WMO/IOC/UNEP/ICSU
GLOBAL CLIMATE OBSERVING SYSTEM (GCOS)

Doc. 2.1.1
(24.II.2012)

4rd GRUAN Implementation-Coordination Meeting (ICM-4)
Tokyo, Japan
5 March – 9 March 2012

Session 2

GRUAN Radiosonde Task Team Review Report on the WMO Report No. 107, WMO/TD-No. 1580, 2011

(Submitted by Larry Miloshevich, Masatomo Fujiwara, Rolf Philipona, and the Radiosonde Task Team)

Summary and Purpose of Document

This document summarizes lessons from the WMO report that are relevant to GRUAN and its goals.

GRUAN Radiosonde Task Team Review Report on the WMO (World Meteorological Organization) Instruments and Observing Methods, Report No. 107, WMO/TD-No. 1580, 2011

WMO INTERCOMPARISON OF HIGH QUALITY RADIOSONDE SYSTEMS,
Yangjiang, China, 12 July – 3 August 2010
J. Nash, T. Oakley, H. Vömel, LI Wei

Larry Miloshevich, Masatomo Fujiwara, Rolf Philipona, and the Radiosonde Task Team

Introduction

The 2010 WMO radiosonde intercomparison in Yangjiang, China evaluated 13 operational radiosondes and working references (page 9), as well as the "scientific" (reference) sensors CFH (cryogenic frostpoint hygrometer) for RH, and Meisei MTR (a fast-response tungsten wire sensor) and Sippican LMS-6 multi-thermistor for temperature. Statistical analysis was used to estimate systematic and random errors, day/night differences, time response and other characteristics of each sensor. Case studies provided a detailed assessment of sensor performance and data processing algorithms and corrections. The WMO report also gives recommendations for sensor improvements aimed at manufacturers, and recommendations on operational procedures, data processing algorithms, and needed documentation. The report is available at: http://www.wmo.int/pages/prog/www/IMOP/publications-IOM-series.html

One of the goals of GRUAN is to promote improvement of radiosonde sensors and performance, and the WMO report contains many general and specific conclusions and recommendations for improvements that are aimed at motivating and providing clarity for manufacturers. Although the WMO report contains summary conclusions about each radiosonde sensor's readiness for GRUAN in Table 12.1 (pages 180-187), the goal of this review document is NOT to pick radiosondes suitable for use by GRUAN. This is partly because the WMO report is only one source of information representing one set of climate and weather conditions, and also for the practical reason that GRUAN sites will most likely continue to use their existing radiosounding equipment, at least until such time as competitors produce superior radiosondes.

This document summarizes lessons from the WMO report that are relevant to GRUAN and its goals. It is organized into the following sections:

1. Recommendations for improving radiosonde T and RH measurements.
2. T and RH corrections used by manufacturers in their standard data products.
3. Other radiosonde measurements (GPS, pressure, altitude, winds).
4. Evaluation of the "scientific" (reference) sensors.
5. Other lessons and recommendations concerning the WMO intercomparison and report.

Note: Italicized comments are direct quotes from the WMO report.

1. Recommendations for improving radiosonde T and RH measurements.

- ***If a radiosonde operating system decides that reported values are not reliable, it should flag out the data and not report anything, rather than let the software invent values*** (page 115)**.** Some manufacturers create ("invent") humidity values in their software by decreasing the values linearly or exponentially with height above the tropopause. Another example is that, by default, the Vaisala data system will interpolate across long periods of "missing data" (e.g., due to poor signal reception). It is recommended that by default interpolation be used for at most very short periods (e.g,, 10 s).
- *One concern is that if test results such as those at Yangjiang were used to adjust the measurements, without understanding the reason for the systematic differences, the corrections become arbitrary and not based on an understanding of the sensing system limitations.* ***It is essential that every effort is made to eliminate design deficiencies in the basic radiosonde sensing systems, rather than covering up flaws by using arbitrary software adjustments****.* (page 207)

A. TEMPERATURE SENSORS:

- ***From these results and those in mid-latitudes, it is unwise to assume that any of the current radiosonde temperature measurements can be reproduced consistently to within 0.1 K whether in the tropics or high latitudes, as might be desirable for climate science.*** *The origins of the uncertainty are probably not in the calibration of the sensors, but in the stability of the radiosonde sensor under different conditions during flight or in the stability of radiosonde signal electronics and processing during flight.* (page 62) Note that the GRUAN requirements tables (GCOS-134) say that the precision (repeatability, random error) should be 0.2 K, the accuracy (systematic bias error) should be 0.1 K in the troposphere and 0.2 K in the stratosphere, and the long-term stability should be 0.05 K.
- **The temperature range of the factory calibration should be down to -100°C**, otherwise there is danger of calibration bias in the tropical UT. Table 4.1.2 (page 24) shows the minimum calibration temperature currently used by each manufacturer.
- **A hydrophobic coating minimizes evaporative cooling errors** and recovery time when a wet sensor emerges from low cloud (page 81). *In many countries, evaporative cooling is an operational problem, because it corrupts the measured temperature structure above low cloud, and the numerical forecasters will not accept long-term operational use of a radiosonde, which does not have a hydrophobic coating or other measure to minimize the magnitude of evaporative cooling.*
- **Some radiosondes have heat contamination from their own sensor support structures** (e.g., pages 76-78), also causing large random errors in the daytime stratosphere. Important design suggestions for manufacturers to check are listed on page 65, including: sensor should be mounted high enough that while swinging air that flows over the top of the radiosonde cannot flow over the sensor; exposure to the air is unobstructed with no unnecessary supports above the sensor; a similar cross-section is presented to the sun as the radiosonde rotates; and the radiation characteristics of the sensor do not change with time because of changes in the manufacturing process or coatings. Mountings of systems with high quality daytime measurements are a good guide.

B. RELATIVE HUMIDITY SENSORS:

Most RH sensors were thin-film capacitance sensors. Vaisala uses dual sensors exposed directly to the air, which are alternately heated to drive off water and ice down to -60°C, after which only one sensor makes measurements. Graw, LMS, InterMet, Huayun and Jinyang used sensors manufactured by E+E; however, several models and ways of implementing E+E sensors lead to different performance. Some radiosonde manufacturers measure the temperature of the RH sensor directly, and some use software corrections to account for the difference between the temperature of the T and RH sensors (discussed in next section). Factors affecting the performance of thin-film capacitance sensors in general are listed on page 92.

- **The temperature range of the factory RH calibration should be down to -100°C**, otherwise there is danger of calibration bias in the tropical UT. See Table 4.1.2 (page 24), where it is seen that some sensors are very far from this minimum calibration temperature. **Measurements are only valid over the temperature and RH range of the calibration**.
- ***Vaisala needs to report relative humidity to a decimal place, rather than as an integer value****, if the observations are to be used in this* (troposphere to stratosphere) *transition region* (page 123). For low RH conditions, a jump of 1% RH produces a non-physical discontinuity in mixing ratio, and furthermore is an unnecessary and substantial random-like error (e.g., ±0.5% RH is ±25% error at RH=2%). Additional precision in the RH data is already present in the "FLEDT" processed data files, but this is not the standard (default) operational data product, and it probably should be the default.
- The Vaisala pulse heating is effective against contamination by water or ice on the sensor, but only down to -60°C. Is it possible to extend the sensor heating to lower temperatures where ice-supersaturated conditions can still be found, given the need for a longer heating recovery time at low pressures? (page 123)
- *The E+E sensor response is very slow at temperatures below -70°C and so the sensors are unable to measure the troposphere/stratosphere transition between 17 and 20 km in tropical conditions, as is also true for all the rest of the QRS radiosondes* (except Vaisala*).* (page 126)
- Large biases in many of the relative humidity sensors come from one or more of the following (page 185): *poor calibration, poor referencing, poor sensor ventilation, hygroscopic material in the cap or around the sensor, faulty software utilising humidity temperature sensor measurements. If the origin of these biases could be fixed then most of the quality problems would be solved for relatively dry conditions.*
- Measuring the relative humidity structure above low-level cloud tops is not easy (page 81-82). The main issues are (1) evaporative cooling causing the RH sensor to be cooler than the ambient temperature (in principle this is not a problem if the temperature of the RH sensor is measured directly), and (2) water contamination by poor ventilation of the sensor protective cap and/or lack of a hydrophobic coating on the cap.
- *Contamination after passing through moist levels is worse at night than in the day, or at least balanced by other errors in the day* (page 187).
- **The choice of equation for saturation vapor pressure (SVP) over liquid water (ew) is a critical concern for RH calibrations at low temperatures**, where RH is defined as the ratio of the water vapor pressure (e) to the saturation value, ew. Figure 8.2.8 (page 139) shows

the ratio of various commonly-used ew equations relative to Wexler (1977), which was chosen not because it is inherently better than any other ew equation, but because it is used by NIST (National Institute of Standards and Technology). Differences are apparent below about -40°C, and reach ~20% at -90°C. See Table 4.1.2 (page 24) for ew equations used by manufacturers. From the WMO report:

*A brief survey among all manufacturers has shown that the equations by **Wexler (1977), Hyland and Wexler (1983), and Sonntag (1994)** are the most common equations. These three equations do not differ significantly over the temperature range of interest . It is therefore **recommended that only these three equations be used** to convert relative humidity over liquid to partial pressure at cold temperatures.*

Note that any equation for ew is only a theoretical construct below about -35°C, as it is not subject to experimental verification below -35°C. It is critical that users of RH data use the same SVP equation used by the manufacturer to compute vapor pressure, mixing ratio, etc. from RH values to avoid errors. Therefore, **manufacturers must advertise the SVP equation that is implicit in their RH calibration**. For example, using Goff (1965) with Vaisala RH measurements that implicitly assume Wexler (1977) will lead to large errors in the vapor pressure and other derived quantities at low temperatures. Note also that, in contrast, equations for SVP over ice (ei) do not differ much; for example, the difference between Hyland and Wexler (1987) and Goff (1965) is only ~0.2%. However, radiosonde RH measurements are typically RH with respect to liquid water, and conversion of measured values to RH with respect to ice is similarly subject to large errors if the wrong equation for ew is used. See discussion in Murphy and Koop (2005), and recommendations in Appendix A of Miloshevich et al. (2006).

Murphy, D., and T. Koop, 2005: Review of the vapour pressures of ice and supercooled water for atmospheric applications. Quart. J. Roy. Meteor. Soc., 131, 1539–1565.

Miloshevich, L.M., H. Vömel, D.N. Whiteman, B.M. Lesht, F.J. Schmidlin, and F. Russo (2006), Absolute accuracy of water vapor measurements from six operational radiosonde types launched during AWEX-G and implications for AIRS validation. J. Geophys. Res., 111, doi:10.1029/2005JD006083.

2. T and RH corrections used by manufacturers with their standard data products.
[Note that this does NOT refer to GRUAN-applied corrections (which is a separate topic), nor to the comprehensive uncertainty analysis for each data point that is encouraged and expected from each manufacturer whose radiosondes are used within GRUAN.]

For climate purposes, it is essential that manufacturers keep a public record of changes to their software, especially the temperature and humidity correction software. (page 207)

A. TEMPERATURE CORRECTIONS:

There are two common software temperature corrections: (1) solar radiation correction, and (2) filtering of spurious heating pulses from sensor supports and/or radiosonde body.

Daytime software radiation correction:

- Radiation corrections at 10 hPa are shown in Table 7.1.2 (page 64), which vary from 0.6 to 2.3°C. The LMS Multithermistor uses an active correction based on solving equations for sensors of different emissivity/absorptivity, which in principle should account for differing cloud and radiation environments, unlike the specified and simple solar heating (as a function of only solar elevation angle and air pressure) assumed by the other sensors. However, indications of a warm bias in the LMS Multithermistor measurements suggest that the active correction needs further study, and the manufacturer needs to disclose details of the correction method and provide an uncertainty analysis before the sensor can be used as a reference sensor.
- Further in-situ or lab experiments are needed that clearly show the errors produced by solar and thermal radiation. If such experiments were done in the past by manufacturers they should report on these experiments in order to help the radiosonde community to better understand the phenomena and to help further investigate the problem.

Pulse Filtering:

- *Many of the manufacturers apply filtering to daytime measurements to take out positive heating pulses (from sensor supports or radiosonde body). This filtering creates additional uncertainty in the reported values in the daytime, and introduced an uncertainty into the systematic bias of about 0.2 K, depending on the precise nature of the fluctuations observed.* (page 64)
- *Daytime radiosonde temperature measurements often have a lot more short-term fluctuations than night-time, so the filtering applied to daytime measurements is not necessarily the same as that used at night.* (page 212)
- *Filtering of spikes should be performed; however, the clear identification of spikes in software is an important factor to document, as improper spike filtering may introduce artificial systematic errors, either by incorrectly fitting real temperature structure, or omitting or smoothing over true contamination spikes which should have been removed. To evaluate measurement uncertainty it is essential to have detailed information how contamination spikes are being removed.* (page 90)
- A different configuration of payload (e.g., single radiosonde, or multiple radiosondes) and different rope/unwinder length, and different parachute and balloon size will create spikes

with different frequency and amplitude. A certain filtering algorithm may work for one situation but not for other situations. Do we prepare several versions of spike filtering software for GRUAN, or do we take this issue into consideration in the uncertainty estimation? In the latter case, the uncertainty would become larger if a radiosonde is flown with a different flight configuration than was used in the software development.

- The best approach is to develop a sensor and sensor mount/support structure (and a payload configuration for dual soundings) which are free from positive/negative pulses.

Time Constant of Response:

- Most radiosonde temperature sensors are estimated to have a time constant of response of 4-5 sec at 10 hPa with 6 m/s ascent rate, although some slower sensors have a response time up to 14 sec. *Graw has applied a correction for slow time constant of response, assuming a time constant of response of about 10 s at 10 hPa. The evidence on which this correction is based will need to be documented for future users.* (page 71, Table 7.1.3) Most data users would probably prefer to correct for time response of temperature sensors themselves, and see less noisy data; however, manufacturers should provide assessments of the time constant of response for their temperature sensors (page 211).

B. RELATIVE HUMIDITY CORRECTIONS:

Manufacturers have only recently begun to apply software corrections to RH measurements for (1) slow sensor response time ("time-lag error"), and (2) solar radiation dry-bias error (solar heating of the RH sensor). Some manufacturers directly measure the temperature of the RH sensor such that the appropriate temperature is used in the calibration function, so this approach is not a correction at all. Other manufacturers use a software correction where the actual temperature of the RH sensor is estimated or an RH adjustment is applied that is taken to be a constant value for given measured conditions, and therefore the corrected value does not consider the effect of the actual cloud and radiation environment on the RH sensor. The types of RH corrections applied by various manufacturers are given in Table D4.1 (page 219). **Further study of solar radiation and time-lag humidity corrections by manufacturers and by GRUAN is needed**.

RESPONSE TIME RH CORRECTIONS:

- **Response times of RH sensors are quite slow at low temperatures** and lead to substantial "time-lag error" in the UT/LS, characterized by smoothed profiles that are increasingly unable to follow changes in RH as T decreases, often requiring several km to transition from high RH values in the UT to low stratospheric RH values. Corrections for slow time response by Vaisala, Graw, and to a lesser extent Intermet generally improve results as judged by comparison to the faster CFH measurements and consistency between Vaisala and Graw. The WMO report (pages 115-119) makes the following points:
    - *More work is required to establish the uncertainty in the corrected Graw and Vaisala observations.* For example, further detailed comparison of corrected measurements with each other and with CFH from the WMO intercomparison dataset would be helpful. Also, comparison of multiple available time-lag correction algorithms for RS92 that are

all based on the same time-constant data would be informative about uncertainty related to the implementation. Application of time-constants that are 2-3 standard deviations from the mean would yield uncertainty estimates related to variability in the response time of individual sensors.

- Some examples show unusual behavior of the Graw results in the tropopause region; however, these instances may involve characteristics other than the time-lag correction (e.g., sensor icing in ice-supersaturated conditions). Therefore, more work is required to establish the reliability of the corrected Graw measurements.
- *If these types of software corrections are to be applied to routine operational measurements of relative humidity in the upper troposphere, it is essential that the procedures are well documented so that users of the data know what is happening and recognises the limitations of the technique.*
- *All sensors of a given radiosonde type must have similar time constants of response to justify using relatively large response time correction, and evidence needs to be supplied to show that this is the case.*
- *In some Yangjiang flights it looks like errors/noise in some radiosonde measurements have been amplified* (by the correction) *instead of providing more reliable atmospheric structure.*

- **Very long time constants may be problematic** for correcting time-lag error, depending on the reliability of the sensing system (page 223). The sensitivity of a correction to "noise" increases as the time-constant increases. Corrections for slow time response can be beneficial for time constants up to 2 or even 3 minutes, but some sensors with time constants near 4 minutes often showed structure above the tropopause that was not real, because other errors such as ice contamination or very small variations in relative humidity are greatly amplified by the correction. ***In Yangjiang, where corrections for such slow response were applied, the result looked reasonable in about 65 per cent of the cases and quite wrong the rest of the time***.

SOLAR RADIATION SOFTWARE RH CORRECTIONS:

Summary conclusions about the software corrections of Vaisala and Graw for solar heating of the RH sensor are given on pages 224-228:

- ***In the long term for GRUAN, it would be better if the temperature of the humidity sensor was measured directly, so that the solar heating correction was unnecessary***. The actual sensor heating can be approximated by, for example, a function of solar elevation angle and air density for clear-sky conditions, but reduction in solar heating by the actual and instantaneous cloud optical depth between the sun and the sensor cannot be represented, leading to either overcorrection or undercorrection if an incorrect cloud assumption is made. Other factors affecting the incident spectrum of solar radiation on the sensor and the sensor cooling due to ventilation are also not considered, such as the integrated aerosol load, effects of radiosonde swinging and spinning, variations in the ascent rate, albedo of the surface and underlying cloud, etc. Another factor concerns the large difference in time constants of response for separate RH and T sensors, such that even a correct adjustment for the temperature of the RH sensor will lead to artifacts related to the different time responses.

Furthermore, a solar radiation correction that correctly addresses all these factors in a statistical sense for a particular location and season will not be correct even statistically for another location or season, which may introduce non-physical geographic and seasonal biases into the record. Therefore, although an RH solar radiation correction is an improvement over past results where day-night differences were large, a theoretical or empirical software correction is a less desirable solution than directly measuring the RH sensor temperature.

- However, despite the undesirability of a simple software correction it is noted that, ***sensing the temperature of the humidity sensor has not yet been implemented successfully by all manufacturers using the technique***.
- The magnitude of the correction is substantial, especially at upper levels, and the accuracy depends on the validity of the model, especially regarding cloud cover. ***The correction is likely to be the limiting factor on daytime systematic bias in relative humidity measurements at upper levels***.
- Based on day-night differences, Vaisala's correction scheme appears to be too large at high humidity for Yangjiang conditions, and probably associated with high cloud cover.
- *Vaisala humidity sensors are directly exposed to solar heating, but in the case of Graw the sensor is covered by an aluminised cap, so it may not be so straightforward to try and build a model of the temperature difference between the humidity sensor and the main temperature sensor*. Characteristics of the temperature difference will differ between these two sensor designs, and furthermore neither considers the actual cloud and radiation environment.

3. Other Radiosonde Measurements (GPS, pressure, altitude, winds)

A. GPS TECHNOLOGY (for geometric and geopotential height, pressure, and winds):

- The equations to calculate geopotential height and pressure from the GPS geometric height should be documented explicitly in the CIMO Guide (John Nash is currently working on this).
- The definition of GPS geometric height is not trivial and the GPS geodesy community and meteorological community might need different geometric height bases. For example, it is mentioned by M. J. Mahoney at NASA, http://mtp.jpl.nasa.gov/notes/altitude/altitude.html, that the current GPS is based on a reference ellipsoid defined by World Geodetic System 1984 (WGS-84).

The same GPS method by different manufacturers gives different results. Differences are not large and only revealed by an extensive intercomparison, and are attributed to these causes:

- Wrong surface altitude.
- Bugs/misunderstanding in the equations.
- Different algorithms to detect the launch time (For example, 1 sec difference causes 0.6 hPa pressure difference near the surface).
- Different algorithms to treat the first few minutes after the launch when the GPS signal reception is rather unstable.
- Different algorithms for horizontal winds to filter out and smooth the oscillations due to rotating pendulum motions of the payload.

The WMO report concludes that there is no need of a pressure sensor anymore for operational GPS radiosondes because GPS has proven to be useful and workable for the geopotential height and pressure measurements (see pages 153-162). At upper levels, GPS-derived height and pressure have superior reproducability and lower random error than direct pressure measurements, which is clearly desirable for climate purposes. However, CIMO accuracy guidelines for pressure must be relaxed to 1.5 hPa near the surface to accommodate the lower accuracy of GPS-derived height and pressure near the surface where GPS signal reception can be poor. Therefore, it may still be desirable to retain an independent pressure measurement.

B. Geopotential Height:

The low errors in the heights from GPS radiosondes eliminate one of the main problems for climate scientists with historical radiosonde measurements in the stratosphere, where the errors from the pressure sensors in height assignments were often producing bigger temperature errors than the errors in the temperature sensor itself. Unfortunately, it is almost impossible to trace the systematic errors in these old radiosondes. (page 157)

- Based on the results from Yangjiang, it is proposed that the CIMO Guide requirements should be revised to 15 m/1.5 hPa at the surface (the current number, 0.5 hPa, cannot be achieved with the GPS method), and 120 m/0.2 hPa at 10 hPa. (page 156-157)
- On the other hand, the GRUAN requirements are: precision 0.01 hPa, accuracy 0.1 hPa, and long-term stability 0.1 hPa (GCOS 134). This means that the GRUAN requirements will never be satisfied with the current GPS technology. We need to be aware of this fact.

C. Pressure

Pressure is probably the most difficult meteorological variable to compare reliably, because an ***error of 1 s in synchronisation can produce a systematic bias of 0.6 hPa near the surface****. However, the results in Fig. 10.3.1 suggest that manufacturers should check how they are processing their data close to the ground, to try to minimise errors in the height and pressure computations when locking to the surface values.* (page 162)

- For this reason and other reasons it is **important to have regular procedures for synchronizing data system clocks with a reliable time standard**, with the goal of <1 s accuracy if this is feasible. Automatic synchronization prior to every radiosonde launch is advised. Note that every time a PC is put to sleep and awakened, >1 s of time error is typical.
- How are errors propagated from GPS geometric height through geopotential height to pressure?

D. Horizontal Winds

Gravity waves are mentioned several times. Signals from Rossby and other types of waves may also be included in the measured profiles.

- Gravity wave signals are useful to investigate the radiosonde measurement performance.
- Radiosonde data sets are very useful for gravity wave studies, so the original vertical resolution data should be archived.
- For climate studies, where the average fields are considered, many soundings that can filter out the wave signals are necessary. This condition will be one of the important factors to decide the sounding frequency at sites.
- *Note: The filtering used by a given radiosonde system may be particularly designed for individual flights on a given length of suspension, so may not always be best tuned for the in flight movement of the multiple radiosonde rig relative to the balloon.* (page 170)
- A different configuration of payload (e.g., single radiosonde, or multiple radiosondes) and different rope/unwinder length, and different parachute and balloon size will create oscillations with different frequency and amplitude. A certain filtering algorithm may work for one situation but not for other situations. How should oscillations due to rotating pendulum motions be filtered out of the raw data for the GRUAN data product? Do we prepare several versions of spike filtering software? Or, do we take this issue into consideration in the uncertainty estimation process, and how does the uncertainty change if a radiosonde is flown with a flight configuration different from the configuration that was expected in the software development? This is also a factor to consider regarding the filtering of heat pulses in the temperature measurements.
- How do we set the vertical resolution for the GRUAN data product? 250 m? 1 km? The uncertainty magnitude and thus the GRUAN requirement numbers would depend on the specified vertical resolution.

Note that the comparisons are made for zonal wind and meridional wind, not for wind speed and direction. This is because the wind direction is less meaningful if wind is weak. The GRUAN requirements are for wind speed and direction; this might need to be changed to orthogonal wind components. Comparison between CIMO Guide requirements and GRUAN requirements:

- CIMO Guide:

  Random error (k=1) 1 m/s in the troposphere, 1.5 m/s in the stratosphere

- GCOS 134 (GRUAN requirements):

SPEED	DIRECTION
Precision 0.5 m/s (tropo), 1.0 m/s (strato)	1 deg (tropo), 5 deg (strato)
Accuracy 0.5 m/s	5 deg
Long-term stability 0.1 m/s, 0.5 m/s	1 deg, 5 deg

4. Evaluation of the "scientific" (reference) sensors, including CFH for RH, and Meisei MTR and LMS Multithermistor for temperature. *[Note that a separate task is underway that will consider detailed uncertainty estimation, operational procedures, and metadata by instrument leaders for CFH, NOAA FPH, and Snow White hygrometer, and in part based on a manuscript being prepared by the GRUAN CIMO intercomparison team.]*

- **Uncertainty assessment needed for temperature reference sensors**. Meisei MTR (a very fast sensor) and LMS Multithermistor (which requires no radiation correction) would be ideal reference sensors if a thorough uncertainty analysis had been performed, but without such an analysis neither sensor could serve as a reference in the statistical analysis. *Full documentation about the uncertainties of the Multithermistor instrument is necessary before it could usefully be used for temperature measurements in GRUAN* (page 188). Therefore, the Vaisala RS92 raw temperature measurements corrected using the GRUAN Lead Centre processing was used as the primary temperature reference because the Lead Centre had done an acceptable uncertainty analysis (page 83, 85). [Note: in the last section of this document it is stressed that documentation and validation of the GRUAN data processing is needed.]
- **The MTR 6 Hz temperature measurements reveal transient effects**, and showed that positive and negative pulses probably mainly result from the bamboo rig (page 87), pointing to the need to consider the payload configuration for multiple radiosonde soundings (Task on Dual Sounding led by Hannu Jauhiainen). However, some operational radiosondes also have heat contamination from their own sensor support structures (e.g., pages 76-78.)

The SSI part of the report looks preliminary and immature compared with the QRS part. How to improve the SSI temperature part of the WMO Intercomparison in the future?

- Increase the type and number of SSI sensors. We need to encourage manufacturers and sensor developers much more!
- Understand the detailed mechanism and issues of each sensor, and discuss how to improve, similar to the work for the operational radiosondes. The GRUAN consistency check is important (page 83-84), but more important is to understand the issues with each sensor.
- The LM-Sippican Multithermistor sonde is very important. How do we get more active involvement from Sippican? Perhaps the forthcoming paper by Schmidlin et al. on the theory behind the similar NASA ATM sensor will encourage LMS to be more forthcoming about how the "solution temperature" is derived, and to provide an uncertainty assessment. Active use of this sensor at some GRUAN sites is desirable to gain experience and may also promote more involvement from Sippican.
- For the MTR, we should pose questions and then make specific test flights designed to answer those questions. For example, what are flight configurations for dual or multiple soundings that are free from the temperature contamination issue?

About the CFH:

- The quality control methods and CFH data processing should be documented.

- "Excessive controller instability" required averaging of data into 25 s bins (page 128), and is probably due to the manufacturer change to EN-SCI. The instability also affected the uncertainty estimation (page 135). This year the manufacturer has further changed to DMT (Droplet Measurement Technologies), and we should watch for any quality changes. Obviously addressing the controller instability is desirable.
- There is an unexplained wet bias of the CFH in the lower troposphere (page 132), possibly associated with the region where the dewpoint rather than the frostpoint is measured (i.e., liquid water on the mirror). Should these values be flagged out until the bias is explained and fixed, to avoid confusion and the incorrect assumption that differences are attributable to the operational radiosondes?
- Consider using "Valved balloon descent" for special stratospheric water vapor measurement intercomparisons in the future to avoid water contamination issues from the balloon train on ascent (see page 128-129). (Note that this method cannot be used with hydrogen gas.)
- Determining the altitude where contamination becomes significant seems somewhat subjective.

5. Other lessons and recommendations concerning the WMO intercomparison and report

The way of defining the "references" is quite complicated, and people outside this community may have trouble understanding the logic that it is not validation but intercomparison, with no true standard to refer to. Even such a method can reveal issues with sensors (in their hardware design and/or software), but is a clearer explanation for the various "references" possible?

About uncertainty assessments:

- For all radiosondes, *a better understanding of the sources of measurement uncertainty is needed, i.e. a better understanding of the sensors, their calibration, their implementation and their analysis. Sensors for which this information cannot be obtained either through disclosure by the manufacturer or carefully designed experiments, cannot be used as reference sensors. It is also necessary to be able to eliminate observations where anomalies occur that are not represented in the uncertainty model.* (page 137)
- *Manufacturers are being asked to work with the GRUAN Lead centre to develop models of their measurement uncertainty, so that these can be reported with the measurements submitted to GRUAN.* (page 187)
- *Software corrections that are applied to basic measurements need to be documented, especially for time constant of response and solar heating, whether temperature or relative humidity. The operational community needs to understand what is happening and to evaluate whether they think the corrections are likely to be reliable*. (page 187)
- For use in GRUAN, it is essential that manufacturers produce records of the basic measurements before any corrections are applied (i.e., raw data). (page 187)
- GRUAN uncertainty estimation standards *need suitable vertical resolution requirements added to present a complete picture. For instance, what vertical resolution is required for relative humidity in the upper troposphere in the tropics, 250 m, 500 m, 1 km? If this is established it would then be possible to establish standards about time constant of response errors, and so judge when a sensor gets too slow that its values should not be reported rather than a correction applied. If vertical resolution could be introduced as one of the categories evaluated, the scoring system in Table 12.1 would be improved for upper troposphere relative humidity in the future.* (page 184)
- Manufacturers are strongly encouraged to document all changes to sensor hardware, manufacturing method, calibration function, and other hardware or software changes that affect the measurements, and to make that information and an estimate of the impact on the data publicly available (e.g., on a website), so that future discontinuities in the data record can be correctly attributed to changes in the radiosonde rather than real climate trends.

About dual sounding procedures and launch metadata collection:

- The collected launch metadata was not sufficient for the Yangjiang campaign. A coordinated recording method is needed, which should include: Bamboo rig specification (including each radiosonde's location), the lengths of the rope from the bamboo rig to the sensors and to the parachute/balloon, unwinder (yes or no; the length), parachute (yes or no; the size), balloon type, the launch detail (e.g., calm ascent, rough launch), etc.

- *The part of the CIMO Guide related to instrument intercomparisons needs to be reviewed and updated. For instance, the design model of support rig frame should be formalized and included in CIMO guide, e.g. proper rig material selection, design method considering various simultaneous radiosondes launching model, suitable rig length and radiosonde suspending height from frame to balloon for preventing from extra heat contamination, along with rope length advice between balloon and support frame, as well as different balloon launching method with respect to various ground wind condition, and standard intercomparison procedure, e.g. launching flow, regulation and standard data processing method.* (page 192-193)

Other messages for GRUAN and/or future WMO intercomparisons:

- A publication is needed that fully describes the GRUAN data processing, including but not limited to the applied corrections (radiation and time-lag) and how they were derived, evaluated, and their estimated uncertainty. Other important details that are often erroneously considered "trivial" include: general treatment of outliers; filtering of spurious heating pulses; and factors related to geopotential height, pressure, and winds from GPS (treatment of first few minutes after launch, reference ellipsoid, equation used, smoothing filter for oscillations due to rotating pendulum motion). All details of the data processing matter.
- Future intercomparison campaigns need a documented procedure to identify and fix minor issues with basic data processing. Several manufacturers struggled with such things as calculating the geopotential height from PTU and again from GPS, and minor bugs in incorrect meta data such as station elevation vs surface GPS elevation. Such things are trivial and simple to fix, but first they must be identified and documented. One possibility is to have a first test launch that is scrutinized heavily before the official intercomparison begins.
- Manufacturers who discovered issues late in the Yangjiang campaign or even afterwards suffered from not being able to test and verify their fixes. A procedure is needed for additional observations, reprocessing and/or special campaigns to address how issues that have been identified and supposedly fixed can be verified. One possibility is that selected and willing GRUAN sites could be offered as testing sites for manufacturers, who need to verify or test fixes they implemented. This could be done with some financial contribution by the manufacturer to the common good, like a trust fund.
- It is important to note that the philosophy of the requirements by CIMO and by GRUAN are different. CIMO sets the requirements by considering the current technological level so that the WMO members can make actual choices of one or two highest-grade radiosonde systems currently available. On the other hand, GRUAN sets the requirements purely from the climate science needs which will be updated by new research; GRUAN does not consider the current technological level when determining the requirements. We need to understand both approaches by CIMO and by GRUAN and understand the strengths and weaknesses of both. For example, GRUAN desires consistent measurements over decades for tracking climate changes, yet methods and measurements will improve over time, so great care must be taken to maintain consistent long-term records.
- WMO and GRUAN should pay attention to balloon technology improvement. If meteorological balloons become able to reach 40 km or higher in the future, the climate science may improve very much with perhaps only a small additional cost.