P1.10 THE EFFECT OF RADIOSONDE INSTRUMENT CHANGES ON CLIMATE TRENDS OF GLOBAL ATMOSPHERIC PRECIPITABLE WATER

Steven R. Schroeder *

Texas A&M University, College Station, Texas

1. INTRODUCTION

While global surface warming since the 1970s is quite well established, efforts by different researchers to determine upper air temperature and moisture trends produce diverging results, whether computed from radiosonde data, satellite data, or reanalyses (Thorne et al. 2005). In the radiosonde record, most earlier instruments have a warm and wet bias compared to modern radiosondes, which are more responsive and better protected from radiative effects. So, climate time series computed from historical radiosonde data have an artificial trend, presumed to be mainly cooling and drying of unknown magnitude, superimposed on the actual climate trend. Surprisingly, the possibility of an artificial trend in data sets which are derived at least partially from radiosondes is still sometimes not acknowledged, as in Amenu and Kumar (2005). This paper summarizes progress toward developing unbiased global upper air climate trends using radiosonde data, focusing mainly on total precipitable water.

The uncorrected global precipitable water trend from 1973 to September 2005 shows substantial moistening in the late 1970s and a relatively steplike drying around 1990, with no noticeable trend since then. The net result is some moistening since 1973 and slight apparent drying after the 1980s. Correcting for the wet bias of earlier instruments is likely to intensify the 1970s moistening and cause a slow moistening trend since the early 1990s, but it is very unlikely to eliminate the dryness of the 1990s relative to the 1980s.

Because major instrument changes take several years to complete over a large region, instrument changes do not substantially affect interannual fluctuations, such as a very large ENSO-related moistening and drying oscillation from 1997 to 2000.

The goals of this project are as follows:

(1) Identify and document all radiosonde models and similar upper air instruments.

(2) Infer complete historical station and instrument metadata, building on all available metadata sources.

(3) Develop temperature and dew point adjustments to make each instrument type equivalent to a common "reference instrument" to compensate for biases.

(4) Apply the adjustments to each sounding and develop climatological and time series averages from the adjusted data.

While the steps to carry out the goals above are not completed, some major findings to date are as follows:

(1) Much more instrument and station metadata is available than has been previously compiled by other researchers. Over 1770 instrument model codes have been assigned so far.

(2) At many stations, old and new instrument types are used for a year or more, in varying percentages, before the transition to a new instrument is completed. Therefore, station histories are quite complex and are fairly difficult to develop because individual soundings (rather than only time series of monthly anomalies) need to be examined.

(3) Because of gradual instrument changes, data discontinuities are blurred, which is the main reason why it is often not possible to distinguish instrument-caused discontinuities from natural variations.

(4) Sometimes, it is possible to infer the existence of an undocumented instrument type by observing consistent data changes at several stations.

(5) While older humidity sensor types (hair, goldbeaters skin, and lithium chloride) have a wet bias, some instruments widely-used before and during the 1970s have a dry bias. Several countries built radiosondes based on 1960s VIZ radiosonde designs, where the carbon hygristor was heated in sunlight and produced too-dry readings.

(6) The most widespread changes in humidity sensor types were concentrated in three periods, the 1960s to mid-1970s (for countries using radiosondes based on United States VIZ designs, which changed from lithium chloride to carbon sensors), about 1975 to 1985 (as Vaisala switched from a hair to a capacitive sensor), and since 2000 (as China, India, and the Russian Federation are transitioning to carbon or capacitive sensors). In between these periods and in areas where these instrument types were not used, there is little artificial trend in the global net moisture bias from instrument changes.

2. DEVELOPING COMPLETE INSTRUMENT AND STATION METADATA

While available radiosonde station and instrument metadata is extensive, it is quite incomplete and often conflicting. This project combines many metadata sources in one location, attempts to validate the metadata for accuracy, and infers missing or inaccurate metadata.

2.1. Data and metadata sources

Archived radiosonde observations have been obtained from 1973 through (currently) September 2005 from NCAR (National Center for Atmospheric Research) Data Set 353.4. Pre-1973 data and some additional data

***** *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.

can be obtained from the Integrated Global Radiosonde Archive (IGRA) project (called Comprehensive Aerological Reference Data Set, or CARDS, until 2004) at the National Climatic Data Center (NCDC)

While the goal of this effort is to infer complete historical metadata, all available metadata sources should be considered. Major metadata compilations include Gaffen (1993, 1996), the *WMO Catalogue of Radiosondes and Upper-Air Wind Systems in Use by Members* (WMO, 2004 and earlier years), the CARDS online station history file (ftp://ftp.ncdc.noaa.gov/pub/data/cards/long_sonde.lst), and 31313 (WMO instrument code) entries in the soundings. All sources are admittedly incomplete, with some errors and inconsistencies.

A systematic literature search has uncovered many additional sources of information about instrument types, procedures, and experiments. Some sources are *Monthly Weather Review*, *Bulletin of the American Meteorological Society* (AMS), other AMS journals, conference preprints, *Journal of Geophysical Research*, *Geophysical Research Letters*, *EOS*, *IEEE Transactions on Military Electronics*, *Journal of Scientific Instruments*, *Electronics*, *Journal of Research of the National Bureau of Standards*, WMO publications (especially Instruments and Observing Methods reports) and station and observing ship catalogs, reports of field experiments and intercomparisons, advertisements in journals, manufacturer brochures and web sites, several dozen foreign journals, civilian and military manuals, radiosonde collections (at Texas A&M University, the Smithsonian, and the National Climatic Data Center), *Meteorological and Geoastrophysical Abstracts*, atlases (for station names and locations), and some books.

2.2. Developing a comprehensive list of radiosonde instrument types

To determine what instrument types were or are used at each location, a comprehensive list of radiosonde types is needed. This effort starts with published lists such as Gaffen (1993), Smith (2002), and the WMO list of instrument codes (which are part of the 31313 group in most operational soundings). The WMO and Smith (2002) lists are undocumented, but this effort has uncovered references to most instruments in those lists.

The scope of the instrument catalog includes all atmospheric in-situ profiling instruments for which references were found, because IGRA should have the capability to be be comprehensive. Besides radiosondes, other categories are nonbroadcasting instruments, dropsondes, rocketsondes, ozonesondes, tethersondes, aircraft profiling instruments, other specialized radiosondes, and wind-only instruments. So far, over 1770 instrument type codes are assigned. Some entries are erroneous or are not actually radiosondes, some are doubtful, some are duplicate names for the same instrument (but entries are combined if possible), and some entries are codes to identify a family of instruments where the specific model is unspecified.

Proliferation of instrument codes results because individual sources of metadata include different information:

(1) The largest number of added codes results from slightly different model permutations in some radiosonde families, particularly VIZ/Sippican and Vaisala. Permutations can include different navigational aids for wind finding (such as Omega, LORAN, or GPS), different transmission frequency bands or modes (AM, FM, or pulse), and instruments with and without a hypsometer, transponder, dereeler, precalibrated sensors, or extra channels for additional sensors. Each disclosed variation is assigned a different code because it is not known in advance if a particular combination of features causes a detectable difference in data characteristics.

(2) Some important changes are not reported as new models but need to be assigned separate codes. Such changes include the new VIZ carbon hygristor in June 1980, changes in ground calibration or data processing, changes in solar or radiation corrections, and changes in formulas which relate ordinates or other broadcast signals to values of meteorological variables.

(3) The instrument type may be obtained from the 5-digit 31313 group, which has been reported in most operational soundings since the late 1990s. The first digit is the solar or radiation correction, the next two digits are the WMO instrument type (but some separate codes are assigned for different combinations of instrument types and ground processing systems), and the last two digits describe the wind finding method (but some rarely-used codes indicate other conditions of the sounding, such as "all systems operating normally."). Some of these 5-digit combinations can be related to a specific radiosonde model, but others cannot. For example, the Vaisala RS80-15N model appears to have been reported with several 31313 codes, and in that case a separate instrument code is assigned here for each regularly-used 31313 group.

(4) Additional proliferation of codes results because some metadata relevant for instruments is not in the 31313 code, such as whether Vaisala RS80 uses an A-Humicap or H-Humicap humidity sensor. Similarly, because the number of 2-digit instrument types in the 31313 group is nearly exhausted, new Indian and Chinese models use their old instrument type codes, but here each new model is assigned a separate code.

2.3. Validating station elevations

Each observation in Data Set 353.4 contains a station ID, latitude, longitude, and elevation (but not the station or ship name) from an operational catalog maintained by the National Centers for Environmental Prediction (NCEP, called National Meteorological Center or NMC until the mid-1990s). Before February 1995, the catalog was infrequently updated, and even now, updates occur only after the actual station change.

Here, one of the first metadata steps for each station is to develop a complete history of surface elevations and dates of changes. By computing the elevation from the first above-surface height, errors of 300 meters or more have been found, with many erroneous elevations between 20 January 1976 and 4 February 1980, and between 4 June 1986 and 30 June 1989. Elevation changes of 5 meters can often be detected to the exact observation, and changes of 1 meter can usually be detected to within a month. An error which varies widely between soundings often indicates that the surface observation is missing. It is not possible to validate elevations using observations with no heights reported. Alduchov and Eskridge (2002) performed a similar analysis, but based on CARDS metadata files, it appears they only checked a small subset of all stations. Operational agencies compute persistent height errors relative to model analyses, and it is possible that some of these are actually surface elevation errors.

The station name is also checked using station lists from WMO, NCAR, CARDS, the United States Air Force, and Gaffen (1996). While station locations cannot be verified directly from observations, many erroneous locations can be corrected, such as a wrong latitude or longitude sign, a reported location which differs from an atlas location, or a transient discontinuity in the path of a ship or Arctic ice island. A change of more than about 0.25° of latitude or longitude, or a change of location with no change in elevation, is suspicious, especially if the name of the station does not change in metadata catalogs. NCEP metadata reports latitudes and longitudes in hundredths of degrees, while the other sources report latitudes and longitudes in degrees and minutes. A fairly frequent error in the NCEP metadata is to not convert minutes to hundredths of degrees, so if the reported location changes from 10.40° to 10.24° and WMO metadata says the station is at 10°24', the station probably did not move and the correct location is most likely 10.40°. (WMO metadata may similarly be erroneous, and a location of 10°75' is obviously wrong.)

In many countries, names of cities have changed to remove colonial influences, or the spellings in English have changed because of updated transliterations of other languages into the English alphabet. For example, many name changes of India cities since the 1970s (Madras is now Chennai, Bombay is now Mumbai, Poona is now Pune, and Calcutta is now Kolkata) are not in metadata catalogs yet. Web data sources provide the most updated information about location names, and also some of the underlying political controversies. Some country names have changed, as well as names of states or other political subdivisions (For example, part of the Northwest Territories of Canada became Nunavut), and these are updated as far as possible. As in Gaffen (1996), the reason for monitoring such changes is that country alliances may provide clues about the suppliers of radiosondes.

A complete elevation history has been prepared for each ship in this data base. While the operational archive shows a station elevation of 0 meters for ships, the launch elevation computed hydrostatically is typically from 5 to nearly 30 meters, and has gradually increased as ships have become larger. Ship names and countries performing weather observations (which may differ from the country owning the ship) have been identified for over 90 percent of the ship IDs, and the country performing the observations has been identified for almost all ships without a known name.

As an example of a detectable station location error, station 10384 (Berlin Templehof Airport) was reported to be in the Southern Hemisphere from September 1975 until January 1976. A probable error is the location of station 99877 in a very remote part of Afghanistan (which started reporting 30 June 2004, and is not in any known catalog). The claimed elevation is 673 meters, but that region has an elevation of 2500 to 3000 meters, and the surface elevation computes to an average of 303 meters. The correct station probably is Termiz, Uzbekistan, which is almost exactly 3° north of the stated location, with an elevation close to 300 meters (Termiz, or Termez, is also station 38927).

2.4. Inferring instrument types and transitions

This project validates reported instrument types and infers unreported instruments and changes. The consistency of hypothesized instrument signals can be best verified by cross-checking all stations. The same procedures can be applied to both land and ship stations. Most ships travel into different climate regions, but this has not made it difficult to identify instrument transitions.

The basic method to develop a complete metadata history involves two processes, performed repeatedly. First, the observations and derived statistics at stations which appear well-documented are examined for common characteristics of instrument types. Second, similar instrument signatures are sought at stations or in time periods which are less well documented.

When seeking consistent signals of each instrument type, an instrument model should have smoothly-varying differences between stations, as expected by climatology. Three types of variables can be examined:

(1) Temperature-related variables, especially at high altitudes and differences between day and night. If solar and radiation corrections are effective, these variables may not distinguish instrument types because of large real stratospheric variations such as transient volcanic warming and the gradual cooling from ozone depletion.

(2) Moisture-related variables at all levels, such as the average dew point depression, largest dew point depression, and lowest relative humidity. Moisture variables are often very effective in distinguishing radiosonde types. "Dew point censoring" as practiced in the United States from 1 April to 1973 to 30 September 1993 (continuing into 2005 at a few military stations) can be detected by at least one dew point depression of 30° C, no other relative humidities under 20 percent, and no dew point depressions above 30° C. This "US-style censoring" indicates the use of VIZ sondes, or possibly Space Data sondes starting in the late 1980s, or Sippican sondes starting in the late 1990s for military stations continuing this data reporting practice.

(3) Statistics of sounding "quality" such as average number of temperature and dew point levels, lowest pressure of the sounding, lowest pressure with wind reported, number of wind levels, and lowest pressure and temperature with dew point reported. These may indicate an instrument change, or a change in radar, ground processing equipment, software, or operating procedures. A discontinuity in temperature or moisture may occur even if the radiosonde does not change. For example, starting to report dew point regardless of temperature instead of only to a temperature around -40°

may cause apparent upper-tropospheric moistening because the colder cases are no longer excluded.

Caution must be used in inferring that a change in a statistical indicator is a change in an instrument. If such a change is an administrative policy, it is likely to be adopted the same day or in a short time by all (or at least many) stations in a country, and should affect all observations after the change is adopted. If a station gradually applies a practice to more and more observations (such as reporting dew points only to a temperature around -40°), it is likely to be introducing a new instrument gradually. If a new practice applies to all soundings starting a certain date, but this date varies considerably from one station to another, that may indicate that each station introduces a new radiosonde after using up all of the old ones. If the dew point ends at random points in different soundings (not consistently to a temperature just above -40° or some other threshold), the station may simply have a defective batch of humidity sensors, or the dew point may be omitted when the humidity drops below a certain percentage (as in certain Indian and Australian radiosondes). Even for the same instrument, the reporting practice may differ between countries, or between agencies such as civilian versus military.

The procedures used here to determine instrument transitions are most effective when some metadata is available for at least some stations in a country. Even if the metadata is incomplete and of suspected quality or timeliness, it provides useful clues that can help narrow down the possible radiosondes.

For example, the data characteristics of the India MK III and Australia Phillips RS4 II (both with lithium chloride hygristors) are quite similar. The dew point is omitted when the humidity drops below 10 to 20 percent so in an area with pronounced wet and dry seasons, the dew point may be reported to temperatures around -40° or colder in the wet season and only to temperatures around freezing in the dry season. Certain stations in Thailand showed those characteristics in the late 1970s and early 1980s, and limited documentation claims that Thailand used VIZ and Australian instruments around 1982. Without this amount of metadata, it is possible that Indian instruments would have been inferred. Based on the data examination, it appears that most stations in Thailand were using VIZ instruments in the late 1970s but not in 1982, because a signature of VIZ instruments at that time was reporting dew points to a temperature around -40° all year, with few relative humidities much below 20 percent or (at some, but not all VIZ stations) the use of US-style censoring.

2.5. Developing a consolidated metadata file

The main output of this portion of the project is a single text file called "RaobMetadata" which is designed to be both human-readable and suitable for computer processing. This is accomplished by placing formatted data at the beginning of lines. Comments either follow formatted data or are on a separate line. If a line violates the format (any nonnumeric character in a numeric field), starts with a semicolon, or is blank, it is a comment line and is ignored by the program. Comments can include the station name, operating agency, a reference or metadata source, data characteristics, reasons for certainty or uncertainty of the metadata, or anything else that helps explain what is observed in the data.

Having formatted metadata, comments, and references in the same file reduces duplication in maintaining the metadata. Even for people who do not plan to use radiosonde data directly in their research, this file should still be useful as documentation of station history and instrument or location changes.

The major portion of the metadata file lists all stations with their location and instrument type, with the date and time when that combination of parameters begins. A new data line is used for any change. Station metadata includes 3 lists, with the following formatted data:

(1) Land stations in numerical order, listing each station with its 5-digit ID, latitude, longitude, surface elevation, instrument type, and starting date and time of this combination of parameters. Temporary field experiment sites are assigned 5-digit IDs, and one "land" station is actually the Ekofisk oil platform (station 01400).

(2) Fixed ship stations in alphabetical order by ship station (for example, the "Ship M" location reported using station IDs 4YM, then C7M, and now LDWR), listing each station with its ID, nominal latitude and longitude, elevation from which the radiosonde is launched, instrument type, and starting date and time of this combination of parameters. In the archived data, a ship station reports an elevation of 0 meters, but here the computed launch elevation is listed. Because a fixed ship can drift slightly, individual observations specify the actual latitude and longitude, not the nominal location which is in this metadata file.

(3) Moving ship (or ice island) stations in alphabetical order by reported station ID, listing each station with its ID (4 to 6 characters), computed launch elevation, instrument type, and starting date and time of this combination of parameters. Moving ships report their location with each observation, so this part of the file does not specify latitudes and longitudes.

The metadata file contains some optional sections which are used to correct errors that are specific to this data set. Operational processing accepts all station IDs that are not a 5-digit number because there is no consolidated list of alphanumeric identifiers, and apparently accepts all 5-digit station IDs which are in surface or upper air station catalogs (For example, there are over 100 observations for station 28552 which are mostly fragments of reports from station 28952). Over 3000 station IDs are simply typing or communication errors. Optional metadata sections rename or reject certain observations or station IDs. One section lists Arctic ice islands and their approximate locations for each month. Observations given other names (often SHIP) near these locations are actually Arctic ice island data. A final metadata section lists land stations reported by alphanumeric IDs which are assigned 5-digit station IDs here. Most of these are temporary field sites, so identifying the station as a ship allows reporting the latitude, longitude, and elevation

3. DEVELOPING ADJUSTMENTS FOR INSTRUMENT TYPES

After substantially complete metadata is developed for all stations, it is possible to systematically search for consistent differences between instrument types and apply adjustments to statistically correct for instrument biases. Steps below are not applied strictly in sequence. For example, preprocessing of archived data was done before preparing the metadata file above, and is repeated and refined as the metadata file is developed.

3.1. Processing of archived observations

All data since 1973 is reprocessed almost every month to improve the evaluation of observations and collection of statistics. The program writes files of processed soundings in a readable (and also computer-readable) format. This file has no comments, but has indicators of the quality of the sounding and computed statistics such as precipitable water (total and in layers), sea level pressure, free-air lapse rate, and hydrostatically-computed surface elevation and heights.

A separate statistical file lists each station ID, the starting date and time of each new location or elevation, and the number of observations and other statistics for each year. This file reveals many problems with the archived data, especially erroneous station IDs.

Each sounding is evaluated using over 30 tests. Many errors reject a sounding, and others only cause rejection of that data value. Errors which might be corrected include temperature sign errors (if reversing a sign makes the temperature realistic for the level and does not create an unrealistic inversion or superadiabatic cooling) and height errors of 500 or 1000 meters (quite often, a 1000-mb height of 0 meters is reported as 500 meters).

Reported 1000-mb heights at high altitudes in the Andes, Antarctic, and Himalayas are often unrealistically high, indicating that the assumed subsurface air column is too cold. Here, subsurface heights and the sea level pressure are computed by projecting the free-air temperature (above the surface inversion, if any) downward using the free-air lapse rate. This gives much more realistic sea level pressures, especially in the Andes where reported heights often indicate a sea level pressure above 1040 mb.

Validation assigns a "quality indicator" to each sounding, which, if a letter, indicates the first reason found for rejecting a sounding. For example, "Z" indicates a sounding with no heights reported, "U" means there is a substantial superadiabatic layer, "c" indicates an 850-mb height that is too low or high, or "P" indicates that the reported surface elevation is not hydrostatically realistic. Rejected soundings are still processed as much as possible to see if the information can be rescued by, for example, correcting the surface elevation. For soundings which are not rejected, the quality indicator is a punctuation symbol, and indicates a minor limitation of the sounding ("_" indicates that no defects are found). For example, ";" indicates that dew points are not reported if the temperature is below about -40°.

3.2. Inferring instrument types and identifying transitions

The most labor-intensive phase of this project is to examine the data record at each station to validate any existing metadata, to determine consistent characteristics of each instrument type, and to infer a complete instrument history where documentation is missing or inaccurate. Other researchers generally do not try to generate a complete history of specific instrument types, but simply determine that a discontinuity is presumably instrument-caused (Lanzante et al. 2003)

The *first* step is to prepare time series and monthly values (counts, averages, and some extremes) of data variables for each station from preprocessed observations produced in the preceding step. A "station" can be a sequence of station IDs where one station replaces another (This is subjectively determined. Some station replacements, such as in the United States in the late 1990s, are too far apart to be considered homogeneous). Including stations with sparse data, ships, wind-only stations, and stations which are probably erroneous, there are over 2400 stations or sequences from 1973 to September 2005. There are 3 files produced for each station or sequence:

(1) A file with one line of statistics (data elements listed below) for each observation.

(2) A file containing monthly averages or totals based on these statistics.

(3) A similar monthly file that uses observations within 3 hours of 0000 and 1200Z to compute statistics. The reason for this file is that soundings around 0600 or 1800Z are often less detailed (for example, terminating at 100 mb or containing only mandatory levels) than those at 0000 and 1200Z, so starting or stopping 0600 or 1800Z observations may produce spurious changes in statistics such as the average number of levels per sounding.

Data elements listed for each observation are station ID, reported latitude, reported longitude, reported elevation, archived and reported instrument type, quality indicator, computed elevation, computed sea level pressure, surface pressure, pressure at top of sounding, lowest pressure with dew point reported, number of temperature levels (*), number of significant temperature levels (*), number of dew point levels (*), coldest temprature in the sounding, coldest temperature with dew point reported, coldest dew point, largest dew point depression (*), lowest relative humidity (*), number of levels with dew point and average dew point depression in 3 layers (800 to 600, 600 to 400, and 400 to 200 mb), number of wind levels by pressure (*), number of significant wind levels (*), lowest pressure with wind reported, and highest height with wind reported. Data elements with (*) are computed only from the surface to 100 mb, and some minor indicators not listed here are also included.

The *second* step is to examine available metadata, these files, and the processed observations, to infer instrument types and transitions. Automated methods to identify discontinuities are unsuccessful. Stations are examined in groups according to likely instrument histories. Histories for stations using Russian (Schroeder 2005), Indian, and Chinese instruments are nearly completed. First, the time series at stations which appear to be well-documented are examined for consistent signals of reported instruments and discontinuities coinciding with reported transitions. Then, time series at stations with little or no documentation are examined for similar signatures of instrument types and discontinuities. These steps are repeated many times, especially as variables are identified that seem to be consistent indicators of instrument types, to refine inferences of instrument types and transition dates. Differences between nearby stations may also help identify instrument transitions.

The inferred instruments and any other station changes are entered in the "RaobMetadata" file as the history is developed, along with any comments such as the apparent consistency of signals for the stated instrument type.

3.3. Designating a "reference" instrument

The basic adjustment philosophy here is to make "absolute" adjustments to a chosen "reference instrument," rather than "relative" adjustments from an earlier period to the latest period. This means that stations which have had no instrument changes are still adjusted to be statistically equivalent to the reference instrument. For example, most Russian stations still use older types of instruments, with goldbeater's skin hygrometers, and adjustments make dew points drier even at stations which have not yet started using newer instrument types.

An ideal reference instrument is correct and unbiased, but no such operational instrument exists. Here, the reference instrument is defined as the average of certain VIZ and Vaisala models, specifically VIZ models 1190 and higher through VIZ B (with a carbon hygristor, but excluding models where the hygristor was excessively heated in sunlight) and Vaisala RS21 and RS80 (except for a variety of RS21 used from the late 1970s to mid-1980s which was excessively dry). These models were widely used from the 1970s to about 2000, so most other instruments can be directly compared with one or more of the reference models.

VIZ tends to be wetter than Vaisala. Recent papers reporting comparisons with dew point hygrometers indicate a growing consensus that VIZ had a moist bias and Vaisala had (and may still have) a dry bias. So, the average of the two instruments may be nearly correct.

It would appear to be desirable to average the latest models (such as VIZ/Sippican B2 and Microsonde II, and Vaisala RS90 and RS92) as the reference, but those models are not yet as widely used as the models mentioned above, so fewer instruments are able to be directly compared with such a reference model. However, those models may be used as the reference eventually. In any case, while current radiosonde models can provide much more detailed atmospheric profiles than earlier instruments, the lost information from smoothed profiles cannot be restored regardless of the chosen reference instruments.

Even the reference models are inhomogeneous because of changes in sensors or their exposure. VIZ used a new carbon hygristor starting June 1980 and the case enclosing the hygristor changed several times. For Vaisala, RS80 is much smaller than RS21, the radiation corrections changed several times, and some stations used the H-Humicap humidity sensor, which initially produced artificially dry readings due to contamination in storage. These differences will need to be investigated further and initial corrections to make the reference instruments homogeneous may be needed.

3.4. Developing temperature adjustments

A goal of this project is to adjust the temperatures and dew points in all archived observations to be statistically unbiased with respect to the reference standard. Other researchers use station-specific corrections to develop homogeneous radiosonde data. That is a legitimate approach because the climate environment is different at each station. Here, the basic principle is to develop adjustments for each instrument type, with a scheme that accounts for differences in environment. Both approaches use the data itself, before and after a transition or otherwise comparing periods using different instruments, to determine adjustments.

While several hundred of the more than 1700 instrument models are expected to be identified or inferred to be used, probably there will be about 50 to 60 distinct types with different temperature or humidity characteristics since 1973, because many of the different models do not have changes to the sensors.

To compare any pair of instrument types, 3 types of comparisons can be made using archived data:

(1) A transition from one type to another at a station, in either order, possibly with a brief gap (or temporary use of another type) between these instruments. The time period of comparison for a station should include an integer number of years (up to 3 years) with each instrument type.

(2) Simultaneous use of the instrument types at nearby stations. The stations should be in very similar environments, such as Berlin and East Germany. The time period of comparison should be as long as possible, but preferably an integer number of years, even if a real climate change is suspected in that period.

(3) Frequent alternations of the instrument types at the same station. Again, the comparison should be as long as possible. Caution must be used if one type is used in daytime and the other is used at night.

A fourth type of comparison, a formal intercomparison with different instruments on the same balloon, is not considered here because such comparisons involve a very small number of radiosonde launches.

For each instrument type, it is ideal if stations are found with transitions or other comparisons directly to or from a reference instrument. Some instruments may not have a direct transition to a reference instrument (for example, "Type A" to "Type B" to a reference model). Before 1973, some instruments may have a longer "chain" of transitions or comparisons to a reference. Some types, such as (possibly) Indian or Japanese instruments, may have no direct "chain" of transitions to or from a reference. Poor-quality (long-distance) comparisons may be required in such cases. The chain of transitions should be as short as possible because the statistical uncertainty of the adjustment rises as the number of adjustments applied increases. (Adjustments applied by other researchers such as in Lanzante et al. (2003) have a similar problem, because each preceding time segment is adjusted to be statistically equivalent to the latest time segment.)

Here, the first step is to develop differences between reference instrument models, and the second step is to develop differences between each instrument type and a reference instrument. The inferred metadata must be substantially complete at this point so all stations with appropriate instrument comparisons can be included. If only stations with the most obvious instrument discontinuities are included, the adjustments will be too large.

Instrument pairs which need to be compared include VIZ models in the reference, Vaisala models in the reference, comparisons of VIZ and Vaisala, and each instrument pair in each chain of comparisons to a reference model. If it turns out that the instruments in any pair have only small and unsystematic differences, those instruments can be considered homogeneous and can be consolidated.

Because an instrument type is assigned to each sounding, a single list of stations and time periods can be used to generate comparisons for each instrument pair. All 3 types of comparisons can be made to estimate the differences between the same two instruments. For example, to compare instruments called “Type A” and “Type B,” some stations could transition from type A to B, some from type B to A, some could alternate frequently between types A and B, and some nearby pairs of stations could use type A at one station and type B at the other.

For each instrument pair, a computer program would use this list and, for each station and time period (and for all stations), would accumulate statistics considering only observations with each specified instrument type, and would ignore soundings not assigned the specified type (or observations where the instrument type is uncertain).

Temperature adjustments are prepared first. The specific statistics to accumulate for each instrument type in a pair are the cumulative distributions of temperatures in specified pressure layers (usually, the surface is not adjusted because surface data should be obtained from permanently-installed instruments) and intervals of sun angle above the horizon. Such statistics should be accumulated for each station and for all stations as a group. Because each station (or adjacent station pair) should have a similar climate environment, if the cumulative temperature distributions are quite similar for both instrument types (with small, unsystematic differences and little difference in the means), then the instrument types should be considered homogeneous, with no adjustments made.

A temperature adjustment (if needed) is the amount to add to the archived temperature, to make the probability distribution the same as for the instrument type adjusted to, within the same interval of pressure and sun angle. Because the reference instruments appear to be well-protected from radiative errors, most adjustments will probably be negative (causing the readings to become cooler). Adjustments are expected to be larger in the stratosphere than nearer the surface, and probably will be small at night.

Even VIZ and Vaisala instrument models in the reference are adjusted because the reference instrument represents the average of the chosen VIZ and Vaisala models. A VIZ observation is adjusted by adding half of the difference from VIZ to Vaisala, and a Vaisala observation is adjusted by adding half of the difference from Vaisala to VIZ. It is possible that both the VIZ and Vaisala series are inhomogeneous enough that small adjustments may be needed to correct some VIZ models (possibly before and after the hygristor change in 1980) to a "VIZ average" and to correct some Vaisala models to a "Vaisala average" before defining the characteristics of the reference instrument. If any VIZ or Vaisala model requires a large adjustment, then that model should be excluded from the reference.

3.5. Developing dew point adjustments

When adjusting temperatures in the step above, dew point depressions are not changed. This changes the relative humidity slightly. For example, with a negative temperature correction (cooling), the relative humidity with a constant dew point depression decreases. However, a major difference between inhomogeneous instrument types is in the response of different humidity sensors. So, a comparison of dew points between instrument types after applying temperature adjustments includes all factors which cause the distribution (including bias) of moisture data to differ.

For each instrument type in a pair, cumulative probability distributions of dew point depressions are obtained in intervals of pressure (with the surface excluded), temperature, and possibly sun angle, using the same lists of stations and time periods which were used to develop temperature adjustments. As with temperature distributions, if the two instrument types are unbiased, or are equally biased, they should show similar probability distributions in each interval.

Even though many instrument types are not expected to need temperature adjustments, most or all instrument pairs are expected to show systematic dew point differences. Dew point depression adjustments are more complex than temperature adjustments because for most older and less sensitive instruments, the difference from the reference instrument is usually a narrow probability distribution (infrequent reporting of dry or nearly saturated conditions) as well as a bias. So, the adjustment for an instrument type is stored as a 3-dimensional array. Specifically, within a pressure layer and temperature interval, an adjustment amount (to be added to the reported dew point depression) is specified for different intervals of reported dew point depression. If intervals of sun angle are included because the humidity sensor is affected by solar radiation, then the adjustment array has 4 dimensions. To widen a narrow probability distribution, the dew point depression adjustment is negative in moist cases (a low dew point depression is decreased further, closer to saturation) and positive in dry cases (a high dew point depression is increased, which lowers the reported relative humidity).

For instrument types which report dry circumstances as a fixed minimum relative humidity (such as 10 or 20 percent) or a fixed dew point depression (the only case encountered is "US-style censoring" which reports relative humidity under 20 percent as a 30° depression), a relative humidity just under the threshold is assigned. For example, based on relative humidity statistics with VIZ instruments just after US-style censoring was discontinued, a censored dew point is changed to the value corresponding to about 17 percent relative humidity, or to a slightly lower humidity as the temperature rises. For an instrument with a relative humidity threshold of 15 percent, cases reporting 15 percent should be assigned a relative humidity of 13 percent, and if the threshold is 10 percent, a 9 percent relative humidity should be assigned. (If the lowest reported relative humidity is lower than 10 percent for some instrument type, a missing relative dew point should be assigned a relative humidity equal to the lower limit because the amount of moisture is small.) Because the assigned relative humidity is based on measurements by this instrument, which has its own bias, the dew point corresponding to this assigned relative humidity is adjusted in the same way as any dew points reported in observations using this type of instrument.

As with temperature adjustments, the adjustment from VIZ to the reference is half the difference between VIZ and Vaisala, and the adjustment from Vaisala to the reference is half the difference between Vaisala and VIZ.

At the end of this step, the temperatures and dew points at each level of each observation are adjusted to be statistically unbiased with respect to the hypothetical reference instrument, although lost detail in a smoothed profile cannot be restored.

4. PRELIMINARY GLOBAL PRECIPITABLE WATER VARIATIONS SINCE 1973

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.

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

(1) Observations around 0000Z (2100 to 0300Z) are weighted half to the day before 0000Z and half to the day starting 0000Z. 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 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.

(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 32-year climatology from 1973 to 2004, 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 2004, 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.

4.2. Observed global precipitable water variations since 1973

Monthly and annual grids and time series are produced from the unadjusted radiosonde data almost every month 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 2005.

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.

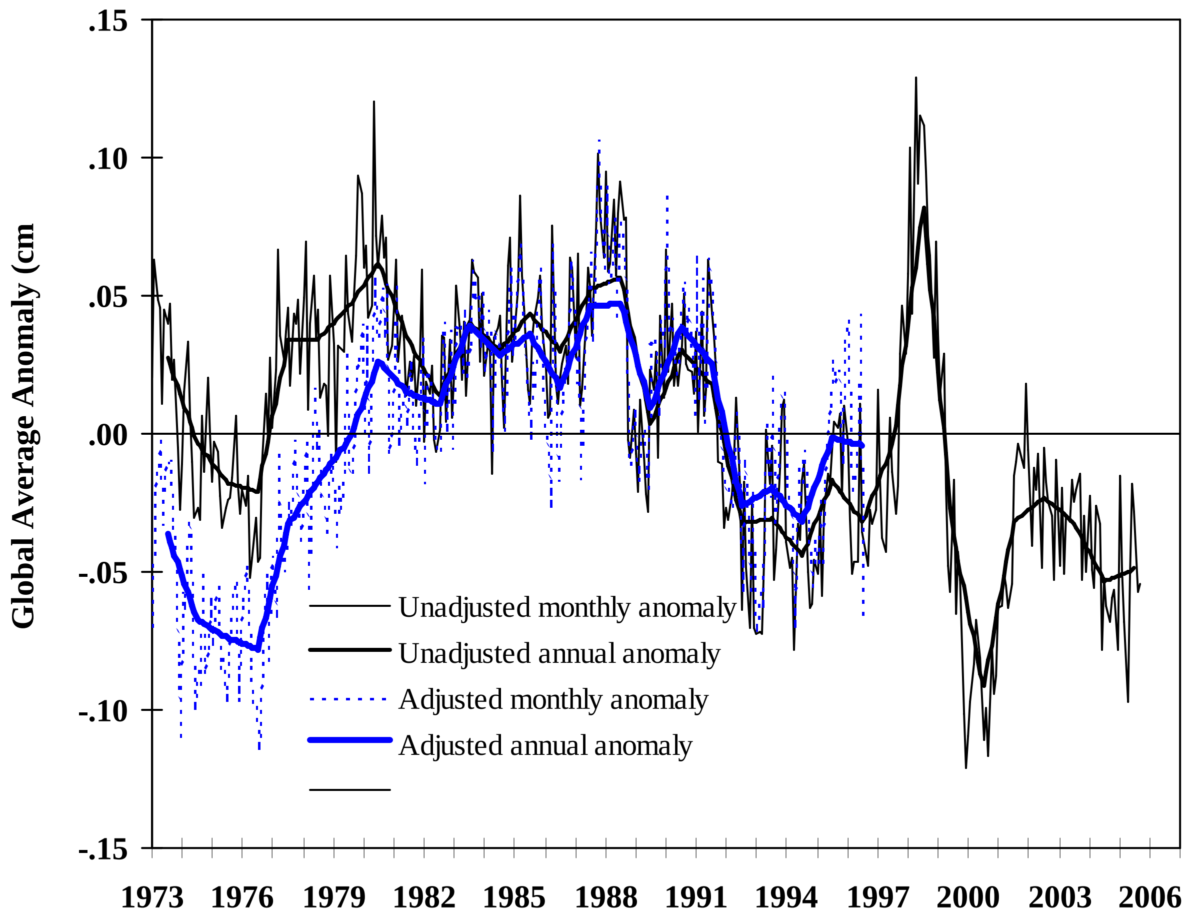

FIGURE 1. Monthly (thin lines) and annual (thick lines) global area-averaged anomalies of precipitable water. The unadjusted values for January 1973 to September 2005 are relative to the 32-year climatology for 1973 through 2004. The adjusted values for January 1973 to July 1996 are based on a previous project and are relative to the average for January 1973 to July 1996.

5. 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.051003.

Acknowledgements

This project is 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.

Amenu, G. G., and P. Kumar, 2005: NVAP and Reanalysis-2 global precipitable water products: Intercomparison and variability studies. *Bull. Amer. Meteor. Soc.*, **86**, 245-256. Doi:10.1175/BAMS-86-2-245.

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

Gaffen, D., 1996: *A Digitized Metadata Set of Global Upper-Air Station Histories* (NOAA Technical Memorandum ERL ARL-211). NOAA Air Resources

Laboratory, Silver Spring, MD, 38 pp.

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.

Schroeder, S. R., 2003: Completing instrument metadata and adjusting biases in the radiosonde record to allow determination of global precipitable water trends. *Preprints, 12th Symposium on Meteorological Observations and Instrumentation*, Long Beach, CA, Poster P1.6, http://ams.confex.com/ams/pdfpapers/54950.pdf.

Schroeder, S. R., 2005: Toward developing unbiased upper-air temperature and moistrue 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.

Smith, D. P., 2002: *Data Documentation for Data Sets SSI 6301 - DSI 6379, NCDC Upper Air Digital Files*. National Climatic Data Center, Asheville, NC, http://www1.ncdc.noaa.gov/pub/data/documentlibrary/tddoc/td6300.txt.

Thorne, P. W., D. E. Parker, J. R. Christy, and C. A. Mears, 2005: Uncertainties in climate trends: Lessons from upper-air temperature records. *Bull. Amer. Meteor. Soc.*, **86**, 1437-1442. Doi:10.1175/BAMS-86-10-1437.

WMO, 2004 and earlier years: *Catalogue of Radiosondes and Upper-Air Wind-Finding Systems*. Geneva, WMO, online through http://www.wmo.ch/web/www/ois/ois-home.htm.

Zaitseva, N. A., 1993: Historical developments in radiosonde systems in the former Soviet Union. *Bull. Amer. Meteor. Soc.*, **74**, 1893-1900.