew point is usually reported only to fairly low levels. Except for stations with unusually poor-quality or incomplete observations, the data characteristics are not completely obscured in different environments. For example, Russian radiosondes in Vietnam and Japanese radiosondes in Antarctica show characteristics close to those in their home countries. Similarly, instruments launched from ships show consistent characteristics even as the ships move through different climate environments.

An additional complicating factor is that each instrument has a distinct, but not necessarily unique, set of characteristics, especially when the different instruments are in the same series (such as Russian MARS-2-1 and MARS-2-2), so it is not always possible to assign an exact instrument type in every period. However, because sensitive variables amplify differences between instruments, if instruments are not distinguishable using such variables, the instruments may actually not be different in variables of research interest. With over 2000 instrument codes, some instrument models may differ only in minor electric circuit details that have no effect on the transmitted or processed data.

5. CONSTRUCTION OF MISSING INSTRUMENT METADATA

In this project, the same time series variables are computed for all stations. Data signatures are very similar whether stations do or do not have metadata. The main support for the association of a certain instrument type with a set of data characteristics is consistency over a large number of stations. Based on this hypothesis, it is possible to infer the instruments which are used at a station or in a period with either no metadata or metadata with suspected errors. The method to construct missing metadata (or to change metadata when it is present but inaccurate) is simply to examine station time series develop data characteristics, to search for dates when a given set of characteristics starts or ends, and to hypothesize which instrument type or types may have resulted in the observed data characteristics. Of course, the candidate instruments must have been produced before the period for which the instrument type is to be identified. While some stations have used instruments which were several to over 10 years old, the number of candidate instrument types is usually fairly small.

The easiest case for inferring an instrument is at a station where reliable metadata is available, but the metadata is a "static event" or "snapshot" stating that a certain instrument type was in use at a certain time. In that case, the task is to determine when the use of that instrument started and stopped. Usually the use of that instrument continued backward and forward in time during the period with the same data characteristics as those observed at the stated time of the static event.

A slightly more difficult case for inferring an instrument is at a station or period with no metadata, but when other stations in the same country and with the same operator have metadata. As an example, the station metadata for the USSR in Gaffen (1996) is quite complete, except for a few omitted stations such as Moscow (27612). While the instrument histories in the USSR show many variations between stations (see section 6), with a small number of candidate instrument types it is not exceptionally difficult to identify the instruments used at Moscow and the constructed instrument history should be a good test case for determining if instruments can be inferred correctly, since the metadata for this station should still exist.

It is slightly more difficult to infer instrument types without metadata if no stations in the same country have metadata covering a period. There may be metadata stating that a certain instrument is used at one time and another instrument at a later time, while the data shows a period with a third characteristic in between the two times. Especially in countries which use radiosondes from multiple regions (such as Southeast Asia), it may be challenging to determine the candidate manufacturers which might have supplied radiosondes in that period.

Probably the most challenging case for inferring radiosonde types is when multiple radiosonde types are used at a station without metadata. Of course, each observation uses only one radiosonde, but there are many ways in which the mixed characteristics can be seen in the data. Ideally, each radiosonde type has a distinct processing and reporting procedure, and there are few defective radiosondes (which would seem to have different characteristics), so each observation can be attributed to a radiosonde type. More often, the radiosondes are distinguished by data characteristics (such as one model being wetter than another), and as mentioned above, a dry day with a "wet" radiosonde may appear drier then a moist day with a "wet" radiosonde, so individual instrument types are not reliably distinguished by this approach. Basically, to more effectively detect circumstances when a station is using more than one radiosonde type in a period, it is necessary to be as familiar as possible with the characteristics of the radiosondes where only one type is used at a time, which is most likely in the country where the radiosondes are produced (or also in large countries using those radiosondes, such as with Vaisala models). It is usually possible to estimate the relative proportions of each instrument used at a station even if the individual observations cannot be reliably attributed to specific instruments.

6. DEVELOPING NATIONAL STATION AND INSTRUMENT HISTORIES

Because China, India, and the Russian Federation are the last major countries which are phasing out the oldest operational radiosonde models (with goldbeaters skin or lithium chloride humidity sensors), their histories were constructed first as a test of the feasibility of using the methods in this paper to validate metadata and construct missing metadata. The instrument history of Japan was also constructed as a part of the effort to develop the instrument history for Taiwan because of a personal communication (Y.-A. Liou, 2003) that Taiwan was using Japanese radiosondes.

Even with somewhat lengthy explanations below, the discussions are only a sample of the detailed evaluations which were performed in those countries, and which will be needed for almost every country which has launched radiosondes.

6.1. Instrument history of Japan

Development of the instrument history of Japan is simplified somewhat because Japan has not changed radiosondes frequently since the late 1950s, with the Meisei RSII-56, RSII-80, RSII-91, and RS-01G being predominant, and with distinct differences between instrument types seen in the data. Also, papers in the *Journal of the Meteorological Society of Japan* have provided some documentation of instrument types and their use, and stations in Japan have routinely used 31313 codes since 8 November 1995. The Gaffen (1996) station history states that stations changed from Meisei S-50M-L to Meisei RSII-56 in 1957 and to Meisei RSII-80 in March 1981, and lists the instrument used at each station in either 1993 or 1994, but the time each station is reported to have opened is too recent (based on the beginning of upper air data for most stations in *Monthly Climatic Data for the World*), and each station is reported to use three nonexistent radiosondes (Vaisala RS 53, Vaisala RS 56, and Vaisala RS 2 56) in 1960. Those models were never produced by Vaisala (Ken Goss, personal communication, 2003), and it is likely that Vaisala was erroneously substituted for Meisei in some preceding document, based on the similarity of RS 56 and RS 2 56 to RSII-56 (Meisei RS 53 is listed in http://www1.ncdc.noaa.gov/pub/data/stnhistory/before.txt).

Japanese station histories also include Syowa, Antarctica (station 89532), which is accurately documented in issues of *Japanese Antarctic Research Expedition Data Report* (e. g., JMA 1964 and 1994) and 22 Japanese weather ships, which are documented partially in some journal papers and field program reports.

Based on checking DS353.4 archived observations back to 1973, in almost all periods stations occasionally reported relative humidities as low as 1 percent, but average relative humidity values and the proportions of soundings with very low relative humidities show discontinuities which are consistent with instrument changes.

First, the Meisei RSII-56, with a hair hygrometer (Ninomiya 1975, p. II), showed moist readings on the average, but even this model occasionally reported humidity values as low as 1 percent. Dew points were reported until the temperature fell below -30° C, and there was no special effort to report a temperature level close to -30° C.

Second, the Meisei RSII-80, with a carbon hygristor (JMA 1994), is much drier than the RSII-56. Steplike drying occurred on or very close to 1 March 1981 at each station, in accordance with the documented change in Gaffen (1996). Also, the lowest temperature with a reported dew point changed to around -40° C instead of -30° C, and the proportion of observations with very low relative humidities greatly increased. The characteristics of both RSII-56 and RSII-80 are confirmed by this documented transition and by the common behavior at a large number of stations.

Third, the Meisei RSII-91, with Vaisala RS-80 sensors in a container very similar to the VIZ 1492 radiosonde (Yagi et al. 1996), was reported in use at most but not all Japanese stations in 1993 or 1994, and at those stations there was further steplike drying relative to RSII-80 and almost always the coldest temperature with a reported dew point was exactly -39.9 or -40.1° C. The timing of the instrument change could be identified quite accurately by the data discontinuity even though available metadata does not state the date of the change. However, Japanese stations reported the 31313 code starting 8 November 1995, and the code is consistent with the data behavior, including at the few stations which continued to use Meisei RSII-80 after 1995. Five stations show a change to reporting the dew point to a temperature almost always exactly -39.9 or -40.1° between 1987 and 1991 without drying until later, but the time of the drying, between 1993 and 1997 at those stations, is likely to be the actual change in the radiosonde (which is confirmed by the 31313 code at two stations that transitioned in 1996 and 1997). Apparently the ground processing was upgraded before the radiosonde was changed, but at most stations the ground unit changed at the time the Meisei RSII-91 was first used.

Various references have been found that mention four variations of Meisei RSII-91, the “Type 93” uncoated and “Type 94” antiradiation coated thermistor (Kitaoka 1997), a new humidity sensor starting about July 1999 which corrected a dry bias but caused a moist bias in cold conditions (Ishihara 2004), and a humidity correction in February 2003 (Ishihara 2004). The thermistor change is not noticeable (only a few stations started using RSII-91 before 1994), but timing of the humidity sensor change and the humidity correction are both identifiable, with no more than an error of about 2 days, at each station using the RSII-91.

The latest model, Meisei RS-01G, is described at the manufacturer web site, http://www.meisei.co.jp/english/product/p0111.htm. The use of this instrument is indicated by the 31313 code, but the data is not otherwise distinguishable from the RSII-91.

Finally, Vaisala RS80 instruments were used in 1995 and 1996 at some stations and ships, and some stations have more recently switched to the Vaisala RS92 or

Sippican MK IIA. These are documented by 31313 codes and are also listed in WMO (2006). In some cases there may be few differences in data characteristics between distinct radiosonde types because in Japan, dew points are not reported at temperatures colder than about -40° regardless of the instrument type.

As mentioned above, the Syowa (89532) history is accurately documented, at least to the nearest month, in issues of *Japanese Antarctic Research Expedition Data Report* (e. g., JMA 1964 and 1994), and each individual sounding is published, so most instrument changes can be determined exactly. This station used instrument types that were special models designed for the Antarctic environment in several periods, as well as regular operational radiosondes in other periods, and all models were described in enough detail that they can be distinguished. Similarly, ships sometimes used special shipboard radiosonde models such as the SCM 404.5 MHz radiosonde and the ES61A Echosonde (Ninomiya 1975, p. II), but 31313 codes since about 1997 now identify the radiosonde types.

A minor exception to the usual Japanese radiosonde history is that a few stations were operated by the United States military after World War II until the 1950s to 1970s. Their histories through 1960 are documented in Ratner (1964), but for the stations which were not turned over to Japanese operation until after 1960, documentation is less available, although it can be assumed (and this can be checked for consistency in periods when data is available) that those stations continued to use American military instruments in that period.

There are few complications with the Japanese station history at least since the early 1970s. Various sources describe quite a few additional Japanese radiosonde models, but these apparently were used mostly experimentally, or at least in observations that were not transmitted on the GTS. One model, the Oki RSII-80, is assigned a WMO instrument code number, but it has not been reported at any station. It is possible that RSII-80 is simply a specification, and Meisei and Oki both manufactured radiosondes to that specification.

6.2. Instrument history of China

The China instrument history through 1990 is described in detail in Gaffen (1993, 1996), with only one noticed error, that Vaisala RS12 radiosondes were used at 19 stations as early as 1952, while according to Rossi (1957), that model was first produced in 1957. It is possible that those stations first used the Vaisala RS11 and switched to the RS12 after it became available.

There are two minor complications in verifying the Chinese instrument history. First, the operational GZZ-2 has been referred to by many names, including Shanghai Radio Model 23, Shang/M, and 59-701. Second, at all stations the surface level and significant levels are not in the DS353.4 archive until 1974 or 1975, and many stations have similar sketchy reports from 1997 to 2002. The original reports may have been more complete, based on occasional reports with a considerable number of levels in the IGRA data even in the 1960s. The difficulties are not a significant problem because the GZZ-2 has apparently been unchanged since its introduction in 1964 (Zhai and Eskridge (1997) say that the data is homogeneous from 1970 to 1990), and stations changing to the GTS-1 radiosonde did not have incomplete reports before the introduction of the new radiosonde.

An additional potential difficulty is that Chinese stations do not use the 31313 code. As mentioned below, this has not been a problem so far, but transitions to future instruments may not be identifiable without the use of the 31313 code. However, very few 2-digit codes remain for possible assignment to new instrument types.

For validating and constructing the Chinese instrument history by comparison with the archived data back to 1973, actually only a few transitions need to be identified. These are the transition from GZZ-2 to GTS-1 at most stations and brief use of Vaisala RS80 in the middle 1990s at one station (Wang 2002).

The new Chinese GTS-1 radiosonde is described by Guo (2004), and that paper names stations that introduced the new model in 2002 and 2003. Only the latest WMO radiosonde catalog (WMO 2006) reports two instrument types in China, but the names of the instruments, “Shang/E” and “Shang/M,” are undefined. However, when the data is checked, stations listed as using “Shang/M” have much more moist readings than the stations with “Shang/E” reported, so apparently “Shang/M” is the same as GZZ-2 and “Shang/E” is the same as GTS-1. There are only a few exceptions. First, a few stations reported using “Shang/E” were moist until slightly later than the stated date of March 2006, so the list supplied to the WMO must have included some planned changes, and 3 stations still have moist readings through the end of the latest data in September 2006. Second, one station reported using “Shang/M” showed drying starting 2004, so it must have simply been misidentified. The WMO metadata is manually prepared so having a few errors is not unexpected. Additional support for this identification of the instruments is that the station reported by Guo (2004) as using GTS-1 starting January 2002 showed the same kind of drying at that time, and the stations reported as changing in 2003 showed drying in 2003 or by February 2004, except for 2 stations that closed in 2002. So, in general, the recent Chinese history can be verified even without 31313 codes and the timing of instrument changes at each station can be inferred with high accuracy. It appears that Chinese stations tend to change instruments at the beginning of a month, with few exceptions, and almost always use only the new instrument type after making the change, instead of switching back and forth for a long period.

Hong Kong has had a different radiosonde history, but that history is documented in Gaffen (1993) and in Hong Kong upper air reports. Also, Hong Kong has used the 31313 code since 1994.

Taiwan also has had a different radiosonde history. Both Gaffen (1996) and WMO (2006) claim that Taiwan has used Chinese radiosondes, which is quite unlikely. A complicating factor is that each station in Taiwan is assigned two station numbers, one starting with 46 and the other starting with 58 or 59. Until 2000, many observations were reported twice, using both identifiers. The report starting with 58 or 59 was usually less complete, and apparently was a Chinese retransmission of the original report. In the 1958 station index (US Navy 1958), a note referring to the numbers starting with 58 and 59 states "These index numbers have been assigned to stations in Taiwan by the Chinese Peoples Republic. THEY DO NOT CONFORM WITH THE RECOGNIZED WMO BLOCK AND INDEX NUMBER ASSIGNMENTS." However, the Taiwan station numbers starting with 46 were removed from the WMO station catalog sometime before 1992 (WMO 2006b, 1992 edition), although some observations with station numbers starting with 46 continued until 2000.

The Taiwan radiosonde history is one of the more challenging to develop because only a small amount of specific metadata has been located, Taiwan has used radiosondes from several countries, and some stations have routinely used more than one radiosonde in the same period. Fortunately, most Taiwan stations used the 31313 code starting in 1992, although it appears to be a manual entry in the observation due to a large number of typographical errors, and some periods where the codes appear to be used incorrectly. However, most of the data characteristics are consistent with other countries using the same instruments.

The 1 October 1977 Air Weather Service Master Station Catalog reports that United States military radiosondes were used at almost all stations in Taiwan, and the data at 5 out of 6 stations (minimum 10 percent relative humidity) is consistent with other stations using VIZ radiosondes without "censoring" in the 1970s. At the other station, the relative humidity was occasionally 1 to 3 percent and the dew point was reported to a temperature around -30° C, which matches behavior of the Japanese Meisei RSII-56. At that station, additional drying and reporting of the dew point to a temperature around -40° started in November 1981, consistent with the Meisei RSII-80. While the use of Japanese instruments is inferred at that time, Japanese radiosondes have been used in Taiwan more recently, confirmed by 31313 codes and also a personal communication (Y.-A. Liou, 2003).

The 31313 codes in Taiwan have reported the use of Japanese, Vaisala, VIZ, and Atmospheric Instrumentation Research (AIR) radiosondes, with the use of two radiosonde types at most of the stations in some periods. Some unusual behavior has been seen, including dew point "censoring" with AIR sondes at one station from 1993 to 1997 (This is also observed at the South Pole, station 89009). "Censoring" at that station was more frequent with VIZ sondes, and not all observations with AIR sondes showed censoring. AIR sondes are "drier" than VIZ sondes of the middle 1990s, and AIR observations without censoring frequently had relative humidities as low as 0 to 5 percent. It is possible that the AIR observations with censoring were actually VIZ soundings, but even in that case only a few dozen soundings would be misidentified.

6.3. Instrument history of India

The India instrument history is more problematic than that of Japan, especially since 1991 because there is little recent specific documentation that is comparable to the detailed India station instrument histories through 1990 in Gaffen (1993). Also, although there is a WMO instrument code assigned to the India Mark III instrument, stations do not use it consistently and sometimes apparently interpret it to refer to any Indian radiosonde, so the 31313 code has not been helpful in identifying the introduction of new models (some stations use the previous 31313 code, have added a different code in early 2006 indicating an unknown instrument, and also use no 31313 code in some observations, with no consistent difference seen between the instruments yet). In this project, the India history was prepared in late 2004, and it is planned to be reviewed in detail in the next few months.

However, papers in *Mausam* (formerly the *Indian Journal of Meteorology and Geophysics*) have given extensive details of many aspects of the Indian radiosonde models and ground equipment, although they are sometimes vague about if and when a change has actually become operational, or whether the paper is simply reporting on a proposal or a test. Also, some on-line WMO documents have provided recent very helpful information, as discussed below.

Finally, the Indian upper air data is widely perceived to be of poor quality. In the archived data, one noticeable feature is that the altitude to which dew points are reported varies widely and is much closer to the surface (sometimes ending with a temperature above freezing) in the dry season than in the summer monsoon. Further examination indicates that the minimum reported relative humidity is about 15 percent, and drier values are omitted instead of having a "motorboating" (statistical) or "censoring" or similar dry value reported. However, this practice may make it difficult to distinguish unreported minor instrument variations, because the average lowest temperature with a reported dew point shows fairly consistent changes at a large number of stations. Counterintuitively, more relative humidity data would be reported with a slowly responding hygrometer because the sensor does not decrease to a reading under 15 percent in a dry atmosphere at as low an altitude as with a more responsive sensor.

Only a few of the station and instrument change dates in the India microfiche in Gaffen (1993) differ from dates in various *Mausam* papers. Some interim changes to the India Meteorological Department (IMD) Mark III instrument, mentioned in Mausam, do not have dates listed in the Gaffen (1993) history because they are not considered to be model changes. However, the Mark III and earlier instruments do not seem very distinguishable even in sensitive variables, so they may be quite close to homogeneous.

The most major change in the India network is currently occurring, which is the transition to the IMD Mark IV, which has a carbon instead of lithium chloride hygristor (WMO 1999). While WMO documents do not list which stations are using the new radiosonde, certain stations showed drying with occasional relative humidities as low as 1 percent, sometimes accompanied by a significant increase in the number of reported temperature levels and possibly an increase in the number of wind levels. However, some stations showed increases in temperature or wind levels with no drying, or drying with no increases in temperature or wind levels. It appears that stations are replacing radiosondes or ground equipment independently, since the old and new radiosondes are designed to work with old or new ground equipment. Especially at the stations with no improvement in the number of temperature levels, the signature of the new radiosonde is somewhat ambiguous and it was hypothesized that at least some of the stations used a mixture of old and new radiosonde models. A recent document (Bhatia 2006) lists the transition month at 11 stations, and based on inferred instruments in late 2004, the transition month was identified correctly at 8 stations, at 2 stations the transition was identified 2 months later than the stated month, and at the last station the stated transition did not occur until 2005 and the station was inferred to still be using the old instrument. So, even in this ambiguous situation, all of those transitions were correctly identified. While this source claims that all stations have changed to the new instruments by 2006, some stations show no drying by September 2006. It is possible that those stations are using only small quantities of the IMD Mark IV radiosondes.

Bhatia (2006) also mentions two radiosondes currently being developed. A solid state pressure sensor for the IMD Mark IV was tested starting March 2006, with the goal of reducing the pressure variability relative to model initialization. Also, a digital radiosonde (the name of that model is not mentioned) is in early development. Without distinguishable 31313 codes, it may be difficult to detect transitions to these instruments when they become operational, unless they cause the quality or quantity of data (number of data levels) to increase consistently.

6.4. Instrument history of the Russian Federation

The station history for the former USSR (station numbers starting with 2 or 3, plus Russian ships and Antarctic stations) was developed in 2004 (Schroeder 2005), but it will be reviewed and updated in the near future. Here, a very short summary of the development of that history is given.

First, the history of instrument changes given in Gaffen (1996) is mostly consistent with data characteristics in archived data since 1973. Section 4b discussed some of the characteristics of the major instruments, which were the A22, RKZ, MARS, and MRZ families. They show differences in reporting characteristics, but little difference in data characteristics, mainly because there is only a minor difference in sensors. While the A22 used a bimetal thermometer and the other radiosondes use a thermistor, the A22 radiation corrections were apparently effective, since stations using the A22 showed temperatures at 100 hPa averaging no more than 0.5° C warmer near noon than at night, almost identical to the differences with later radiosondes. As long as the instrument differences are actually minor, a misidentification of instruments has only small differences in trends.

The instruments appear to be almost exactly distinguishable at many stations because the data characteristics show no significant wind levels (except possibly the tropopause and fastest wind levels) with A22, and dew points are reported only to a temperature around -40° with MARS. The few problematic transitions would then be from RKZ directly to MRZ, and A22 to RKZ or MRZ in a period when the archived data does not include significant wind levels. Even then, the A22 seems to have slightly fewer temperature levels than more recent instruments. Changes in instruments within a family are also not distinguishable, such as MARS-2-1 (using the Meteorit-1 radar) and MARS-2-2 (using the Meteorit-2 radar). In the station histories, such transitions were fairly rare. Even if the MARS variety is not specified, MARS-2-1 and RKZ-2 use the same radar, and MARS-2-2 and RKZ-5 use the same radar.

Since the middle 1990s, there has been a proliferation of new instruments introduced in the Russian Federation, in addition to continued use of the MARS and MRZ instruments, with 11 WMO instrument codes now assigned: 27 (AVK-MRZ), 28 (Meteorit Mars-2-1), 29 (Meteorit Mars-2-2), 53 (AVK-RF95), 58 (AVK-BAR), 68 (AVK-MRZ-UAP), 69 (AVK-BAR-UAP), 75 (AVK-MRZ-ARMA), 76 (AVK-RF95-ARMA), 88 (MARL-A-MRZ), and 89 (MARL-A-BAR). By September 2006, codes 58, 68, 69, and 89 have not been seen in the data. AVK, Meteorit, and MARL-A are radars or radiotheodolites. MRZ, Mars, RF95, and BAR are radiosondes, but no documentation has been located for the BAR radiosonde. UAP and ARMA are apparently ground processing systems. Some stations in the Russian Federation outside Russia use Vaisala radiosondes, which are not included in the above list. Fortunately, most stations have reported the 31313 codes regularly since 1996, so potentially the instrument history of Russian Federation stations can be constructed exactly, within the limits of the 31313 code.

There are two limiting factors in the 31313 codes in the Russian Federation that do not affect other countries:

First, at stations reporting WMO code 27 (AVK-MRZ), some stations have been moist for the entire period of record and other stations have been moderately dry for part of their period of record, with dryness similar to stations reporting code 75 (AVK-MRZ-ARMA). So, code 27 appears to be associated with two different radiosonde models. The moist model is MRZ-3A, which has been used since 1987, but the moderately dry model is called MRZ-3AM and uses a DVR capacitive humidity sensor (Balagurov et al. 1998, 2002).

Second, the inference that the use of the MARS radiosonde is associated with dew point reporting to a temperature above -40° is based on examination of time series at stations with documentation, and is consistent at more than 100 stations. Alternatively, Gaffen (1993) says that stations using MARS with manual processing report dew points only to a temperature above -40°, but it would be strange if stations with manual processing were not converted to automated processing over a period of decades. Also, the observed pattern at some stations of irregularly reporting dew points to temperatures above -40°, and to or near the top of each sounding, is difficult to explain with the alternative assumption. If automated processing is introduced at a station, all observations should be automated (with dew points reported to or near the top) after a short transition period. When the station history is reviewed, the possibility will be rechecked that stations now reporting MARS instruments with dew points reported to or near the top of the sounding are using the correct WMO instrument code.

7. PROPOSED DATA ADJUSTMENT PROCEDURES

Schroeder (2006, sections 3.3 to 3.5) discusses proposed data adjustment procedures in some detail. With complete metadata, other researchers should also apply their data adjustment procedures to the data. The success of the adjustments should help validate this metadata. If an adjustment makes a discontinuity worse at a station in some period, then the inferred instrument is likely to be incorrect in that period.

The proposed adjustment procedure in this project is only sketched out here because of the more detailed discussion in Schroeder (2006):

First, a hypothetical "reference instrument" needs to be chosen. This will probably be the average of recent Sippican and Vaisala instruments, or possibly the average of the slightly older Vaisala RS80 and VIZ A and B series, because for those instruments, almost all instrument types can be compared and adjusted to a reference instrument using a short "chain" of adjustments.

The second step is to identify the specific cases where each distinct instrument type can be compared to another instrument type, aimed at developing short "chains" of comparisons to a reference instrument. Suitable comparisons are (1) a transition from one instrument type to the other at a station, in either order, possibly with a short gap but with a considerable number of observations of each type, (2) frequent transitions back and forth from one instrument type to the other at a station, as long as each instrument is identified accurately, and (3) simultaneous use of the two instrument types at two nearby stations, such as Berlin and nearby East Germany, or Hong Kong and nearby China. Formal intercomparisons are not in this list because the number of soundings is small. For a comparison of the two instrument types, all types of comparisons can be used, at as many stations as possible involving the two instruments, and preferably involving an integer number of years at each station to include the entire seasonal cycle. If the reference instruments are certain VIZ and Vaisala models, the adjustments of those VIZ and Vaisala models to the reference (which is the average) is based on transitions between VIZ and Vaisala or simultaneous use of VIZ and Vaisala at adjacent stations. For other instruments, the ideal comparison is at stations transitioning from that model to VIZ or Vaisala (or from VIZ or Vaisala to that model), or simultaneous use of that instrument and VIZ or Vaisala at adjacent stations. If such a comparison is not available, the shortest "chain" of comparisons ending with a reference instrument is used. The "chain" of comparisons does not need to be complete at any one station. For example, if the "chain" is comparisons of "Type A" to "Type B" and "Type B" to a reference, the "Type A" and "Type B" comparison can involve a different set of stations than the other comparison.

Third, temperatures are compared at each station and for the group involved in each comparison. The procedure is to develop tables of the cumulative probability of each temperature value, stratified by pressure intervals and possibly sun angle. For example, if "Type A" is slightly warmer than "Type B" at some atmospheric level with some sun angle, the cumulative probability distribution at each temperature is lower for "Type A" than for "Type B". The cumulative probability tables are checked to see if there are substantial differences between stations within each category of stratification, other then the expected climatic differences in the environments. For all stations and time periods as a group, if distributions do not show systematic differences, the two instrument types will be considered to be the same. It is possible that many instrument types will not need adjustments to be comparable at lower levels.

Fourth, temperature adjustments will be developed using the cumulative distributions. To adjust from one instrument type in the reference to the hypothetical "reference instrument" which is the average of both models, the adjustment will be half of the difference from one type to the other. For example, if certain VIZ and Vaisala models are in the reference, then VIZ is adjusted by half the difference from VIZ to Vaisala and Vaisala is adjusted by half of the difference from Vaisala to VIZ. For other instrument pairs to be made equivalent, the adjustment equals the temperature difference from one model to the other.

Fifth, the temperature adjustments are applied to each observation according to its identified instrument type. If the instrument is one of the reference types, the adjustment is from that instrument type to the reference average. For any other instrument type, adjustments are applied in sequence from that instrument type and ending with the adjustment to the reference. Of course, much evaluation of the adjusted temperature data will need to be performed to check if the adjustments reduce or eliminate discontinuities. If data discontinuities are not reduced, or are made worse, either there is an error with the adjustment or with the identified instrument. Because of the process of applying varying numbers of adjustments with different instrument types, it is not straightforward to determine statistical error bars for the adjustments either for the result of a particular adjustment or to determine if the global or regional climate trend at any level is overcorrected or undercorrected.

Sixth, after the temperature adjustments are considered satisfactory, the dew point depressions are compared for the same instrument pairs using the same sets of stations. For the input to this step, the reported dew point depressions should stay unchanged even as temperature adjustments were applied. So, if an instrument was made colder, the dew point is made colder by the same amount. This makes that instrument slightly drier, but all instruments are considered statistically equivalent in their temperature behavior at the end of step 5 and this step starts the process of making the moisture behavior comparable. To make environments as similar as possible, the cumulative probability distributions are stratified by atmospheric layer, temperature interval, and possibly sun angle. Probably all instrument pairs will require dew point adjustments (the differences will not be deemed negligible), but in some cases the differences will not be considered significant as sun angle varies.

Seventh, the differences between instrument pairs are used to develop a "chain" of adjustments, first for each instrument model in the reference to the reference average, then for each other instrument type in a pair to the other instrument type, using the same procedure as in step 4. Note that differing special procedures will be needed to develop appropriate adjustments in the case of "censoring," a minimum limiting humidity, and omitting the humidity if it is below a certain percentage.

Finally, the dew point depression adjustments are applied to each reported level in each sounding, whether a single adjustment from a reference model to the average, or a sequence of adjustments from any other instrument and ending with the "reference instrument" average. As in step 5, extensive examination of station time series and other statistics will be needed to establish confidence in the adjustments and the inferred instruments. It is possible that an incorrect inferred instrument is revealed by a worsened dew point discontinuity even if the temperature adjustment appears satisfactory.

8. LIMITATIONS OF CONSTRUCTED METADATA AND ADJUSTED DATA

While the procedures discussed here appear to be sound, there are some limitations, both in the area of constructing metadata and in developing and applying adjustments. However, as mentioned briefly in the last section, if an instrument is inferred incorrectly, that period at that station is likely to show an uncorrected (or even worsened) discontinuity, and the inferred instrument needs to be reevaluated. Multiple iterations may be needed to establish confidence in both the metadata and the adjusted data. Even with the need for such reevaluations, it should be evident that the inference of instrument types is not a random process, especially because a considerable amount of metadata is available.

In the area of validating or inferring instrument metadata, some limitations are as follows: Usually, an inferred instrument type is somewhat generic, such as Vaisala RS80, and it is often not feasible to identify the exact model (such as RS80-15G) without some specific metadata. Poor quality observations may make it difficult to distinguish instrument types, but sparse observations with good quality often can be accurately attributed to specific instrument types. If multiple instruments are in use, it may be difficult to identify the two (or more) instrument types involved, and it usually is not possible to exactly attribute each observation to a specific instrument type, unless there is some regular pattern to the use of the different instruments. Sometimes, without documentation it is difficult to distinguish a policy change from an instrument change, but it is unlikely that a station frequently alternates between policies but it is likely that a station frequently alternates between instruments, and a policy change is likely to be implemented at a large number of stations on the same day while an instrument change is rarely simultaneous at many stations.

Overall, when attempting to identify instrument types, if the instruments are not clearly distinguishable in sensitive variables, the differences in more stable variables of research interest should be correspondingly smaller, so the effect of such a misidentification (such as inferring that Vaisala RS80 is used when RS90 is actually used) on trends should be small, and probably not systematic on a global or regional basis.

In the area of developing and applying data adjustments, some limitations are as follows: Adjustments cannot restore missing information, such as adding vertical detail to archived smoothed profiles. Adjustments cannot and should not attempt to fully correct differences from an unresponsive sensor (For example, if the lower limit of a humidity sensor is 20 percent, while another sensor can detect relative humidity as low as 2 percent, it is not realistic to reduce the dew point reported by a sensor of the first type by about 30° C in such a case to be statistically equivalent to the second type of sensor). Finally, adjustments cannot correct a sampling error, where an observation is not made because it is beyond the capability of some component of the system.

9. OBSERVED UNADJUSTED PRECIPITABLE WATER FROM 1973 TO 2006

Because unadjusted and adjusted soundings have the same format (except the adjusted data will have both original and corrected metadata, so the original sounding can be traced), climatology and statistics can be readily developed from either unadjusted or adjusted data.

Time series of the same variables used to infer instrument types are first produced to compare to unadjusted time series. The time series can be used to evaluate the inferred instruments, because a discontinuity is often made worse if an inferred instrument type is incorrect. Detailed data examination may also show that some instrument types need to be split into two or more types. After repeating preceding steps until the inferred instruments appear satisfactory, climatology and other statistics can be prepared.

9.1. Grids and climatology of precipitable water

While the approach to develop grids and statistics is the same for all variables, this research is focused on atmospheric moisture trends, so grids and climatology so far have been prepared only for total precipitable water.

Daily 2.5° grids of the desired variable are produced first. Spacing of 2.5° is appropriate because few areas have more closely-spaced stations. The global grid has

73 rows and 144 unique columns, each 2.5° by 2.5° in latitude and longitude, centered on each latitude and longitude divisible by 2.5° (so the top and bottom rows are actually only 1.25° tall). The first column covers longitudes from 178.75° E eastward to 178.75° W, and column 73 covers longitudes from 1.25° W eastward to 1.25° E. There is also a column 145 which is the same as column 1, and column 144 is immediately west of column 1 or 145, from 176.25° E to 178.75° E.

Gridding procedures are not complex and are not described in detail here. Some special considerations are as follows:

(1) Observations around 0000 UTC (2100 to 0300 UTC) are weighted half to the day before 0000 UTC and half to the day starting 0000 UTC. Daily grids include the weighted number of observations as well as the value of the variable in each grid box.

(2) For variables such as total precipitable water, where the surface elevation affects the column amount, the quantity at a station is adjusted to the average elevation in the grid box. Based on a study of nearby stations at different elevations, the scale height for total precipitable water is about 2.5 km. So, if a station in a valley is 1 km lower than the average elevation in its grid box, the reported precipitable water is multiplied by *exp* (-1/2.5) = 0.67032 to be corrected to the grid box average elevation. (A typical scale height for water vapor in the free air above a location, as reported by other researchers, is 2 km. Water vapor decreases faster with height above a station than in columns of air over locations with higher surface elevations, because the air column above each location contains a boundary layer which tends to be more moist than in the free air at the same altitude above a lower elevation.)

(3) For variables with large diurnal variations such as near-surface temperatures, grids of daily average values may be inappropriate.

(4) In daily grids, empty grid boxes are not filled in.

A monthly average grid is simply prepared by summing the weighted values from the daily grids, and dividing by the sum of the weights. Such a grid is still sparse and empty grid boxes need to be filled in to produce climatological averages. The grid filling process is summarized as follows, with underlying assumptions stated:

(1) If a grid box has at least as many observations as some defined threshold (such as 5 or 10 percent of the days), the grid box value is accepted.

(2) If a grid box has few or no observations, surrounding boxes are searched and their values are accumulated with weights declining with distance (For each observation in this grid box, the weight is 1.0). When the sum of the weights reaches a threshold, the search for data ends and the grid box value is the weighted sum divided by the sum of the weights. The empirical part of this process is that a roughly diamond-shaped area is scanned (farther east and west than north and south from the grid box, except near the poles) because the climate varies less in the zonal than the meridional direction. Also, for variables depending on elevation, other grid box values are adjusted to the elevation of this grid box before weighting, and when filling in a low-elevation grid box, the scan in any direction stops if a grid box with an elevation over 750 meters is encountered. For example, a grid box just off the west coast of South America is not filled in with any data from higher elevations in the Andes or from any elevation east of the Andes.

(3) Each annual average is simply the average of the 12 filled-in monthly grids. A global or regional average of any variable weights the grid boxes by area.

This study computes a 33-year climatology from 1973 to 2005, although any sufficiently long period can be used. A climatology is built in the same way as the monthly grids are built, using a grid for each month of the year, but observations in that month for all years are included before filling in the grid. A grid box is accepted without weighting from surrounding points if the number of observations is at least as large as 5 percent of the number of days in the period. The annual climatology is the average of monthly filled climatology grids.

With the empirical scheme of a diamond-shaped scan to fill in empty grid boxes, the quality of gridding is good even in large data-sparse areas. The quality of the gridded climatology can be evaluated by comparing patterns with known climatological processes. With total precipitable water, the main potential problem area is the eastern Pacific. The ITCZ and SPCZ are moderately well reproduced in the western and central Pacific, but the eastern Pacific ITCZ is broader in meridional width and has a lower peak value of total precipitable water than in satellite climatologies, because of the lack of suitable stations in that area.

Climatological averages for the world or for a region are most likely to be correct when the spatial patterns are reasonable (within the limitations of sparse station coverage) and the values are as accurate as possible at individual locations. With this analysis, the global annual average is 2.514 cm of precipitable water from 1973 to 2005, close to averages obtained by others.

To develop time series of global or regional averages of a variable, with sparse data it is usually best to construct grids of anomalies and then fill in the anomaly grids, from which the spatial averages are computed. This is because a filled-in average of surrounding anomalies is a conservative estimate for an empty box, but a filled-in average of surrounding absolute values can be very extreme at that location. So, climatological grids are actual variable values, but monthly and annual grids are expressed as anomalies. For some variables, monthly grids of percentages of mean values should be constructed and then filled in.

9.2. Observed global precipitable water variations since 1973

Monthly and annual grids and time series are produced from the unadjusted radiosonde data almost every month, using the procedure of section 9.1, to look for unexpected trends and variations, and ensure that there are no data problems.

Figure 1 shows the latest time series of monthly global average precipitable water anomalies, ending September 2006.

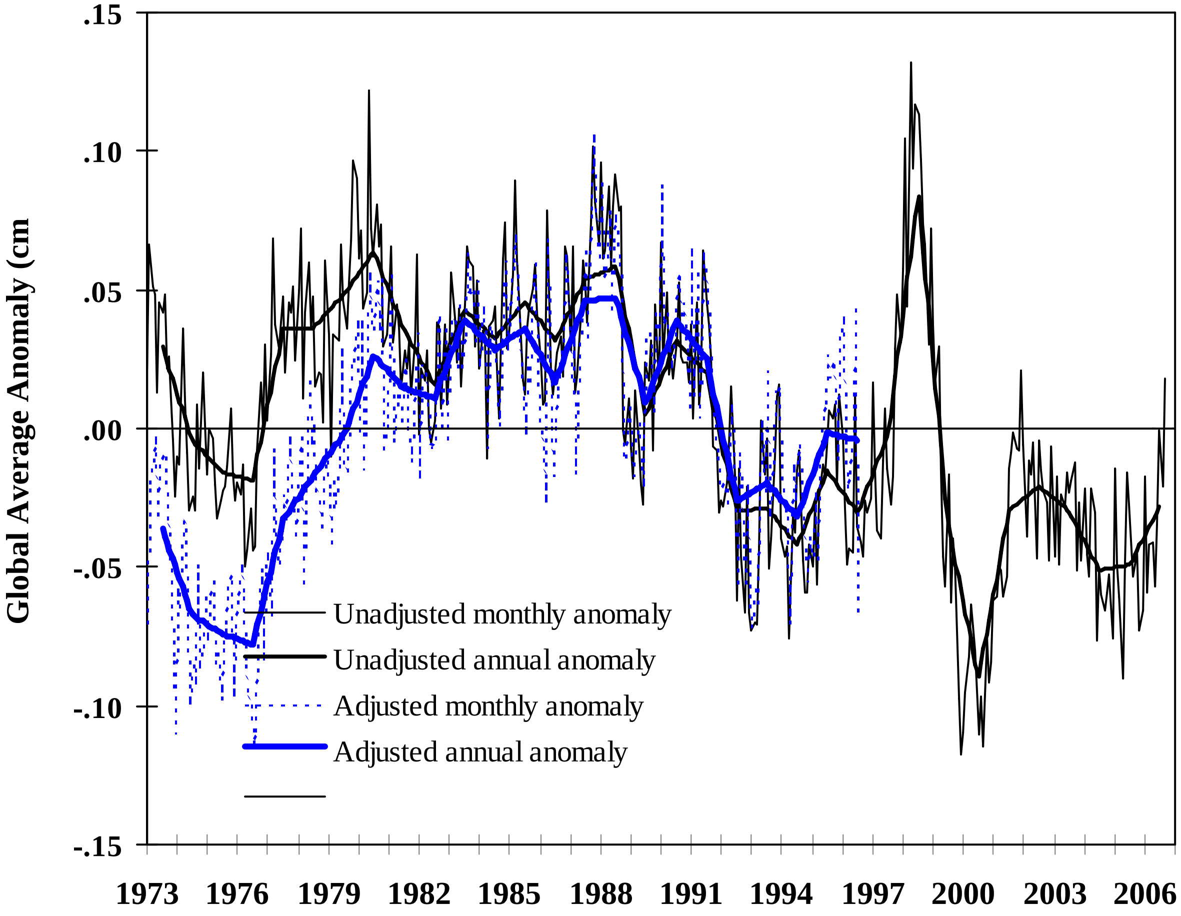

FIGURE 1. Monthly (thin lines) and annual (thick lines) global area-averaged anomalies of precipitable water. The unadjusted values for January 1973 to September 2006 are relative to the 33-year climatology for 1973 through 2005. The adjusted values for January 1973 to July 1996 are based on a previous project and are relative to the average for January 1973 to July 1996.

A previous project (Schroeder 2003) developed very preliminary assessments of instrument types and the resulting adjustments, covering 1973 to July 1996. The adjustments in that effort are much less detailed than in the current project, and no temperature adjustments were made. The time series of monthly anomalies of global precipitable water from that project is superimposed on the unadjusted time series, and shows the approximate effect of instrument adjustments on the global trend in the last few decades.

In Figure 1, the black lines are not adjusted for instrument differences, and the blue lines are computed from preliminary adjustments. Note that adjustments have a gradual effect on the time series because the transition to drier instrument types has not been sudden in any substantial part of the world. The adjustments have little effect on the size of short-term or interannual variations, such as the effects of El Niño.

With either unadjusted or adjusted data, the basic trend of global precipitable water shows nearly-steplike changes coinciding with documented climate shifts. Starting from 1973, the initial period was dry, the period from the late 1970s to about 1990 was moist, and the period since then has been generally dry, except for a very large moistening and drying from 1997 to 2001. The 3 climate regimes are more distinct in tropical averages (30° N to 30° S, not shown), with the latest dry regime starting in the tropics from 1988 to 1989.

The moistening in the late 1970s is intensified by the instrument adjustments, since a moistening trend was occurring at the same time that much of the world was transitioning to drier instrument types. The rate of change to drier instrument types slowed since the late 1980s, so it is unlikely that the adjusted data will completely eliminate the dryness of the last 15 years relative to the 1980s. The final transition to drier instrument types in the Russian Federation, India, and China (which has recently begun in all of these countries) will cause some additional drying in the global averages. After those transitions are completed, there will still be fluctuations in future adjustments, but they should be of smaller magnitude and they should not have a "one-way" (exclusively drying) effect on the global averages.

During the transition from the 1997-98 El Niño to the following La Niña, 1998 was exceptionally moist from the eastern Pacific into the Caribbean and in much of the Indian Ocean, and very dry in much of the western Pacific. The basic pattern in the Pacific and Indian Oceans was similar in 1983 as the 1982-83 El Niño decayed. During the persistent La Niña of 1998-2002, 2000 was drier than usual almost globally except from Australia to east of Japan, with a 7 percent decline in annual average precipitable water from 1998 to 2000. After adjusted data is prepared, the pattern of 2000 should be compared with previous dry periods such as 1974 to 1976, and possibly before 1973.

Even with the incomplete state of instrument metadata and adjustments, Figure 1 shows that it is unlikely that there has been a consistent global moistening trend during the persistent global warming since the early 1980s. However, moistening in the late 1970s was large enough that the dry period since the early 1990s is more moist than the 1970s dry period. The main goal of this research is to quantify the moisture trend more accurately. It should then be more feasible to investigate physical mechanisms and feedbacks involved in both interannual and decadal moisture changes, and to relate these shifts to the ongoing global warming trend.

10. DATA AVAILABILITY

While this 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.061001.

Acknowledgements

This project has been partially supported by NOAA Office of Global Programs grant NA04OAR310125.

References

Alduchov, O. A., and R. E. Eskridge, 2002: Determination of radiosonde station elevation from observational data. *J. Applied Meteor.*, **41**, 357-362.

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.

Balagurov, A., A. Kats, and N. Krestyannikova, 1998: Implementation and results of the WMO radiosonde humidity sensors intercomparison (Phase I laboratory test). A. Van Gysegem, ed., *Papers Presented at the WMO Technical Conference on Meteorological and Environmental Instruments and Methods of Observation* (TECO-98) (Instruments and Observing Methods Report No. 70, WMO-TD-No. 877). WMO, Casablanca, Morocco, 181-184.

Balagurov, A., N. Krestyannikova, and V. Bogov, 2002: Results of humidity sensor intercomparisons of new Russian radiosondes. *Papers Presented at the WMO Technical Conference on Meteorological and Environmental Instruments and Methods of Observation* (TECO2002) (Instruments and Observing Methods Report No. 75). WMO, Bratislava, Slovakia, online at http://www.wmo.ch/web/www/IMOP/publications/IOM-75-TECO2002/Posters/P1.2(01)Balagurov.doc.

Bellue, D. G., B. F. Boyd, W. W. Vaughan, T. Garner, J. W. Weems, J. T. Madura, and H. C. Herring, 2005: Weather support to the Space Shuttle - An historical perspective. *3rd Presidential History Symposium*, American Meteorological Society, http://ams.confex.com/ams/pdfpapers/85962.pdf.

Bhatia, R. C., 2006: *Status of Indian Radiosonde and Future Plans for Improving the Quality of Upper Air Radiosonde Data*. Available at http://www.oco.noaa.gov/docs/GCOSUAWII/pres_Bhatia_RC_Status_India n_Radiosonde.ppt.

Blake, D., 1933: Remarkable low humidities aloft over San Diego, Calif. *Mon. Wea. Rev.*, **61**, 170.

Brousaides, F. J., 1975: The radiosonde hygristor and low relative humidity measurements. *Bulletin of the American Meteorological Society*, **56**, 229-233.

Collins, W. G., 2001: The operational Complex Quality Control of radiosonde heights and temperatures at the National Centers for Environmental Prediction. Part I: Description of the method. *J. Applied Meteor.*, **40**, 137-151.

Collins, W. G., 2001: The operational Complex Quality Control of radiosonde heights and temperatures at the National Centers for Environmental Prediction. Part II: Examples of error diagnosis and correction from operational use. *J. Applied Meteor.*, 40, 152-168.

Dashiell, B. F., 1938: Aerological observations for 1938. *Monthly Weather Review*, **66**, 424.

DuBois, J. L., R. P. Multhauf, and C. A. Ziegler, 2002: *The Invention and Development of the Radiosonde*. Washington, DC, Smithsonian Institution Press, 78 pp.

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. *Journal of Climate*, **19**, 53-68.

Elliott, W. P., R. J. Ross, and W. H. Blackmore, 2002: Recent changes in NWS upper-air observations with emphasis on changes from VIZ to Vaisala radiosondes. *Bulletin of the American Meteorological Society*, **83**, 1003-1017.

Endlich, R. M., 1954: A note on pressure at the tropopause. *Bulletin of the American Meteorological Society*, **35**, 131-132.

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. *Journal of Climate*, **17**, 4171-4179.

Gaffen, D. J., 1993: *Historical Changes in Radiosonde Instruments and Practices*. WMO/TD-No. 541 (Instruments and Observing Methods Report No. 50), 128 pp.

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.

Gregg, W. R., 1918: Aerology No. 5, Free-air data at Drexel Aerological Station: January, February, March, April, May, and June, 1917. *Monthly Weather Review Supplement No. 10*, 101 pp.

Guo, Y. T., 2004: *Preliminary Quality Analysis for Chinese L-Band Radar - GTS1 Radiosonde Upper-Air Observation System*. Geneva, WMO, 7 pp. (available at http://www.wmo.ch/web/www/IMOP/meetings/Upper-Air/Systems-Intercomp/Doc3-4(4).pdf)

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

JMA, 1964: *Antarctic Meteorological Data Obtained through the Japanese Antarctic Expedition, Aerological Data at the Syowa Base*. Tokyo, Japan Meteorological Agency, 91 pp.

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.

Kalnay, E., and 21 coauthors, 1996: The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, **77**, 437-471.

Karl,T. R., S. J. Hassol, C. D. Miller, and W. L. Murray, eds., 2006: *Temperature Trends in the Lower Atmosphere - Steps for Understanding and Reconciling Differences*. United States Climate Change Science Program and the Subcommittee on Global Change Research, Washington, DC, 164 pp. (available at http://www.climatescience.gov/Library/sap/sap1-1/finalreport/default.htm).

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

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, 1997: Worldwide radiosonde temperature corrections. *Preprints, Eighth Symposium on Global Change Studies*, Long Beach, CA, 142-145.

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

Maedlow, E., and E. Pantzke, 1966: Beitraege zur aerologie der stratospharer ueber Berlin in jahre 1965. *Meteorologische Abhandlungen*, **52 (14)**, 1-2.

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. *Journal of Climate*, **19**, 2094-2104.

Ratner, B., ed., 1964: *History and Catalogue of Upper Air Data for the Period 1946-1960*. Key to Meteorological Records Documentation 5.21, US Weather Bureau, Washington, DC, 352 pp.

Ross, R. J., and W. P. Elliott, 1996: *Tropospheric Precipitable Water: A Radiosonde-Based Climatology*. NOAA Technical Memorandum ERL ARL-219, Silver Spring, MD, Air Resources Laboratory, 132 pp.

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.

Ross, R. J., and D. J. Gaffen, 1998: Comment on “Widespread tropical atmospheric drying from 1979 to 1995” by Schroeder and McGuirk. *Geophysical Research Letters*, **25**, 4357-4358.

Rossi, V., 1957: A laboratory investigation on the radiation error of the radiosonde thermometer. *Geophysica*, **6**, 91-105.

Schroeder, S. R., 2003: Completing instrument metadata and adjusting biases in the radiosonde record to allow determination of global precipitable water trends. *Preprints, 12th Symp. On Meteorological Observations and Instrumentation*, Long Beach, CA, Poster P1.6.

Schroeder, S. R., 2005: Toward developing unbiased upper-air temperature and moisture trends from historical radiosonde data: Validating and completing Russian radiosonde history. *Preprints, 16th Conference on Climate variability and change*, San Diego, CA, Poster P1.15, http://ams.confex.com/ams/pdfpapers/83854.pdf.

Schroeder, S. R., 2006: The Effect of Radiosonde Instrument Changes on Climate Trends of Global Atmospheric Precipitable Water. *Preprints, 18th Conference on Climate variability and change*, Atlanta, GA, Poster P1.10, http://ams.confex.com/ams/pdfpapers/103891.pdf.

Schroeder, S. R. and J. P. McGuirk, 1998: Widespread tropical atmospheric drying from 1979 to 1995. *Geophysical Research Letters*, **25**, 1301-1304.

Schroeder, S. R., and J. P. McGuirk, 1998. Reply. *Geophysical Research Letters*, **25**, 4359-4360.

Schwartz, B. E., and M. Govett, 1992: *A Hydrostatically Consistent North American Radiosonde Data Base at the Forecast Systems Laboratory, 1946-Present*. NOAA Technical Memorandum ERL FSL-4, Boulder, CO, NOAA/ERL/FSL, 81 pp.

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.

US Navy, 1958: *Weather Station Index, Hydrographic Office Publication No. 207*. Washington, DC, US Navy, 462 pp.

US Weather Bureau, 1941: *Instructions for Airway Meteorological Service (Circular N, Fifth Edition)*. Washington, DC, USWB, 203 pp.

US Weather Bureau, 1957: *Circular P, Manual of Radiosonde Observations (WBAN), 7th Edition, 1 June 1957*. Washington, DC, USWB.

Wade, C. G., 1991: Improved low humidity measurements using the radiosonde hygristor. *Preprints, 7th Symposium on Meteorological Observations and Instrumentation*, pp. 285-290.

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. Atmospheric Oceanic Tech.*, **19**, 981-1002.

Wang, X. H., 2002: Vaisala cooperates with the China Meteorological Administration. *Vaisala News*, **158**, 21.

WBAN (Weather Bureau - Air Weather Service - Naval Weather Service), 1968: *Manual for Radiosonde Code (WBAN) (Third Edition)*. Washington, DC, 69 pp.

World Meteorological Organization (WMO), 1999: *Committee on Instruments and Methods of Observation, Working Group on Ground-Based Upper Air Observing Systems, New Delhi, India, 6 - 10 December 1999, Final Report*. Available at http://www.wmo.int/web/www/IMOP/reports/1999-2002/CIMO-WG-UpperAirM-1999.pdf

World Meteorological Organization (WMO), 2006a and earlier years: *Catalogue of Radiosondes and Upper-Air Wind-Finding Systems*. WMO, Geneva, Switzerland (latest version available at http://www.wmo.ch/web/www/ois/ois-home.htm).

World Meteorological Organization (WMO), 2006b and earlier years: *Observing Stations*. WMO Publication 9, Volume A, WMO, Geneva, Switzerland (latest version available at http://www.wmo.ch/web/www/ois/ois-home.htm).

Wright, J. M., Jr., ed., 1997: *Federal Meteorological Handbook No. 3, Rawinsonde and Pibal Observations*. FCM-H3-1997, NOAA, NWS, Office of the Federal Coordinator for Meteorology, Washington, DC (available at http://www.ofcm.gov/fmh3/text/default.htm).

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.

Zaitseva, N. A., 1993: Historical developments in radiosonde systems in the former Soviet Union. *Bulletin of the American Meteorological Society*, **74**, 1893-1900.

Zhai, Panmao, and Robert E. Eskridge, 1997: Atmospheric water vapor over China. *J. Climate*, **10**, 2643-2652.