82 IMPLICATIONS OF COMPLEX UPPER AIR HISTORIES ON APPROACHES TO ADJUST ARCHIVED RADIOSONDE DATA FOR INSTRUMENT DISCONTINUITIES

Steven R. Schroeder *

Texas A&M University, College Station, Texas

ABSTRACT

Radiosondes provide the longest time series of meteorological measurements above the surface, but discontinuities from frequent instrument changes at all stations contaminate atmospheric climate trends. As sensors have become more sensitive and better protected from radiative errors over time, the erroneous trend is hypothesized to be in the direction of artificial cooling and drying (although individual discontinuities can differ). Metadata describing upper air stations, instruments, and dates of changes, is incomplete and sometimes inaccurate, Trends adjusted for instrument discontinuities are questioned because researchers make adjustments while many instrument transitions and their dates are not known.

This study is part of a longer-term project to develop complete radiosonde metadata and instrument adjustments based on the completed station instrument histories (Schroeder 2009). The main finding from histories developed so far is that many station instrument histories are very complex, with dozens of transitions including many alternations between 2 to 8 instrument types, and few long periods with the same instrument type.

The complex station histories have implications for data adjustment methods. First, frequent alternations of instruments blur the discontinuity from a transition, sometimes for up to several years. Second, some adjustment methods are not designed to handle transitions less than about 2 years apart. Third, many of the instrument discontinuities are real but small and are not likely to be detected by automated discontinuity detection methods, but still have a noticeable cumulative effect on trends. Finally, the usual adjustment method, which equalizes the mean value of a variable for a specified averaging period before and after a discontinuity (so it can be called the "segment mean matching" method), inevitably removes a portion of the real trend, so when stations have many discontinuities, the adjustments can remove a large percentage of the real trend.

This study focuses on the last issue because the removal of some of the actual trend is not well publicized in accounts of upper air data adjustment projects. Any time series with artificial discontinuities for which overlapping data is not available is subject to the same type of unintentional trend removal. This paper estimates the theoretical proportion of a long-term trend that is likely to be removed by a "segment mean matching" adjustment method and shows some simple illustrations.

1. INTRODUCTION

Two requirements for proper data adjustments to allow correction of erroneous trends are complete and accurate metadata (including a list of locations and instruments with dates at all stations) and a suitable data adjustment method. Available metadata (such as from the Integrated Global Radiosonde Archive, http://www.ncdc.noaa.gov/oa/climate/igra/index.php) is incomplete and sometimes incorrect, but researchers still develop adjusted data series by using various methods to attempt to detect undocumented discontinuities. When discontinuities are detected, the usual adjustment method, which can be called "segment mean matching," equalizes the mean values in a specified averaging period before and after each discontinuity.

Previous work in this project (Schroeder 2007, 2009, 2010) has found that complete and quite accurate station histories of instrument changes can be constructed by systematically searching station time series of sensitive variables that amplify differences between instrument types. Consistent data signatures in multiple variables are seen when the same instrument type is used at individual stations, with smooth variations with level and season, and discontinuities indicate instrument changes. While a signature of an instrument type is derived using data from stations where metadata appears reliable, the same signatures are seen at stations or in periods when metadata is questionable or not available. So, these signatures can be used to validate available metadata, correct errors, and construct missing metadata.

* *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.

Data and metadata sources and the procedures to develop complete metadata are described in more detail in the papers referenced above. To summarize the metadata, this project started with the metadata of Gaffen (1993, 1996), and adds a large amount of additional information from many other sources such as WMO upper air station catalogs (http://www.wmo.ch/pages/prog/www/ois/volume-a/vola-home.htm). The added metadata clarifies many periods where information is not available in the Gaffen metadata files, but also reveals additional inconsistencies that need to be resolved.

Station metadata elements besides instruments (station ID, name, location, elevation, periods of station operation, and the date of each change) are validated first since station changes can cause data discontinuities. Accurate surface elevation histories can be constructed for all stations with temperature soundings and reported heights, but validation of other station metadata elements is more limited.

Many current or recent station locations can be accurately identified using online high-resolution aerial or satellite photos, but some upper air stations do not have a distinctive enough appearance to be confidently confirmed. Most past locations cannot be directly validated (but some are labeled on topographic maps), and it is more difficult than originally thought to associate past reported locations with computed elevation changes because catalogs have a very large number of location disagreements. However, most location errors are probably minor (no more than several km), and the effect of location or elevation errors on global upper air trends should be negligible.

After validating station information as much as possible (although these steps are not precisely sequential; information found later can be used to make corrections), time series of sensitive variables are examined at each station to develop complete instrument histories. Most sensitive variables are moisture-related, and many of those variables are not of climate interest (such as the lowest humidity reported in each sounding, or the lowest temperature for which a dew point depression is reported). However, the purpose of this procedure is solely to identify instrument types, and developing data adjustments is performed after the instrument histories are produced.

In general, available metadata is sufficient to narrow down the number of candidate instrument types, so an inferred instrument type in a period without specific metadata is not likely to be seriously wrong. The main uncertainty is distinguishing instrument types in the same family, such as MARS-2-1 and MARS-2-2. However, in most cases there is no difference in sensors within a family, and the differences are internal electronics that have no detectable effect on data.

Of course, inferred metadata will not be perfect, but the initial instrument assigned should not be significantly different in data characteristics from the actual instrument because if instruments are hard to distinguish using sensitive variables, differences between the instruments are even smaller in variables of climate interest such as temperatures at specific pressure levels. A significant incorrect identification of an instrument should be seen as an uncorrected discontinuity in the adjusted data. Over time, such data checks and comparisons with other sources (such as satellite retrievals) should allow the inferred instrument histories to become more and more accurate.

The main finding from station histories developed so far is that many station histories are very complex with frequent closely-spaced transitions and only a few periods where the same radiosonde type is used for several years or more.

For example, in the Russian Federation, at least 37 out of about 180 stations with long data records were found to have 40 to over 200 transitions from 1973-2006, 49 out of 122 stations with extensive data reported 4 to 8 radiosonde types in 2007-2009, and at least 15 radiosonde models are currently in use from 2007 to the beginning of 2010.

Some of the station instrument history complexity arises because many previously-undocumented upper air instrument types have been identified in literature searches. Some radiosonde types commonly thought to be homogeneous have had changes made during production that have effects on data. For example, the Japanese Meisei RSII-91 has had 4 varieties: The original 1991 version, a new thermistor in 1994, a new humidity sensor in 1999 to correct a dry bias (but the change was later noticed to cause an intermittent moist bias), and a software change in 2003 to correct the moist bias (Ishihara 2004). As of early 2010, the Texas A&M University metadata file includes a documented list of over 2600 upper air atmospheric profiling instrument types.

Because station instrument histories tend to be quite complex with many transitions, many more adjustments need to be made to each radiosonde station time series than in projects published so far. While several implications of complicated instrument histories on data adjustment methods are mentioned in the Abstract, this paper focuses only on the fact that the usual data adjustment method, the “segment mean matching” method to equalize the means of a variable in specified averaging periods before and after each discontinuity, inevitably removes some of the real trend as well as the discontinuity. With a large number of adjustments at many stations, the "random walk" errors of the adjustment process grow rapidly going back in time, some transitions are so closely spaced that this method cannot be used to derive the adjustments, and a large portion of the real trend is expected to be removed.

Section 2 discusses the process of developing and applying the "segment mean matching" adjustment method in simplified mathematical terms and computes the expected proportion of the real trend removed by this adjustment process. Section 3 shows simple examples using idealized and realistic (but simulated) data of the effects on a trend, first when there is one discontinuity to be adjusted, and second when there are three discontinuities. Section 4 proposes some possible ways to derive instrument adjustments besides equalizing the means for periods before and after each instrument change.

2. PARTIAL TREND REMOVAL BY "SEGMENT MEAN MATCHING" DATA ADJUSTMENT METHOD

An observed data time series of any type contains variations at many time scales. Here, the observed values are decomposed into a constant value, long-term trend, instrument-type bias, cyclical and irregular variations, and "residual" day-to-day or instantaneous variations, as follows:

$$y\,(t) = A + B\,(t) + C\,(t) + D\,(t) + E\,(t), \qquad (1)$$

where

t = time, from $t = 0$ to T

A = constant

$B\,(t)$ = underlying long-term trend

$C\,(t)$ = instrument-type biases, assumed 0 after the last instrument transition

$D\,(t)$ = cyclical and irregular variability, such as ENSO and the annual cycle

$E\,(t)$ = "residual" day-to-day or instantaneous fluctuations (depending on the measurement interval)

In the most general case, it is assumed that $y\,(t)$ is measured continuously, although real data is usually measured at discrete and possibly irregular intervals. Depending on the length of the total time interval T, in different circumstances the separation of data values into components of the underlying long-term trend, cyclical and irregular variability, and residual fluctuations can use different definitions. At least, here the underlying trend is generally assumed to have a fairly simple form, but it does not have to be linear or even monotonic during the whole time interval. Finally, an instrument-type bias is assumed to be steplike during the period of that instrument type, but it does not actually have to be constant.

Here, a very simplified analysis is performed by assuming that A, $D\,(t)$, and $E\,(t)$ are filtered out (set to 0), that the long-term trend is linear (starting with 0), and that there is only one instrument transition with a constant bias c before the transition time t_1.

So, the true (unbiased) data series contains only the long-term trend, or

$$y_{TRUE} = B\,(t) = bt, \quad t = 0, \ldots, t \qquad (2)$$

Defining the magnitude of the trend as the ending minus starting value of the long-term trend component, the trend has a magnitude of bT.

The observed series with a constant bias c (it does not matter if c is negative or positive) before the transition is

$$\begin{aligned} Y_{OBS} &= B\,(t) + C\,(t) \\ &= bt + c, \quad t = 0, \ldots, t_1 \\ &\quad\; bt, \qquad\; t = t_1, \ldots, T \end{aligned} \qquad (3)$$

Where the averaging periods have length T_a both before and after the transition time t_1 (it is assumed that the transition is not too close to either endpoint), the means in the "after" and "before" averaging periods are as follows:

$$\begin{aligned} Y_{AFTER} &= \overline{bt}, \quad t = t_1, \ldots, t_1 + T_a \\ &= [\, bt_1 + b\,(t_1 + T_a)\,]\;/\;2 \\ &= bt_1 + (b\,/\,2)\,T_a \end{aligned} \qquad (4)$$

$$\begin{aligned} Y_{BEFORE} &= \overline{bt + c}, \quad t = t_1 - T_a, \ldots, T_a \\ &= [\, b\,(t_1 - T_a) + c + bt_1 + c\,]\,/\,2 \\ &= bt_1 + c - (b\,/\,2)\,T_a \end{aligned} \qquad (5)$$

The adjustment is $C'\,(t) = Y_{AFTER} - Y_{BEFORE}$, added to the observed data values in the period up to t_1:

$$C'\,(t) = [\, bt_1 + (b\,/\,2)\,T_a\,] - [\, bt_1 + c - (b\,/\,2)\,T_a\,]$$

$$= bT_a - c, \quad t = 0, \ldots, t_1$$
$$0, \qquad t = t_1, \ldots, T \qquad (6)$$

Note that the adjustment should be equal to the negative of the bias, or $-c$, but the amount of the trend in the length of an averaging interval is added. The erroneous amount added becomes smaller as the averaging period T_a is shortened, but when real data is used, it is not feasible to have a short (or zero-length) averaging period.

The adjusted series is then

$$Y_{ADJUST} = Y_{OBS} + C'(\mathrm{t})$$

$$= bt + c + bT_a - c = bt + bT_a, \quad t = 0, \ldots, t_1$$
$$bt, \qquad t = t_1, \ldots, T \qquad (7)$$

The final long-term trend is then equal to the starting minus ending value of the long-term trend component of Y_{ADJUST}, or $bT - bT_a$. This means that the proportion of the long-term trend removed by this adjustment is equal to the ratio of the averaging period to the total length of the time series, or T_a / T. For example, if a 40-year trend is computed, and the averaging periods before and after a discontinuity are each 2 years long, the adjustment tends to remove 5 percent of the true long-term trend.

Note also that the proportion of the long-term trend removed does not depend on the size of the discontinuity in this idealized example.

Of course, the example above is extremely oversimplified and the adjustment derived above would immediately be recognized as not properly correcting the discontinuity. However, the simple mathematical development above generalizes to a realistic data series where the true trend is unknown because of superimposed variability at all observed time scales.

First, the long-term trend component $B(t)$ does not need to be linear. In that case, the long-term trend component from the average of the "before" to the average of the "after" period is removed by the adjustment process. The generalization is that if one adjustment is made per station for a large number of stations, with transitions occurring at random times at different stations, the expected proportion of the long-term trend removed from an appropriate average over all stations is still T_a / T.

Second, the averaging period does not need to have the same length before as after the transition, possibly due to sparse data either before or after the transition. However, both averaging periods should contain an integer number of cycles (such as years) so the climate circumstances before and after the transition are matched as well as possible. The expected amount of the long-term trend removed is still the long-term trend component from the average of the "before" to the average of the "after" period, or T_a / T if T_a is the average length of the averaging periods.

Third, there can be a gap between the "before" and "after" averaging periods, such as if the station closed or used a different instrument type between the averaging periods of the instruments being compared. In that case, the expected amount of the long-term trend removed is the long-term trend component between the midpoints of the "before" and "after" averaging periods.

Fourth, if an individual time series is shorter than the total study period, an adjustment removes the same number of years of the true trend as at a station with a full-length data record. While this is a greater proportion of the total trend at a station with a short record, when this station and others are appropriately averaged to develop a long-term trend for a region or the world, the expected proportion of the total trend removed by this adjustment is the same (T_a / T) as for a station time series with the full length.

Fifth, some methods determine an adjustment as the average difference between averaging periods before and after the same transition type at multiple stations, with individual station transitions possibly occurring at different times. The adjustment may be determined using some but not all stations with the same transition type (for example, excluding stations with a transition too close to the beginning or end of the time series for satisfactory averaging on both sides of the transition), and then applied to all stations with that transition. Such a composite adjustment may undercorrect or overcorrect discontinuities at some stations, but the expected proportion of the long-term trend removed is the same as in the case where separate adjustments are determined at each station.

Sixth, suppose that a station time series has more than one transition, and that each transition is separated by an interval at least as long as an averaging time. The calculations to derive and apply adjustments are performed for each transition from the latest to the earliest, and the effects of all adjustments on the period up to the first transition are additive. So, if n adjustments are performed and all of the averaging periods have length T_a, the total proportion of the trend expected to be removed is the sum of the proportions removed by each adjustment, or nT_a / T. In the extreme case of transitions separated by intervals of T_a (for example, at times 2, 4, 6, . . . , 18 in a 20-year record with a 2-year averaging time before and after each transition), only $1 / n$ of the original trend is expected to remain after the adjustments are applied.

Finally, to detect discontinuities in real data, the data is not usually filtered except possibly to express data values as anomalies, so $D(t) + E(t) \neq 0$. While the average of $D(t) + E(t)$ may be 0 over the entire data period, if an averaging period before or after a transition coincides with an extreme event, the mean of these data components can be significantly different from 0. Since the computed adjustment is the difference between the "before" and "after" periods, which includes all components of variability, an individual adjustment can be very far from correcting a specific instrument-related discontinuity. However, the generalization over a large number of cases is that the expected proportion of the long-term trend removed is still the long-term trend component between the midpoints of the "before" and "after" averaging periods.

A separate issue is that if the time of an adjustment is not a correct instrument transition time, making an adjustment still removes the same proportion of the real trend while leaving the instrument discontinuity either partially corrected or not corrected. It is likely that many published upper air adjustment projects have some spurious adjustments to remove natural discontinuities because, without complete metadata, some incorrect breakpoints are assigned.

Considering the idealized calculations and generalizations to real data above, the expected proportion of the real trend removed by the "segment mean matching" method can be summarized in the following 3 hypotheses:

Hypothesis 1: *When a discontinuity in a time series is adjusted by equalizing the means for a specified period before and after the discontinuity, the expected proportion of the real trend removed is*

(midpoint of averaging period before - midpoint of averaging period after) / (total time series length).

Hypothesis 2: *If there is more than one discontinuity, the expected proportion of the real trend removed by all adjustments is the sum of the amounts removed by each adjustment.*

Hypothesis 3: *If the time of a discontinuity is not identified correctly, the same proportion of the real trend is removed and the artificial discontinuity is either partially or not corrected.*

3. SIMULATED EXAMPLES OF PARTIAL TREND REMOVAL

Figures 1 to 4 show simulated examples of data with a linear trend and one discontinuity. In each figure, the left portion is an idealized example containing only an underlying trend and possibly a discontinuity and an adjustment, with no fluctuations, as in equations (1) to (7). The right portion shows simulated random data added to the underlying trend, with the same procedure used to develop and apply adjustments.

The left portion of Fig. 1 shows the underlying trend and the right portion shows a simulated data series with the trend incorporated. The amount of lag correlation included is probably appropriate if the data represents monthly averages of a typical atmospheric variable.

In Fig. 2, a bias of -0.4 unit is added to the first half of each data series from Fig. 1. The underlying trend then becomes 1.4 unit in 100 time periods, or 1.0 – (-0.4). If the trend is determined from the slope of a least squares line, the slope can differ from 1.4 unit per 100 time units. For example, a least squares line fitted to the biased line in the left half of Fig. 1 has a slope of 1.8 units per 100 time units.

Note that in the right portion of the figure, neither the unbiased or biased time series shows an obvious discontinuity around time 50. This specific simulation appears more likely to have a discontinuity around time 64. While the imposed discontinuity is 0.4 unit (or about 40% of a hypothetical monthly standard deviation of this variable), in temperature averages at various levels most actual instrument discontinuities are probably smaller than 0.4 of a typical monthly standard deviation.

Fig. 3 illustrates the process of deriving the adjustment, where it is assumed that the instrument discontinuity has been correctly determined to occur at time 50. With an averaging period length T_a = 20 units in this example, the “before” average from time 30 to 50 in the “true” trend with a discontinuity (the left half of Fig. 3) is Y_{BEFORE} = 0.0 and the “after” average from time 50 to 70 is Y_{AFTER} = 0.6, giving an adjustment c = 0.6 – 0.0 = 0.6. The “segment mean matching” adjustment then adds 0.6 to data values before time 50.

In the right half of Fig. 3, the adjustment is computed in the same way. The “before” average is Y_{BEFORE} = -0.0538 and the “after” average is Y_{AFTER} = 0.5461, giving an adjustment c = 0.5461 – (-0.0538) = 0.5998. This simulation was chosen to have an adjustment as close to 0.6 as possible, but in 320 simulations the computed adjustments were

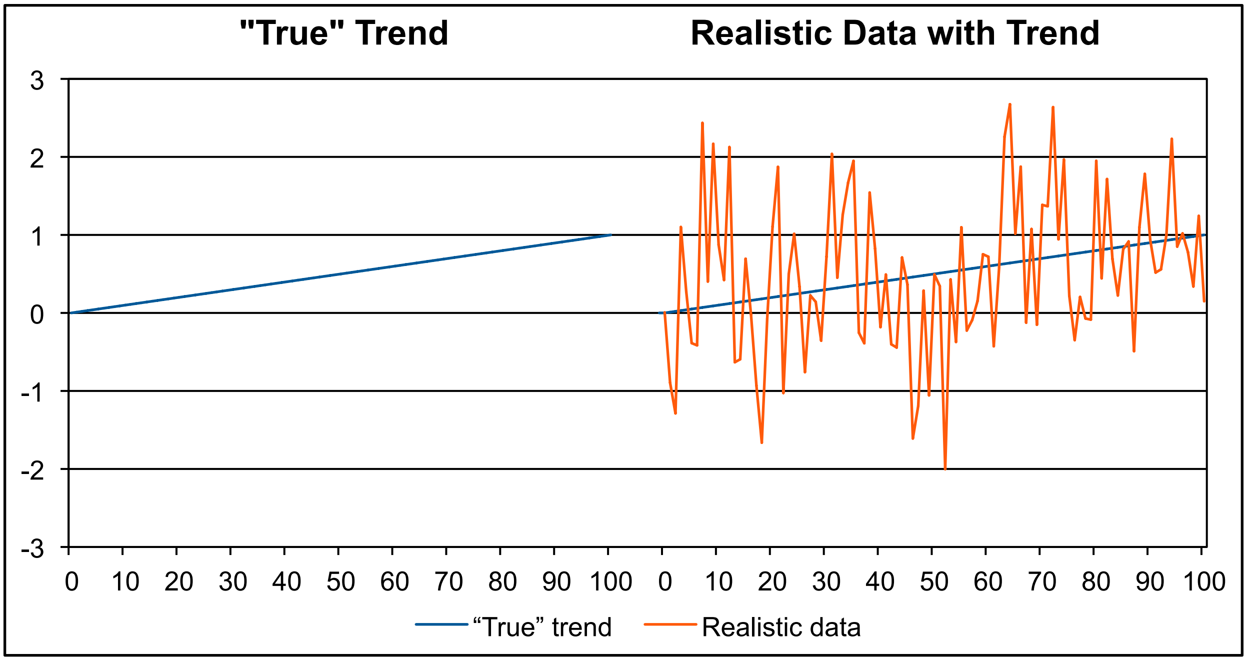

Fig. 1. A very simple example of a linear trend. The left portion shows a hypothetical trend of 1unit in 100 time units. The right portion shows random fluctuations superimposed. The fluctuations were drawn from a normal distribution with a standard deviation of 1.0. Then, each value after the first is 0.2 * (previous value) + 0.8 * (new random number) to provide a moderate amount of serial correlation. Then, each value is multiplied by *sqrt* (1 / [1 – 0.8]) to increase the standard deviation back to approximately 1.0, and finally the linear trend is added.

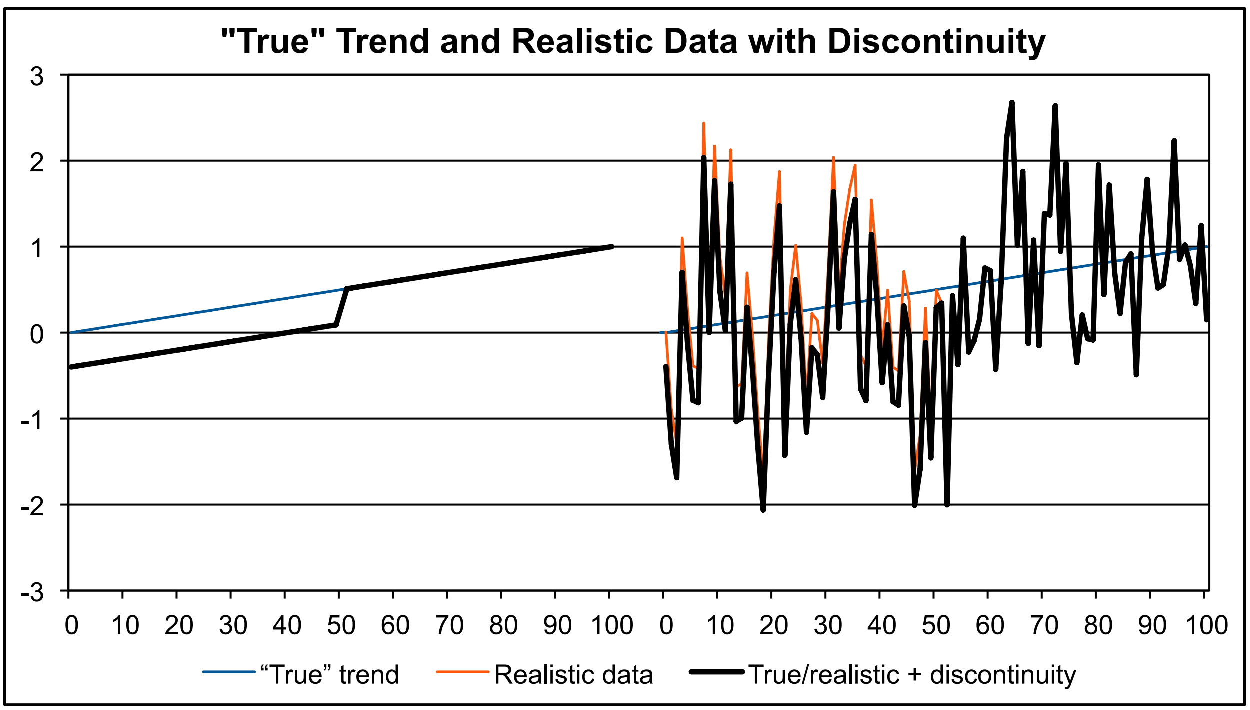

Fig. 2. Data of Fig. 1 with a bias of -0.4 unit added from times 0 to 50. The thick lines show the biased values, the blue lines show the unbiased trend, and the red line in the right portion shows the original unbiased “realistic” data from Fig. 1.

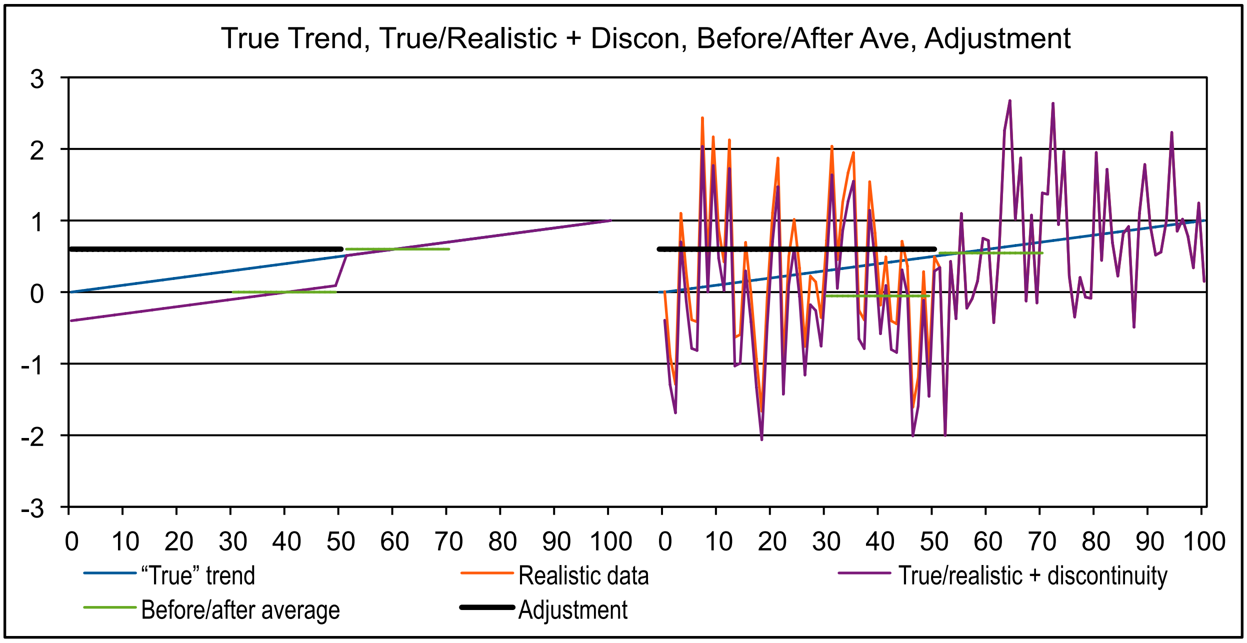

Fig. 3. Development of the adjustment for the discontinuity. The purple line shows the biased time series (corresponding to the thick line in Fig. 2). The green lines show the "before" and "after" averages in averaging periods of 20 time units before and after the discontinuity. In the left half, the average from time 30 to 50 is 0.0, and from time 50 to 70 it is 0.6, so the adjustment (thick line) is the after minus before average (0.6 – 0.0 = 0.6) added to data values before time 50. In the right half, the average from time 30 to 50 is -0.0538 and from time 50 to 70 is 0.5461, so the adjustment is 0.5461 – (-0.0538) = 0.5999, also to be added to data values before time 50.

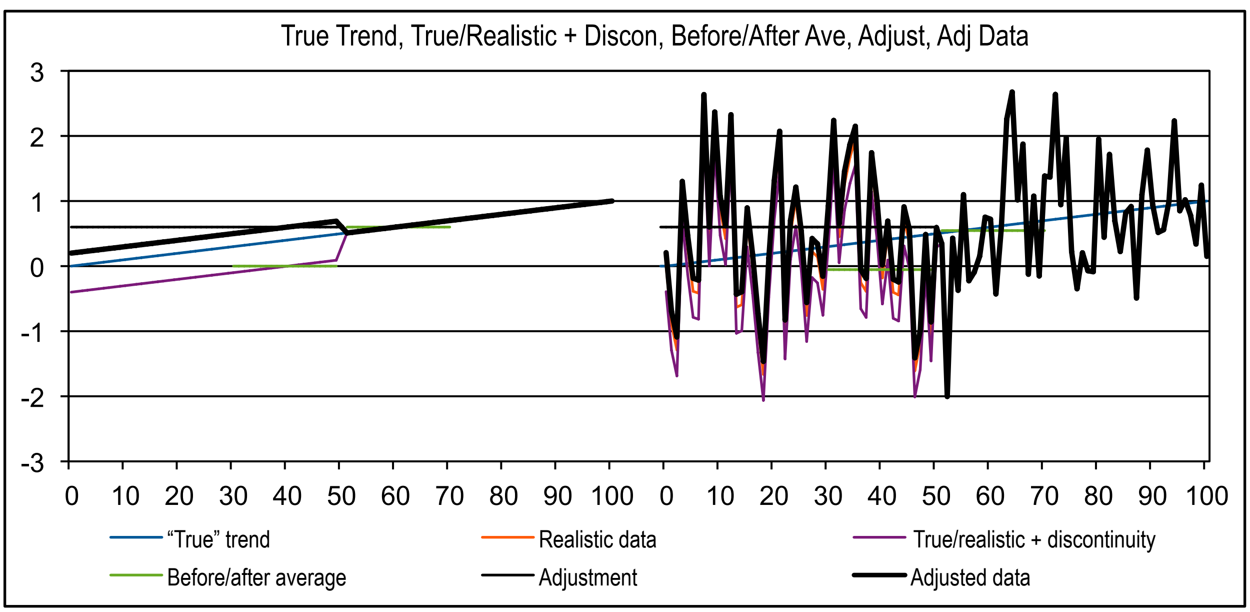

Fig. 4. Application of the computed data adjustment. The blue, red, purple, and green lines have the same meanings as in previous figures. The thick line is the adjusted time series. In the left half of the chart, it is obvious that the adjustment of 0.6 is 0.2 unit too high. With an ending value of 1.0 and a starting value of 0.2, the trend is 1.0 - 0.2 = 0.8, so 20 percent of the underlying trend has been removed. In the right half, it is not obvious that the adjustment is similarly too high and no discontinuity is obvious around time 50 in either the biased or adjusted series.

as small as -0.5704 or as large as 1.5399. The reason for the large variation is the range of values in the averaging periods before and after the discontinuity date. In real data, an extreme event such as a volcanic eruption may occur near the time of an instrument discontinuity, so an adjustment would project that anomaly to some extent into the entire period before the instrument change.

The thick lines in Fig. 4 show the adjustments added to the lines in Fig. 3. In the left half of the chart, it is obvious that the discontinuity is overcorrected. Since the starting value of the adjusted time series is 0.2, the trend becomes 1.0 – 0.2 = 0.8 unit per 100 time units. (A least-squares line would have a slope of only 0.6.)

In the right half of the chart, it is not obvious whether the discontinuity is corrected or overcorrected (or undercorrected) since the original data series does not have an obvious discontinuity. Extreme adjustments in some simulations (not shown) obviously introduce a discontinuity.

Figures 5 to 8 correspond to Figs. 1 to 4 but have 3 discontinuities at times 25, 55, and 75, and use a different simulated random series added to the unbiased linear trend and the trend with discontinuities. Note that, unlike Fig. 1, the simulated random series in the right half of Fig. 5 does not show any obvious discontinuities, although it does have some quite large variations.

In Fig. 6, the absolute biases are -0.4 unit from time 0 to 25, -0.58 from time 25 to 55, and +0.15 from time 55 to 75, so the discontinuities are -0.18 unit at time 25, +0.73 at time 55, and -0.15 at time 75. The trend including the biases is the same as in Fig. 2, or 1.4 units per 100 time periods, because the starting and ending data values are the same. Again, the thick line in the right half of Fig. 6 is the unbiased data simulation with the same biases added as in the left half of Fig. 6. As in Fig. 2, the discontinuities are not obvious (even though the discontinuity at time 55 is quite large), so it is assumed that the instrument transition dates are determined by other methods.

In Fig. 7, the adjustments are derived in the same way as in Fig. 3, from the latest to the earliest discontinuity, and the cumulative adjustment at any time is the sum of adjustments for all discontinuities after that time.

For the trend with 3 discontinuities (the left half of Fig. 7), around time 75 the "before" average from time 55 to 75 is Y_{BEFORE} = 0.80 and the "after" average from time 75 to 95 is Y_{AFTER} = 0.85, so the adjustment is c_1 = 0.85 – 0.80 = 0.05, to be added to data values from times 55 to 75.

Around time 55, the "before" average from time 35 to 55 is Y_{BEFORE} = -0.13 and the "after" average from time 55 to 75 is Y_{AFTER} = 0.80 . Note that the "after" averaging period for the second discontinuity is the same as the "before" averaging period for the third discontinuity. The discontinuity is 0.80 – (-0.13) = 0.93, so the adjustment (added from time 25 to 55) is $c_2 = c_1$ + 0.93 = 0.05 + 0.93 = 0.98.

Around time 25, the "before" average is Y_{BEFORE} = -0.25 and the "after" average is Y_{AFTER} = -0.23 so the discontinuity is 0.02. The adjustment added to data values before time 25 is $c_3 = c_2$ + 0.02 = 1.0.

Note that, due to the inclusion of the long-term trend in the averaging periods around each discontinuity, each discontinuity is considered to be 0.2 unit too high, so the final adjustment is 1.0 (added to times 0 to 25) instead of 0.4.

For the realistic simulation with the same 3 discontinuities added, shown in the right half of Fig. 7, around time 75 the "before" average from time 55 to 75 is Y_{BEFORE} = 0.2468 and the "after" average from time 75 to 95 is Y_{AFTER} = 0.9661, so the adjustment is c_1 = 0.9661 – 0.2468 = 0.7194. This adjustment is to be added to data values from times 55 to 75.

Around time 55, the "before" average from time 35 to 55 is Y_{BEFORE} = -0.2487 and the "after" average from time 55 to 75 is Y_{AFTER} = 0.2468. The discontinuity is 0.2468 – (-0.2487) = 0.4955, so the adjustment (added from time 25 to 55) is $c_2 = c_1$ + 0.4955 = 0.7194 + 0.4955 = 1.2149.

Around time 25, the "before" average is Y_{BEFORE} = -0.0345 and the "after" average is Y_{AFTER} = -0.2501 so the discontinuity is -0.2156. The adjustment added to data values before time 25 is $c_3 = c_2$ – 0.2156 = 0.9994.

The simulated "realistic" case in Figs. 5 to 8 was chosen to have a cumulative adjustment (added before time 25) as close to 1.0 as possible. In 320 simulated cases, the cumulative adjustment varies greatly from -0.6126 to 2.0928.

In the left half of Fig. 8, the adjustments are added to the true trend with biases, and it is obvious that much (specifically 60%) of the true trend has been removed. The starting value of the adjusted line is 0.6 and the ending value is 1.0, so trend has been reduced from 1.0 to 0.4 units per 100 time units.

In the right half of Fig. 8, the adjustments are fairly large and positive, so all data values before time 75 are increased by from 0.7194 to 1.2149 units. Again, the adjusted line contains little obvious trend. However, other simulations can show a substantial trend in either direction and some of them show noticeable discontinuities, due to possible anomalies occurring during one or more

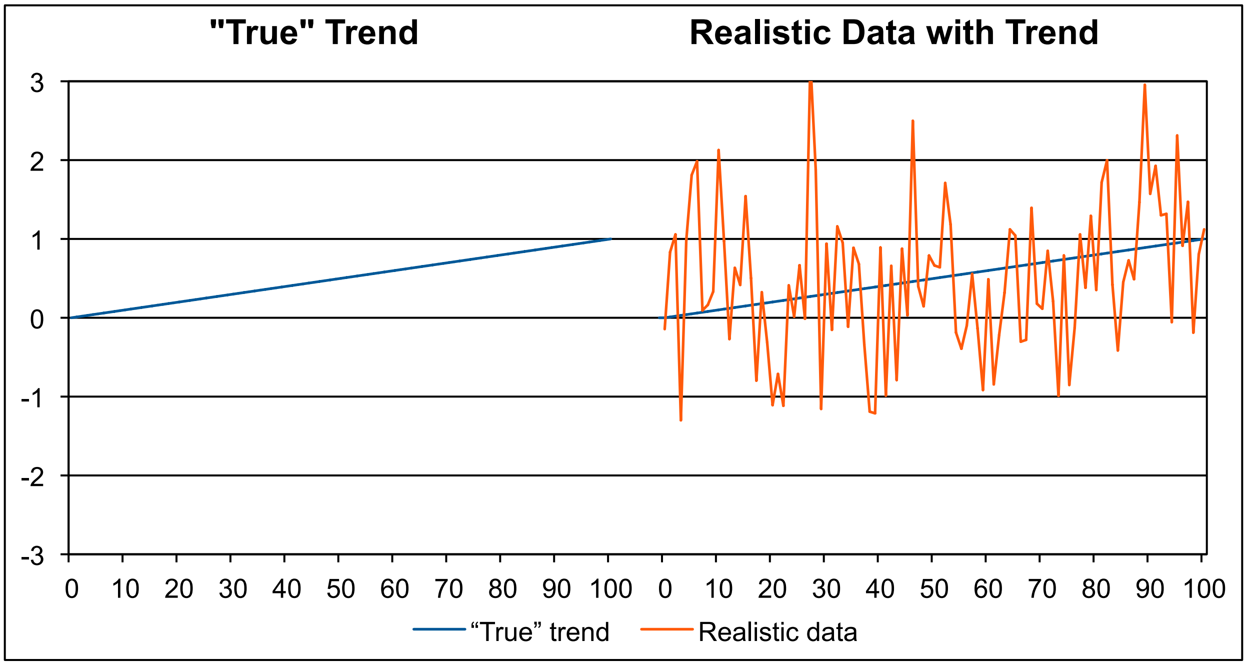

Fig. 5. The blue line in each half of the figure shows the same idealized linear trend as in Fig. 1 with a magnitude of 1.0 unit in 100 time units, and the red line in the right portion is a simulated random time series computed in the same way as described with Fig. 1, but with a different set of random numbers.

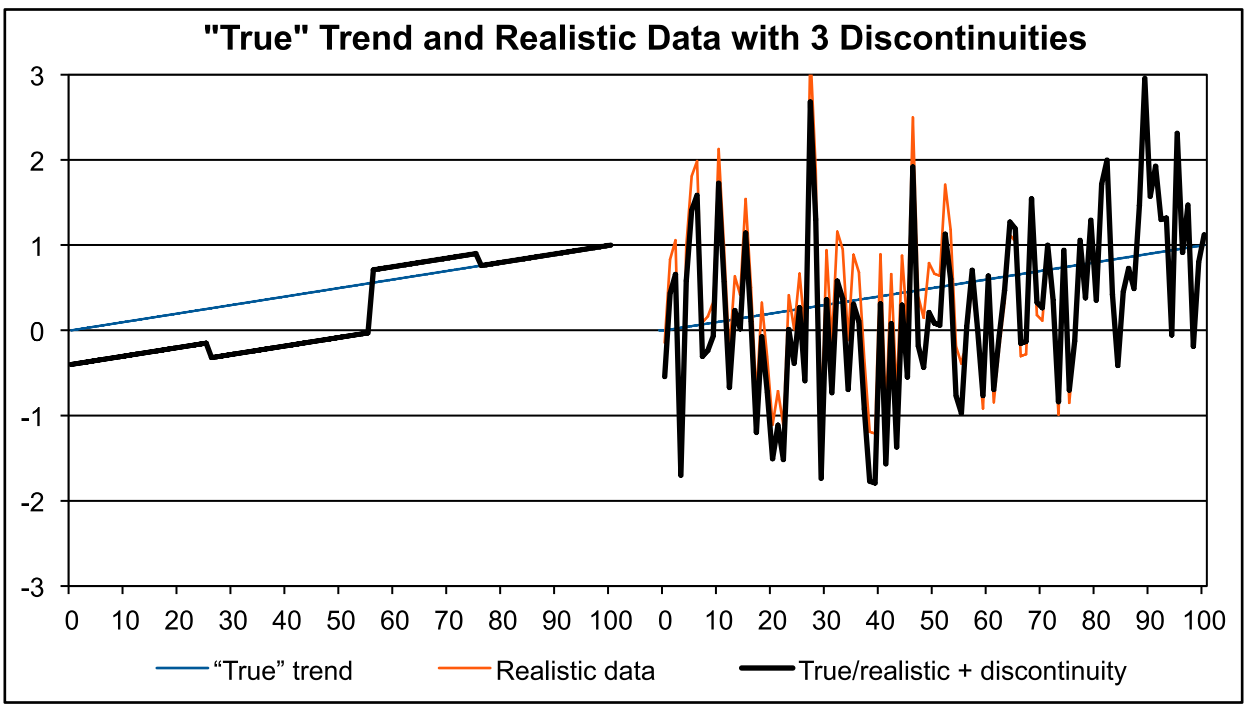

Fig. 6. Data of Fig. 5 with three biases of -0.4 unit added from times 0 to 25, then -0.58 to time 55, then +0.15 to time 75. The thick lines show the biased values, the blue lines show the unbiased trend, and the red line in the right portion shows the original unbiased "realistic" data from Fig. 5.

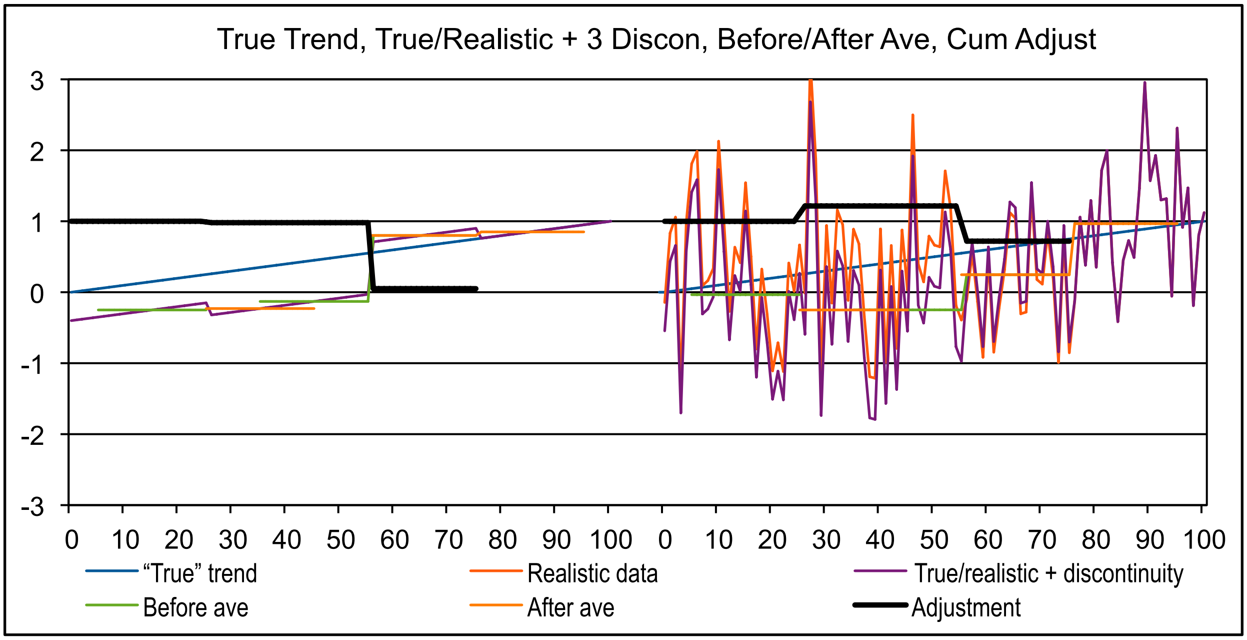

Fig. 7. Development of the adjustment for the discontinuities. The purple line shows the biased time series (corresponding to the thick line in Fig. 6). The green and orange lines show the "before" and "after" averages in averaging periods of 20 time units before and after the discontinuity. The computation of each adjustment (from the latest to the earliest discontinuity) is described in the text. The thick line shows the cumulative adjustment, which is the latest adjustment (to be added to values from times 55 to 75), then the sum of the latest 2 adjustments (for times 25 to 55), and finally the sum of the three adjustments (to be added to data up to time 25).

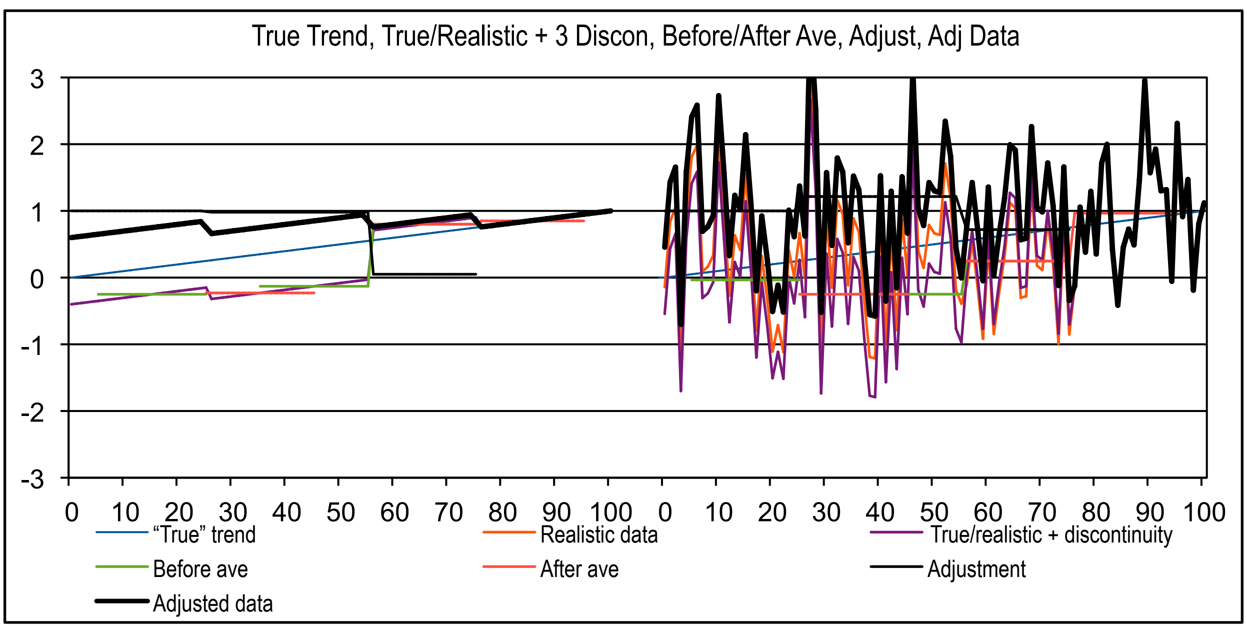

Fig. 8. Application of the computed data adjustment. The blue, red, purple, green, and orange lines have the same meanings as in Fig. 7. The thick line is the adjusted time series. In the left half of the chart, it is obvious that the adjustment of 1.0 before time 25 is 0.6 unit too high. With an ending value of 1.0 and a starting value of 0.6, the trend is 1.0 – 0.6 = 0.4, so 60 percent of the underlying trend has been removed. In the right half, it is not obvious that the adjustments are similarly too high because there is no obvious discontinuity in either the biased or adjusted series, but the adjusted series has no obvious trend at all.

of the averaging periods. In this hypothetical example, the frequent transitions cause much of the entire time (all times from 5 to 95) to be in an averaging period, so almost any anomalous period will affect at least one instrument adjustment.

In most cases, published upper air data adjustments still show nonnegligible trends. This probably occurs because researchers make no more than a few adjustments per station, and because subjectivity is used to decide when to make an adjustment. Almost all researchers use published incomplete metadata, and possibly obtain a few additional metadata events by personal communications. In periods where metadata is not available at a station, this subjectivity understandably leads to caution in deciding that a variation needs to be adjusted, so the final adjusted data tends to produce trends that are similar to the expected trend.

4. PROPOSED ALTERNATIVE COMPARISONS FOR MAKING ADJUSTMENTS

The usual "segment mean matching" method of deriving a data adjustment to compensate for a discontinuity has been shown to be likely to remove a significant portion of the real trend when applied to archived radiosonde data because of the large number of instrument changes that have occurred. Hovever, an individual adjustment can be spuriously large in either direction if an extreme phase of ENSO, a volcanic eruption, or other unusual event occurs within an averaging period before or after the transition. For example, Japanese radiosonde transitions from Meisei RSII-56 to RSII-80 occurred within about 2 years after the eruption of El Chichon and many of the station transitions from RSII-80 to RSII-91 occurred within 2 years after the Pinatubo eruption.

The reason the "segment mean matching" method is most commonly used to make data adjustments is that the ideal comparison method, which is to use more than one instrument type in the same environment for an extended period, is not performed by operational radiosonde stations. Radiosonde intercomparisons are made occasionally, but involve no more than a few dozen flights per instrument type, so they are not numerous enough for developing adjustment factors to be applied in a wide range of operational situations.

However, if an instrument adjustment is to be derived for each instrument transition type (such as from the Russian MARS to MRZ instruments), the complexity of actual instrument histories can usually provide data comparisons that are less vulnerable to removal of a significant proportion of the actual trend. Such comparisons are nearly identically affected by the long term trend, and they are also nearly identically affected by unusual events, so the derived adjustments tend to filter out the effects of such anomalies. To make one instrument type statistically equivalent to a "target" instrument, both instruments are compared in carefully matched circumstances, and the adjustment (added to readings from the instrument to be adjusted) is the target instrument reading minus the reading of the instrument to be adjusted.

The first alternative comparison type is when both instrument types frequently alternate at the same station for an extended period. Both instruments should then experience the long-term trend (and anomalous events) almost identically. The averaging period could be the period of alternating use of the instruments, but instead of "before" and "after" averages, averages are computed for soundings with each instrument type. Of course, the instrument type of each sounding needs to be known. In addition, each instrument needs to be used approximately equally in all environments, or may need to be time-weighted for consistent representation of different environments. For example, the comparison is degraded if one instrument type is used only in daytime and the other is used only at night.

The second alternative comparison type is use of different instruments at nearby stations for an extended period. Again, both instruments should experience the long-term trend and anomalous events almost identically. Of course, the stations need to be in the same climate environment as much as possible, such as when the stations are nearly directly on an east-west line. Examples of possible station pairs are Berlin and Lindenberg, Germany (about 1958 to 1983), Hong Kong and nearby China, and quite a few stations in east-west lines in Russia or the United States. In the United States, there have only been limited periods when adjacent stations were likely to use different radiosondes, but in the Russian Federation, it has almost always been common for adjacent stations to use different instruments. Berlin and Lindenberg allow comparison of certain Russian and East German instruments with contemporary VIZ models, and Hong Kong provides the only nearby comparison of British (or Vaisala since the early 1980s) instruments with Chinese radiosondes. Comparisons involving different stations are not as exact as comparisons at the same station because the climate environment is not exactly the same, but for some instruments this type of comparison is the only type available. In addition, the difference between stations in the same period may be smaller than the likely difference in the long-term trend between "before" and "after" averaging periods.

If neither of these comparisons is available, it is sometimes possible to find stations that change instruments in the opposite order. Stations have relatively frequently changed from VIZ to Vaisala or from Vaisala to VIZ. If "segment mean matching" is used for each individual comparison, the VIZ data averages are obtained in the "before" averaging period at stations changing from VIZ, and in the "after" averaging period at stations changing from Vaisala, so in the composite average of differences between those particular VIZ and Vaisala models, the embedded effects of the long-term trends should largely cancel out. Depending on how the transitions in both directions are spread out in time, this comparison type may or may not filter out the effects of ENSO and other anomalous events.

Finally, some adjustment methods detect discontinuities by comparing a data time series with an independent data source such as a satellite retrieval or a model estimate. If the independent data source is rigorously related to the radiosonde variable to be adjusted, and is itself free from artificial discontinuities and trends (a satellite retrieval rarely meets either requirement), the adjustment (added to the radiosonde data before the discontinuity) is the change in the difference of the radiosonde variable from the external variable after minus before the discontinuity. Such an adjustment should largely be unaffected by the long-term trend. However, the data sources are likely to be affected differently by anomalous events. For example, aerosols from a volcanic eruption may affect the brightness temperature of a satellite retrieval, with effects varying with frequency.

5. SUMMARY

Long-term climate data sets such as archived radiosonde observations usually contain biases caused by factors such as changes in instruments. The most common data adjustment method is to equalize the mean value of each variable in a specified averaging period before and after each instrument transition or similar discontinuity. This method can be called "segment mean matching" and is the most common method because overlapping measurements with both instrument types are almost never performed.

While investigating the "segment mean matching" adjustment method, it was realized that in theory, part of the underlying trend is removed in the process of adjusting for each discontinuity.

Specifically, the expected proportion of the total long-term trend removed by a data adjustment is the ratio of the time between the midpoints of the averaging periods before and after the discontinuity to the total length of the time period. With an averaging period 2 years before and 2 years after a discontinuity and a 40-year time series, an adjustment would remove an average of 5 percent of the actual trend. The effect of multiple adjustments is additive. Since many radiosonde stations frequently change instrument types, potentially much of the actual trend could be removed (At some stations, instrument changes are so frequent that many periods between transitions are to short to produce a satisfactory average either before or after the transition).

In addition, each adjustment is affected by climate anomalies (such as ENSO phases) during an averaging period, so an individual adjustment can be spuriously large (or small) and ends up projecting the climate anomaly into the entire time series before that discontinuity.

The observed complex station instrument histories can provide opportunities for instrument comparisons that are less affected by both the long-term trend and anomalous events occurring near the instrument transitions. However, not all adjustment methods can use these comparisons. Such comparisons include frequent alternations of different instruments at the same station, long-term use of different instruments at adjacent stations, changes between the two instrument types in the reverse order at some stations, and comparisons of radiosonde data with the same variable from a different source such as a model.

Acknowledgements

This project has been partially supported by NOAA Office of Global Programs grant NA08OAR4310686.

References

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 online at http://www.ncdc.noaa.gov/oa/cab/igra/papers/gaffen1996.pdf and http://www1.ncdc.noaa.gov/pub/data/stnhistory]

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

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., 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, http://ams.confex.com/ams/pdfpapers/146252.pdf.

Schroeder, S. R., 2010: Validating and completing global historical radiosonde metadata by examination of sensitive variables (in preparation).