DEVELOPING HIGH-QUALITY FIELD PROGRAM SOUNDING DATASETS

BY PAUL E. CIESIELSKI, PATRICK T. HAERTEL, RICHARD H. JOHNSON, JUNHONG WANG, AND SCOT M. LOEHRER

A new procedure is presented which should reduce the time and effort necessary to correct and quality control rawinsonde observations from field experiments, which are often plagued by differing sonde types, biases, errors, and other issues.

For nearly half a century, field experiments have been conducted from the tropics to the polar regions involving intensive observations of atmosphere, ocean, and land processes over selected locations of interest. In a recent article, Johnson et al. (2012) identified more than 50 past field experiments with upper-air sonde datasets (see their Fig. 2, which shows a map depicting the shape, size, and location of many of these sonde networks). Field program sondes are characterized by high temporal (3–6 h) and vertical (5–10 m) resolutions and enhanced accuracy, made possible by the large suite of instruments deployed in the experiment, which allows for cross calibration. In comparison, sondes from operational networks only have data at standard and significant levels with 250–500-m and 12–24-h resolution. Upper-air observations from field experiments have numerous applications: environmental condition descriptions, which provide a context for validating and understanding other observations (radar, aircraft, satellite, etc.); heat and moisture budget analyses from which the properties of convection can be inferred (Yanai et al. 1973); large-scale advective tendency computations that force cloud-resolving and single-column models, which aide in the improvement of model parameterization schemes (Wu et al. 2000); and model initialization for reanalyses and case studies. In addition, these datasets provide the basis for constructing high vertical resolution, model-independent analyses for model validation purposes. Because of their usefulness, accuracy, and resolution, field program sonde datasets are in high demand.

Developing a research-quality sonde dataset that is able to meet the scientific objectives of the experiment begins with high-quality observations. This fact was eloquently stated by Ooyama (1987) in his Global Atmospheric Research Program (GARP)

AFFILIATIONS: CIESIELSKI AND JOHNSON—Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado; HAERTEL—Department of Geology and Geophysics, Yale University, New Haven, Connecticut; WANG AND LOEHRER—National Center for Atmospheric Research, Boulder, Colorado
CORRESPONDING AUTHOR: Paul E. Ciesielski, Department of Atmospheric Science, Colorado State University, Fort Collins, CO 80523
E-mail: paulc@atmos.colostate.edu

The abstract for this article can be found in this issue, following the table of contents.
DOI:10.1175/BAMS-D-11-00091.1

In final form 23 August 2011

Atlantic Tropical Experiment (GATE; see Table 1 for experiment details) sounding analysis paper when he asserted the following, "After having explored every possible avenue to extract 'facts' from the observational data, the author cannot hide his empathy with Bernard Trevisan (alchemist, 1406–1490) who uttered with his last breath his conviction, '**To make gold, one must start with gold**'" (p. 2501). Several recommendations for obtaining high-quality sonde observations are listed in the "Data quality assurance" sidebar. The goal of this paper is to describe a procedure for refining these sonde observations into legacy datasets.

To realize the full benefits of field program sonde data, postprocessing efforts should be directed toward creating an easily readable dataset in which rigorous and well-documented quality controls have been applied. Such efforts are often challenging and time consuming, since the geographic expanse, operational requirements, and international aspect of many experiments result in a variety of sonde and ground station types being used, from which a multitude of instrumental problems and data formats arise. For example, during the 1974 GATE, there was a problem with the wind-finding systems on some of the research vessels (Kuettner and Parker 1976), which led to a lengthy, complicated sounding postprocessing effort. In the more recent African Monsoon Multidisciplinary Analysis (AMMA) field program, four different sonde types were used (Nuret et al. 2008),

Table 1. List of field experiment acronyms along with expanded name, dates of primary field operations, and approximate number of upper-air sonde observations processed.

Acronym	Experiment name	Dates	Approximate number of sondes
AMMA	African Monsoon Multidisciplinary Analysis	Jun–Sep 2006	6,600
GATE	Global Atmospheric Research Program (GARP) Atlantic Tropical Experiment	15 Jun–15 Sep 1974	2,000
NAME	North American Monsoon Experiment	1 Jul–15 Aug 2004	3,000
SCSMEX	South China Sea Monsoon Experiment	1 May–30 Jun 1998	23,000
TiMREX	Terrain-Influenced Monsoon Rainfall Experiment	15 May–25 Jun 2008	2,300
TOGA COARE	Tropical Ocean Global Atmosphere Coupled Ocean–Atmosphere Response Experiment	1 Nov 1992–28 Feb 1993	14,000

DATA QUALITY ASSURANCE

Developing a high-quality sounding dataset in a field program requires both *quality assurance* and *quality control* of the data. Quality assurance involves the preparations that are made prior to taking observations that ensure the measurements are properly taken. In regard to sounding data, these preparations should include the following: 1) To the extent possible, the launch site should be representative of the larger-scale environment (e.g., it should avoid hot asphalt surfaces or sheltered areas between large buildings). 2) The geolocation of the launch site should be precisely determined (e.g., the starting elevation of a sonde launch impacts geopotential height computation at all levels). 3) Because accurate surface measurements can help identify sonde biases, surface data from collocated (both vertically and horizontally), well-calibrated surface instruments should be used. 4) Newly calibrated, noncontaminated sondes should be used, because older sondes (>1 yr old) tend to have larger biases. 5) For Vaisala RS92 sondes, fresh desiccant should be kept in the ground check chamber; and the desiccant should be replaced if RH > 1% (Miloshevich et al. 2009). 6) The sonde should be adequately ventilated prior to launch (to equilibrate sonde sensors to ambient conditions). 7) For proper ventilation of the sensors during ascent, use enough gas in the balloon to achieve an ascent rate of 4–5 m s^{-1}. 8) Collocate other instruments to provide independent, redundant measurements to identify biases, such as using a ground-based GPS system to obtain independent estimates of PW (Wang and Zhang 2008). 9) If multiple sonde types and ground station systems will be used, then sonde intercomparison launches (such as that shown in Fig. 4) made prior to and during the course of the experiment are helpful for identifying platform biases. 10) Sonde manufacturers and instrument developers should be encouraged to record engineering and housekeeping data, as well as other metadata, in the raw data files; such information can be quite useful in monitoring instrument performance, investigating bad data, and potentially correcting it. 11) Operator judgment should be exercised with the option of delaying or postponing a launch when releasing sondes into strong convective storms (for safety reasons, and to avoid balloon icing issues and poor representation of large-scale environment).

which presented a serious challenge for data quality control.

An important issue in producing research-quality sonde datasets, often related to multiple sonde types in the same experiment, is dealing with different sensor biases. This situation is particularly true for the sonde humidity sensor. For example, using ground-based GPS measurements as an independent estimate of precipitable water (PW), Wang and Zhang (2008) examined the biases for 14 different sonde types from 169 stations. They found significantly different biases among the sensor types, with capacitive polymer types having a mean dry bias of −6.8%, and the carbon hygristor and goldbeater skin types having a mean moist bias of 3%–5%. If uncorrected, these biases can adversely impact the analyses using the moisture field. For example, in Tropical Ocean and Global Atmosphere Coupled Ocean–Atmosphere Response Experiment (TOGA COARE), a dry bias in the Vaisala RS80A and RS80H sondes used near the equator and a moist bias in the VIZ sondes used to the north resulted in the reanalysis fields misdiagnosing the region of maximum precipitation by several 100 km (Ciesielski et al. 2003). Identifying and reducing these biases is an important step in producing high-fidelity datasets that are able to achieve the scientific objectives of the experiment. For TOGA COARE sondes, this process took nearly a decade to complete (Wang et al. 2002). Being aware of such issues will hopefully expedite future efforts to reduce such biases.

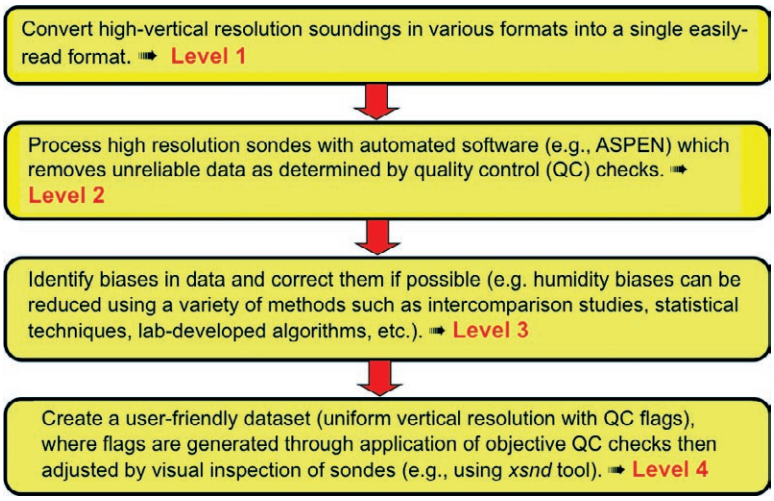

Fig. 1. Flowchart showing various stages for developing a research-quality sonde dataset.

Creating a corrected, quality-controlled (QCed) dataset in a timely fashion is important to facilitate use of the dataset while interest is still high and funding is available. Based on our experiences in developing QCed sonde datasets for several field programs [e.g., TOGA COARE, South China Sea Monsoon Experiment (SCSMEX), North American Monsoon Experiment (NAME), Terrain-Influenced Monsoon Rainfall Experiment (TiMREX); see Table 1 for number of sondes processed], we have designed a general procedure for creating user-friendly, bias-reduced, QCed sonde datasets. This procedure represents an extension of that described by Loehrer et al. (1996), which was used to process more than 14,000 soundings for TOGA COARE. This paper provides examples of how to approach the various stages of the QC procedure, with links to the software tools to implement them. Figure 1 presents a flowchart outlining this procedure with four stages (or levels) of data processing. "Level 1 and 2 processing: Creation of a common data format and automated quality processing" describes the processing in stages 1 and 2, which includes conversion to a single data format and removal of egregious data. In "Level 3 processing: Bias identification and reduction," the identification and reduction of sonde biases are considered. The final step in developing a research quality dataset is described in "Level 4 processing: Creation of a user-friendly, visually inspected product" along with some examples in the appendix of a software tool for visually adjusting QC flags. Some concluding remarks are offered in the "Summary."

LEVEL 1 AND 2 PROCESSING: CREATION OF A COMMON DATA FORMAT AND AUTOMATED QUALITY PROCESSING.

The first step in processing field program sonde datasets should be to convert the soundings from the various data formats created by the different data systems into a single, easily utilized format, which we refer to as level 1 (L1)[1] processing. In creating an L1 data product it is important that no information be dropped from the original data files. Data systems that do not provide certain information should fill these data fields with missing values rather than

[1] The level 1, 2, and 3 nomenclature used here differs from that introduced during the First GARP Global Experiment (FGGE), namely, levels I, II, and III, which referred to raw, processed real-time, or near-real-time and analyzed datasets, respectively.

limit the number of fields to those that are common among all systems. At a minimum these files should contain the basic fields (pressure, height, temperature, dewpoint temperature, and winds) and geolocation information as a function of time. The geolocation information is especially important in mesoscale field programs where balloon drift distances are a non-negligible fraction of station spacing. It is particularly important that metadata be carried along as far as possible through the various processing stages. We recommend that metadata include the sonde type and ground system, the exact time of sonde launch, and the sonde serial number, which provides information on the manufacture date, and can prove helpful in developing corrections.

These efforts should be followed by level 2 (L2) processing in which the high vertical resolution soundings are passed through a series of automated QC algorithms to systematically detect bad values. An example of such a tool, the Atmospheric Sounding Processing Enviroment (ASPEN) developed by the National Center for Atmospheric Research's (NCAR's) Earth Observing Laboratory (EOL) can be run in a Windows- or UNIX-based environment. In addition to removing egregious data points based on several QC checks (gross limit, vertical consistency, etc.), ASPEN filters the winds, computes geopotential height, smooths pressure, and writes out the processed high-resolution data in one of many convenient formats. The ASPEN software was designed to operate as automatically as possible while allowing the user to have some control over the QC process. This software and its complete documentation can be obtained online (see www.eol.ucar.edu/isf/facilities/software/aspen/aspen.html).

At a minimum, level 1 and 2 processing should always be performed, which, compared to the later processing stages, requires less time and effort.

LEVEL 3 PROCESSING: BIAS IDENTIFICATION AND REDUCTION. Raw sonde data contain both random and systematic errors (Parker and Cox 2007). Automated QC algorithms, such as ASPEN, are efficient at identifying large random errors (e.g., outliers as seen in Fig. 6 or excessive noise levels[2]). In contrast, systematic errors or biases in the data are more difficult to identify and remove, particularly in the humidity field. However, minimizing these biases represents an important step toward producing a high-quality dataset.

With recent improvements in radiosonde technology and ground station software, biases in temperature, pressure, and winds are generally quite small, and their measurements are of suitable quality for both weather and climate research. While temperature biases are dominated by radiation errors (Luers 1997), which maximize in the upper part of the profile, most sonde manufacturers implement radiation corrections in their software to minimize such errors. In the latest World Meteorological Organization (WMO) radiosonde intercomparison study in Yangjiang, China (Nash et al. 2011), the temperature differences in the troposphere among the 11 different sonde types tested were generally ±0.2°C. Special care must be taken in analyzing sonde data obtained from ships, which may contain errors in the boundary layer temperature field resulting from contamination from the ship structure (Yoneyama et al. 2002; Ciesielski et al. 2010). Pressure sensors are quite accurate with systematic pressure biases for various radiosonde types within ±0.8 hPa, with the largest differences near the surface, and random errors within 1 hPa (Nash et al. 2011). Geopotential height (z) is derived from the pressure and virtual temperature variables via the hypsometric equation, and its accuracy is related to errors in these fields, which are quite small. In the 2011 study (Nash et al. 2011), the range of z differences in the troposphere among the different sonde types tested was ±10 m. Radiosonde GPS-derived winds are subject to two errors: noise as a result of balloon pendulum motion and noise in the GPS system itself. To minimize these effects, all manufacturers use filtering to smooth winds. Results from the 2010 intercomparison study (Nash et al. 2011) show that the examined GPS sonde types had a wind bias of generally ≤ 0.2 m s^{-1} with random errors of ≤ 0.1 m s^{-1}. In systems using no GPS-derived winds (e.g., radiothedolite tracking), these errors are considerably greater.

While sonde manufacturers are continually striving to improve the accuracy of humidity sensors, water vapor retrieval continues to be the most problematic variable measured by upper-air sondes. Humidity differences among the various sonde types in the recent WMO study (Nash et al. 2011) generally ranged from ±4% at warmer temperatures (20°–30°C) increasing to ±12% at temperatures colder than –20°C. In their comparison of sonde and GPS PW measurements, Wang and Zhang (2008) found much larger differences for PW < 40 mm, with a wide range of

[2] Excessive noise in a sonde data field represents large point-to-point scatter. Causes for this noise include either faulty sensor or improper sensor calibration, electrical interference, difficulty achieving theodolite lock for wind computation, or intense updraft/downdraft couplets in strong convection, etc.

biases for different sensor types in these drier conditions (see their Figs. 4 and 5). The accuracy of sonde RH retrievals differs between measurement technologies, sonde models, and manufacturers, as well as temporally within the same model, resulting from changes in hardware, manufacturing, and calibration (Miloshevich et al. 2009). In "Identification of sonde humidity bias," we consider how sonde humidity errors can be identified and corrections can be validated using independent estimates of water vapor measurement. Because of the multiplicity of causes for humidity biases, a generic correction for all sonde types is not possible. Thus, in "Reduction of humidity biases," we review a few of the more popular and robust methods for correcting humidity biases. We note that the corrected data are probably improved but most likely are not correct in the sense that some bias presumably remains, especially when corrections are not based on reference instruments.

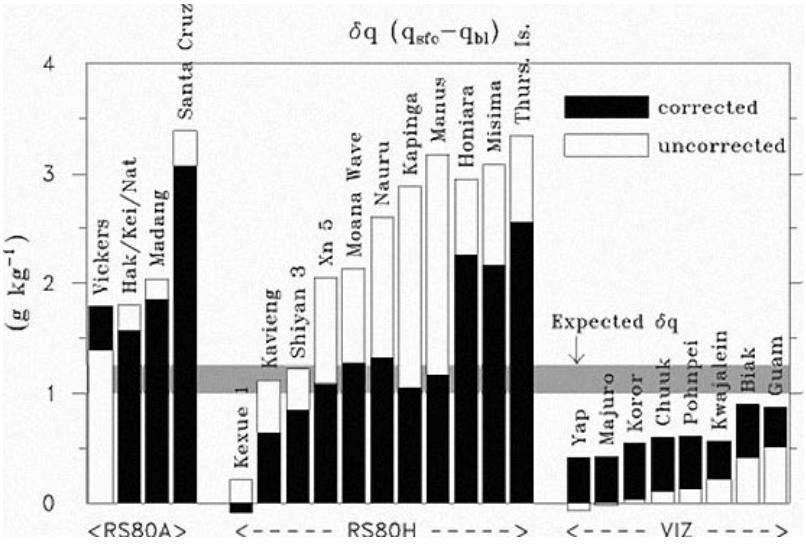

Fig. 2. Mean specific humidity (q) difference (g kg^{-1}) between surface and the mixed layer (δq) for 23 sites and three sonde types used in TOGA COARE (from Ciesielski et al. 2003) computed from uncorrected data (top of white bars) and corrected data (top of black bars, except for Kexue I in which case the correction resulted in a negative δq). The expected range of δq over the warm pool is denoted (horizontal gray bar). Values above this range indicate a sonde dry bias, values below it a moist bias. Sites are grouped by sensor type.

Identification of sonde humidity bias. Because the surface values in a sounding are independently measured by surface meteorological sensors, a simple means of identifying sonde humidity biases is to examine the moisture differences between the surface and first sonde value immediately above the surface.[3] Alternatively, one could examine the specific humidity difference δq between the surface and the boundary layer mean, as was used to identify the sonde humidity biases in TOGA COARE (Wang et al. 2002). An example of this type of analysis is shown in Fig. 2 from Ciesielski et al. (2003) for the three sonde types (Vaisala RS80A, Vaisala RS80H, and VIZ) used in TOGA COARE. According to Monin–Obukhov similary theory, δq should range between ~1.0 and 1.25 g kg^{-1} over the warm pool region (Zipser and Johnson 1998). In this case, the analysis in Fig. 2 shows a dry bias in the Vaisala sondes and moist bias in the VIZ sondes. Though this type of analysis focuses on the lower levels, moisture in the vertical column is heavily weighted toward the lower troposphere.

While use of surface data can provide an indication of low-level sonde biases, independent estimates of PW are valuable for determining total column biases. Independent estimates of PW can come from several sources, including ground-based microwave radiometer (MW) retrievals (Cady-Pereira et al. 2008), satellite-based MW retrievals over the oceans, and ground-based GPS estimates over land. The accuracy of GPS and microwave PW retrievals is ~1–2 mm (Bock et al. 2007; Wentz 1997), making them an excellent source for identifying sonde biases. Figure 3 shows a comparison of the mean sonde PW to that from independent estimates for different types of Vaisala sondes used in TiMREX.[4] This analysis shows a slight dry bias in the RS92 sondes and a large dry bias in the RS80A sondes. Ideally, the independent measurements should be collocated with the sonde location, but here GPS estimates were used if they were within 50 km of the sonde site. Also, in Fig. 3 the GPS estimates were adjusted to account for elevation differences between the sonde and GPS sites (see Ciesielski et al. 2010 for details).

[3] A key to using this indicator is that the surface data should come from a well-calibrated instrument collocated with sonde release site (see "Data quality assurance").

[4] TiMREX was conducted in the vicinity of Taiwan from 15 May to 25 June 2008.

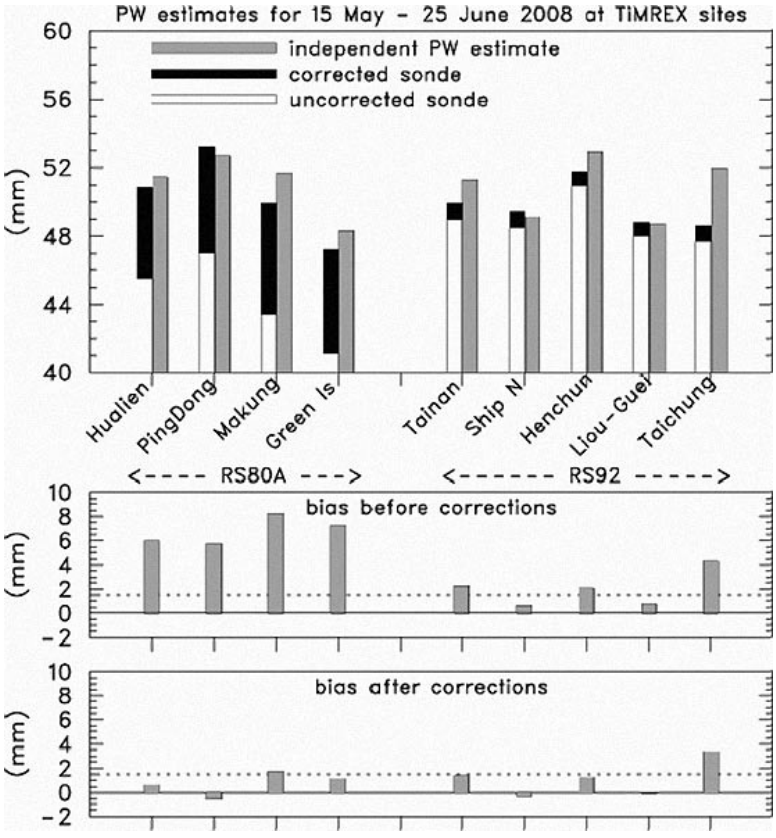

Fig. 3. (top) Comparison of PW from sonde-independent estimates [either the Advanced Microwave Scanning Radiometer for Earth Observing System (EOS; AMSR-E) retrievals over oceans or the nearest GPS site over land (gray)], uncorrected sonde (white), and humidity-corrected sonde (black) for the TiMREX special observing period. Also shown is the PW bias (independent estimate minus the sonde) (middle) before and (bottom) after corrections, with positive values denoting a dry bias. Analysis is for Vaisala RS80A (four leftmost) and RS92 (five rightmost) sondes. (middle, bottom) The 95% level above which biases (i.e., difference in means) are significant (horizontal dashed lines) are shown.

Because GPS data have a high temporal resolution (0.5-h resolution), it is possible to examine the diurnal cycle of the bias as in Wang et al. (2008) and Ciesielski et al. (2010).

Reduction of humidity biases. Several methods have been used for reducing sonde humidity biases. A particularly effective method involves performing a series of intercomparison launches wherein a reference sonde (e.g., chilled-mirror hygrometer) is flown on the same platform as the sonde one desires to correct. For example, every 4–6 yr the WMO conducts such intercomparison studies, which examine the major sonde types used around the globe. The objective of these launches should be to sample the atmosphere as broadly as possible (day–night, wet–dry, etc.) to obtain a statistically significant population of samples. These intercomparison launches form the basis for developing a correction scheme. This approach was used to compute a daytime correction for the Vaisala RS92 sondes (Vömel et al. 2007; Yoneyama et al. 2008). In these studies, a polynomial is fit to a mean bias difference vertical profile to define a correction that is a function of pressure and solar zenith angle.

A second method for reducing humidity biases is to use the data from the intercomparison launches (Fig. 4) with a statistical technique referred to as cumulative distribution function (CDF) matching method. This method, which assumes homogeneity for encountered conditions, creates a correction table that is a function of temperature and humidity by matching the CDFs of the problem sondes to those of the reliable sonde over several different temperature ranges. Typically, separate day–night correction tables are created. This approach appears to be quite robust in removing humidity biases if the intercomparison sample size spans the full range of atmospheric conditions (Ciesielski et al. 2010). While this approach reduces biases between different instrument types, caution must be exercised in its application so that unrealistic geophysical signals are not introduced. For example, using uncorrected Vaisala RS92 sondes as a reference to adjust other sonde types will introduce a daytime dry bias into these other sondes.

A variant of the CDF method was used in adjusting the humidity in the AMMA (Nuret et al. 2008) and NAME (Ciesielski et al. 2009) sondes. In these studies the bias-correction tables were based not on statistics of intercomparison launches but on sondes launched at different times of day, as in AMMA, or contemporaneously from nearby locations, as in NAME. For example, in the AMMA case, the statistics of Vaisala RS92 sondes (used as the reference) launched at 0000, 0600, 1200, and 1800 UTC were matched to the statistics from Vaisala RS80A sondes launched at intermediate hours (i.e., 0300, 0900, 1500, and 2100 UTC). In this manner a correction table was generated for

Fig. 4. Taiwanese scientists prepare for an intercomparison sonde launch in Banchiao, Taiwan, on 15 Apr 2008. In this launch, a trirod structure was utilized to compare three sondes types (Meisei, Graw, and Vaisala RS92).

adjusting the RS80A sondes to the standard of the RS92 sondes (Nuret el al. 2008). An additional correction was then needed to account for the biases in the RS92 sondes (Miloshevich et al. 2009).

Possibly the most straightforward approach to developing a bias correction is by conducting laboratory tests in a controlled environment on the problem sondes. This procedure was applied to several different batches of Vaisala RS80A and RS80H sondes to understand and develop methods to correct the humidity biases seen in the TOGA COARE data (Fig. 2). Based on analysis from tests involving heat treatment of the sondes to determine sources of contamination and sensor accuracy, six different corrections were developed to address various measurement errors in the RS80 sondes (Wang et al. 2002). These corrections accounted for factors such as sensor aging, temperature dependence, sensor contamination, sensor arm heating, and errors in the basic calibration model and ground check procedures. Recognizing the critical role of low- and midlevel moisture in regulating deep convection (e.g., Crook 1996), Ciesielski et al. (2003) examined the impact of these corrections on convection using the Raymond–Blyth (Raymond and Blyth 1992) buoyancy sorting cloud model to compute the convective mass flux both with and without the humidity corrected data. Figure 5 shows that the mean convective mass flux peak for the TOGA COARE period shifts from 8°N in the uncorrected analysis to just south of the equator using the corrected data, which agrees better with the diagnosed vertical motion and observed rainfall for this period. These results suggest that the intensity and location of convection would differ significantly in model simulations with humidity-

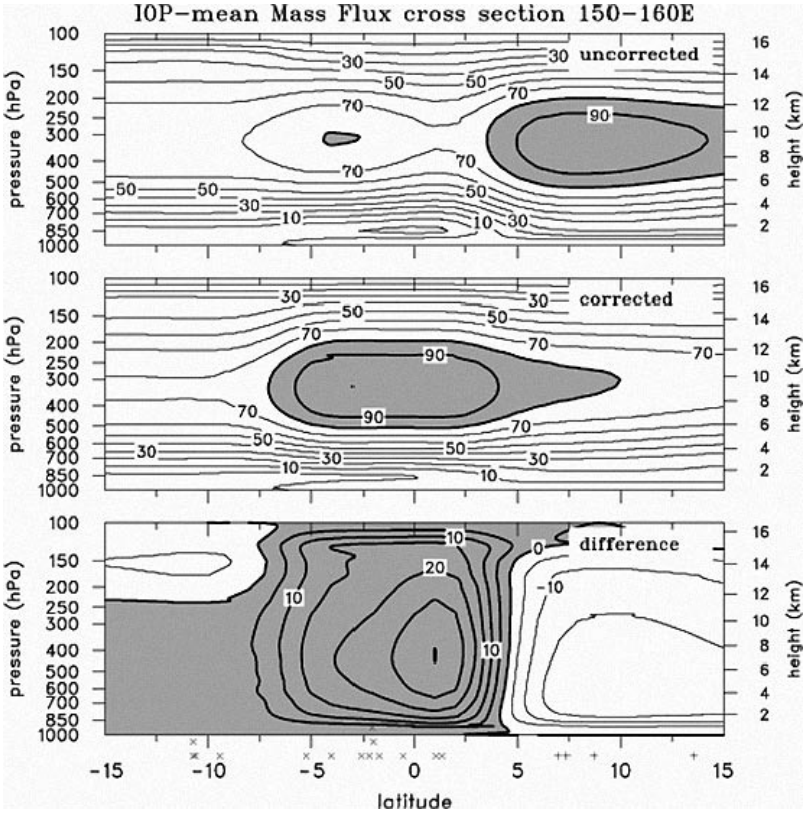

Fig. 5. Mean zonally averaged convective mass flux (normalized by max value) between 150° and 160°E computed over the TOGA COARE period. Computed with (top) uncorrected data, (middle) corrected data, and (bottom) their difference. (top, middle) Values >80 units and (bottom) values >0 units are shaded. (bottom) Symbols at base indicate the latitudes of the Vaisala (x) and VIZ (+) sites used in creating this plot (from Ciesielski et al. 2003).

corrected data, and that the difficulties that the reanalysis products had in reproducing the observed rainfall during TOGA COARE were likely due to the humidity biases (Guichard et al. 2000; Ciesielski et al. 2003).

Lab tests were also conducted on a sample of 70 Vaisala RS80A sondes following TiMREX in which a large dry bias was observed at sites using this type of sonde (see Fig. 3). These test revealed that the dry bias was caused by an unusually high level of contaminants on the humidity sensor, which reduced their ability to absorb water. These sondes were given a heat treatment to burn off the impurities, similar to what is done with the ground check chamber used to precondition Vaisala RS92 sondes. The sondes were then recalibrated and a set of revised calibration coefficients was generated to correct the dry bias in the RS80A sondes. The bottom two panels of Fig. 3 show the PW bias before and after correction. Clearly the dry bias was substantially reduced at these sites, such that the mean PW bias of the corrected sondes is less than 2 mm, or within the accuracy of the independent estimates. Examining various measures of convection, Ciesielski et al. (2010) found that use of the corrected sondes gives a much different perspective on the characteristics of convection during TiMREX. For example, at the RS80A sites, use of the corrected humidity data increases the mean CAPE by 500 J kg^{-2}, decreases the mean convective inhibition (CIN) by 80 J kg^{-2}, and increases the midlevel convective mass flux by greater than 70%.

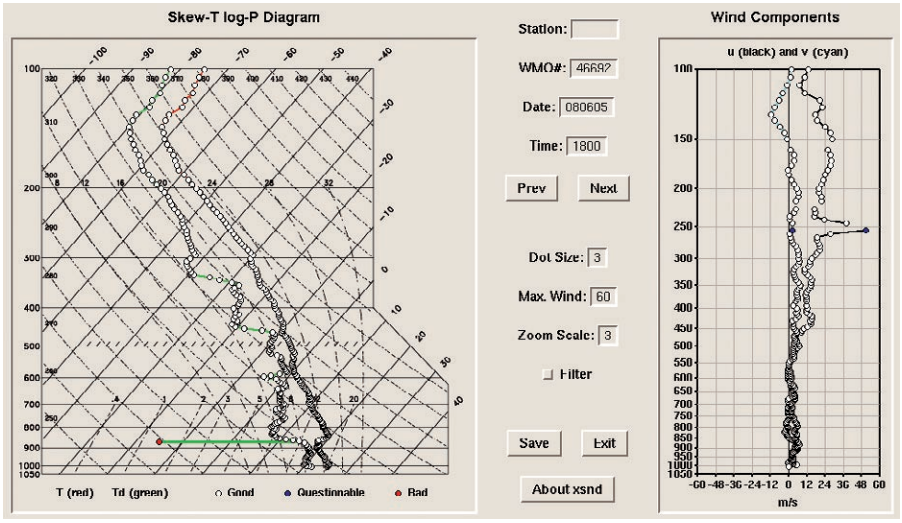

Fig. 6. Example of an xsnd display screen with (left) sonde thermodynamic data displayed as a skew *T*–log*p* plot [temperature curve (red), dewpoint curve (green)] and (right) winds displayed as vertical profiles of wind components [zonal wind curve (black) and meridional wind curve (cyan)]. Color of dots indicates their quality (white: good, blue: questionable, and red: bad). (middle) Sonde information (WMO No., date, and time) and xsnd options (zoom scale, dot size, max wind scale, and filter flag) are displayed. To toggle between adjacent sondes in a file, one clicks on the "previous" and "next" buttons. See text for additional details.

LEVEL 4 PROCESSING: CREATION OF A USER-FRIENDLY, VISUALLY INSPECTED PRODUCT. Once the high-resolution data have been processed through an automated QC program, such as ASPEN, and corrected, if necessary, we recommend creating a more "user friendly" version of the sonde dataset with values at uniform vertical resolution and QC flags assigned to each variable, providing a measure of the data's reliability. For many applications sonde data at uniform vertical resolution (either height or pressure) is more convenient to work with. Uniform pressure coordinates have the advantage of dividing the atmosphere into layers of equal mass, while uniform height coordinates provide much higher resolution at upper levels. Isentropic coordinates, while popular in applications involving the upper atmosphere, have difficulties in the boundary layer where the potential temperature may not be monotonic. In the examples in this section and the appendix, the sondes have been interpolated to a uniform 5-hPa resolution. In the level 4 (L4) stage of processing, suspicious data should be identified through the application of both objective QC tests, as in Loehrer et al. (1996), and a subjective adjustment of QC flags by visual inspection. By flagging suspect data values, the reliable data are easily retrievable, with the users deciding which level of quality is acceptable for their analyses. Additional details on interpolating the data to a uniform vertical resolution, objective tests for assigning QC flags, definition of QC flags, and Fortran programs to perform these tasks can be found online (www.eol.ucar.edu/projects/sondeqc/). Interpolation software is provided to produce either a uniform pressure or height coordinate dataset at whatever resolution one requires.

Following the objective QC flag assignment, each sounding should be visually inspected. This processing stage, while tedious, is necessary to ensure a research-quality dataset because subtle errors in sonde data are often difficult to identify with objective procedures. To facilitate this processing, we have developed a software tool that allows one to visually examine vertical profiles of thermodynamic and wind variables up to 100 hPa. This software called xsnd was written in Tcl/Tk, an easy-to-learn scripting language that runs under UNIX, Windows, or Macitosh environments. In our application of xsnd, all of the sondes for one site are combined chronologically into one file.

Using xsnd provides a means to easily "buddy check" the data, that is, visually compare the sondes that are adjacent in time and in close proximity to each other for continuity of features. An example of how xsnd displays a sounding is shown in Fig. 6, with white dots for good data, blue dots for questionable data, and red dots for bad data. By constructing the file with sondes in chronological order, xsnd allows one to toggle between adjacent soundings by clicking on the "next" and "previous" buttons. In the example shown in Fig. 6, the L4 objective QC checks have identified a few points in this sounding as being either questionable or bad. Examining adjacent sondes confirms this assessment and reveals that additional winds values near 250 hPa should also be marked as bad. Executing a second xsnd session for a nearby site allows one to check for spatial continuity in fields. Suspect data are marked as being either questionable, by clicking once on a data point (the dot color changes to blue), or bad, by clicking a second time on the data point (the dot color changes to red). This action has the effect of changing the quality flags in the L4 data file but not the data value itself. A zoom feature is also provided to expand an area of interest in the plot. Within this zoom window, points can be marked either individually or collectively with a single mouse click. Changes to the QC flags are updated in the file when the "save" button is clicked. Two examples of using the xsnd tool to identify suspicious temperatures and winds are presented in the appendix.

To illustrate the value of L4 processing, we show in Fig. 7 an assessment of the number of sondes at each site in TiMREX containing either questionable or bad data. Here the quality of the thermodynamic (temperature T and dewpoint T_d) and wind variables are considered separately. In general, sites using Meisei sondes with radiotheodolite wind tracking (Banchiao, Tainan, and Dongsha) had the highest frequency of suspect data. We note that even at the sites with more than half of the sondes containing suspect data, only a few percent of the levels are actually affected. In other words, most sondes with suspect data also have reliable data as well. By flagging suspect data values, the useful data in these sondes are easily retrievable, with the users deciding the level of quality that is acceptable for their analyses.

Because adjusting quality flags with a visual editor such as xsnd is a subjective procedure, this step should be performed by a single analyst to ensure a measure of consistency in the subjective decisions that are being made. Furthermore, this analyst should be knowledgeable about the range of atmospheric conditions in the region of interest, such that reasonable decisions are made in adjusting QC flags. Having a standard for good data (i.e., a sonde site in the region of interest with consistently high-quality observations) is quite

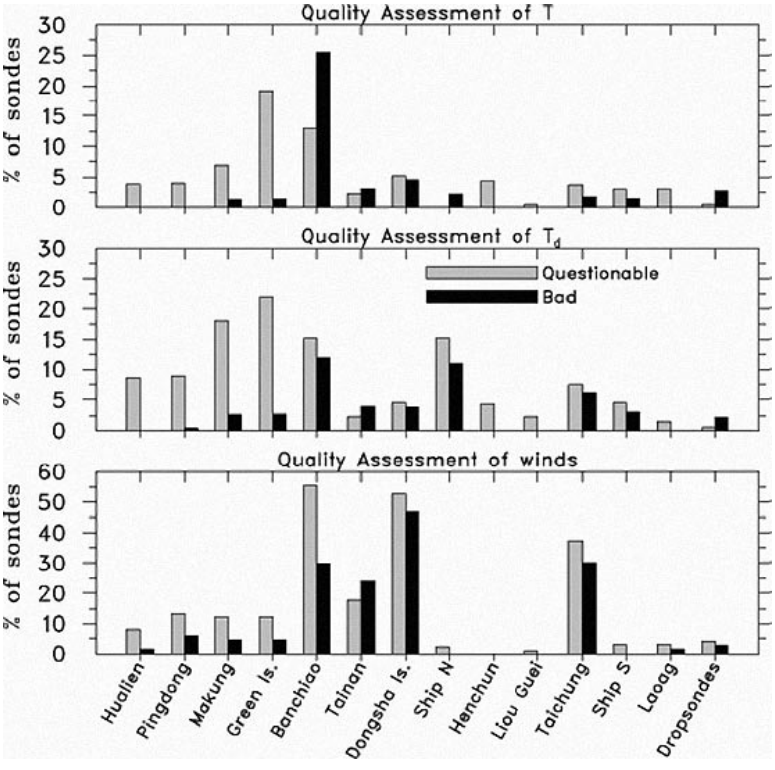

Fig. 7. Assessment of the percentage of sondes at each site in TiMREX that contain questionable (gray bar) or bad (black bar) L4 data (from Ciesielski et al. 2010).

useful for such purposes. Because QC flags, not data values, are adjusted, the user ultimately has the option to either accept or reject the quality assessment placed on the data.

SUMMARY. In this paper we describe a general procedure for producing research-quality sounding datasets in a timely fashion. As outlined in Fig. 1, this procedure includes four stages (or levels) of data processing. In stages 1 and 2, which at a minimum should always be performed, the high vertical resolution sondes are put into a common data format and are processed through a series of automated QC algorithms to systematically detect and remove bad values. In our experience ASPEN, a freely available software tool developed by NCAR EOL, is well suited for stage 2 processing. In stage 3, sonde biases, which are most prevalent in the humidity variable, are identified and reduced if possible. This step can be crucial in meeting the scientific objectives of the experiment. While a number of possible approaches for reducing humidity biases are discussed in "Level 3 processing: Bias identification and reduction," unfortunately no generalized software is available to handle this problem, which is often quite specific for any particular field campaign. Finally, in stage 4 a user-friendly dataset is created (i.e., uniform vertical resolution with QC flags on each variable). In this stage is it helpful to visualize the data graphically to identify errors that the automated QC algorithms missed. One approach for doing this uses xsnd, a visual sonde editor, which allows one to visually inspect the sonde and subjectively adjust the QC flags.

We encourage those involved in future field programs with upper-air sonde networks to make use of this procedure and the freely available software tools described herein, and to pay close attention to quality assurance practices as outlined in the sidebar on this subject. Bad sonde data can be the result of poor operating procedures, flawed recording software, or even faulty or outdated instrumentation. Identifying the source of persistently bad data can lead to improvements, which will minimize such problems in future studies. Furthermore, we recommend that a naming convention similar to that described in the "Suggested dataset naming convention" be adopted to help creators and users of these datasets specify precisely which level of processed data they are using in their research. Along these lines, it is imperative that all processing steps be thoroughly documented and made available to the dataset users.

SUGGESTED DATASET NAMING CONVENTION

Level 0 (L0)—Raw, original data
Level 1 (L1)—Data in uniform, digital, easily readable format
Level 2 (L2)—Processed with automated QC checks to remove bad data (native resolution)
Level 3 (L3)—Corrected (if possible and necessary, native resolution)
Level 4 (L4)—Uniform vertical resolution with QC flags
L4u—Uncorrected L4 version based on L2
L4c—Corrected L4 version based on L3

Recently a project has been initiated (Johnson et al. 2012) to collect legacy upper-air sounding datasets from atmospheric field programs into a central archive, beginning with 1956 and 1958 Marshall Island datasets (Yanai et al. 1973) to the present. One of the motivations for this project is that, with the passage of time, there is a danger that some of these sounding datasets either will be difficult to find or will languish on obsolete media that cannot be read. After all of the datasets have been located and archived, plans are underway to create a user-friendly, quality-controlled version of each dataset. Quite likely the procedure outlined in this paper will be used for this purpose.

All software tools described in this paper, additional details on the QC procedure, and examples of recommended data file formats can be found online (www.eol.ucar.edu/projects/sondeqc/).

ACKNOWLEDGMENTS. This research has been supported by the NOAA Grant NA09OAR4320074 and NSF Grants ATM-0500061 and ATM-0966758. Development of the xsnd software tool was supported by NASA funding under Grants NNX07AD35G and NNX10AG81G. We thank Wayne Schubert for the several insightful discussions, Richard Taft for his efforts to improve xsnd, and Brian McNoldy for his assistance in creating several of the figures. Comments from three anonymous reviewers were greatly appreciated.

APPENDIX: EXAMPLES OF XSND TOOL TO FLAG SUSPICIOUS DATA. In certain instances automated QC routines have difficulty detecting suspicious sonde data, which are much easier to identify with a visual editor. In this appendix two such examples are presented. First, an example of using xsnd to identify some suspicious temperature data is shown in Fig. A1. In this example the sounding contains a superadiabatic lapse rate at the top of a cloud layer. This unrealistic feature, which generally occurs in sondes that are poorly ventilated, is due to wetbulbing, that is, the wetting of the thermistor in the cloud layer and the subsequent excessive cooling by

evaporation or sublimation once the sonde exits the cloud. Automated QC software, such as ASPEN and our L4 checks, are not sophisticated enough to detect wetbulbing effects because often the temperature gradients involved are not that unusual. Rather, it is the location of these strong gradients (i.e., at the top of saturated layers) that renders the data suspicious. With a visual editor, such as xsnd, an experienced analyst can readily identify and mark the suspicious data points. Clicking on the bad points and selecting the xsnd filter option allows one to see how the sounding looks with these points removed (lower-right zoomed image in Fig. A1).

A second example shown in Fig. A2 illustrates how suspicious winds can be marked using xsnd. In this sounding, the L2 processing identified and removed some bad winds near 850 hPa, as evidenced by the data gap at this level. Our L4 objective QC checks identified only one level near 850 hPa with bad winds; however, some suspect winds near this layer clearly remain, which is confirmed by examining adjacent sondes. This again highlights the shortcomings of automated QC checks and the need to visually inspect each sonde. The right-side panels in Fig. A2 show how the unreliable winds can be zoomed and marked. Then, by selecting the filter option, we can view the wind profile with the suspect points removed.

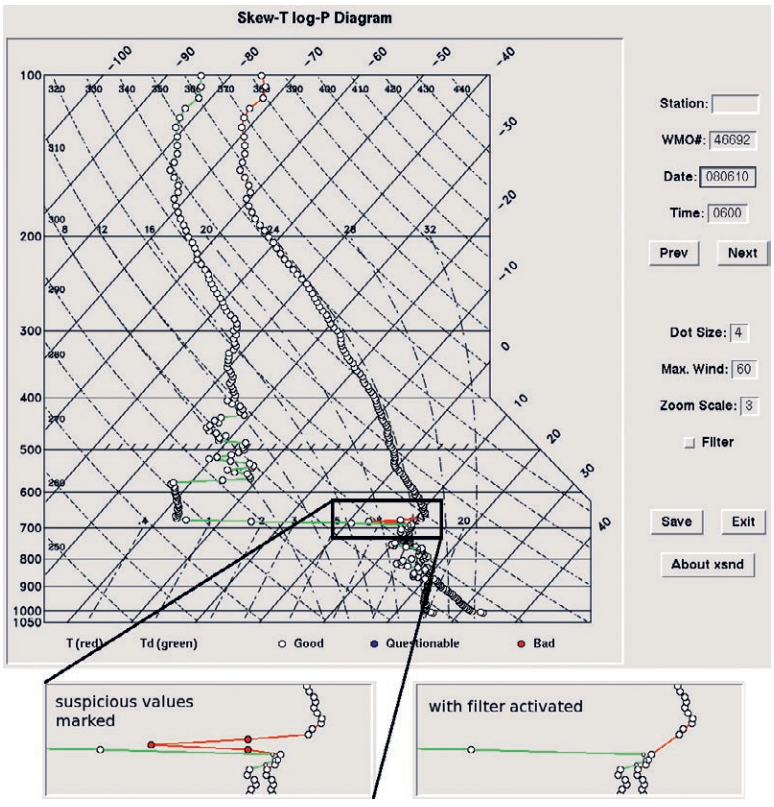

Fig. A1. Example of using xsnd to adjust QC flags in a L4 sounding. (bottom left) In this example, a region of suspicious temperatures is zoomed and three temperature values are marked as bad by clicking on points twice. The zoom feature is activated by placing the cursor over the region of interest and clicking with the right mouse button. (bottom right) By activating the filter option, the data profile is shown with bad points removed.

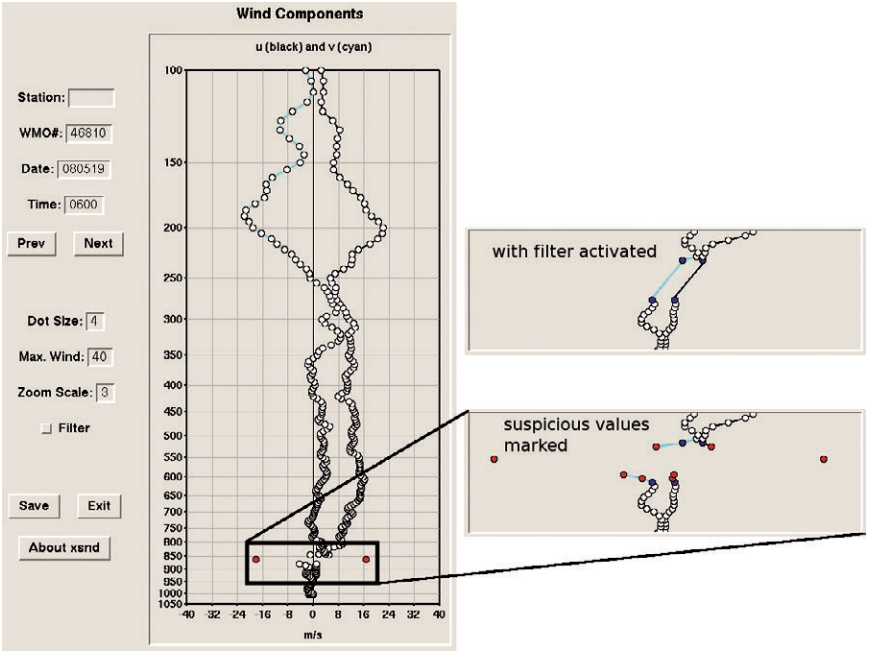

Fig. A2. Example of using xsnd to adjust QC flags in a L4 sounding. Objective tests identified only one level with bad winds. (bottom right) In this example, the layer of suspicious winds was zoomed and several additional winds were marked as being either questionable or bad. (top right) The wind profile is shown with bad points removed.

REFERENCES

Bock, O., M.-N. Bouin, A. Walpersdorf, J. P. Lafore, S. Janicot, F. Guichard, and A. Agusti-Panareda, 2007: Comparison of ground-based GPS precipitable water vapour to independent observations and numerical weather prediction model reanalyses over Africa. *Quart. J. Roy. Meteor. Soc.,* **133**, 2011–2027, doi:10.1002/qj.185.

Cady-Pereira, K. E., M. W. Shepherd, D. D. Turner, E. J. Mlayer, S. A. Clough, and T. J. Wagner, 2008: Improved daytime column-integrated precipitable water vapor from Vaisala radiosonde humidity sensors. *J. Atmos. Oceanic Technol.,* **25,** 873–883.

Ciesielski, P. E., R. H. Johnson, P. T. Haertel, and J. Wang, 2003: Corrected TOGA COARE sounding humidity data: Impact on the diagnosed properties of convection and climate over the warm pool. *J. Climate,* **16,** 2370–2384.

——, ——, and J. Wang, 2009: Correction of humidity biases in Vaisala RS80-H sondes during NAME. *J. Atmos. Oceanic Technol.,* **26,** 1763–1780.

Crook, N. A., 1996: Sensitivity of moist convection forced by boundary layer processes to low-level thermodynamic fields. *Mon. Wea. Rev.,* **124,** 1757–1785.

Guichard, F., D. Parsons, and E. Miller, 2000: Thermodynamic and radiative impact of the correction of sounding humidity bias in the tropics. *J. Climate,* **13,** 3611–3624.

Johnson, R. H., S. F. Williams, and P. E. Ciesielski, 2012: Legacy Atmospheric Sounding Data Set Project. *Bull. Amer. Meteor. Soc.,* in press.

Kuettner, J. P., and D. E. Parker, 1976: GATE: Report on the field phase. *Bull. Amer. Meteor. Soc.,* **57,** 11–27.

Loehrer, S. M., T. A. Edmands, and J. A. Moore, 1996: TOGA COARE upper-air sounding data archive: Development and quality control procedures. *Bull. Amer. Meteor. Soc.,* **77,** 2651–2671.

Luers, J. K., 1997: Temperature error of the Vaisala RS90 radiosonde. *J. Atmos. Oceanic Technol.,* **14,** 1520–1532.

Miloshevich, L. M., H. Vömel, D. N. Whitman, and T. Leblanc, 2009: Accuracy assessment and correction of Vaisala RS92 radiosonde water vapor measurements. *J. Geophys. Res.,* **114,** D11305, doi:10.1029/2008JD011565.

Nash, J., T. Oakley, H. Vömel, and W. Li, 2011: WMO intercomparison of high quality radiosondes. WMO Rep. 107, 238 pp.

Nuret, M., J.-P. Lafore, O. Bock, F. Guichard, A. Agusti-Panareda, J.-B. N'Gamini, and J.-L. Redelsperger, 2008: Correction of humidity bias for Vaisala RS80-A sondes during the AMMA 2006 observing period. *J. Atmos. Oceanic Technol.,* **25,** 2152–2158.

Ooyama, K. V., 1987: Scale-controlled objective analysis. *Mon. Wea. Rev.,* **115,** 2479–2506.

Parker, D. E., and D. I. Cox, 2007: Towards a consistent global climatological rawinsonde data-base. *Int. J. Climate,* **15,** 473–496.

Raymond, D. J., and A. M. Blyth, 1992: Extensions of the stochastic mixing model to cumulonimbus clouds. *J. Atmos. Sci.,* **49,** 1968–1983.

Vömel, H., and Coauthors, 2007: Radiation dry bias of the Vaisala RS92 humidity sensor. *J. Atmos. Oceanic Technol.,* **24,** 953–963.

Wang, J., and L. Zhang, 2008: Systematic errors in global radiosonde precipitable water data from comparisons with ground-based GPS measurements. *J. Climate,* **21,** 2218–2238.

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

Wentz, F. J., 1997: A well calibrated ocean algorithm for special sensor microwave imager. *J. Geophys. Res.,* **102,** 8703–8718.

Wu, X., M. W. Moncrieff, and K. A. Emanuel, 2000: Evaluation of large-scale forcing during TOGA COARE for cloud-resolving models and single-column models. *J. Atmos. Sci.,* **57,** 2977–2985.

Yanai, M., S. Esbensen, and J. H. Chu, 1973: Determination of bulk properties of tropical cloud clusters for large-scale heat and moisture budgets. *J. Atmos. Sci.,* **30,** 611–627.

Yoneyama, K., M. Hanyu, S. Sueyoshi, F. Yoshiura, and M. Katsumata, 2002: Radiosonde observation from the ship in the tropical region. JAMSTEC Rep., Vol. 45, 31–39.

——, M. Fujita, N. Sato, M. Fujiwara, Y. Inai, and F. Hasebe, 2008: Correction for radiation dry bias found in RS92 radiosonde data during the MISMO field experiment. *SOLA,* **4,** 13–16, doi:10.2151/sola.2008-004.

Zipser, E. J., and R. H. Johnson, 1998: Systematic errors in radiosonde humidities: A global problem? Preprints, *10th Symp. on Meteorological Observations and Instrumentation,* Phoenix, AZ, Amer. Meteor. Soc., 72–73.