8B.6A HOMOGENIZING THE RUSSIAN FEDERATION CLIMATE RECORD BY ADJUSTING RADIOSONDE TEMPERATURES AND DEW POINTS FOR INSTRUMENT CHANGES

Steven R. Schroeder *

Texas A&M University, College Station, Texas

ABSTRACT

The archived radiosonde record is less useful than desired for identifying the causes of ongoing climate changes in the last few decades, or in some cases for even determining the direction of the temperature or moisture trend, because of frequent instrument changes at all stations. In general, newer radiosonde models are more sensitive and are assumed to be better protected from radiative errors than early models, so on the average an artificial cooling and drying trend is hypothesized to be superimposed on the true climate trend. Researchers attempt to adjust radiosonde data to compensate for instrument changes, but the adjustments are questionable due to incomplete and inaccurate knowledge of the instrument types and the times when instruments changed at each station.

Previous work in this project has found that complete and very accurate station histories of instrument changes can be constructed by systematically searching time series of variables that are very sensitive to different instrument types. These variables show common data characteristics of the same instrument type used at individual stations, and discontinuities indicate instrument changes. Station histories for the Russian Federation, developed back to 1973, appear to be very reliable with consistent signals at over 200 stations. At least 15 radiosonde models are currently in use in 2007 to the beginning of 2009. Station histories are usually complicated because of frequent alternations of up to 7 instrument types at the same station.

It is a much more feasible task to determine differences between instrument types and to develop instrument corrections when reliable station histories are available. This project will test various methods to determine differences between instrument types and develop adjustments, first for reported temperatures and then for dew points. The primary proposed method is "histogram matching," where cumulative reported probability distributions of reported data values are developed for common circumstances involving each distinct instrument type. To make one instrument statistically equivalent to another "reference" instrument type, the data value with the observed percentile is replaced by the data value with the same percentile in observations obtained with the "reference" instrument.

For a proper transformation, the data circumstances for the two instrument types involved in each comparison must be equivalent. Probability distributions will be developed in categories stratified by pressure level and sun angle, and also temperature in the case of dew point comparisons. However, either the long-term trend or short-term fluctuations such as ENSO or volcanic eruptions can affect the probability distributions and contaminate the adjustments. For example, if adjustments are based on the differences in probability distributions before and after an instrument change at a station, the adjustments remove part of the long-term trend or project short-term climate anomalies into other time periods. Averaging the adjustments over a large number of stations making the same transition at various times only partially relieves this problem. To determine if more satisfactory adjustments can be developed, this project will examine other comparisons such as alternating use of the instruments at a station, simultaneous use of the instruments at nearby stations, or transitions in both directions between instruments.

1. INTRODUCTION AND BACKGROUND

Climate models are used to attribute the observed historical climate trend to factors such as the buildup of greenhouse gases, but their trends tend to differ from the observed trends above the surface obtained from radiosondes. In general, radiosondes show less warming than expected in the lower troposphere and more cooling than expected in the stratosphere. Every radiosonde station with a long record has experienced instrument changes due to technological improvements such as faster-responding sensors and better protection from radiation. On the average, ongoing improvements are hypothesized to add an erroneous cooling and drying trend to the actual climate trend.

Many researchers have attempted to determine the actual trend by applying adjustments for each distinct instrument type to the data. However, metadata describing the nature and timing of instrument changes at each station, as well as other relevant information such as station location and elevation changes, is incomplete and sometimes erroneous. So, adjustments are applied to remove discontinuities in the data without knowing whether the discontinuities being removed are caused by instrument changes or possibly by natural fluctuations that seem unusually large or abrupt.

As background, this section contains 4 summaries of data sources, metadata sources, adjustment methods used by other researchers, and how this project develops complete metadata. Section 1.1 summarizes sources of current and archived historical radiosonde data, including a discussion of a planned future transition from the text-based reporting codes that have been used throughout the entire history of transmitted observations to a binary code. Section 1.2 summarizes sources of recent and historical metadata considered in this project, including a

---

* *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.

discussion of the instrument code in the 31313 group of reported soundings. Section 1.3 summarizes reported data adjustment approaches that other researchers have used to deal with the problem of instrument changes. Section 1.4 summarizes the methods used in this project to develop complete station and instrument metadata from archived observations.

1.1. Upper air observation data sources

The primary data source used in this project is National Center for Atmospheric Research (NCAR) Data Set 353.4 (DS353.4, available from http://dss.ucar.edu/datasets/ds353.4), which contains global radiosonde observations from 1973 to February 2007 transmitted over the Global Telecommunications System (GTS) and processed by the National Centers for Environmental Prediction (NCEP, formerly National Meteorological Center (NMC) before 1994). This data set contains almost all upper air data received by NMC or NCEP, including radiosondes from land, ship, and mobile stations, dropsondes, and some wind profiler observations. In some periods, DS353.4 omits significant temperature or wind levels from some countries or regions, and of course it does not include any observation that is not received in time for use in forecasting. Also, an observation with a numeric station identifier (ID) is omitted if the ID is not in an NMC/NCEP station file (see section 1.2) because the location is not known. However, overall DS353.4 has been found to be the most comprehensive available radiosonde data archive covering the period since 1973.

DS353.4 stores the data in an ASCII text format described by *Office Note 29* (ON29), which is available at the DS353.4 web site. NCEP provided files to NCAR in the ON29 format until early 2000, but since 2000, has provided files only in BUFR (Binary Universal Form for the Representation of meteorological data) format (see section 1.1a). NCAR performed the conversion to the ON29 format starting early 2000, but has not produced DS353.4 files after February 2007 because of difficulties they have had with the BUFR to ON29 conversion.

After February 2007, this project has continued to examine radiosonde data using additional sources of soundings. Some of the programs to express all data sources in a common format for analysis are not completed so not all needed analyses have been performed.

First, NCAR has recently made Data Set 351.0 (DS351.0, the source for DS353.4) available back to 31 March 2000 at http://dss.ucar.edu/datasets/ds351.0, along with BUFR downloading software. NCAR usually updates this data set each Friday with data through the preceding Saturday. Here, data back to January 2007 was downloaded for a brief overlap with DS353.4 in January and February 2007. Instead of attempting to reproduce the ON29 format, this project modified the BUFR software to extract the station ID, latitude, longitude, and elevation (these are from the NCEP metadata file, which is discussed in the next section), and the "raw report string" (variable RRSTG in the files), which is the original text report from the station. The text reports are extracted because most data problems are easier to identify and fix in the original text format than in a further-processed format. Examples of problems that can often be fixed are repeated or missing characters, and the usual human or random mistakes that occur when typing or transmitting a message.

NCEP also prepares an additional upper air data set, which NCAR archives as Data Set 337.0 (DS337.0, http://dss.ucar.edu/datasets/ds337.0). DS337.0 contains the final BUFR-formatted output files from the PREPBUFR process. PREPBUFR files store all stages of NCEP initialization quality control, so if quality control corrects a data error by changing a reported data value, the file contains a "stack" of multiple data values at the same level which record the changes made as a result of various checks (such as a hydrostatic check of reported heights). This project does not use DS337.0 because to identify instruments, it is best to use the data values originally reported by the stations, which are most easily retrieved from DS351.0.

In addition to DS351.0, several other sources of files of original radiosonde reports can be obtained. Upper air data files from Florida State University are in the most convenient format for processing and are archived for the past few days at http://www.met.fsu.edu/index.pl/wxdata/reports. Hourly dropsonde, pibal, and radiosonde files for each day starting 29 July 2008 have been downloaded and are concatenated into a single file for each day. The files are distributed by Unidata, and the same data in a slightly more bulky form (including a large number of reports from commercial aircraft) is archived by the State University of New York at Albany for about a year at http://www.atmos.albany.edu/weather/data1/archive/upperair and at Iowa State University back to 25 January 2005 at http://mtarchive.geol.iastate.edu (hourly files are in daily directories such as http://mtarchive.geol.iastate.edu/2007/10/07/text/upa). In addition, binary files at Texas A&M University formatted for GEMPAK, containing parts of the raw reports, were extracted from late July 2007 to early August 2008. The FSU files have been extracted as the primary source since 29 July 2008. Other sources have been used only if FSU data is missing and to fill in 2 missing dates in DS351.0 in 2007.

The data files above (DS351.0, FSU or other text files since 29 July 2008 and covering the 2 missing days in DS351.0, and the GEMPAK files from July 2007 to August 2008) are processed into a common format. Each sounding report stores the station ID, latitude, longitude, elevation, sounding nominal hour and date, launch time in hours and minutes, 31313 instrument code (see section 1.2), the source (currently "DS3510", "GTSdat", or "GEMPAK"), and the original text report. Dummy values are given if a particular variable is not available. For example, the latitude, longitude, and elevation are not available unless the data source is DS351.0, and the 31313 instrument code is not available if the report does not contain the 31313 group.

The next step in this project is to develop software to sort the common format files and eliminate duplicates, and then to extract the sounding data from the text reports into a uniform data format. Also, the DS353.4 extraction programs need to be revised to produce the same output format, and then the analyses can be performed in the same way on all data starting 1973.

The Integrated Global Radiosonde Archive (IGRA) project (Durre et al. 2006; http://www.ncdc.noaa.gov/oa/climate/igra/index.php; also see Gaffen 1996) combines DS353.4 from 1973 to October 1999 with 10 other data sets (including NCDC processing of GTS reports starting in January 2000). IGRA has many observations back to 1963 and a few observations as early as 1938 to 1946, but DS353.4 still accounts for over half of IGRA. IGRA is the most popular data source used by climate researchers because it is arranged as time series of soundings at individual stations, and because the data format is easy to read or process.

However, IGRA has some significant limitations. First, newly-established stations, including stations with new ID numbers which replace nearby closed stations, are added infrequently, so the decline in the the number of active stations since the peak in the early 1990s has been exaggerated. Second, IGRA does not contain data for ships (including fixed ships) or dropsondes, so the amount of oceanic data is significantly reduced. Third, IGRA omits stations or periods where the data is considered to have questionable quality, such as almost all dew points through 1970, almost all Chinese stations from 1973 to 1990, and most United States dew points from 1990 to 1992. However, Haimberger (2007) considers the data omitted from IGRA to be quite usable. Finally, IGRA data does not include the 31313 instrument codes reported by stations.

This project will eventually include additional sounding data sources, including field program and ozonesonde soundings that were not entered into the GTS but have been posted online, and recently-digitized soundings that existed only as paper records before data recovery efforts started in the last few years.

As mentioned above, soundings have historically been transmitted as text. Current reporting code formats are defined in the WMO Manual on Codes (WMO No. 306, available at http://www.wmo.int/pages/prog/www/WMOCodes/ManualCodesGuides.html) and have been nearly unchanged since 1968. A sounding is transmitted with up to 8 parts (with starting characters such as TTAA through TTDD and PPAA through PPDD) because data for lower levels can be transmitted even while the radiosonde is still collecting data in the stratosphere.

Reported soundings show evidence of random communication errors (missed characters or larger portions of reports, concatenation of part of one report to part of another report, or insertion of random nonalphanumeric characters), multiple modes of transmission (such as occasional html instead of plain text), human attempts at making corrections (such as the phrase “CHECK TEXT NEW ENDING ADDED” at the end of a truncated report), and human coding or typing errors (such as adding 40 or 60 instead of 50 to the date to indicate that wind speeds are reported in knots, interchanging characters, indicating missing data by \ instead of /, or repeatedly not filling in the same part of a code group). Frequently, only some parts of a coded sounding are received (such as TTBB and not TTAA), so the full detail of the reported sounding profile cannot be reconstructed. So, it is evident that the system of coding and transmitting data has a far higher error rate than would be expected if the system is fully automated, and the system is not fully automated even in the United States (“Fully automated” does not mean that there is no human intervention such as correction or deletion of erroneous values, but that the system checks for feasible data values and proper formatting of all manual inputs and changes when transmitting or receiving a message).

1.1a. BUFR and CREX data formats

The reason for discussing the observed data errors above is that WMO hopes that the weather data communication system will transition from text to BUFR in the next few years. As described in Dey, ed. (2002), a BUFR message starts with the ASCII characters “BUFR” and ending with “7777”, but all numeric data in the message is stored as nonnegative binary integers. Therefore, it would be extremely tedious for a human to attempt to read a BUFR observation, so a BUFR decoder is required.

BUFR converts most variables to SI units (such as temperatures reported in Kelvin), multiplies or divides some values by a power of 10 to provide the desired precision, and can subtract a “reference value” or offset so no values are negative or to reduce the number of bits required to store the full range of data values. For example, a wind direction from 000 to 360 degrees (361 possible values) can be stored in 9 bits. Variables are concatenated, so a variable may not start at a byte boundary.

The “identification” part of a message identifies the type of data and the number of observations included in the message. The “description” part contains 6-digit binary integers that refer to code tables in WMO Publication 306 and define the type of data, measurement units, scaling and other data compaction, and numbers of repetitions such as the number of reported data levels in a sounding. The “data” part is formatted according to the description section. The BUFR format specification was first approved in 1988 and has been unchanged since 1995, and each version is backward-compatible with earlier versions.

Using BUFR, it is easier to add or delete a data element, or to change its precision, than in the present WMO code table system because every BUFR message contains a description section that applies to the observations in that message. Also, new data values (such as new radiosonde instrument types) are added to the data description tables and old values are not deleted.

In theory, a BUFR decoder designed for the 1995 version, along with the most recent copy of the data description tables, can read any BUFR file ever produced. Other advantages of BUFR are that the file size is reduced as long as several or more observations are included in one BUFR message (a message with only one observation is slightly longer than a single report in text format because of the length of the description), conversion of integers to floating point values for processing is faster than converting ASCII to numeric values, and an observation can include metadata (such as station location and elevation, instrument type, serial number, and even quality control information) and report data values to a higher precision than usual as long as the description part of the observation is properly stated.

As of December 2008, NCEP does not receive radiosonde soundings in BUFR format (J. Ator, NCEP, personal communication, 2008), although it receives BUFR files containing large-volume array-structured data such as satellite and wind profiler data. NCEP converts the soundings into BUFR-formatted files, and uses the BUFR files for local operations and for archiving at NCEP and exchange with other agencies such as NCAR. NCEP performs the conversion from text to BUFR by storing the reported data values in BUFR format (as mentioned above, until October 1999, NCEP stored the reported data values in ON29 format). The BUFR format is quite compact, is likely to be adaptable into the indefinite future, and in addition, permits NCEP to store additional variables along with the data from the sounding itself for later retrieval and research. Specifically, DS351.0 contains station locations and elevations from the NCEP metadata file, quality control indicators for the data values, and also the original text reports (variable RRSTG). The original text reports are stored in their original ASCII form (but the 8 bits per character do not need to start on a byte boundary), so they compose about half of the BUFR radiosonde data files because they are not compressed. The facts that every sounding has an original text string, and that the radiosonde temperature variables conform to the WMO text reporting convention that negative temperatures are reported to odd tenths of °C and positive temperatures are reported to even tenths, confirm that NCEP does not yet process radiosonde data in BUFR format.

The reason why NCEP does not yet receive radiosonde soundings in BUFR format is that in practice, it is challenging for individual stations to produce observations in BUFR format. Apparently, a major obstacle is that human intervention and correction of errors in a BUFR sounding is almost impossible. First, each station needs a computer with software capable of running the BUFR encoding process. A general-purpose BUFR encoder and decoder, developed by NCEP, can be downloaded from the DS351.0 web site. It contains 209 subroutines, many of which must be extensively modified if this system is to be used to validate input data at a station before writing the sounding in BUFR format. In addition, the NCEP BUFR decoder operates relatively slowly on a workstation, probably because of the large amount of overhead involved in operating a general-purpose decoding program. Finally, each station needs direct access to a communication line with high reliability to transmit the file without corruption (at some stations, the observer hand-carries a paper copy of the coded sounding to another office so the observation can be manually typed into a communication system accessible to GTS). Since even NCAR has had some difficulties with the decoding system, the observed human and communications error rate in the present text-based system suggests that reliable coding and transmission of BUFR reports will be difficult to attain in many areas such as Africa and South America.

Dey, ed. (2002) also describes an alternate general code format called CREX (Character form for the Representation and EXchange of data) which may need to be implemented as an interim transition to BUFR. A CREX message is entirely in ASCII and is not substantially more difficult to read than the current WMO text formats, except for the added complexity caused by including the complete sounding in a single message (instead of separate parts such as TTAA and TTBB), along with additional data elements. CREX uses the same data descriptions and tables as BUFR as far as possible, so a program to convert between CREX and BUFR should be quite compact. One specific anticipated improvement that is relevant for this research is that either CREX or BUFR could include full station and instrument metadata with each observation, which would make it possible to automatically generate and update station history tables.

1.1b. The global radiosonde network in 2007 and 2008 based on archived data

A preliminary evaluation of the data from 2007 to 4 January 2009 was performed using DS351.0, GTS files starting 29 July 2008, and GEMPAK-formatted GTS files from July 2007 to August 2008. Wrong station IDs were eliminated and all dropsondes were called "DRPSND" (the original text dropsonde reports have no station ID, and NCEP constructs a station ID from parts of the dropsonde sounding). In this period, there were 1089 numeric land IDs (plus 4 numeric IDs assigned here for low level airplane observations in California) and 54 alphanumeric IDs (DRPSND as a single dropsonde identifier, 48 ship IDs, and 5 temporary field experiment sites on land).

Out of these 1147 different stations, 89 have observations only in DS351.0, and an additional 36 stations appear very seldom in GTS files, so apparently NCEP has access to additional data streams that are rarely or never available through Unidata. The stations only in DS351.0 are mainly in South Asia, Africa, South America, United States military stations, and NATO military stations in the Middle East.

However, 14 stations in GTS data are not in DS351.0 because they are not in the NCEP upper air station catalog (see section 1.2), but 11 of these stations have been identified elsewhere, and only 3 stations (17516, 74005, and 74006) have not been found in any current or historical data catalog by January 2009.

Since NCEP changes some data values during initialization, there was a concern that data values in DS353.4 and DS351.0 could contain corrections from NCEP instead of the original data reported by the stations. The reason for desiring to use the original data is that when a station changes an instrument type, a discontinuity occurs in the average data values if the instrument characteristics differ. However, if NCEP applies corrections, NCEP is not likely to know about a changed instrument type until after the station has changed the instrument. Since they apply the wrong correction until they update the catalog, an adjusted data base would contain many additional discontinuities that indicate when NCEP changes a correction, not when the station changes its instrument, and it would be impossible to produce accurate metadata from examining the data.

Fortunately, based on comparisons with other data, it appears that DS353.4 and DS351.0 contain the original data values transmitted by the stations, except for a period from 1997 to 1999 when some temperature and dew point values in DS353.4 were changed by 0.1° in either direction (this was noticed because some unusual values were found, such as temperatures below freezing reported as even tenths of degrees). While NCEP has produced the files that NCAR archives as DS351.0 since early 2000, NCEP apparently used BUFR internally in the late 1990s, and it is possible that these erroneous values in DS353.4 were rounding errors in the conversion from Celsius to Kelvin and back. Since these errors are very small and are in both directions, they do not interfere with detection of discontinuities.

1.2. Upper air observation metadata sources

The WMO text reporting format for land stations with numeric IDs (report types starting with TT or PP) does not provide for reporting of station locations or elevations, so that information must be provided by a catalog. Other sounding types (ship reports start with UU or QQ, dropsonde reports with XX, and mobile land station reports with II or EE) specify the location to the nearest 0.1° of latitude and longitude, which is not sufficient accuracy for some purposes. Only mobile land reports specify the surface or launch elevation.

Some major sources of current and historical information are as follows. This is not a complete list of the sources that have are available. Some catalogs copy from each other, so the original source of a particular station ID is often unknown.

(1) The *WMO Catalogue of Radiosondes and Upper-Air Wind Finding Systems in Use by Members* is available at http://www.wmo.int/pages/prog/www/ois/ois-home.htm and should be the official source of detailed station and instrument metadata for every operational upper air station. However, it is updated infrequently (in January 2009, the latest edition is dated July 2007) so recent changes are not included. That edition lists 1525 stations, and the data set described above contains observations from 1025 stations, so 500 listed stations had no observations in this data set. Also, the data set contains observations from 63 stations with numeric IDs that are not listed in this catalog, and not all of those stations are either new (starting operations after July 2007) or closed (ending observations before July 2007).

The same web address also provides a global WMO surface and upper air station list (Publication 9A) through the "WMO No. 9, Volume A - Observing Stations" link. This list is updated nearly every week and provides for separate entries for upper air stations, but out of 11,743 station entries in the 8 December 2008 version, there are only 93 separate upper air station entries (giving 11,650 catalogued numeric IDs assigned), not all at a different location or elevation than the surface observing site. Also, of the 63 stations with upper air observations not found in the July 2007 version of the WMO upper air catalog, 42 station IDs do appear in the latest Publication 9A, and of these, 18 stations have some reference to upper air observations, including irregular or suspended operations.

WMO Operational Information Newsletters (http://www.wmo.int/pages/prog/www/ois/Operational_Information/Newsletters/) report changed or new stations, but countries often do not report metadata to WMO so these newsletters are an incomplete guide to changes since the last upper air catalog edition.

(2) As mentioned above, NCEP omits observations from stations that are not in the NCEP station catalog. A gap in DS353.4 or DS351.0 when one station closes and is replaced by a nearby station with a different ID may reflect a delay in NCEP updating of the station catalog, rather than a real shutdown of observations.

NCEP catalog files are available at http://www.nco.ncep.noaa.gov/pmb/codes/nwprod/dictionaries/, and the files of interest are sonde.land.tbl, sonde.ship.tbl, lsfc.tbl, and metar.tbl. These may not be the latest versions used internally by NCEP (in January 2009, all files are dated May 2007 or earlier) because DS351.0 contains some observations (with locations) from stations that are not in the May 2007 files. Apparently NCEP only checks the sonde.land.tbl file before deciding to include an observation in DS351.0 because some stations which are in GTS data but are omitted in DS351.0 are listed in other NCEP catalog files.

(3) Each sounding is supposed to report a data group starting with 31313 that includes a 5-digit instrument type code (called the "31313 code" here), which in theory allows the construction of a complete instrument history. The 31313 group has been reported in some United States soundings since 2 March 1989, by many countries since early 1992, and now by almost all countries except India and China (Indian soundings often include a 31313 code of "/////").

The 31313 code is a 1-digit solar and radiation correction code (WMO-No. 306, Vol. I Part B, BUFR Table 0 02 013, or Code Table 3849), a 2-digit radiosonde and ground equipment code (BUFR Table 0 02 011 or Code Table 3685), and a 2-digit tracking (wind finding method, such as radar or GPS) or sounding status code (BUFR Table 0 02 014 or Code Table 3872). The most recent radiosonde type list is at http://www.wmo.ch/pages/prog/www/WMOCodes/Operational/CommonTables/BufrCommon-11-2007.pdf. All of the initial 2-digit radiosonde and ground equipment codes from 01 to 99 have been assigned, while BUFR allocates 8 bits, allowing values up to 255. New radiosonde codes are allowed for text-based reporting by reassigning only obsolete 2-digit codes, and adding 100 to the reassigned 2-digit code. For example, the original instruments with code 10 (VIZ Type A) are obsolete, so a sounding after 2007 with instrument code 10 implies a new instrument type (Lockheed Martin Sippican LMS5) with a full BUFR code value of 110 (and the full BUFR code value for the old instrument would be 010 for uniqueness).

The 31313 code has limitations which make instrument identification inexact.

(a) The solar and radiation correction code is generic (such as "country solar correction") so the 5 different Vaisala RS80 corrections are not distinguished.

(b) The radiosonde type does not specify varieties such as Vaisala RS80 A-Humicap and H-Humicap, which had differing dry biases from packaging contamination (Wang et al. 2002). Usually, stations report code 90 ("unknown instrument") if no code is assigned, but at least one informal code reassignment occurred when French stations reported code 34 (Czechoslovakia Vinohrady) starting November 2004 to refer to Vaisala RS92 with a STAR ground station (confirmed in the 2006 WMO upper air catalog).

(c) The wind finding code is also generic and sometimes a code value such as "systems operating normally" is reported instead.

(d) Finally, tables do not give any references or further explanations about code values, so codes are sometimes used incorrectly. For example, some countries report the code for "no wind finding method," but interpret this to mean that the radiosonde does not determine the wind, while the wind is actually obtained by radar tracking of the balloon.

(4) The IGRA metadata file at http://www.ncdc.noaa.gov/oa/cab/igra/ index.php?name=metadata is intended to be a comprehensive global radiosonde station and instrument history for the IGRA stations back to the beginning of upper air soundings at each station. It is a text file based on the Gaffen (1996) metadata file, which was based on documentation at NCDC, 5 WMO upper air catalogs from 1965 to 1993, the Gaffen (1993) project which analyzed responses from about 40 countries to a questionnaire, and some personal communications. Since the purpose of the Gaffen (1996) project was to show what information was and was not available, the station histories are incomplete when documentation is not available. Station location information is even more limited than instrument information because the date of a station move is rarely documented in published sources. The current available IGRA metadata file was prepared in early 2006 based on limited updates received at NCDC since 1996, so only a few stations have history information from the last 12 to 15 years.

(5) Copies of the Master Station Catalog from the US Air Force Air Weather Service (now Air Force Weather Agency) from 1967 to 1999 are in NCAR Data Set 900.0 (http://dss.ucar.edu/datasets/ds900.0). Two consolidated files were prepared by combining multiple editions, and displaying each different line for each station name. The consolidated files are helpful for determining approximately when stations moved or changed their name. As with other catalogs, there is an unknown lag time from when a station change occurs and when it is recorded in the catalog. The reporting format allows for separate surface and upper air observing sites, and a general radiosonde type code is listed at many stations. The instrument type list is helpful for indicating when certain instrument models were in use in the late 1970s through the 1980s, and this list is the only documentation that confirms the existence of some radiosonde models.

(6) The British Atmospheric Data Centre has a catalog of upper air stations at http://badc.nerc.ac.uk/data/radiosglobe/world.html, including an extensive list of "unknown" stations, for which at least the elevation is not known. The United Kingdom Meteorological Office archive of global radiosondes back to 1997 is one of nearly 200 data sets, including many field experiment data sets, archived at http://badc.nerc.ac.uk/data/ (The global radiosonde data is stored under the "Met Office - Global Radiosonde Data" link).

(7) Metadata described above has little or no information about ships that have launched radiosondes. Ship soundings generally have reported the call sign (usually 4 or 5 characters) of the ship from which the radiosonde was launched. A recent list which includes ship names, countries of registration, and call signs is available from http://www.wmo.int/pages/prog/www/ois/ois-home.htm at the "WMO Publication 47 - International List of Selected, Supplementary, and Auxiliary Ships" link. Older editions from 1955 to 2003 are available at http://icoads.noaa.gov/metadata/wmo47/. In the past, knowing the country of the ship was a useful clue to the type of radiosonde used, but all of the ship stations in since 2007 in the data described in section 1.1 regularly report the 31313 instrument code. In addition, many ship observations are now made using Automated Shipboard Aeronautical Programme (ASAP) launchers, which are containers that can easily be moved from one ship to another, and most ASAP launchers now use a generic call sign for that particular launcher (such as ASDE07 or ASUK01, where the first 2 characters indicate ASAP, the next 2 characters are a 2-letter country abbreviation or EU for a general European unit, and the last 2 numbers indicate the sequence number of the ASAP unit owned by that country or agency). Since the radiosonde type is regularly reported, the remaining metadata variable that needs to be obtained elsewhere is the launch height, which can vary as the ASAP container can be placed at a different elevation above sea level on each voyage. The launch height can be computed hydrostatically by computing the thickness downward from the first reported above-surface height.

(8) While most of the large-scale lists of radiosonde stations or instrument types are in the sources above, many pieces of information about individual stations, groups of stations, field programs, instrument types, operational procedures, and errors and methods of correcting these errors are found in papers, reports (including military research and operational reports), field program documentation, and web sites (including web sites of radiosonde manufacturers, government agencies, and research organizations that operate upper air equipment).

1.3. Indirect methods to identify and correct biases

Determining unbiased trends in data is a two-step process. The first step is to construct or obtain complete metadata, and the second step is to develop data adjustments, but many researchers address the second step even though the first step is still quite incomplete. Most projects summarized below use Gaffen (1996) or the IGRA metadata as their source of radiosonde station and instrument metadata.

Documented methods used to attempt to determine climate trends include the following:

(1) 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.

(2) Make no adjustments based on the argument that the number of discontinuities in each direction seem close to equal (Schroeder and McGuirk 1998). Even if the discontinuities largely do cancel out in a specific data set, this is not a rigorous approach and the trends will probably be distorted in some way.

(3) Adjust for a single change or small number of known instrument changes (Parker et al. 1997). While that paper adjusts for specific instrument changes around Australia, the lack of adjustments elsewhere obviously would not correct erroneous trends in those areas.

(4) Use an automated algorithm to detect and remove discontinuities that meet a specified threshold (Gaffen et al. 2000). This approach was found to remove essentially all trends, regardless of the tuning of the level of sensitivity.

(5) Remove detected discontinuities only if they approximately coincide with known transitions (Gaffen et al. 2000, Lanzante et al. 2003). Trends from this approach appear mostly reasonable, but there is no assurance that the adjustments are correct for the "right" reason because they are subjective and because it is known that there are many gaps in available metadata.

(6) 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 2008) have used GZZ-2 radiosondes (based on the Russian A-22 design from the mid-1950s) since 1964.

(7) Compute the trend using "first differences," or the difference from one year to the next in each month, after deleting data at individual stations around each suspected discontinuity (Free et al. 2004). The first differences are area-averaged, and are then summed over time (the first difference in the first year is zero) to produce a time series which retains the trend of the accepted data. This method does not produce a time series for any individual station, and the decisions to delete data are subjective.

(8) Compare radiosonde minus satellite tempera- ture retrieval time series, where radiosonde data is averaged in altitude bands corresponding to the weighting function of each satellite channel (Randel and Wu 2006, Christy et al. 2007). Various approaches are used to make comparisons and detect and compensate for discontinuities. Of course, the satellite record is also not homogeneous, so discontinuities when satellites change need to be accounted for.

(9) Adjust daytime temperatures to be equivalent to night temperatures (Sherwood et al. 2005) so the time series of day minus night temperatures has no trend. This approach assumes that radiative errors at night are zero, and does not correct errors other than radiative heating. At 50 and 300 hPa, the authors find daytime cooling relative to night temperatures, with the largest trends in longitudes where observation 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. Christy et al. (2007) use this approach by separately comparing day and night radiosonde soundings with satellite retrievals.

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

(11) Use radiative theory to construct adjustments based on sensor shapes and materials. Luers and Eskridge (1998) developed theoretical "temperature correction models" of the radiation and lag errors of major radiosonde types. Durre et al. (2002) applied these corrections to archived data where transitions appear well-documented and found that many discontinuities became larger. It is possible that the main reason for the apparent failure of the corrections is that the originating station or country adjusted the observations before transmitting them. The original adjustments may have undercorrected or overcorrected the actual instrument errors, and the residual errors which Durre et al. (2002) attempted to correct might not have characteristics that resemble the modeled radiation and lag errors.

(12) Simultaneously estimate trends, change points, and natural variability using a procedure called "iterative universal Kriging" or IUK (Sherwood 2007). In theoretical tests, no variation of the sensitivity parameter used in this method resulted in a hit rate above 45 percent (a "hit" is detection of an instrument-caused discontinuity), and almost all variations had a false detection rate above 50 percent (a "false detection" is considering a natural variation to be caused by an instrument change). Certain sensitivity settings still tended to produce nearly-correct trends because the errors from undetected transitions and false detections tended to cancel out.

(13) Apply a statistical test to radiosonde minus model initialization ("background forecast") temperatures at each station to identify and adjust for discontinuities (Haimberger 2007). As with Thorne et al. (2005), undocumented change points are detected and adjusted.

(14) Develop adjustments for a radiosonde type based on laboratory tests (Wang et al. 2002). This paper quantifies 6 errors of the Vaisala RS80-A and RS80-H radiosondes as a function of the reported relative humidity and the radiosonde age (years since manufacture). While relevant tests could in principle be performed for other radiosonde types, in practice it would be difficult or impossible to perform the needed tests on radiosondes no longer in production, and even the most

extensive laboratory tests cannot fully detect and quantify all in-flight biases.

Since adjustments are mostly subjective, a method is considered "successful" if the trend is close to the expected magnitude or at least is close to results from another approach. For example, Lanzante (2007) concludes that “adjustment seems to unambiguously improve agreement with the climate model.” However, these trends are still questioned, and it is not known whether the errors are undercorrected or overcorrected even if the adjustment methods tend to produce the “expected” results. Actually, almost any method can produce an “expected” result. As stated in Sherwood (2007), if a method uses only the radiosonde data to derive adjustments, the adjustments tend to remove some of the observed trend, while if a method compares radiosonde data to a separate data source (such as satellites or model output), the adjusted radiosonde trend tends to approach the trend of the separate series.

If complete metadata makes the nature and timing of all transitions known, methods such as those listed above can be applied much more confidently because false detections would be nearly eliminated. The main remaining potential error is that if a natural variation occurs at or near the time of an instrument change, some of the natural variation might be considered to be part of the effect of the instrument change. If an adjustment method develops corrections for instrument types rather than only for individual stations, that problem is diminished only if the instrument transitions occur at different times at individual stations.

1.4. Summary of approach to produce complete metadata

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

The usual approach to fill in gaps in historical metadata or to obtain updated information is to request the relevant information from the agencies making radiosonde observations. This is a slow process, as indicated by the difficulties experienced by WMO in attempting to obtain timely updates from each country.

An alternate approach is to develop complete metadata by detecting discontinuities in the archived data itself. Hypothetically, instrument changes can be identified by data discontinuities, but other researchers have not achieved much success with this approach because in variables of research interest, primarily temperatures at specific levels, individual discontinuities caused by instrument changes are often not obvious enough to be confidently distinguished from natural variations, even though the cumulative effect of all discontinuities is clearly large enough to greatly contaminate all trends.

This research more successfully completes metadata by combining both approaches. Specific techniques are described in detail in Schroeder (2007).

First, the references in the IGRA metadata file and in Gaffen (1993, 1996) lead to further references that contain a large amount of additional station location and instrument metadata, which is being consolidated into a single metadata file.

Second, instrument-related discontinuities are quite visible in especially sensitive variables that amplify differences between instruments. Other researchers have not studied these variables because they have little or no research interest by themselves. At stations with accurate metadata, consistent data characteristics involving combinations of sensitive variables can be attributed to specific instrument types, and discontinuities (usually involving multiple variables at the same time) indicate instrument transitions. Similar signals and discontinuities at a station or in a period without metadata allow inference of the use or change of a specific instrument type. The same methods are used to examine time series of the same variables at all stations, with or without documentation. So, the available metadata is validated based on consistency with the data, erroneous metadata is detected, and missing metadata is constructed.

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

While some metadata must be available as a starting point to identify instrument types, with even partial metadata it is almost always 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 that might have been used is greatly narrowed down. If this method does not correctly identify the exact instrument type at a station, the error is because the wrong instrument type is similar to the correct instrument, and because an error does not affect other stations or periods.

The discussion above is mostly concerned with developing instrument signals to identify changes in instruments. In addition, limited checking of location and elevation metadata is possible.

Locations are verified primarily by intercomparisons of different catalogs. For many stations, discrepancies between metadata sources are so numerous that it is not likely that the actual location has been identified, even to the limited accuracy stated in catalogs (the nearest minute or .01° of latitude and longitude). In the past, accurate locations were not needed for either forecasting or climate purposes because the radiosonde drifts with the wind away from the launch location, and because an upper air observation is considered to be representative of a large area. However, forecast models now have high enough resolution that a slightly misplaced sounding may have a noticeable effect on the forecast, and some models can properly account for both the time lag and position drift as a radiosonde ascends. So, catalogs are gradually emphasizing more accurate station locations. This means that many station location changes appearing in catalogs are probably corrections of locations rather than actual moves.

Recent station locations often can be verified using online satellite photos at web sites such as Wikimapia (http://www.wikimapia.org). In some countries, an upper air station tends to have a distinctive appearance, such as the 4.5-meter dome at most United States stations that originally housed a tracking radar. Some countries document their station locations with enough precision that some or all stations can be located on the online satellite photos. Most of the current online satellite photos were probably taken between 2004 and 2006, although individual photos are undated.

Surface elevations can be checked quite accurately. A complete radiosonde observation can be used to compute the surface elevation hydrostatically by computing the thickness of the layer from the first reported above-surface height to the reported surface pressure. A coded observation reports the surface pressure to the nearest hPa, so elevations computed from individual observations fluctuate by several meters, but an annual average should be accurate within 1 meter in almost all cases.

Strictly speaking, the surface elevation is the radiosonde release height in geopotential meters from sea level (C. Bower, NOAA NWS, personal communication, 2008). This should be the actual release height, not a historical elevation for climate continuity, because otherwise other levels would need to be adjusted to avoid distorting the reported lapse rate profile.

An isolated different computed elevation is probably caused by an error in the surface pressure or reported height, or by a missing level in the data, or by a wrong station ID. However, if the computed elevation varies consistently within a range of a few meters, the average elevation is the elevation of the surface level of the sounding. If the computed elevation differs from the metadata elevation, the metadata elevation is assumed to be wrong. An elevation change of 5 meters or more can usually be identified to the exact observation, while a change of 1 or 2 meters or so can usually be identified to within a few weeks.

It was hoped that a change in elevation would allow determination of a horizontal move, but often the metadata elevations do not match observed elevations accurately enough to confirm the locations associated with each actual move. For example, a computed elevation might change from 23 to 28 meters at an accurately determined time, but metadata in that period may show different reported locations with elevations of 25, 22, and 26 meters. The station might have moved from the 25-meter location to the 26-meter location at that time, or from the 22-meter to the 26-meter location. An additional complicating factor is that WMO catalogs allow a station to specify 2 elevations for each site, and computed elevations often do not correspond closely to either reported elevation. So, the computed elevations in the metadata produced in this project are quite accurate, including the dates of changes, but locations are often questionable.

Development of complete metadata using this approach is still in progress, but instrument histories are nearly complete back to 1973 or earlier in Japan, Australia, the Russian Federation, China, and India.

2. SUMMARY OF FORMER SOVIET UNION RADIOSONDE HISTORY

In this paper, stations investigated from the former Soviet Union include land stations with WMO IDs from 20000 to 39999 (204 stations with temperature data in DS353.4 or DS351.0 since 1973, 127 of which were operating since 2007), 33 ships operated by these countries (only 1 ship still operates in 2008), 6 stations in Antarctica (2 are operating since 2007) and 6 drifting Arctic ice islands between 1973 and 1991. The 204 land stations in the former USSR include 12 cases where a station moved a short distance and changed its ID number, giving 216 different station IDs. The number of stations has increased slightly from the minimum in the late 1990s and early part of this decade, with 7 stations resuming soundings in 2007 or 2008 (20046, 20292, 23933, 31770, 32215, 34172, and 37860) after a gap of 4 years or more, 1 new station (26435) starting in 2006, and 1 ship (UFTA, or Volgoneft-131) starting in 2008.

Other countries in eastern Europe, southeastern Asia, and Cuba have used instruments from the former Soviet Union, but these stations have not yet been checked in much detail, so they are not included in this analysis. This analysis also excludes stations making only pibals and stations that closed before 1973.

The Russian instrument history is often considered problematic (e. g., Lanzante et al. 2003), but data signals indicate that very reliable instrument histories can be developed.

Gaffen (1996) has detailed instrument histories for almost all Russian land stations, but a few station histories are omitted. Major instruments from 1973 to the late 1990s are the A22, RKZ, MARS, and MRZ series. They are described in Zaitseva (1993) and specimens are in the NCDC Weather Museum. A22 used a bimetal thermometer, other models use a thermistor, and all models use a goldbeater's skin humidity sensor that appears unchanged except for its position.

Temperarure and RH differences between series are small due to minor sensor differences and apparently effective corrections. Consistent reporting policies can distinguish series, but models in a series, such as MARS-2-1 and MARS-2-2, are indistinguishable. A22 soundings report no significant wind levels except some tropopause and fastest wind levels, few significant temperature levels, and DPD usually at all levels. Elevation changes are fairly frequent with transitions from A22. Most RKZ soundings report some significant wind and temperature levels. MARS soundings report DPD only to a temperature around -40° so they are often exactly identifiable. MRZ is almost identical to RKZ. Examples of the degree of consistency of some of the signals are discussed below.

At 202 USSR stations with extensive DS353.4 data, Gaffen (1996) lists a change to MARS at 147 stations but not at 50 stations, and 5 stations have no stated history. At stations listing a MARS transition, DPD reporting to a temperature about -40° occurs in most or all soundings at 129 stations (starting within 1 month of the listed date at 87 stations and within 2-6 months at 19 stations), and in a few soundings at 14 stations (within 1 month of the listed date at 6 stations and within 2-6 months at 5 stations), but does not occur at 4 stations (implying no use of MARS). The earliest DPD reporting to a temperature about -40° was in August 1983, matching the MARS introduction year in Gaffen (1993, p. 76). These patterns appear sufficient to identify DPD reporting to a temperature about -40° as a signal of MARS. Most stations transition gradually based on frequent alternations of reporting policies.

At 55 stations with no reported MARS transition or no instrument history, this DPD reporting policy begins between about 1984 and 1992 in most soundings at 23 stations and in a few soundings at 3 stations (at these stations, the use of MARS is inferred), but does not occur at 29 stations (implying no use of MARS radiosondes).

Gaffen (1996) lists transitions to MRZ at 98 stations, 96 of which show signatures of MRZ (DPD reporting at all levels, and some significant wind levels). No Russian instrument changes after July 1991 are reported, but 69 other stations have signals of transitions to MRZ (42 transitions after July 1991). A direct transition from RKZ to MRZ should be hard to confirm due to similar data signatures, but only 9 stations report that transition. Based on data examination, 7 of those stations changed to MARS and then MRZ, 1 station was closed 12 years and reopened using MRZ, and the last station appears to have transitioned from A22 to MRZ. Most stations frequently alternate reporting policies, indicating gradual transitions to MRZ.

Reporting policies are consistent in over a million soundings, so the probability that signals are coincidental and not associated with instrument changes is negligible. If reporting DPD to a temperature about -40° had been only a policy, it could have been implemented within a few days in the whole network, and a station would be very unlikely to switch reporting policies repeatedly.

In some cases, changes in computed surface elevations help support a transition from A22 to another instrument. For example, the elevation at Almaty (36870) alternates frequently from September 1984 to January 1985. At Ostrov Preobrazheniya (21504), the elevation changes from 35 m to 60 m in 0000 UTC soundings on 1 December 1990 and in all soundings on 1 April 1991. In both cases, a site with the same ID but a different elevation and radiosonde type opens, and the existing A22 site continues to operate for several months before closing. At 36870, Gaffen (1996) says A22 changes to MARS in January 1985, but DPD reporting continues at all levels, so RKZ radiosondes are inferred (not MRZ, which was not introduced at any station until mid-1986). At 21504, Gaffen (1996) says A22 changes to MRZ in December 1990, which agrees with data signatures.

While Gaffen (1996) and the IGRA metadata file have no Russian metadata since the mid-1990s, most stations in the Russian Federation have routinely reported 31313 codes since between 1995 and 1998. Only one station (36003) does not report 31313 codes in 2008.

Stations reporting 31313 codes for MARS or MRZ since about 1997 were checked for consistency with the hypothesized DPD reporting policies. Signals are "consistent" if DPD is reported to a temperature about -40° with MARS or at all levels with MRZ. At 26 stations using MARS, only 2 stations generally and 2 stations briefly are inconsistent. At 125 stations using MRZ, 13 stations generally and 2 stations briefly are inconsistent, and only 7 stations are inconsistent after 2003. At 13 stations using both MARS and MRZ, signals are always consistent at 9 stations, inconsistent for a few MARS cases at 2 stations, and inconsistent for MRZ at 2 stations (only until 2003 at 1 station). In Gaffen (1993, p. 76), manually processed MARS reports DPD to a temperature of -40°, implying that MARS with automated processing reports DPD at all levels. No reason for DPD reporting to a temperature of -40° with MRZ has been found. Discrepancies might simply be incorrect 31313 codes. In any case, confidence is over 90% that MARS and MRZ are identified correctly.

Recent Russian radiosonde history is complex, with 11 codes for Russian-manufactured instruments in use: 27 (AVK-MRZ), 28 (Meteorit Mars-2-1), 29 (Meteorit Mars-2-2), 53 (AVK-RF95), 58 (AVK-MRZ*), 68 (AVK-RZM-2), 69 (MARL-A or Vektor-M-RZM-2), 75 (AVK-MRZ-ARMA), 76 (AVK-RF95-ARMA), 88 (MARL-A or Vektor-M-MRZ), and 89 (MARL-A or Vektor-M-MRZ*). MRZ, Mars, RF95, RZM-2, and MRZ* (BAR before 2006) are radiosondes. AVK, Meteorit, MARL-A, and Vektor-M are radars or radiotheodolites, each of which is compatible with a specified set of ground equipment. ARMA is an optional workstation to upgrade the AVK radar system (Kats and Grinchenko 2006).

RF95 has Vaisala sensors with quite dry readings (see section 3), and is not widely used because of cost (code 53 has not appeared in soundings after June 2007, and only a few stations report code 76). RZM-2 (codes 68 and 69) has not yet been used in large quantities, and MARS (codes 28 and 29) is still used only at a few stations. The first reports of code 88 were in September 2005 and codes 58, 68, 69, and 89 were first reported in February 2007. Code 58 is very widely reported in late 2008 and early 2009. Based on early observations with each new instrument type, MRZ* is as moist as MRZ, and RZM-2 is only slightly drier than MRZ, but with a substantial variation in dryness between stations.

In addition, 2 stations (22113 and 22217) occasionally reported instrument type 09 (no radiosonde) in early and mid-2008, which is incorrect because the observations are complete. Also, 5 stations in the former Soviet Union outside Russia report using instrument types 71 (Vaisala RS90 with DigiCORA I or II), 79 (Vaisala RS92 with DigiCORA I or II), and 80 (Vaisala RS92 with DigiCORA III).

The complete 31313 instrument code has 5 digits. For all instruments except Vaisala, the only correct 5-digit codes have radar or radiotheodolite windfinding with ranging (the last 2 digits are 03). Vaisala soundings have a wind finding method of 06 (LoRAN) or 08 (GPS), except for the only operating ship, which reports last 2 digits of 26 (an unassigned value).

The solar and radiation correction method code for Vaisala soundings is 4 (automatic solar and infrared radiation correction). Other radiosonde types except MARS usually report a correction of type 5 (automatic solar radiation correction), but RF95 often reports a correction code of 4, probably because it uses Vaisala sensors. For MARS instruments, correction type 7 (country solar radiation correction) is reported, or type 0 (no correction) in some MARS-2-1 soundings, or type 3 (CIMO solar radiation correction, where CIMO is the WMO Commission on Instruments and Methods of Observations) in some MARS-2-2 soundings. Some MRZ (instrument type 27) soundings also report corrections of type 0 or 3.

As mentioned in section 1.2, the codes for solar and infrared radiation corrections are not specific, so it is possible that a code number (such as 5) does not indicate the same correction for all radiosonde models, or for all soundings with the same instrument model, or for all soundings with the same ground equipment. Different odd-numbered codes might not really indicate different radiation corrections because codes 3, 5, and 7 mean “CIMO,” “automatic,” and “country” solar and infrared radiation codes, respectively. Each of these labels could be applied to the same radiation correction since they are named but are not specifically defined in the code table.

To summarize, these stations have reported 15 instrument type codes and 22 different 5-digit codes since the beginning of 2007. The most widely used codes are 52703 and 55803. Other codes used slightly less often are 57503, 58803, and 58903. The remaining codes (02703, 32703, 02803, 72803, 32903, 72903, 55303, 56803, 56903, 47603, 57603, 50903, 47106, 47908, 48006, 48008, and 48026) are not widely reported in the former Soviet Union, but all or almost all reports at an individual station may use these codes.

Differences in data behavior between instruments have been found in other countries where any digits of the 31313 codes differ, so these 22 codes need to be investigated as up to 22 different types of radiosondes. Sometimes the differences are unexpected, such as if a radiosonde with GPS windfinding (last 2 digits 08) is drier than a radiosonde of the same type with LoRAN windfinding (last 2 digits 06). As mentioned in section 1.2, instrument codes are not precise so in that case models with different windfinding methods may also have different humidity sensors.

Many stations have very complex instrument histories, based on the 31313 codes reported. Since 2007, 9 stations have reported 5 different codes, 5 stations (22113, 22217, 23955, 26063, and 28275) reported 6 different codes, and 1 station (23804) reported 7 different codes. Before stations reported 31313 codes, complex histories were still inferred based on frequent alternations between reporting policies that distinguish instrument types. Even though the transition from A-22 that occurred at some stations in the mid-1970s is based on farily weak signals, there are indications that the A-22 transition did not occur suddenly at most stations. So, any researcher who hopes to homogenize the radiosonde record needs to use an adjustment method that can accommodate more than one instrument type used at each station for possibly several years.

One challenge in developing histories is that 2 different MRZ models use code 27 (or code 75 if the AVK ground system is augmented with an ARMA workstation). Some stations are moist, but other stations are moderately dry as early as August 1996, so a separate instrument was inferred before documentation was found. Later documentation confirmed the 2 different models (Balagourov et al. 2002). The moist model is MRZ-3A, used since 1986. The moderately dry model is MRZ-3AM with a DVR capacitive humidity sensor. As mentioned in the next section, some stations using the RF95 radiosonde may also have reported code 27 before WMO assigned codes 53 and 76 to RF95.

A specific "problematic" example of Russian station histories should be discussed. Lanzante et al. (2003) check 11 USSR stations and assign a warming breakpoint in 1979 and cooling in 1987 from 250-700 hPa at 5 stations including Pechora (23418). In Gaffen (1996), Pechora changes from A22 to RKZ5 in November 1976 and to MARS in November 1984. In DS353.4, significant wind levels start on 8 September 1978, indicating that RKZ5 begins almost 2 years later than listed. DPD reporting to a temperature about -40° starts in some soundings 24 May 1984 (indicating mixed RKZ5 and MARS), and in all soundings 20 November 1984 (MARS only). The 1979 breakpoint in Lanzante et al. (2003) is the first full year of inferred use of RKZ5, but no instrument discontinuity is found in 1987. At the other 4 stations with the same breakpoints, only Turuhansk (23472) has a similar history, introducing RKZ5 on 1 January 1978, MARS on 14 February 1986, and some MRZ starting 23 October 1990. Omsk (28698) introduces RKZ5 in 1972, MARS on 27 March 1986, and some MRZ starting 7 April 1987. The other 2 stations use A22 until late 1990 and have no instrument changes in 1979 or 1987.

Because the discontinuity is not related to station instrument histories, it is likely to be a natural variation. While Lanzante et al. (2003) check only 11 stations, checking all stations in the former Soviet Union should further confirm whether their apparent discontinuities are natural variations, especially if the “discontinuities” are somewhat gradual and if they have a coherent spatial structure.

To summarize the Russian upper air history, Gaffen (1996) is mostly but not entirely complete or accurate. Sensitive variables can be used to construct nearly exact station histories, which are often quite complex due to frequent alternating instrument types. Because of the alternating instrument types, all of the adjustment procedures summarized in section 1.3 would be expected to have difficulties in determining the timing and magnitude of instrument-related breakpoints.

3. PRELIMINARY ASSESSMENT OF INSTRUMENT CHARACTERISTICS

Only limited analyses have been performed to compare characteristics of radiosondes used in the former Soviet Union, especially for radiosonde types (MRZ* and RZM-2) that were first reported in February 2007. The preliminary analyses did not show significant temperature differences between Russian models, probably because all models starting with the RKZ series have used a thermistor on an outrigger. The A-22 series, designed in the 1950s, used a bimetal thermometer in a duct, but it did not show significant temperature differences in the lower stratosphere even as the sun angle varied, so the radiation correction applied at the time that model was used appeared to be effective. However, more extensive analyses are planned.

Because temperature differences seemed small, these preliminary analyses performed analyses of humidity data as if there were no temperature differences. When corrections are specifically developed as in section 1.4, temperature adjustments need to be developed and applied first, but it is not likely that the temperature adjustments will change the basic changes in humidity characteristics of each instrument type.

As documented in Zaitseva (1993), differences within an instrument family after 1967, such as between RKZ-2 and RKZ-5 or between MARS-2-1 and MARS-2-2, did not involve changes in sensors, but only changes in electronics needed to send signals to different tracking radars. In addition, the goldbeater's skin humidity sensor was the same in all models from A-22 through MRZ, and the MMT-1 rod thermistor was the same in all models from RKZ-2 through MRZ. The differences between these models should therefore be small, but not necessarily zero because the sensors are mounted differently on each model. Almost no documentation has been found for the MRZ* and RZM-2 models except for a statement that they are called "MRZ clones" in an unlabeled presentation from mid-2006 at http://www.oco.noaa.gov/docs/GCOSUAWII/pres_Russia_UAW.ppt, and that the MRZ* has a goldbeater's skin humidity sensor while the prototype RZM-2 has a capacitive humidity sensor, which is larger than the Vaisala humidity sensor, and the RF95 has both a Vaisala RS80 bead thermistor and capacitive A-Humicap.

Section 2 has mentioned instrument characteristics in terms of data reporting, primarily the number of significant temperature and wind levels, and whether the dew point is or is not reported at levels with a temperature below about -40°. The humidity characteristics of the major instrument types are summarized as follows:

(1) A-22 tends to report humidity to the top of the sounding, with very few levels not reporting dew point depressions. The humidity values tend to be quite moist, rarely below 15 to 20 percent, even in the stratosphere, and humidity usually does not decrease much with height, indicating an unresponsive hygrometer. The average dew point depression is usually smaller at 300 than at 500 or 700 hPa.

(2) RKZ also tends to report humidity to or near the top of the sounding, but stops reporting dew point depressions before the top of the sounding slightly more often than with A-22. The average dew point depression is usually smaller at 300 than at 500 or 700 hPa, but when RKZ replaces A-22, the lowest relative humidity often drops slightly. Still, the most reliable indicators of introduction of RKZ are usually indirect, such as an increase in the number of reported wind levels.

(3) MARS is usually identifiable exactly in every sounding because it tends to stop reporting the dew point at levels where the temperature is colder than -40°. Since at the upper levels with reported dew point depressions, the colder cases are excluded, the average relative humidity at such levels may be lower than with A-22 or RKZ, but the average dew point depression is usually nearly the same at 300, 500, and 700 hPa.

(4) If MRZ is used after MARS, the use of MRZ is usually identifiable exactly by a resumption of dew point depression reporting to or near the top of the sounding. However, if MRZ follows RKZ or A-22, or if only a few MARS radiosondes are used with frequent alternations between MARS and either RKZ or MRZ, the timing of the transition from A-22 or RKZ to MRZ may not be obvious. MRZ reports significant wind levels (distinguishing MRZ from A-22), and usually reports about 10 percent more temperature levels than earlier models (distinguishing MRZ, with some uncertainty, from RKZ). MRZ is about as moist as MARS, and is sometimes but not always slightly drier than RKZ or A-22.

(5) MRZ-3AM is an inferred variety since some stations reporting 31313 instrument codes of 27 or 75 have intermittently been moderately dry since 1997. It is unlikely that this instrument is or was used at any station without having a 31313 code reported, so the challenge is to identify the use of MRZ-3AM instead of MRZ-3A when the instrument code is 27 or 75. On the average, a station using MRZ-3AM frequently reports a relative humidity below 10 percent, with fairly frequent dew point depressions above 30° C. Dew points around the tropopause are frequently about -80 to -85°, compared to frequently -70 to -75° with MRZ-3A. The average dew point depression from 700 to 300 hPa is usually about 2° larger than with MRZ-3A. However, the drying is probably not consistent enough in every observation to assume that individual observations can be assigned to MRZ-3AM or MRZ-3A.

(6) RF95 was initially introduced in 1997, but separate codes (53 and 76) were not assigned until 1999, so at some stations reporting instrument code 27 (MRZ), it appears that there were 3 distinct levels of dryness. Instrument codes 53 and 76 appear to be reliable indicators of the use of RF95. With RF95, dew point depressions average about 2 to 3° larger than for MRZ-3AM (or about 5° larger than for MRZ-3A), with some dew points around -90 to -95° C near the tropopause, and some dew point depressions larger than -40° C.

(7) MRZ* was introduced recently enough (in February 2007) that it is almost always identified by a reported 31313 instrrument code of 58 or 89. Based on examination of the initial reports in February 2007, MRZ*

is not consistently drier than MRZ, although further analysis is needed to confirm the relative behavior of MRZ* and MRZ throughout the year.

(8) RZM-2 was also introduced in February 2007, and at most stations the initial number of reports is small. However, these reports tend to be similar to the “moderately dry” MRZ-3AM radiosonde, with some humidity values from 1 to 5 percent. Again, further analysis is needed to confirm the behavior of RZM-2 throughout the year.

4. PROPOSED INSTRUMENT ADJUSTMENT PROCEDURE

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

The approach proposed here differs fundamentally by developing adjustments for each distinct instrument type instead of for each station. The technique is "equiprobability transformation" (Eskridge et al. 1995) or “histogram matching” (Horn and Woodham 1979), which replaces data values to transform the cumulative probability distribution observed by each instrument to the same percentiles of the distribution observed by a chosen "reference" instrument. For example, if the observed temperature is the 14th percentile, the 14th percentile observed by the reference instrument is substituted for the temperature observed by this instrument (actually, an amount is added to or subtracted from the observed temperature to transform it to the specified percentile). This makes the readings statistically equivalent to the reference instrument. Variations in environment, including differences between stations, are accounted for by stratifying probability distributions by pressure layer, sun angle, and (for dew point adjustments) temperature interval.

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

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

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

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

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

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

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

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

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

In step 4, many issues are the same as in step 3. Probably all instrument pairs will require dew point adjustments (the differences will not be deemed negligible), but for some instrument pairs, the variations in differences with sun angle might be insignificant. In this step, an additional issue is how to develop cumulative probability distributions when moisture data is not always reported. Special procedures are needed if the dew point is not reported when the temperature or humidity is below a certain value, or dew points are randomly missing or end at random levels, or if dew point "censoring" (such as reporting an artificial 30° C dew point depression to indicate that the relative humidity is under 20 percent) or statistical humidity reporting (such as reporting a statistical average humidity, according to the reported temperature, when the electrical signal from the humidity sensor indicates an unresponsive sensor) is used.

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

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

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

5. DATA AVAILABILITY

While global metadata is still being prepared, incomplete versions of various files are being made available periodically at the Texas A&M University Atmospheric Sciences FTP site by anonymous ftp at ftp.met.tamu.edu. The files are in directory /data/ftp/pub/ schroeder. For files with different versions, the latest version contains the largest number, such as rg5.f. For some files, the version number is the date in the form YYMMDD, such as RaobMetadata.090104.

Acknowledgements

This project has been partially supported by NOAA Office of Global Programs grant NA08OAR4310686 and by contracts with the NOAA Applied Research Center for Data Set Development through the National Climatic Data Center.

References

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

Balagourov, 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. [Available online at http://www.wmo.int/pages/prog/www/IMOP/publications/IOM-75-TECO2002/Posters/P1.2(01)Balagourov.doc]

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

Dey, C. H., Ed., 2002: Guide to WMO Table Driven Code Forms: FM 94 BUFR and FM 95 CREX. WMO, Geneva, 2 vol., 57 and 137 pp. [Available online at http://dss.ucar.edu/docs/formats/bufr/]

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

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

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

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

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

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

Horn, B. K. P., and R. J. Woodham, 1979: Destriping Landsat MSS images by histogram modification. *Computer Graphics and Image Processing*, **10**, 69-83.

Kats, A., and V. Grinchenko, 2006: Operational testing of the Russian upper-air AVK radars using tracking of the sun radiation. Papers Presented at the WMO Technical Conference on Meteorological and Environmental Instruments and Methods of Observation (TECO2006). Instruments and Observing Methods Report No. 94, WMO, Geneva. [Available online at http://www.wmo.int/pages/prog/www/IMOP/publications/IOM-94-TECO2006/P3(21)_Kats_RussianFed. pdf]

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

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

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

Parker, D. E., M. Gordon, D. P. N. Cullum, D. M. H. Sexton, C. E. Folland, and N. Rayner, 1997: A new global gridded radiosonde temperature data base and recent temperature trends. *Geophysical Research Letters*, **24**, 1499-1502.

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

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

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

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

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

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

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

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

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

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