1125 DERIVING COMPLETE UPPER AIR STATION HISTORIES USING SENSITIVE DATA VARIABLES – AN ESSENTIAL STEP IN HOMOGENIZING THE ATMOSPHERIC CLIMATE RECORD

Steven R. Schroeder *

Texas A&M University, College Station, Texas

ABSTRACT

Archived in-situ upper-air soundings can be used to compute climatic atmospheric temperature and moisture trends, in some areas for the past 100-125 years, and are still essential to calibrate global high-resolution satellite observations. However, all data must be adjusted to correct instrument biases before computing temporal and spatial averages and trends. Complete metadata is needed, including station name and ID, location (latitude, longitude, elevation), radiosonde and supporting equipment types, processing that can affect reported data, and dates and times when any metadata element changes. However, such metadata is incomplete and often erroneous, so data homogenization still produces uncertain trends.

This report summarizes an ongoing project to develop complete validated historical and current upper-air station and instrument metadata. No other similar efforts systematically check the metadata for consistency with the archived data and other potential information sources.

This project has obtained a large amount of additional metadata, mostly from informal sources, and focuses on validating station elevations (and locations), and on validating instrument types and inferring the instrument changes if this metadata is erroneous or missing. For all soundings reporting the surface pressure and the height at a pressure level near the surface, the surface elevation used by the station can be computed hydrostatically, and a station move usually coincides with an elevation change. Instrument characteristics in data are determined at stations where the instrument type is known. Similar data characteristics at other stations usually indicate the use of that instrument type, and changed data characteristics usually show when a different instrument type is used.

Distinct characteristics of an instrument type are most often found by examining data that reveals extremes of sensor performance. Humidity-related variables usually show the largest differences. "Conventional" climate variables such as average temperature or dew point depression at certain pressures, or lapse rate in a layer, are contaminated by instrument biases but may not differ enough to reliably distinguish instrument changes. Variables such as the lowest and highest reported relative humidity (RH) in a sounding, highest reported dew point depression (DPD) or lowest dew point, frequency of RH < 10% or RH > 95% (or exactly 30° DPD), number of reported temperature levels, or lowest temperature or lowest pressure with a reported DPD, have little or no climate use but tend to identify a different instrument type (or sometimes a different processing technique), so they are called "sensitive variables".

Results show that many stations have used more than one instrument type in the same period, that it is common for a station to use an instrument type for a short period (sometimes less than a year), and in a few cases, an otherwise unidentified instrument was used at multiple stations. This paper does not discuss homogenization techniques to adjust for biases after complete breakpoint histories have been developed, but any of these situations can cause problems for most homogenization approaches, particularly those that develop adjustments by quantile matching based on several years before and after an identified breakpoint.

INTRODUCTION

Radiosonde stations have operated worldwide since 1957 when an Antarctic network was set up. Some stations have almost daily radiosonde soundings into the stratosphere back to the early 1930s. Earlier upper air observations used meteorographs attached to balloons or kites (starting in the 1890s) and airplanes (starting 1917). So, global atmospheric climate trends can be calculated back to 1957, and some areas have nearly continuous upper air data for 100-125 years.

Earlier instrument sensors were generally larger and less sensitive than modern sensors, leading to two major errors: (1) Radiation errors, where a sensor is heated by sunlight and cooled by radia-

---

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

tion to space. (2) Lag errors, where a sensor responds slowly to changing conditions, so as the balloon rises, readings are an average of conditions below the radiosonde, causing mostly warm and moist biases. (A dropsonde makes measurements while it descends, generally into warmer and moister air, so it has opposite biases. A radiosonde descends after the balloon bursts, but descending data is usually ignored because the radiosonde may be too far downwind from the launch point for reliable reception, and before GPS was used to track radiosondes, the location was not accurately known.) Surface observing instruments have similar errors, but a surface instrument can be calibrated in place. A radiosonde can be calibrated in a chamber with adjustable pressure, temperature, and illumination, but it is not possible to simulate all conditions that affect the accuracy of readings. Many soundings have manufacturer corrections applied, but such corrections were developed from limited tests so data biases remain. Because instrument improvements have gradually reduced radiosonde errors, the global trend is contaminated by general artificial cooling and drying.

This paper and the accompanying poster summarize issues with the steps above to validate and develop complete upper-air station metadata. The steps include obtaining and integrating as much archived data and metadata as possible, validating instrument metadata using the archived data, and identifying historical and current station locations as accurately as possible. All metadata is consolidated into a single Texas A&M University metadata file ("TAMU file"), which is referred to repeatedly, but is not discussed in thorough detail, in this paper.

DATA AND METADATA SOURCES

Data sources. Most radiosonde datasets with continuous updates simply archive incoming operational Global Telecommunications System (GTS) reports. For developing bias corrections, it is best to start with original transmitted soundings. Most transmitted soundings have manufacturer temperature corrections applied, and a change in the correction scheme is a breakpoint even when the radiosonde itself is unchanged. For example, Vaisala documents at least five radiation correction schemes that were used with RS80 radiosondes (1982, 1986, 1995, 2005, and 2010). However, a station would not implement a new correction until the appropriate software version was installed, and the date of the software change at each station is rarely announced.

Some online archived datasets are simply soundings received at a major forecasting center such as NCEP, ECMWF, and JMA. This project mainly uses NCAR Datasets 353.4 and DS351.0, which are sounding files from NCEP for January 1973 to February 2007 (DS353.4) and starting October 1999 (DS351.0). The soundings are reformatted, but generally otherwise unchanged. No efforts are made to restore reports missing due to a network or processing outage, and also soundings from a new station are omitted until NCEP adds the station to its catalog, which usually occurs with a lag of 8 to 36 months.

Unedited raw GTS reports are archived by some recipients, but the only currently publicly available archives found are at Iowa State University (https://mtarchive.geol.iastate.edu, then click on year, month, day, "text", and "upa" to obtain hourly files, with data starting January 2000) and the University of Wyoming (http://weather.uwyo.edu/upperair, then "Soundings", then enter "Text: Raw" and the year, month, range of dates, and a station ID to get up to a month of soundings from one station at a time, with raw soundings starting 1997. Decoded BUFR soundings are available from http://weather.uwyo.edu/upperair/bufrraob.shtm starting about 2018, but only one sounding can be requested at a time).

Probably the most extensive available archived dataset is the Integrated Global Radiosonde Archive version 2 (IGRA2), from NCEI at https://www.ncdc.noaa.gov/data-access/weather-balloon/integrated-global-radiosonde-archive. It consolidates many archived datasets, and as of 13 February 2020, has 2788 stations with 48,735,863 soundings. However, it does not extract the "31313" instrument codes discussed below.

There are many files of radiosonde and other soundings that were not transmitted on the GTS, mostly from field experiments, and are too numerous to list here. Some data was specially processed, since a field experiment often uses a new radiosonde model in unusual environments, where bias issues may be noticed for the first time, and some soundings are archived with no corrections at all (this is sometimes called "research mode") to allow testing of alternative corrections. In addition, "data rescue" efforts have digitized soundings that either were never transmitted (such as from remote stations where communications was unreliable), or were published on paper (such as USA kite soundings from about 1918 to the early 1930s), or were simply never published but were found in various archives worldwide (Humphreys 1929).

It should also be mentioned that forecast centers noticed biases differing between stations in transmitted soundings, even though many stations had applied manufacturer corrections for many years, so in 1964 NCEP (National Meteorological Center or NMC at that time) started applying additional corrections to soundings according to reported instrument types. Most daytime corrections were negative, implying that existing corrections were too small and readings were still too warm (Finger et al., 1965). A cumulative list of corrections at http://www.nco.ncep.noaa.gov/pmb/codes/nwprod/obsproc_prep.v3.8.0/sorc/prepobs_cqcbufr.fd/radcor.f dated Feb 2013 (but not found online as of 2019) shows that many instruments had corrections at upper levels with high sun angles of 3 to >10°C, while corrections used in 2008 were usually ~2°C at the most (many corrections were positive, implying that those manufacturer corrections were too large), and few corrections were 3 to ~4.5°C. Because these adjustments are applied with uncertain knowledge of the actual radiosonde types, datasets with forecast center additional corrections such as NCAR DS337.0 should not be used to detect instrument discontinuities, since many of the discontinuities are generated by the forecast center itself. For this reason, IGRA (and probably all other radiosonde archives) never used such datasets.

Required metadata. Complete station and instrument histories should include the following metadata elements: Station name and ID, location (latitude, longitude, and elevation), radiosonde and supporting equipment types, processing that can affect the reported data (such as pre-transmission radiation corrections), and dates and times when any metadata element changes. However, due to historical limited communication capacity, individual soundings have reported minimal metadata, usually only the day of the month and time (UTC hour), station ID, location for mobile soundings only (ship soundings do not report the elevation so it is assumed to be 0 m, but current commercial container ships often launch from 40-50 m), and (recently) condensed instrument information. So, the missing metadata must be obtained elsewhere.

The condensed instrument information includes variable a4, introduced 1968 in upper wind soundings, and a 5-digit group introduced 1991 in the 31313 section (so it is called the “31313 code”; the following 5-digit group includes the actual UTC launch hour and minute). The 31313 code has gradually been reported by more and more stations, and since 2015 has been reported by >95% of soundings.

Variable a4 describes only the generic wind tracking method (initially, only pressure instrument for height but unstated horizontal tracking, optical theodolite tracking, radiotheodolite tracking, radar tracking, or pressure instrument failed. Later, Omega, Loran, wind profiler, and satellite [basically GPS] tracking methods were added).

In the 31313 code, digits 2 and 3 summarize the radiosonde and ground system type (Digit 1 is a very generic radiation correction code, and the variable in digits 4 and 5 essentially duplicates variable a4 as code values 00 to 08, and values 09 to 99 are rarely used and uninformative for identifying instruments). The radiosonde type (digits 2 and 3, WMO Code Table 3685 or BUFR Table 0 02 011) is very useful, especially when the same station reports different codes in a period (some Russian stations report up to 8 models because many models are designed to work with the same ground station), but still is inadequate for identifying instruments accurately because of the limited number of codes, although an obsolete code can be reassigned to a newer instrument. As an example, code 27 has been used by the Russian MRZ radiosonde with AVK ground system, which was first operational in 1986. The same code is still in use, but has been interpreted as applying to the MRZ “family” (radiosondes reporting this code include MRZ-3A, MRZ-3AT, MRZ-3AM, early RF95, and early Ukraine PAZA-22M) and some radiosondes used the MARL-M phased array tracking system instead of AVK. Humidity sensors include goldbeater’s skin for the early radiosondes, a Russian capacitive sensor for MRZ-AM, Vaisala RS80 Humicap for RF95, and probably a Ukrainian capacitive sensor for PAZA-22M. In general, any code used more than 5 or 10 years with no announced model change probably is not an unchanged instrument. For example, the multiple Vaisala RS80 radiation corrections mentioned above would cause several discontinuities equal to the differences between the correction tables.

Unreported metadata is traditionally obtained from catalogs. WMO maintains official catalogs, but metadata entries are reported by country coordinators, and often are incorrect or not updated for decades.

The official station IDs for radiosonde fixed land stations (as well as surface synoptic stations) since 1949 have been 5-digit numbers. This project uses the official IDs, and continues their use backward before 1949, with a few exceptions. The first exception is when the same station ID is used for considerably different station locations in the same period. In most cases, one of the stations uses the same ID unofficially (such as an operational radiosonde station, and a separate ozonesonde station that does not transmit its observations on the GTS, or a historical dataset that retroactively assigned the WMO ID to a supplemental station at a different station). If an operational station uses 2 sites with frequently alternating observations, this situation can be dealt with by listing separate "events" for the same station ID whenever the location changes, with the appropriate date, time, and location when each change occurred. In other cases, the TAMU file assigns a different 5-digit ID to each duplicate station except the official operational station, notes the original duplicate assigned ID and its source, and if the data from the duplicate station is integrated with files containing the operational soundings, each sounding needs to be relabeled with the reassigned ID. Also, since stations are frequently moved, opened, and closed, the same station may have different IDs at different times, of an unused ID may be reused by a different station (in some cases, >1000 km away). The TAMU file attempts to keep track of the multiple IDs and multiple uses of the same ID, but these lead to complicated station histories in some cases.

Soundings from ships and other mobile sources (including dropsondes) are allowed to have any alphanumeric (letter and number) ID up to about 7 characters in the official reporting formats. The TAMU file uses the reported IDs in most cases with no change.

Before 1949, there was no global ID system, but 3-digit "International numbers" were assigned starting in 1930 for synoptic reports, with separate lists for different continents or some other regions. When a list was nearly completely assigned, additional lists were added, with the first list called "primary", the second list called "supplementary", and a third list (if needed") called "arbitrary". This system, of course, became unwieldy, so it was replaced with the global 5-digit scheme above. Before 1930, other abbreviations were used with 2 or 3 digits or 2 or 3 letters, but turned out to be inadequate with increased data exchanges.

Some historical datasets use a 5-digit WBAN (Weather Bureau-Army-Navy) ID that originated approximately 1945 but was also quickly assigned to earlier stations. It was originally geographical with the starting 3 digits indicating 10 x 10° blocks (first digit 1 to 8 for 90 x 90° octants, or 9 or 0 as overflow) and up to 99 stations in a block (last 2 digits 01-99, 00 not used). If all numbers are assigned, the first digit becomes 9 and last 2 digits 01-99 are assigned, but these stations are mixed from the 8 octants. With the next overflow, the first digit becomes 0. With a further overflow, the first digit becomes 5, 6, or 8 because these octants are mostly oceans. In addition to becoming very non-geographical, this 5-digit scheme has the same limitations of the 5-digit WMO scheme. The WBAN IDs are used in quite a few historical datasets by NCEI and NCAR.

Surface (hourly) METAR and earlier Airway observations have used airport codes assigned by various aviation organizations as station IDs. Comparable observations not from airports have used the same scheme. Airport codes were originally 2 or 3 letters, but all were 3 letters starting 1948 (or a character could be a number starting 1958). These IDs were also used for radiosonde (or airplane) soundings through 1948. METAR reports use a 4-character code that (at least in the USA and Canada) is K (in the USA) or C (in Canada) followed by the previous airport code. These codes were never used for GTS synoptic reports, but the University of Wyoming site shows METAR codes on the maps of radiosonde sounding locations, while 5-digit WMO codes are used in actual retrievals of soundings.

Some metadata archives, especially for field experiments with observations not transmitted on the GTS, use customized ID codes. If such observations are added to a dataset such as IGRA2, their IDs must be converted to the scheme used in the new dataset. IGRA2 also has its own 11-character scheme, with a 2-letter country code, a 1-letter source code, the original ID (5 digits or a variable number of characters) in the rightmost characters, and other positions filled with zeros.

Finally, WMO introduced the WIGOS (WMO Integrated Global Observing System) ID system to replace all WMO-coordinated station systems starting in 2016. The station catalog is online, called OSCAR (Observing Systems Capability Analysis and Review Tool) at https://oscar.wmo.int/OSCAR/index.html#/. This system appears to be under development, since there are very few listed upper wind observation stations. An upper air station ID with a 5-digit WMO ID (such as 45678) would usually be converted to WIGOS ID 0-20001-0-45678, or possibly 0-20000-0-45678 (a ship would use a second group of 20003, 2004, or 20007 and the last group would be the ship reporting ID. In the long run, a new radiosonde station established at a location with any type of existing station ID could potentially use the existing WIGOS ID.

Transition to BUFR (Binary Uniform Format for the Representation of meteorological data) soundings. Radiosonde soundings have been transmitted for many years in code formats that are mostly 5-digit groups, with the current format used since 1968 with only minor changes. As mentioned above, due to very limited transmission capacity when the codes were designed (in some cases, observations were transmitted by manual Morse code). By the 1980s, there was significant concern that the volume of observations would greatly exceed transmission capacity, even though more modern automated circuits were being installed worldwide. So, BUFR was discussed by WMO and initially designed in 1984 to both permit reporting in more detail (including additional metadata), and to compress the transmitted data in binary form. The previous codes are referred to as the Traditional Alphanumeric Code (TAC) format, and BUFR is also referred to as Table-Driven Code Format (TDCF). Overall, BUFR observations are defined using tables that extended the previous WMO code tables that were already used by TAC. Some tables would define either code values or simply variables with definitions and units (such as temperature in K with a stated precision). Other tables would define operations and functions (such as the number of reported pressure levels). Other tables would define the structure of a report, such as the format of a radiosonde observation of a certain type, as a template containing a sequence of BUFR variables and operations that list the metadata of the observation, followed by the observation data. However, because it was defined before the internet led to a huge expansion of circuit capacity, along with protocols to automatically handle large volumes of data, BUFR is still in early implementation stages in 2020 because of its customized structure that makes encoding (preparing the observation messages) and decoding (reading the received observations) difficult.

The variables in the 31313 section mentioned above are just a few of the metadata variables that, in principle, could be included in BUFR soundings to more completely define the instrument, processing, and other metadata. In addition, location data (latitude, longitude, and elevation) can be defined with high precision, and could be reported with every data level so the complete path of the sounding can be reported in space and time. Additional metadata variables already defined include radiosonde computational method, radiosonde ground receiving system, balloon manufacturer, type of balloon, type of balloon shelter, type of pressure sensor, type of temperature sensor, type of humidity sensor, type of surface observing equipment, and geopotential height calculation. However, most of these variables are defined with only 4 to 6 bits (16 to 64 allowed values), so they are not adequate already, and some of the tables contain values only appropriate for technology as of the 1990s and earlier. Data examination shows that generic entries such as “capacitance humidity sensor” or “chip thermistor” are still inadequate to describe data characteristics and biases because different sensors of the same type may have much different lags and radiation errors, as well as response characteristics. WMO documents show that it has been a very intensive task to maintain the radiosonde type table (Code Table 3685), and similar efforts would be needed for each other table. The “radiosonde computational method” variable is so far completely undefined, but from reports describing sounding preparation and operational procedures (such as military technical manuals), the description of steps involved in a sounding would need to fill a book, and internal software details can have significant effects on the data accuracy, including biases.

Some metadata sources. All of the data sources mentioned above are accompanied by a station list, and most lists include station name and location information. There are many other historical station catalogs associated with datasets not listed above, with customized formats, too numerous to list here. However, the sources of the station information are rarely stated, and it is also rare to state multiple locations for a station with dates of moves. Most often, only the latest location for each station is given, even if the dataset contains many decades of data. When the same station is found in multiple datasets, many discrepancies are found. A large error such as an exact error of an integer number of degrees of latitude or longitude, or a wrong hemisphere, is simply a typographical error.

The primary catalog for surface (synoptic) and upper air stations was WMO Publication 9A (“WMO-9A”) starting 1953, listing each station name, ID, latitude, longitude, elevation, and brief summaries of observation types and schedules. Separate surface and upper air station locations can be listed, but less than 10 percent of radiosonde stations have separate entries, even when the radiosonde station is many km away from the synoptic station. The actual date of a change is rarely stated.

WMO has also published a separate Catalogue of Radiosondes and Upper-Air Wind Systems in Use by Members (“WMO-UA”; the oldest one found is from 1957), listing radiosonde stations, locations and elevations, and radiosonde types. Titles of the catalog, publication formats, and additional information (such as radiation corrections) vary. In general, earlier catalogs report broad instrument types at many stations. Also, while the date of update for each station entry may be stated, these dates rarely coincide with the actual dates of station moves or instrument changes. By hydrostatic computation of the surface elevation from heights reported in soundings, there were 5 station moves or elevation changes at the 92 NWS radiosonde stations between 1998-2006 (verified by William Blackmore, NOAA, personal communication, 2007), but the 6 WMO upper air catalogs in that period listed 139 moves or elevation changes, including 4 of the 5 actual changes.

Both WMO-9A and WMO-UA have been replaced by OSCAR, but it appears that much of the information in WMO-9A and WMO-UA is not incorporated in OSCAR.

There has been only one published systematic effort to identify complete global radiosonde station and instrument histories, which was built from available WMO catalogs (1965, 1977, 1982, 1986, and 1993; some of these were unpublished manuscripts), a survey in 1990 with replies received from 49 countries, and a limited number of additional unpublished documents and personal contacts (Gaffen 1993, 1996). That effort obtained no information after 1995, and was compiled systematically but not checked for consistency, to show a sample of the available information and to also show how much information was still needed. The largest effort to compile upper air soundings into a consistent format is probably the Integrated Global Radiosonde Archive version 2 (IGRA2), which is updated daily and is documented at and accessed from https://www.ncdc.noaa.gov/data-access/weather-balloon/integrated-global-radiosonde-archive. Its metadata file is based on the Gaffen (1996) metadata file, but with a quite limited number of updates, and no updates after 2013. The IGRA2 data and metadata are available through https://www1.ncdc.noaa.gov/pub/data/igra/, and the station and instrument historical metadata file is in the history directory, file igra2-metadata.txt. As with the original Gaffen (1996) project, the metadata is still not checked. Also, no systematic effort has been made to add the reported instrument types (31313 codes) to the soundings, when they are available.

Issues concerning station location definitions. While WMO does not specify an official definition of an upper-air station location, the most logical definitions are that a radiosonde station location should be the launch release site, and an atmospheric profiler station location should be the center of the transmitting antenna (where the air volume close to the surface is sampled). For a manually-launched radiosonde, the balloon inflation building is a logical station location, because the radiosonde is usually launched downwind in any direction depending on the wind. The launch tube is the most appropriate station location for an automatic launcher. When the WMO standard was to state the location to the nearest degree and minute, no distinction was usually needed between the surface and upper-air observation locations. For various reasons (such as to avoid air traffic), an upper-air station with the same WMO ID may be many km from the surface (synoptic or METAR) station, at a considerably different elevation. With locations to the nearest second, all upper air stations should have separate surface and upper-air station entries, but recent WMO catalogs still have separate entries for fewer than 10% of radiosonde stations.

While surface data for a radiosonde sounding is usually provided by permanent surface instruments, nearby but far enough away that they do not obstruct the launch of the balloon, the upper-air station location is not the surface instrument location but is the radiosonde release location because the radiosonde path through the air is downwind of the release location. The surface instruments for a radiosonde station are usually not the instruments used for synoptic and METAR reports because some data elements (such as precipitation and visibility) do not need to be measured for a sounding.

The surface instrument location should fully represent the radiosonde launch environment, so if the radiosonde is launched from a roof, the surface instruments should also be on the roof, rather than placing the surface instruments on the ground nearby to sample the whole boundary layer. This is the NWS practice (Carl Bower, NOAA, personal communication, ~2009), because NWS normally establishes a rooftop launch site only if the area is too congested for an unobstructed surface station or radiosonde launch site.

It should be mentioned that before GPS was commonly used, locations were based on continental or regional grids such as the North American Datum. The World Geodetic System 1984 (WGS-84) with recent updates is globally consistent to <1 m, but differs by up to 700 m from regional datums. Early station locations, especially on remote islands, were often not well surveyed so this project attempts to locate historical and current stations using online satellite photos such as Google Earth. Some historical stations from over 100 years ago have been accurately located, but buildings at most early stations either no longer exist or no photos or adequate descriptions have been found.

WMO catalogs stated locations to the nearest minute of latitude and longitude until 2010, when some catalogs were reformatted to allow locations to the nearest second. However, as of 2020, most stations have not updated their entries. For stations stating locations to the nearest second, online satellite photos (such as Google Earth) show that many locations are not weather observing sites, such as an official airfield location, which is the midpoint of a runway. In many cases, the reason for an incorrect location cannot be determined.

Some catalogs state locations in decimal degrees, and a common error is incorrectly converting between (for example) minutes and hundredths of degrees.

This project states locations in decimal degrees up to 4 decimal places, depending on the reported or determined location (when read by a computer program, blank digits such as "12.34__" are filled as zeros ["12.3400"]). Because 0.0001° is ~10 m of latitude (or longitude in the tropics), this is a "practical" accuracy limit, especially for manual launches, which probably vary by 20 to <100 m at most sites as wind direction varies.

Issues concerning station elevation definitions. WMO catalogs define 3 relevant station elevations, reported in 2 variables: H_p is the elevation corresponding to the reported surface observation pressure (this may be a past elevation, with pressure adjusted to that elevation for historical continuity). H is the average ground elevation in the vicinity, or the radiosonde release elevation for a radiosonde sounding, normally ~1.2 m above the ground or surface from which the radiosonde is launched (the elevation clarification has been in WMO instructions starting April 2001, but these instructions conflict if one catalog entry covers a surface and upper air station). H_a is the official aerodrome elevation (usually the highest elevation on the runway) if the station is at an airfield. A WMO catalog uses one variable "H/H_a" to report a single elevation. There is no ambiguity for a surface station because the station is either at or not at an airfield. However, if the same entry covers a surface and upper-air station, frequently the surface elevation computed hydrostatically from the soundings does not match any reported elevations.

It should also be noted that all catalog elevations should be stated in geopotential meters (GPM) because all pressure conversions (such as to sea level pressure, SLP) are computed hydrostatically based on GPM. Most elevation definitions do not mention that elevations should be GPM.

VALIDATING AND INFERRING INSTRUMENTS

Some previous preliminary discussions of aspects of developing metadata and instrument bias adjustments are in Schroeder (2007, 2008, 2009, 2010). The discussion below is mostly taken from the accompanying poster, without the figures that are in the poster.

Useful data variables. After collecting data and all available metadata and organizing data into appropriate files, time series of basic and "sensitive" variables are prepared, including monthly and annual averages and time series of individual soundings. As summarized in the next subsection describing processing steps, a systematic search is performed to look for data discontinuities that indicate the timing of instrument changes. The search is performed starting with annual averages, then monthly averages, and finally individual soundings to try to identify the exact timing of the change.

A "sensitive variable" is a variable that is most likely to show steplike differences when a radiosonde model changes. Any such unnatural data breakpoint is likely to be a radiosonde change. A change in the ground station, other radiosonde equipment (even the balloon), operational or computational procedures, or software is treated as an instrument change because these factors can affect readings reported in a sounding. A few examples of such changes are as follows: (1) A new radiation correction changes readings by the difference between the old and new corrections. (2) Since the balloon is heated in sunlight or cooled at night by radiation to space (also, water, frost, or snow on the balloon can increase the humidity after the balloon is above the clouds), and the radiosonde swings across the balloon wake as it ascends, lengthening the line from the balloon to the radiosonde allows the radiosonde to swing in a larger arc with less time in the balloon wake, which reduces the bias caused by the artificially warm, cold, or moist balloon wake. (3) From April 1973 to September 1993, NWS stations reported relative humidity <20% as an artificial dew point depression of 30°C, which was drier than the humidity sensor at that time could actually measure.

Basic sounding variables are pressure, height (of certain pressure levels), temperature, and humidity (reported as dew point from 1949-1967 and dew point depression since 1968), and also wind direction and speed. These variables are reported at mandatory levels (currently surface and 1000, 925, 850, 700, 500, 400, . . . hPa) and significant levels (including the tropopause and any other levels where data values differ noticeably from a straight-line interpolation of the reported levels). In a BUFR sounding, these variables can be reported at every level, along with additional variables at every level such as elapsed time since launch, latitude, longitude, and height.

Sensitive variables are derived from basic variables and the particular variables to check may need to be customized for different suspected or reported instrument types, since the nature of discontinuities depends on processing and other practices, as well as the basic sensor responses.

Examples shown in the poster focus on typical sensitive variables at Japanese stations mostly before soundings from that country reported the 31313 code. In these examples, the instrument changes are inferred almost completely from the time series of archived soundings by looking for similar breakpoints at each station. If the 31313 codes are consistently reported, the task is to simply validate that each reported instrument type has a set of consistent data characteristics over multiple stations and the entire period of use of that instrument type at each station, and discontinuities coincide with the times when the reported 31313 code changes. However, the same variables are checked and the discontinuities should be similar, whether the station does or does not report the 31313 codes.

Another example of such a situation is when all stations in a region or country start reporting 31313 codes on approximately the same date, but the instrument transition occurs over several years until all stations have changed. For example, China started reporting 31313 codes at the end of Sep 2012, but different stations had 3 codes (31 = Taiyuan GTS1-1, 32 = Shanghai GTS1, 33 = Nanjing GTS1-2, and all used the GFE(L) tracking radar) although sudden drying indicating the new radiosonde started as early as Dec 2001, with a few stations per year switching to the new radiosondes. In Huang et al. (2010), the Shanghai model was introduced first and the Taiyuan and Nanjing varieties made to the same specification were introduced starting Sep 2009.

Because sensitive variables represent extremes of sensor responses, many discontinuities are so large and abrupt that they cannot be natural changes. Most sensitive variables have no real climate use (as mentioned below, some reported values are so extreme, such as dew point (DP) below -110° C, that they cannot be correct), so they have generally not been studied. Examples of sensitive variables are as follows:

* Number of reported levels in transmitted soundings. A sudden increase may indicate improved processing, such as changing from manual computations to a calculator or computer.

* Lowest reported (or computed) relative humidity (RH) or highest dew point depression (DPD, reported since 1968), and also highest reported (or computed) RH. Since it was difficult to produce very dry RH in a calibration chamber, some RH algorithms were tuned to never compute RH below a value such as 10 or 20 percent. Slow-responding sensors often would report RH considerably below 100% (or ice saturation) in clouds, would not detect thin dry (or moist) layers, and also would not measure dry conditions in the upper troposphere and stratosphere. In the upper troposphere and above, the humidity sensor simply became unresponsive and continued to report the last viable humidity value, often 30 to 50%, with very little change to the top of the sounding. The first radiosonde with a capacitive humidity sensor (Vaisala RS21) had a variety starting 1978 that frequently reported humidity <1% (code values permitted dew point depressions up to 49° C), while after much further development, the current Vaisala RS41 more often reports a “few” percent relative humidity in the stratosphere.

* Coldest reported dew point. Some instruments have dry biases at upper levels, seen by reported dew points < -100 °C, which should never occur at radiosonde altitudes. At the South Pole, Air-5A radiosondes used from Apr 1999 to Aug 2002 often reported DP < -110 C at all levels including the surface in winter, due to a defective response at temperature <-40°C.

* Lowest pressure, highest altitude, or coldest temperature with reported RH (or DPD). Due to the above-mentioned RH cold insensitivity, for many radiosondes it was a policy to not report RH or DPD below a chosen temperature or pressure (such as -40°C or 200 hPa), and some radiosondes were wired to disconnect the RH sensor at a temperature or pressure threshold.

* Lowest reported temperature, or temperatures at certain levels such as 500 hPa, 100 hPa, or the tropopause. Usually these variables are not sensitive enough to detect discontinuities in individual sounding data, but monthly averages may show discontinuities caused by factors such as an introduced or changed radiation correction.

* Number of reported wind levels, or maximum height or lowest pressure with wind data. A discontinuity does not necessarily indicate a change in the radiosonde itself, unless the wind reporting is changed due to replacing the entire ground system. In some cases, wind data was obtained by a separate balloon launch, not necessarily at the time or location of the radiosonde launch, so wind observation changes could be independent of radiosonde changes.

Summary of steps to infer complete station *instrument histories*. After preparing the data and collecting and organizing all accessible metadata, the process of validating and inferring the station histories is as follows. In general, detailed studies should be made for a country or group of stations that is likely to have a similar instrument history at multiple stations. In that way, the data signals can be compared over a region to see if the data signals vary smoothly with the climatic environment and over the year.

(1) Compute station elevation histories. This step hydrostatically computes the surface elevation using the reported surface pressure and all levels up to the first above-surface height. TAC sounding formats report pressure to the whole hPa, so individual computed surface elevations vary by about ± 4 m, but in most cases an elevation change of ~±1 m can be detected in averages. An elevation change is likely to coincide with a station move and possibly an instrument change. As shown to the right, additional information may identify station locations corresponding to the computed surface elevations. Even if locations are not accurately determined, elevation needs to be considered due to effects on bulk variables such as precipitable water, or boundary layer characteristics, that may falsely imply an instrument change.

(2) After determining the elevation (and hopefully location) history, finding instrument changes is a repetitive two-stage process:

(2a) Examine annual and monthly averages of appropriate variables to identify reported (or hypothesized) instrument changes and their approximate timing.

(2b) Search sensitive variables in individual soundings for steplike discontinuities that may show instrument transitions to the exact or almost exact observation.

Both steps should consider whether a fluctuation may have a natural cause, such as ENSO or a large volcano. The repetitive nature of this process means that multiple stations with the same hypothesized instrument types need to be carefully compared, possibly repeatedly, because an instrument type should show a high degree of commonality in characteristics at each station. Differing data signals in individual soundings may indicate frequent alternations between instrument types.

Sometimes, evidence of an unreported instrument type with different characteristics can also be found, even when the same instrument code is reported. For example, Russia MRZ (code 27 or 75) is ordinarily moist (goldbeaters skin humidity sensor), but reported dry conditions at some stations in the late 1990s. From personal communications (A. Kats, WMO, 2010), these were experimental radiosondes with MRZ electronics and more modern sensors, including RF95 (using Vaisala RS80 temperature and humidity sensors) and MRZ-3AM (with a domestic capacitive humidity sensor). So, the use of an unreported radiosonde often can be detected, but the identity of the instrument model may not be determined without further research.

Some examples. The examples shown in the poster focus on Japan prior to the beginning of reporting the instrument type code in soundings, which started at most Japanese stations around 1995. There were only three basic Meisei models involved,

RSII-56 (mostly used 1957-1981), with a bimetal thermometer and a hair hygrometer.

RSII-80 (mostly used 1981-1992), with a white coated rod thermistor and a carbon hygristor.

RSII-91 (mostly used 1992-2009), with a bead thermistor and a capacitive humidity sensor.

More recent Meisei instruments are always indicated by 31313 codes, and include RS-01G, RS-06G, RS-11G, and IMS-100, so they are not discussed here. According to personal contacts with Meisei exhibitors at AMS annual meetings, a newer model sometimes uses the same temperature and humidity sensors as the previous model.

The poster shows about 20 Japanese operational stations, including an Antarctic station in the early 1990s.

For the transition from RSII-56, the change from a hair hygrometer to a carbon hygristor causes drying, so the transition is most obvious in certain moisture variables. RSII-80 would send RH signals down to a temperature around -40° C while RSII-56 stopped sending RH signals at around -30° C, so the change is quite obvious in monthly averages of the coldest temperature with a reported DPD. Two military stations continue using RSII-56 for about a year after the other stations transitioned to RSII-80. The drying with RSII-80 is also fairly obvious in monthly averages of the lowest RH reported per sounding, but the month of transition is not obvious due to the large annual cycle, which has a maximum in summer in southern Japan and in winter in northern Japan.

The transitions to RSII-80 can be identified exactly by plotting individual soundings (in practice, the chosen variables are simply computed and stored in tables that do not generally need to be plotted). Two stations are shown, including one civilian station, transitioning in early March 1981. At station 47401 (in northern Japan), after one sounding with RH reported to about -40°, the next sounding reports RH to temperature about -30°, and then all soundings report RH to temperature about -40°. This indicates that one leftover RSII-56 was launched after the first RSII-80 launch. However, in countries where a similar signal indicates a radiosonde change (such as Russia), the RH sensor was unreliable enough that it often failed prematurely. In that case, stopping reporting RH at a moderately warm temperature is not as reliable an indicator of use of an older radiosonde type as in Japan. Also with RSII-80, soundings at station 47401 often reported RH close to 0% and DPD close to 49° C, although some soundings were nearly saturated so individual soundings were not always absolutely attributed to RSII-80 by the lowest RH or highest DPD.

At station 47580 (military), through much of April 1982, soundings often alternated between stopping reporting DPD at temperatures near -30 and -40°, so RSII-56 and RSII-80 were used approximately alternately and the instrument type of each sounding could be identified with high confidence. Starting 19 April 1982 at 0000 UTC, all soundings used RSII-80, based on reporting DPD to temperature close to -40°. Because station 47580 is in central Japan, it reported drier RH than station 47401 in spring with both radiosonde types, but RSII-80 was still drier than RSII-56, although the RH profile could not reliably distinguish the instrument type of all soundings.

The examples of transitions to RSII-91 show only plotted individual soundings. At all stations, both RSII-80 and RSII-91 reported RH to a temperature around -40°. It is not obvious on the plot, but due to more advanced software with RSII-91, the coldest level with a reported DPD is more precisely -40° (but due to the reporting format, temperatures <0° C are reported only with odd tenths of degrees such as -39.7, -39.9, or -40.1°). However, RSII-91 reports RH close to 0% much more frequently than RSII-80, and nearly saturated profiles are also more frequent with RSII-91 due to higher sensitivity. It is possible that if the soundings reported the measured DPD to the top of the sounding, RSII-91 would report a very dry RH (and high DPD) in the stratosphere, but RSII-80 would often report a fairly moist stratosphere due a less responsive humidity sensor. At Syowa (Antarctica), since the RH is reported with respect to water, RH is rarely close to 100% with both radiosondes because saturation in ice clouds corresponds to RH considerably <100% with respect to water.

DETERMINING STATION LOCATIONS

Methods. It is a much more subjective process to determine radiosonde locations (and station moves) than elevations or instrument types. The main reason is that station locations are mostly still reported to a precision of 1 minute of longitude and latitude, which identifies an area roughly 1 km square, but in many cases the reported locations are wrong.

Usually, site photos or detailed descriptions are required to identify a radiosonde launch site. While satellite photos with considerable detail are available with high resolution (usually to the level where lines identifying parking spaces are visible) online at Google Earth covering almost every land location with some human activity, modern technology does not require an observing site to have any distinctive equipment, and in some cases, an observing site is in a quite congested urban environment. However, some countries do continue to generally have distinctive features at least for surface observing sites.

China usually arranges a surface observing site in an orderly way with several traditional instrument shelters and paths to the shelters and other observing instruments. In Russia (and other countries of the former USSR), quite a few stations have been in the same locations since at least the 1950s, with a fairly common design for the weather office, and usually (but not always) a dome covering the tracking radar. In the United States, almost all stations also have a dome covering the tracking radar, and since the late 1980s or early 1990s, almost all radiosonde sites are at an NWS Weather Forecast Office, for which site photos are usually available. Only a few NWS sites launch from rooftop locations (Albany, Key West [because the island could be completely flooded in a hurricane], Tallahassee, and Tucson). All NWS radiosonde launches were manual until recently, but Vaisala Autosonde stations were installed starting 2017 at all Alaska radiosonde stations. While the USA generally uses the Automated Surface Observing System for surface observations, the equipment is relatively compact, but it still is usually visible at the Google Earth resolution and the locations at sites with WMO IDs are usually stated to the nearest second. However, radiosonde sites are rarely at the ASOS locations because ASOS is usually installed in the midst of runways. These are just some examples where a radiosonde station may be accurately located because of some distinctive appearance, and in some cases the stations have been located even without accurate location metadata.

In many other cases, a radiosonde station, even with a nearby surface observing site, does not have a distinctive appearance. If a station has a dome covering tracking equipment and the station is in a city or other location where most buildings have water tanks on the roof, the radiosonde tracking dome is often not distinctive. Similar, around a military base or even some airports, there may be several other radars with domes that are used for other purposes. Also, much modern radiosonde equipment is quite small and highly portable. For receiving the data, a quite small antenna is sufficient for receiving the transmission from a GPS radiosonde because the receiving antenna does not need to accurately track the direction and angle of the radiosonde. Also, with a small radiosonde, the balloon can start at <1 m diameter and can be inflated in almost any room with a door to the launch site.

Historical radiosonde sites may not now be identifiable simply because the buildings may no longer exist. Especially at commercial airports, expansion over the last few decades has obliterated many older facilities. Similarly, many former urban sites were at locations where the buildings no longer exist, although some stations up to over 100 years old have been accurately located because the buildings are historic enough that they have not been demolished or made unrecognizable.

Some examples. Only a few examples of accurately located current or historical stations are shown in the poster, and they are briefly summarized here.

The Qingdao (54857) location in the 1980s was located by a chance communication with Eckart Dege, who visited the location in 1988 and took some photos of a radiosonde launch. The exact site no longer exists, but triangulation from the photos can locate the site in Google Earth due to distinctive buildings in the vicinity, including a nearby surface observing site. Radiosondes are still launched nearby, and it is assumed that the current launches are from very near the surface observing plot because there is almost no surrounding open and unobstructed land. Two smaller photos show Google Earth photos of other sites, Nagqu (55299), and Pyongyang, North Korea (47058), both of which have weather stations with surface observing sites of the Chinese pattern.

Lindenberg (10393 or 09393) has made upper air observations of several types nearly continuously since 1905, and site photos and other documentation allow accurate determination of the launch sites. The first observations were from kites, and the original kite house has been restored at its original location. Also, the 1905 balloon inflation building is still used, where kites were stored and kite balloons were inflated for soundings when the wind was too weak to lift a kite. In addition, a second balloon inflation and kite storage building was built in 1936 for use when wind was strong from the east, because it was hazardous to use the original kite location, and the restored second balloon building was found in Google Earth photos. Also, its elevation matches radiosonde elevations from Oct 1993 to Jun 1994, so the 1905 balloon building may have been renovated the, with soundings made from the 1936 balloon building temporarily.

The Sterling, VA (72403) example is given because the NWS Forecast Office and sounding site moved twice since 2001 because of expansion of the Dulles Airport runway system. Google Earth photos show each of the three sites while they were in operation, the first site when it was abandoned and then later completely demolished, and the second site is now abandoned but not demolished. In addition, there is another radiosonde launch location near the third site, but that is an NWS test site that does not perform the operational launches.

BRIEF IMPLICATIONS FOR BIAS REMOVAL

This paper is not directly concerned with radiosonde homogenization methods, but it is desirable to correct individual soundings for biases so climate trends can be as accurate as possible at all scales from global down to individual stations.

There are quite a few documented methods to attempt to homogenize long radiosonde data series. A recent method is in Wang et al. (2020), which examines temperature means and variance (using individual soundings) to detect change-points. The process of accepting changepoints is partly automated, with considerable manual input because it is desirable that detected changepoints are supported by available metadata, but metadata is known to be incomplete and inaccurate. The bias corrections are basically determined by differences for several years before and after each accepted changepoint from the 20CR and JRA55 reanalyses. The differences do not have to be fitted to any standard distribution by using quantile matching.

This method should be moderately effective in correcting for artificial (instrument-related) biases without removing or distorting natural (real) trends and variations as long as the reanalyses themselves are not significantly contaminated by the changing biases in the input radiosonde data. (The 20CR model does not ingest any upper air data, but simulates the whole atmosphere by a dynamically consistent model that generates upper air conditions that are consistent with the input surface data.)

However, there are still some limitations in the bias removal by the Wang et al. (2020) method in all cases, and alternatively in a period when comparison data is not available, based on the conclusions of this research.

First, in the case where unbiased comparison data, such as from 20CR, which ingests only surface data, and if it is a valid conclusion that the input surface data is not biased, the changepoints may not be correctly identified when a station uses an instrument type for a short period (so there is not an adequate time before and after the change-point for a comparison), when the station uses multiple instrument types in the same period, or if a real discontinuity is not detected because it is small (unless such discontinuities average out to zero, but it is common for a detected bias to be gradually fixed by several small adjustments, so the small discontinuities may be in the same direction). Wang et al. (2920) find an average of 6.8 changepoints per station in a series ~60 years long, and it is probable that is an underestimate of the true number of changepoints that cause changes in the biases (they also estimate that 3.2 changepoints are changes in the variance only, meaning that only about 3.6 changepoints per station affect the mean long-term trend). In that case, many changepoints may be missed, and the cumulative bias of the missed changepoints is probably not negligible.

Second, if radiosondes are compared to satellite data, a satisfactory comparison time series cannot generally extend before 1979. In that case, when there is no comparison time series, quantile matching or a similar comparison at each station for a period before versus after a changepoint inevitably removes some portion of the long-range trend. This is discussed in Schroeder (2010). For example, if the periods 5 years before and after a changepoint are compared, the “before” period is the average centered on 2.5 years before the changepoint and the “after” period is centered 2.5 years after the changepoint, so the difference removes the instrument bias (hopefully) and probably 5 years of the actual trend.

THE TEXAS A&M UNIVERSITY METADATA FILE

The station and instrument histories, along with a large amount of supplementary analyses, are documented in a single text file. This file is quite incomplete as a part-time project, but as of February 2020, the Texas A&M University metadata file includes a documented list of 3515 upper air atmospheric profiling instrument codes including 1704 radiosonde codes. In most cases, each instrument code is a different permutation of any factors that can cause the instrument to be considered distinct, but this effort does not necessarily determine that the different varieties (for example, the only difference might be a different frequency band for data transmission) have distinctly different characteristics such as biases. Part of the documentation includes a description of how each instrument records or transmits data, and measures pressure, temperature, humidity, and wind. Currently there are lists including 24 recording/transmitting methods, 43 pressure measuring methods, 55 temperature measuring methods, 70 humidity or moisture measuring methods, and 56 wind measuring methods including 44 non-Navaid and 12 Navaid

In addition, there are 6068 numeric (fixed land location) upper air stations, 42 fixed ship station locations, and 653 mobile stations including ships, for a total of 6763 stations. (These “station” counts include name changes and distinct locations.)

Finally, the station histories list 54383 events, where each event is the establishment of a station, a change in the instrument type, a changed station location, or a station closure or at least the beginning of a substantial break in the data.

Acknowledgements

This project has been partially supported by TAMU funding supplied from multiple NASA and NSF research grants.

References

Finger, F. G., H. M. Woolf, and C. E. Anderson, 1965: A method of objective analysis of strato-spheric constant pressure charts. *Mon. Wea. Rev.*, **93**, 619-638.

Gaffen, D. J., 1993: *Historical Changes in Radio-sonde Instruments and Practices*. WMO Instruments and Observing Methods Report No. 50, WMO/TD-No. 541, 128 pp. + 4 microfiche [available online, except for microfiche, as of 2020 at https://library.wmo.int/doc_num.php?explnum_id=9592].

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 online in 2020 at ftp://ftp.ncdc.noaa.gov/pub/data/images/gaffen 1996.pdf. The station history file is at ftp://arlftp.arlhq.noaa.gov/web/climate/metadata/ascii/stationh.txt. The complete set of text files is obtained at ftp://arlftp.arlhq.noaa.gov/, then log in as a guest, and when the directories appear, click on "web", then "climate", then "metadata" (README summarizes the files), then "ascii". The list of radiosonde types is radioson.txt and the station history file is stationh.txt].

Huang, B., Y. Ma, and W. Yao, 2010: Indirect radiosonde intercomparison using First Guess fields. http://www.wmo.int/pages/prog/www/IMOP/publications/IOM-104_TECO-2010/P3_6_Huang_China.doc viewed Nov 2010.

Humphreys, W. J., 1929: Meteorology's frozen assets (Presidential address, AMS, New York City, 27 December 1928). *Bull. Amer. Meteor. Soc.*, **10**, 57-65.

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, https://ams.confex.com/ams/pdfpapers/119792.pdf.

Schroeder, S. R., 2008: Adjusting archived radiosonde data using complete validated and inferred radiosonde metadata to compute unbiased atmospheric temperature and moisture trends. *20th Conference on Climate Variability and Change*, New Orleans, LA, Poster P4.1, https://ams.confex.com/ams/pdfpapers/130838.pdf.

Schroeder, S. R., 2009: Homogenizing the Russian Federation upper air climate record by adjusting radiosonde temperatures and dew points for instrument changes. *21st Conference on Climate Variability and Change*, Phoenix, AZ, Paper 8B.6A, https://ams.confex.com/ams/pdfpapers/146252.pdf.

Schroeder, S. R., 2010: Implications of complex upper air histories on approaches to adjust archived radiosonde data for instrument discontinuities. *22nd Conference on Climate Variability and Change*, Atlanta, GA, Poster 82, https://ams.confex.com/ams/pdfpapers/163013.pdf.

Wang, J., C. Zhou, and A. Dai, 2020: A new method to homogenize atmospheric radiosonde daily temperature data. *33rd Conference on Climate Variability and Change*, Boston, MA, Paper 10A.3, https://ams.confex.com/ams/pdfpapers/366666.pdf.