GCOS
Reference
Upper-
Air
Network

GRUAN Report 2

GRUAN Radiosonde Task Team Review Report on the 2010 WMO Radiosonde Intercomparison

LARRY MILOSHEVICH, MASATOMO FUJIWARA, ROLF PHILIPONA, AND
THE RADIOSONDE TASK TEAM

Publisher

GRUAN Lead Centre

Number & Version

GRUAN-RP-2
Rev. 1.0 (2012-12-20)

Document Info

Title:	GRUAN Radiosonde Task Team Review Report on the 2010 WMO Radiosonde Intercomparison
Topic:	Intercomparison
Authors:	Larry Miloshevich, Masatomo Fujiwara, Rolf Philipona, and the Radiosonde Task Team
Publisher:	GRUAN Lead Centre, DWD
Document type:	Report
Document number:	GRUAN-RP-2
Page count:	21
Revision / date:	1.0.0 / 2012-12-20

Abstract

The 2010 World Meteorological Organization (WMO) radiosonde intercomparison in Yangjiang, China (21°50' N, 111°58' E) has been reported in detail by Nash et al. (2011). Here we review results of the intercomparison and implications for the Global Climate Observing System (GCOS) Reference Upper Air Network (GRUAN). Seven GRUAN scientists were involved in the intercomparison, and part of its objective was to provide support and advice to GRUAN.

Italicized comments in this review are direct quotes from Nash et al. (2011). **Bold text** generally reflects our emphasis. Page numbers are from Nash et al. (2011), which is available at:

http://www.wmo.int/pages/prog/www/IMOP/publications-IOM-series.html.

Editor Remarks

Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement by the GRUAN.

There are some specific formats in the digital form of this document (pdf document) which can help to navigate easy through the document: click-able links, e.g.

- Bookmarks in pdf document → complete table of contents
- External links → open linked website or show/download linked document, e.g. http://gruan.org

Table of Contents

1 Introduction

The 2010 World Meteorological Organization (WMO) radiosonde intercomparison in Yangjiang, China (21°50' N, 111°58' E) has been reported in detail by Nash et al. (2011). Here we review results of the intercomparison and implications for the Global Climate Observing System (GCOS) Reference Upper Air Network (GRUAN). Seven GRUAN scientists were involved in the intercomparison, and part of its objective was to provide support and advice to GRUAN. *Italicized comments* in this review are direct quotes from Nash et al. (2011); **Bold text** generally reflects our emphasis; and page numbers are from Nash et al. (2011), which is available at:

http://www.wmo.int/pages/prog/www/IMOP/publications-IOM-series.html

The intercomparison evaluated 10 operational Quality Radiosonde Systems (QRS) and 2 working references (page 9), as well as several research-grade Scientific Sounding Instruments (SSI). The SSI sensors are candidate reference sensors and include the cryogenic frostpoint hygrometer (CFH) for RH, and the 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. Nash et al. (2011) also give recommendations for sensor improvements aimed at manufacturers, and recommendations on operational procedures, data processing algorithms, and documentation that are needed by GRUAN.

One goal of GRUAN is to promote improvement of radiosonde sensors and performance, and Nash et al. (2011) contains many general and specific conclusions and recommendations for improvements that are aimed at motivating and providing a clear forward vision for manufacturers. Although Nash et al. (2011) summarizes conclusions about each radiosonde sensor's readiness for GRUAN in Table 12.1 (pages 180-187), the goal of this review is NOT to pick radiosondes suitable for use by GRUAN. This is partly because the 2010 intercomparison 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 suppliers produce superior radiosondes.

This review summarizes lessons from Nash et al. (2011) that are relevant to GRUAN and its goals. It is organized into the following sections:

- Recommendations for improving radiosonde temperature (T) and relative humidity (RH) measurements
- T and RH corrections used by manufacturers in their standard data products
- Other radiosonde measurements (GPS, pressure, altitude, winds)
- Evaluation of the "scientific" (potential reference) sensors
- Other lessons and recommendations concerning the WMO intercomparison and report

2 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, Vaisala data systems 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)

2.A Temperature Sensors

- ***From these*** (subtropical Yangjiang) ***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, 2009) 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.
- **A robust analysis of the sensor calibration uncertainty should extend down to -100 °C**, otherwise there is danger of calibration bias in the tropical UT. Measurements are only valid over the temperature range of the calibration uncertainty analysis. Table 4.1.2 (page 24) shows the current minimum valid temperature of the sensor calibration for each manufacturer; most are 10-20 °C warmer than the recommended -100 °C.
- **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.* (page 82) *Also, it is recommended that white paint is not used on radiosonde temperature sensors for both operational and climate purposes.* (page 75)
- **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 (page 65) include:
  1) mount the sensor high enough that, while swinging, air flowing directly over the top of the radiosonde does not reach the sensor;
  2) remove unnecessary supports above the sensor to ensure unobstructed exposure to the air;

3) expose a similar sensor cross-section to the sun as the radiosonde rotates; and
4) minimize changes to the radiative characteristics of the sensor when the manufacturing process or coating is changed (and revise correction tables or algorithms accordingly).

Mountings of systems with high quality daytime measurements may provide a good guide for radiosondes with poorer quality daytime measurements. Any changes made to sensors or to algorithms or correction tables should be thoroughly documented.

2.B Relative Humidity Sensors

Most RH sensors on the intercompared radiosondes were thin-film capacitance sensors. Vaisala RS92 has dual sensors exposed directly to the air, which are alternately heated to drive off water and ice down to -60 °C, but for the remainder of the flight only one sensor makes measurements. Graw, LMS, InterMet, Huayun and Jinyang all use sensors manufactured by E+E; however, differences in E+E sensor models and their implementation led to different performance. Some radiosonde manufacturers measure the temperature of the RH sensor directly, while others 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.

- **A robust analysis of the RH sensor calibration uncertainty should extend down to -100 °C**, otherwise there is danger of calibration bias in the tropical UT. **Measurements are only valid over the temperature and RH range of the calibration uncertainty analysis.** Table 4.1.2 (page 24) shows the current minimum valid temperature of the sensor calibration for each manufacturer, which in many cases is much warmer than -100 °C.
- ***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, quantized jumps of 1 %RH produce non-physical discontinuities in mixing ratio and create unnecessary and substantial random-like relative errors (e.g., ±0.5 % RH is ±25 % error at RH=2 %). Additional significant figures in the RH data are already present in the "FLEDT" processed data files, but this is not the standard (default) operational data product.
- The Vaisala method of alternately heating two RH sensors 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 cycle to lower temperatures where ice-supersaturated conditions can still exist, given the longer post-heating sensor recovery time required 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 (e_w) is a critical concern for RH calibrations at low temperatures**, where RH is defined as the ratio of the water vapor partial pressure (e) to e_w. Figure 8.2.8 (page 139) shows the ratio of various commonly-used e_w equations relative to Wexler (1977), which was chosen not because it is inherently better than any other e_w equation, but because it is used by NIST (National Institute of Standards and Technology). Differences become significant below about -40 °C, and reach ~20 % at -90 °C. Table 4.1.2 (page 24) lists the e_w equations used by different manufacturers.

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.* (page 140)

Note that any equation for e_w is only a theoretical construct below about -35 °C, as it is not subject to experimental verification below -35 °C. To avoid unnecessary errors it is critical that users of RH data use the same SVP equation used by the manufacturer to compute water vapor partial pressure, mixing ratio, etc. from RH values. Therefore, **manufacturers must disclose the exact SVP equation used in their RH calibration**. For example, using Goff (1965) instead of Wexler (1977) with Vaisala RH measurements at low temperatures leads to large errors in the vapor pressure and other derived quantities. Note also that, in contrast, the various equations for SVP over ice (e_i) 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 reported with respect to liquid water, so their conversion to RH with respect to ice is similarly subject to large errors if the wrong equation for e_w is used. See discussion in Murphy and Koop (2005), and recommendations in Appendix A of Miloshevich et al. (2006).

3 T and RH corrections used by manufacturers with their standard data products

Here we do not consider GRUAN-applied corrections (which is a separate topic), nor the comprehensive uncertainty analysis for each data point that is expected from each manufacturer whose radiosondes are used in 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)

3.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.

3.A.1 Daytime software radiation correction

- Radiation corrections at 10 hPa are shown in Table 7.1.2 (page 64); these range 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 corrections for other sensors that assume simple solar heating (e.g., a function of only solar elevation angle and air pressure). However, Nash et al. (2011) found an altitude-dependent warm bias in the daytime LMS Multithermistor measurements relative to the adopted reference measurements (GRUAN-corrected Vaisala RS92), suggesting that systematic uncertainties in both the Multithermistor measurements and the RS92 radiation correction need further study (page 85 and Figure 7.2.4). LMS needs to disclose details of the solution temperature calculation as well as conduct a thorough uncertainty analysis before the Multithermistor could be considered as a reference sensor. F. Schmidlin suggests, based on test data from the similar NASA ATM radiosonde and on recent comparisons of the ATM and LMS Multithermistor, that two important sources of uncertainty need to be characterized:

  1) the absorptivity of the actual sensors, and

  2) the manufacturing variability of sensor dimensions, coating thickness, etc.

- Further in-situ or lab experiments are needed that clearly and consistently demonstrate the errors produced by solar and thermal radiation. If such experiments were performed in the past by manufacturers they should be reported to help the radiosonde community better understand these errors.

3.A.2 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 tether/unwinder length, and different parachutes and balloon sizes will create spikes with different frequencies and amplitudes. A certain filtering algorithm may work for one situation but not for others. Are several versions of spike filtering software needed for GRUAN, or should the GRUAN data processing algorithms generate uncertainty estimates that account for these differences? In the latter case, measurement uncertainties would necessarily be larger for radiosondes flown in different configurations than were used to develop the filtering algorithm(s).
- The best approach is to develop a sensor and sensor mount/support structure (and a payload configuration for dual soundings) that are free from positive/negative pulses.

3.A.3 Time Constant of Response

- Most radiosonde temperature sensors are estimated to have a response time of 4-5 s at 10 hPa with 6 m/s ascent rate, although some slower sensors have a response time up to 14 s. *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 lags of temperature sensors themselves. Manufacturers should provide assessments of the time constant of response for their temperature sensing systems (page 211).

3.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 (caused by solar heating of the RH sensor). Some manufacturers directly measure the temperature of the RH sensor and use it in the calibration function, so this approach to addressing solar heating of the RH sensor is not a correction at all. Other manufacturers apply a software correction that is based on an estimate of the sensor heating that is assumed to be a constant value for given measurement conditions. In these cases the corrected RH value does not account for 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.**

3.B.1 Response Time RH Corrections

- **Response times of RH sensors are quite long 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. The sensors often require several km to transition from high RH values in the UT to low stratospheric 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-response CFH measurements. The following points are made by Nash et al. (2011) on pages 115-119:

- ◦ *More work is required to establish the uncertainty in the corrected Graw and Vaisala observations.* For example, further detailed comparisons between Graw and Vaisala corrected measurements and with CFH measurements from the WMO intercomparison dataset would be helpful. Also, comparison of multiple versions of 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 of the correction. 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 relative to Vaisala and Snow White in the tropopause region; however, these instances may involve factors other than the time-lag correction such as sensor icing in ice-supersaturated conditions. Therefore, more work is required to establish the reliability of the corrected Graw measurements, especially in difficult environmental conditions.
- ◦ *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.***

3.B.2 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 or assumed cloud conditions, but the effect on solar heating of the actual and instantaneous cloud optical depth between the sun and the sensor cannot be represented, leading to potentially large errors in the correction. 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/or any underlying cloud, etc. Another concern is the large difference in response times of the RH and T sensors, where estimates of the temperature of the RH sensor based on the T sensor measurements may be poorly synchronized due to the different time lags. Furthermore, a solar radiation correction that correctly addresses all these factors for a particular location and season may not be correct for another location or season, and this may introduce geographic and seasonal biases into the record. Therefore, although an RH solar radiation correction is an improvement over raw measurements where day-night differences are large, a theoretical or empirical software correction is much less desirable than direct measurements of 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 radiation 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 solar radiaton correction appears to be too large at high humidity **for Yangjiang conditions**, and is 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.

4 Other Radiosonde Measurements (GPS, pressure, altitude, winds)

4.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. As mentioned by M. J. Mahoney (NASA) at http://mtp.mjmahoney.net/www/notes/altitude/altitude.html, current GPS measurements are defined relative to the World Geodetic System 1984 (WGS-84) reference ellipsoid, and these height values are typically converted inside a GPS receiver into heights above mean sea level using a more accurate geoid model, currently Earth Gravitational Model 1996 (EGM96).

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

- Wrong surface altitude (initialization).
- Bugs/misunderstanding in the equations or their software implementation.
- Different algorithms detect launch at different times. For example, 1 s time difference causes a 0.6 hPa pressure difference near the surface if data from different sensors are matched on the basis of time.
- Different algorithms to produce GPS altitude values during 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.

Nash et al. (2011) conclude that there is no longer a need for an onboard pressure sensor 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 reproducibility 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.

4.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 altitude and pressure requirements be revised to 15 m/1.5 hPa at the surface (the current 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 pressure requirements are: precision 0.01 hPa, accuracy 0.1 hPa, and long-term stability 0.1 hPa (GCOS 2009). This means that the GRUAN requirements will never be satisfied *at the surface* with current GPS technology.

4.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.* (page 162) If measurements from two radiosondes on the same balloon are compared on the basis of time relative to launch, then the launch detection algorithms on the two data systems must choose the same launch time within a fraction of a second.
- *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 160)
- How are errors propagated from GPS geometric height through geopotential height to pressure? An analytical solution for the error propagation is difficult to deduce from the equations. The uncertainty related to errors from GPS height, as well as e.g. temperature measurement, may be approximated using simulations and statistical methods.

4.D Horizontal Winds

Gravity waves are mentioned several times by Nash et al. (2011). Signals from Rossby and other types of waves may also be included in the measured profiles.

- Gravity wave signals are useful for investigating 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)
- Different payload configurations (e.g., single radiosonde, or multiple radiosondes) with different rope/unwinder lengths, parachutes and balloon sizes will create oscillations with different frequencies and amplitudes in the wind data. A certain filtering algorithm may work for one payload configuration but not for others. How should oscillations due to rotating pendulum motions be filtered out of the raw data for the GRUAN data product? Do we prepare different versions of spike filtering software for different payload configurations? Or, should we estimate how the uncertainty changes if a radiosonde is flown with a payload configuration different from that used to develop the filtering algorithm(s)? Similar considerations apply to filtering out heat pulses in the temperature measurements (Section 2.A above).
- 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 orthogonal zonal and meridional winds, not for wind speed and direction, because the wind direction is less meaningful at low wind speeds. The GRUAN requirements are for wind speed and direction; this might need to be changed to orthogonal wind component vectors. Comparison between CIMO Guide requirements and GRUAN requirements:

- CIMO Guide:

	Speed [m/s]	**Direction [deg]**
Random error (k=1)	1.0 m/s (tropo), 1.5 m/s (strato)	–

- GRUAN requirements (GCOS 2009):

	Speed [m/s]	**Direction [deg]**
Precision	0.5 m/s (tropo), 1.0 m/s (strato)	1 ° (tropo), 5 ° (strato)
Accuracy	0.5 m/s	5 °
Long-term stability	0.1 m/s (tropo), 0.5 m/s (strato)	1 ° (tropo), 5 ° (strato)

5 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 to consider detailed uncertainty estimation, operational procedures, and metadata by instrument leaders for CFH, NOAA FPH, and Snow White hygrometer, 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 thorough analysis of their uncertainties had been performed, but without such documentation neither sensor can now be classified as a reference sensor. *Full documentation about the uncertainties of the Multithermistor instrument is necessary before it could usefully be used for temperature measurements in GRUAN* (page 188). Temperature measurements by the Vaisala RS92 and corrected by the GRUAN Lead Centre (based on extensive uncertainty analyses) were used as the primary temperature reference (page 83, 85). In the last section of this review we stress 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 resulted 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 Nash et al. (2011) seems 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 it, similar to the work for the operational radiosondes. The GRUAN consistency check is important (page 83-84), but more important is the need to fully understand the issues with each sensor.
- The LM-Sippican Multithermistor sonde is very important to the progress of radiosonde temperature measurements. How do we get more active involvement from Sippican? A paper in preparation by Schmidlin et al. on the theory behind the similar NASA Accurate Temperature Measuring (ATM) radiosonde sensor may encourage LMS to divulge how their "solution temperature" is derived, and to provide a quantitative uncertainty assessment. Active use of this sensor at some GRUAN sites is desirable to gain experience, and may also stimulate 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 flight configurations for dual or multiple soundings are free from the temperature contamination issue?

5.A About the CFH

- The quality control methods and CFH data processing should be documented.
- “Excessive controller instability” required averaging of data into 25 s (125 m) bins (page 128), and probably resulted from the manufacturer change to En-Sci. The instability also affected the uncertainty estimation (page 135). This year the manufacturer has further

changed to Droplet Measurement Technologies (DMT), 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 pages 128-129). However, this method cannot be used with hydrogen-filled balloons without significant modifications to the valve hardware.
- Determining the altitude during balloon ascent where water vapor contamination in the stratosphere becomes significant seems somewhat subjective without high-quality, contamination-free measurements during descent for comparison.

6 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 that, having no true standard to refer to, results can be portrayed only as intercomparisons, not validations. While intercomparisons can reveal issues with sensors (in their hardware design and/or software), is a clearer explanation of how various sensors were identified as and served as “references” possible?

6.A 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) These will need to be thoroughly vetted by GRUAN with analyses of data from multiple sensors launched on the same balloon.
- *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 performance 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 methods, calibration methods and response functions, 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.

6.B About dual sounding procedures and launch metadata collection

- The collected launch metadata were 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 weather details (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 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 conditions, and standard intercomparison procedure, e.g. launching flow, regulation and standard data processing method.* (pages 192-193)

6.C 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 the conversion of GPS data into geopotential height, pressure, and winds (especially how the first few minutes of data after the launch are treated, the 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 errors in, and homogenize the basic data processing. Several manufacturers struggled to calculate the geopotential height from pressure, temperature, and humidity. There were also 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 with results that are closely scrutinized for calculational and methodological errors before the official intercomparison begins.
- Routine procedures should be established for synchronizing instrument data system clocks to a reliable time standard, with the goal of maintaining <1 s accuracy if this is feasible. This is especially important for comparing radiosonde measurements to other measurements such as lidar on the basis of absolute time (as opposed to comparing two radiosondes on the same balloon on the basis of time relative to launch). Automatic time synchronization is recommended, perhaps daily or hourly, or prior to every radiosonde launch. Note that >1 s of time error is typical when a PC is put to sleep and awakened, so synchronization after PC sleep is advised.
- 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 to permit verification of such fixes through additional observations, data reprocessing and/or special campaigns. One possibility is that selected and willing GRUAN sites could be offered as testing sites for manufacturers who require post-intercomparison verification of their fixes. This could be done with some financial contribution by the manufacturer to the common good, like a trust fund, or through direct payment to a GRUAN site for their expertise, time and efforts in performing the tests.
- It is important to note that the philosophy of the requirements by CIMO and by GRUAN are different. CIMO sets the requirements so that the WMO members can make actual choices from the radiosonde systems currently available. 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, even when they exceed the abilities of current technologies. We need to understand both CIMO and GRUAN approaches to meaningfully evaluate the strengths and weaknesses of each. For example, GRUAN desires consistent measurements over decades to track climate changes, yet the methods and sensors employed will undoubtedly change 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 available that can reach 40 km or higher in the future, the climate science may improve very much with perhaps only a small additional cost (assuming that the sensors deployed on the balloons can produce the same or better quality data at 40 km that is currently obtained at 30 km).

7 References

Global Climate Observing System (GCOS), 2009: GRUAN Implementation Plan 2009-2013. GCOS-134: WMO/TD No. 1506. [pdf]

Goff, J.A., 1965: Humidity and moisture: saturation pressure of water on the New Kelvin scale, A. Wexler, Ed., Reinhold, 289.

Hyland, R. W. and A. Wexler, 1983: Formulations for the Thermodynamic Properties of the saturated Phases of H2O from 173.15 K to 473.15 K, ASHRAE Trans, 89(2A), 500-519.

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.

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.

Nash, J., T. Oakley, H. Vömel, and LI Wei, 2011: WMO Intercomparison of high quality radiosonde observing systems Yangjiang, China, 12 July – 3 August 2010. World Meteorological Organization Instruments and Observing Methods, Report IOM-107, WMO/TD-No. 1580. [pdf]

Sonntag, D., 1994: Advancements in the field of hygrometry, Meteorol. Z., N. F., 3, 51-66. Vaisala, available at http://www.vaisala.com/Vaisala%20Documents/Vaisala%20News%20Articles/VN184/VN184_16_WMOIntercomparisonofRadiosondeSystemsinChina.pdf, 2010.

Wexler, A., 1977: Vapor pressure formulation for ice, Journal of Research of the National Bureau of Standards-A. 81A, 5-20.