

[Review comments in red, author response in black]

Reviewer #1

The title is rather long and likely a better title can be found. It does not need to be so specific. Suggestions from my side are: "Managing the transition from Vaisala RS92 to RS41 radiosondes within the GRUAN: a progress report" or "Fully traceable switch from Vaisala RS92 to RS41 radiosondes within the GRUAN: a progress report". I think one should either write GRUAN or "Global Climate Observing System Reference Upper-Air Network", since there will be few readers who know GCOS but do not know GRUAN.

Thank you for the suggestion. We will change the title to "Managing the transition from Vaisala RS92 to RS41 radiosondes within the Global Climate Observing System Reference Upper-Air Network (GRUAN): a progress report"

There is some redundancy in stating the importance of characterizing the errors and differences. For example, the description of differences with the transition handling at Tateno (1130-136) could be omitted.

We would prefer to keep the lines 130-136 because they summarize the differences in the scale of the approaches between:

- a) an instrument change at a single site and in contrast,
- b) the GRUAN approach to maintain the consistency of the data records at multiple sites within a network, which includes a coordinated network-wide effort, an extensive laboratory characterization, and the call on expertise from various fields that are represented in the network.

[Review comments in red, author response in black]

Reviewer #2

Abstract

Lines 22 – 24: See my comments on section 5.2: corresponding modifications are needed here.

Agreed. The text will be changed to read

"[...] with uncertainties of < 0.2 K for the temperature and $< 1.5\%$ RH for the humidity sensor."

1 Introduction

Lines 32 – 35: The sentence "The network consists ... in the processing chain." Is very long, complicated and difficult to follow for a reader.

The sentence will be rephrased and restructured to increase readability.

Line 36: "the only GRUAN data products" Should this be " the only sources of the GRUAN data products"?

For clarity the sentence will be changed to start with

"[...]the GRUAN data products for the Vaisala RS92 and the Meisei RS-11G radiosondes are the only [...]"

3 Change management

Line 192: "each other's uncertainty coverage factor"

the coverage factor only specifies the confidence level at which an uncertainty is given; therefore this text should probably be "the uncertainties of the data at the 95% confidence level"

We agree with this comment and the sentence will be changed accordingly.

Lines 207 - 208: "consider the metrological quantification aspects that are necessary in the current case."

such as? to which aspects are referred here?

Aspects such as type B (instrumental) expanded uncertainties (absolute for each radiosonde), relative uncertainty in the comparison, traceability of involved instrumentation, and assessment of the comparability with independent ancillary results.

These are referred to here since this is the key aspect in adopting a metrological approach, where other reference instruments are adopted to make the evaluation traceable to standards and with a documented uncertainty budget.

The manuscript will be modified to include these aspects listed above.

5.2 Results of the laboratory characterization

There is a clear shortcoming in Section 5: This paper emphasizes several times the metrological aspects but no uncertainty values are given for the calibration systems in Section 5.

Thank you for pointing out this omission. We will include information about the uncertainty of the references for the temperature and relative humidity.

The uncertainty of the reference Pt100 temperature sensor is 0.04K (1 sigma), certified by DAkkS (the national accreditation agency in Germany).

The uncertainty of the relative humidity of the air over the reference salt mixtures is at most 0.5%RH, as reported by Greenspan (1977) doi:

10.6028/jres.081A.011.

A reference to Greenspan (1977) and a statement about the calibration uncertainty of the reference temperature sensor will both be included in the updated manuscript.

Line 324: “from its width it can be inferred that the calibration uncertainty is smaller than 0.1 K.”

o Such conclusion cannot be drawn without knowing the uncertainty of the reference. (Basically, you may possibly conclude that the equivalence between your reference and Vaisala's reference is within 0.1 K but the link to the SI realisations is still missing.)

A link to SI realisations is provided by the certification by DAkkS. The calibration uncertainty of the Pt100 0.04K (1 sigma).

For clarity, Figure 4 will be updated to show the percentiles that mark the boundaries of 1-sigma and 2-sigma ranges. The boundaries show that the calibration uncertainty of the RS41s temperature sensor is within 0.2K (1 sigma).

The text will be changed to reflect this.

Lines 328 - 329: “which means that the uncertainty of the humidity calibration is smaller than 1 %RH.”

o Same comment as for temperature above.

The uncertainty of the relative humidity of the air over the reference salt mixture at room temperature is at most 0.5%RH, as reported by Greenspan (1977).

Table 4 will be updated to include the uncertainty in the relative humidity of the air over each reference salt mixture.

Taking into account the uncertainty of the RH references and the histogram distributions at each tested RH level (discussed further in the following points), we will change this statement to “[..]the uncertainty of the humidity calibration is smaller than 1.5 %RH.”

o Moreover, the uncertainty of the reference (both at Vaisala and Lead Centre) is significantly at different RH-levels; therefore analysing them together in single histogram is misleading.

In Figure 5, histograms will be added for the measurements at 0, 11, 33, 75, 100 %RH. We will discuss the calibration uncertainty of the radiosonde's RH sensor at each of these levels.

o Asymmetry of the histogram indicates probably some systematic differences between Vaisala and Lead Centre at some measurement points.

The asymmetry of the histogram in Figure 5 is due to an error in the analysis, where we used incorrect reference values at 33%RH. We want to thank the reviewer for this comment, which prompted us to re-check our analysis of the measurements. The incorrect reference values at 33%RH caused a shoulder in the histogram at around -0.75%RH. With the correct reference values, this

shoulder disappears and the histogram distribution becomes more symmetric, although a slight asymmetry remains.

This remaining asymmetry is caused by the positions and shape of the histograms at different RH levels, which does indeed show discrepancies between the reference salt mixtures and the calibration by Vaisala. A statement to this effect will be added in the manuscript.

o Finally, for me it looks strange that in Fig. 4 all visible data fall in claimed uncertainty but here a significant amount of data fall out from the range of claimed uncertainty.

The updated Figure 4 shows that ± 0.5 K covers more than 2 sigma of the distribution of the differences between the RS41 temperature sensor and the Pt100 reference. In fact only 32 of 1406 measurements (2.2%) are outside the ± 0.5 K range.

We agree that due to the non-Gaussian shape of the distribution ± 0.1 K does not cover 1 sigma, and that ± 0.2 K is more appropriate. The text will be changed accordingly.

6 Metrology

Lines 400 - 414:

o It's unclear to me why these two last paragraphs of this Section are needed here: I suggest focusing them in facilities/services/results developed/obtained at national metrology institutes that can be benefited in the transition from RS92 to RS41 (and in radiosonde measurements in general)

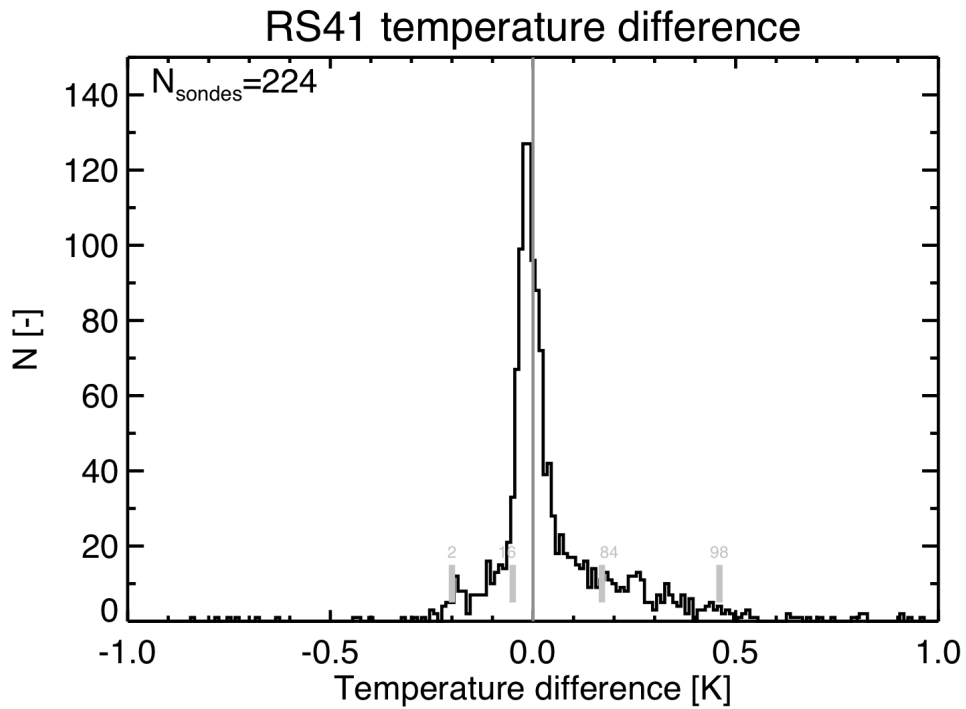
The last two paragraphs will be removed, but we would like to keep a sentence referring to the Meteomet project, that concerns the investigation of environmental factors on the uncertainty of meteorological measurements and therefore is relevant to the paper.

10 Summary and outlook

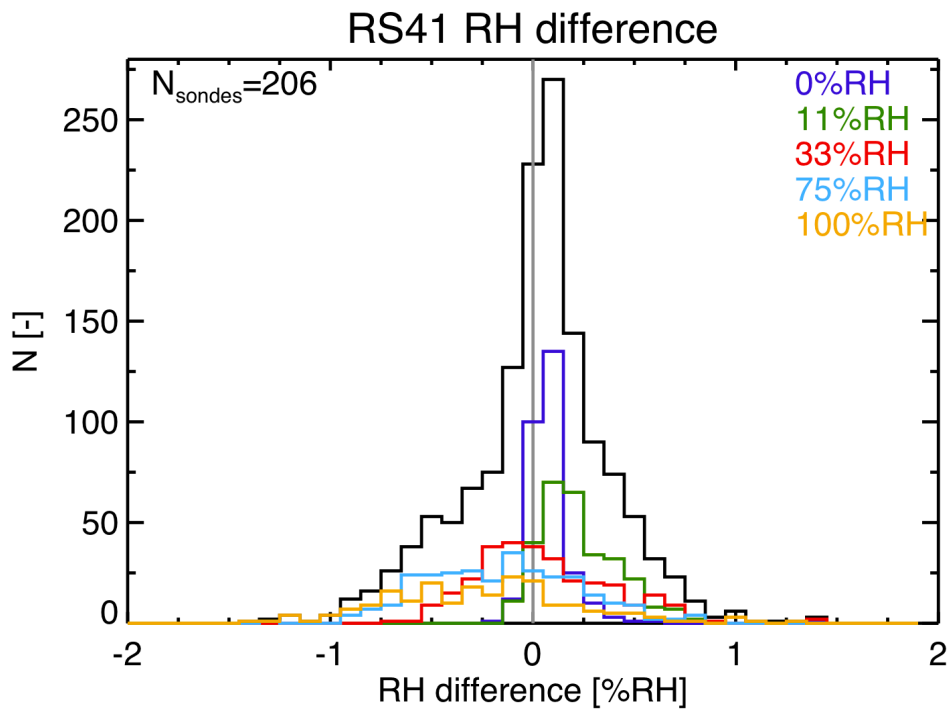
Lines 670 - 672 "2. Comparison with external references show calibration uncertainties of < 0.1 K for temperature and $< 1\%$ RH for the humidity sensor."

o Ref. to my comments in Section 5.2, I don't agree with this sentence.

Agreed. The text will be changed to read uncertainties of < 0.2 K for temperature and $< 1.5\%$ RH



Updated Figure 4



Updated Figure 5.

[internal comment for the co-authors] In the previous version of the plot, the data for 33%RH were shifted towards -75%RH due to an error in the analysis.

This has now been corrected. The peaks of the histograms at the various RH levels don't coincide, which is due to differences between the Vaisala calibration and the reference salts. The reference salts are pretty good (uncertainty $\sim 0.2\%RH$) so I think the differences at 75 and 100%RH are due to Vaisala's calibration. This RH-dependent error is consistent with SHC tests we performed with the RS92.

Manuscript prepared for Geosci. Instrum. Method. Data Syst.
with version 5.0 of the L^AT_EX class copernicus.cls.

Date: 14 May 2020

Progress in managing the transition from the RS92 to the Vaisala RS41 as the operational radiosonde within the GCOS Reference Upper-Air Network

Ruud J. Dirksen¹, Greg E. Bodeker², Peter W. Thorne³, Andrea Merlone⁴, Tony Reale⁵, Junhong Wang⁶, Dale F. Hurst^{7,8}, Belay B. Demoz⁹, Tom D. Gardiner¹⁰, Bruce Ingleby¹¹, Michael Sommer¹, Christoph von Rohden¹, and Thierry Leblanc¹²

¹GRUAN Lead Centre, Deutscher Wetterdienst, Meteorologisches Observatorium Lindenberg, Am Observatorium 12, 15848 Tauche/Lindenberg, Germany

²Bodeker Scientific, Alexandra, New Zealand

³Maynooth University, Maynooth, Ireland

⁴INRI, Turin, Italy

⁵NOAA/NESDIS, Washington DC, USA

⁶Department of Atmospheric and Environmental Sciences, State University of New York, Albany, USA

⁷Global Monitoring Laboratory, NOAA Earth System Research Laboratories, Boulder, CO, USA

⁸Cooperative Institute for Research in Environmental Sciences, University of Colorado, CO, USA

⁹Howard University, Washington D.C., USA

¹⁰National Physical Laboratory, Teddington, UK

¹¹ECMWF, Reading, UK

¹²JPL, Pasadena, USA

Correspondence to: R. J. Dirksen (Ruud.Dirksen@dwd.de)

Abstract. This paper describes the GRUAN-wide approach to manage the transition from the Vaisala RS92 to the Vaisala RS41 as the operational radiosonde. The goal of the GCOS Reference Upper-Air Network (GRUAN) is to provide long-term high-quality reference observations of upper air Essential Climate Variables (ECVs) such as temperature and water vapor. With GRUAN data being used for climate monitoring, it is vital that the change of measurement system does not introduce inhomogeneities in to the data record. The majority of the 27 GRUAN sites were launching the RS92 as their operational radiosonde, and following the end of production of the RS92 in the last quarter of 2017, most of these sites have now switched to the RS41. Such a large-scale change in instrumentation is unprecedented in the history of GRUAN and poses a challenge for the network. Several measurement programmes have been initiated to characterize differences in biases, uncertainties and noise between the two radiosonde types. These include laboratory characterization of measurement errors, extensive twin sounding studies with RS92 and RS41 on the same balloon, and comparison with ancillary data. This integrated approach is commensurate with the GRUAN principles of traceability and deliberate redundancy. A two-year period of regular twin soundings is recommended, and for sites that are not able to implement this burden sharing is employed, such that measurements at a certain site are considered representative of other sites with similar climatological characteristics. All data relevant to the RS92-RS41 transition are archived in a database that

will be accessible to the scientific community for external scrutiny. Furthermore, the knowledge and experience gained about GRUAN's RS92-RS41 transition will be extensively documented to ensure traceability of the process. This documentation will benefit other networks in managing changes in their operational radiosonde systems.

Preliminary analysis of the laboratory experiments indicates that the manufacturer's calibration of the RS41's temperature and humidity sensors is more accurate than for the RS92; with uncertainties of < 0.102 K for the temperature and < 1.5 %RH for the humidity sensor. A first analysis of 224 RS92-RS41 twin soundings at Lindenberg Observatory shows nighttime temperature differences < 0.1 K between the Vaisala-processed temperature data of the RS41 (T_{RS41}) and the GRUAN dataproduct for the RS92 ($T_{RS92-GDP2}$). However, daytime temperature differences in the stratosphere increase steadily with altitude, with $T_{RS92-GDP2}$ up to 0.6 K higher than T_{RS41} at 35 km. RH_{RS41} values are up to 8 % higher, which is consistent with the analysis of satellite-radiosonde collocations.

30 1 Introduction

The Global Climate Observing System (GCOS) has instigated the GCOS Reference Upper-Air Network (GRUAN) (Bodeker et al., 2016) to perform reference quality measurements of upper-air Essential Climate Variables (ECVs) (Bojinski et al., 2014). ~~The network consists of a range of national contributions of high-quality observing facilities undertaking observations in a systematically similar manner that assures traceability to System of Units (SI) or community accepted standards, with full quantification of the uncertainties arising from every step in the processing chain.~~ The network consists of a range of national contributions of high-quality observing facilities that undertake observations in a systematically similar manner. The employed method of observation and subsequent data processing assures traceability to System of Units (SI) or community accepted standards, with a full quantification of the uncertainties arising from every step in the processing chain. To date, the GRUAN data products for the Vaisala RS92 and the Meisei RS-11G radiosondes are the only ones ~~GRUAN data products~~ currently being routinely produced and disseminated to the user community (Dirksen et al., 2014; Kizu et al., 2018). Several additional products, including those from other radiosonde models, ozone-sonde and a variety of remote sensing techniques, are at various stages of maturity and should be available for analysis in the near future. Readers interested in further details of GRUAN, its development and its operation, are encouraged to read Bodeker et al. (2016) or to visit the GRUAN website at <https://www.gruan.org>. A map showing sites participating in GRUAN is shown in Figure 1. Currently, RS92 data from GRUAN sites are available via anonymous ftp from NOAA's National Centers for Environmental Information at <ftp://ftp.ncdc.noaa.gov/pub/data/gruan/processing/level2/RS92-GDP/version-002>.

In ~~the course of~~2014, Vaisala announced the introduction of the RS41 as the RS92's successor, and that the production of the RS92 radiosonde would be terminated by the end of 2017. Before

its end of production, Vaisala's RS92 radiosonde was also widely used outside of GRUAN, with a global market share of approximately 30 % (including at least daily launches at sites on every continent). Its performance was among the best of the commercially available radiosonde models (Nash et al., 2011). Until recently, the majority of the 27 GRUAN sites employed the RS92 (listed in Table 1 and shown in Figure 1), which effectively made the RS92 GRUAN's backbone in terms of upper-air sounding. Any change in instrumentation in a GOS (Global Observing System) network not only presents potential data continuity concerns, but it will also pose a challenge to a far broader community of users. Radiosounding data form a key input to NWP systems and human forecasts such that any differencechange in performance has potentially large impacts. The challenge is to ensure continuity of operations without negative scientific or financial ramifications. The manual on the GOS states that "Changes of bias caused by changes in instrumentation should be evaluated by a sufficient period of observation (perhaps as much as a year) or by making use of the results of instrument intercomparisons made at designated test sites" (Section 2.2.2.13 of WMO, 2017).

One of the key potential benefits of a tiered networks design (Bodeker et al., 2016) is the dissemination of information derived from a subset of top-tier reference quality sites down to the geographically broader lower-tier network sites. Figure 2 schematically depicts a three-tiered upper-air observing system architecture, with a 30-40 station GRUAN network providing reference observations for more extensive networks such as the GCOS upper-air network (GUAN). Reference network sites serve as the long-term anchor points that comprehensively characterize the atmospheric column with the highest quality measurements currently feasible. The base of the system is the entire global upper-air observing system, serving a wide variety of purposes, primarily weather prediction, and including the operational radiosonde network, aircraft and satellite observations, etc., and embracing model-assimilated upper-air datasets and reanalyses. The 177-station (as of March 2019) GUAN is a subset of the operational radiosonde network that, in the late 1990s, committed to long-term, consistent observations, but does not deploy any special instruments for high-quality climate observations as GRUAN does.

In the case of the transition from RS92 to other radiosonde types, the lessons learnt from GRUAN activities to manage the transition may benefit those GUAN sites faced with the same challenge as well as other sonde stations from the remainder of the GOS. Furthermore, by undertaking an intensive characterization of the transition, GRUAN can assist not just the climate community but also other communities such as numerical weather prediction (NWP)/forecasting through active dissemination of the resulting analyses of any effects of the transition. Such an approach requires visibility of the change management process and associated analysis results by the community through transparency of the GRUAN transition results, which will provide the confidence required to encourage uptake of the results.

This paper is a first step in this process and serves to highlight the activities undertaken to date and planned future activities that GRUAN (including the sites, Lead Centre, working group and

90 various task teams) intends to carry out in support of the change management. Community feedback, which may strengthen the planned activities, is strongly encouraged. In particular we would eagerly welcome support in:

- Provision of available data from RS92 + RS41 twin launches (i.e., both on the same balloon) from non-GRUAN sites to expand the geographical characterization of any biases between
95 RS92 radiosondes and their successors.
- Offers of assistance from the expert user community in the synthesis of twin launch data from multiple sites.
- Suggestions of additional analyses that could be performed.

The remainder of this paper outlines various aspects of the change strategy, such as a network-wide
100 approach including burden-sharing, the application of ancillary data, the metrological perspective on the change, the role of documentation and the creation of a scientific database which stores all measurement data which are relevant to the change. Furthermore, it reports on the progress to date and current plans to diagnose biases between RS92 and RS41 radiosondes and provides some initial results based on analysis of 224 twin soundings that have been performed at the Lindenberg
105 Observatory. The data [from](#) the laboratory experiments that were used for the plots in this paper are stored in a permanent repository with digital object identifier (DOI) 10.5676/GRUAN/dpkg-2019-1 (Dirksen and Sommer, 2019).

2 The challenge

For any long-term measurement series, inevitably, the challenge of change management is certain
110 to arise. This could be either through choice when improved or more efficient means of making the measurements become available, or [throughby](#) necessity as an instrument and its manufacturer support become unavailable. Regardless of whether by choice or necessity, the central challenge is to manage the transition in such a manner that the adverse effects on a long-term measurement series' continuity and homogeneity are minimized and any uncertainties propagated by the transition
115 are well understood. Sufficient data, and associated metadata, are required both to perform an initial analysis and to permit future reprocessing and reanalysis of the data.

Several National Weather Service (NWS) organizations implement change management strategies that are executed within their organization. An example of such an activity is the Radiosonde Replacement Plan for the NOAA/NWS (see Peterson and Durre (2004)).

120 While GRUAN does its best to avoid unnecessary changes in its operating protocols, it is not immune from the challenges of change management. Change management has been a key component of the network's goal to constantly strive to make the best possible measurements of the relevant

ECVs. As such, GRUAN has never been envisioned as a network where the underlying instrumentation shall remain forever static. Prior to this forced transition away from RS92 sondes, only one
125 GRUAN site (Tateno, Japan) had undergone a change in their radiosounding system, namely the switch from the Meisei RS2-91 to Vaisala's RS92 radiosonde in 2009-2010 (Kobayashi et al., 2012). This site subsequently switched from RS92 to Meisei RS-11G in 2015-2017.

In case of the RS2-91 to RS92 switch, the change management included a series of twin launches during all four seasons. The study highlighted the importance of undertaking a sustained programme
130 (i.e. >1 year) of coincident soundings by old and new instrumentation to understand any seasonality of biases. Seasonally dependent biases may arise from changes in the measured ECVs and/or annual cycles of covariates such as radiation effects which are particularly important at those sites where launches systematically occur at or near dusk/dawn. The second switch, RS92 to RS-11G, involved 52 twin soundings over a period of two years (Kobayashi et al., 2019).

135 There are some crucial distinctions between this precursor analysis at Tateno and the current GRUAN-wide transition from RS92 to RS41 radiosondes. For Tateno:

- At the time, neither the original nor the replacement sonde models had GRUAN data products being processed and provided routinely to the user community.
- The change related to a single instrument at a single site.
- 140 – The update arose from a choice by the site to change instrumentation such that the timetable could be altered as necessary.

Although the matter of broader change management pertaining to simultaneous instrument transitions at multiple sites has been informally discussed on various occasions, e.g. during GRUAN's annual Implementation and Coordination Meetings (ICMs), prior to the RS92 cessation of production,
145 there existed no formal plan for managing such a wide-spread change. This large-scale transition poses a major challenge for GRUAN as a reference network because it must not compromise the continuity, quality, and homogeneity of the data records.

This change of the operational radiosonde at the majority of the GRUAN sites is unprecedented in the history of GRUAN. A network-wide challenge requires a facilitated and coordinated solution
150 if the change is to be successful, and if GRUAN is to succeed. The fundamental challenge is thus to design and deliver a GRUAN-wide strategy for managing and coordinating the near-simultaneous changes of the operational radiosonde at many sites. This strategy should include all aspects of the change management including ~~inter-alia:~~ inter alia:

- network coordination to share the burden,
- 155 – the necessary roles of ancillary measurements coincident with the sonde measurements,
- instrument characterization in the laboratory and field under a wide range of conditions,

- scientific analysis,
- technical documentation, and scientific publication,
- dissemination of results to GRUAN and the broader GOS through GCOS/ World Meteorological Organization (WMO) Integrated Global Observing System (WIGOS) mechanisms,
- sharing information with agencies conducting Near Real Time (NRT) data applications to aid their managing of the change.

Since GRUAN always seeks to promote competition in the marketplace, sites were encouraged to consider all available options for how to proceed with their transition from the RS92, including changing to sondes produced by other manufacturers. Despite the independent decision-making procedure, ~~each~~every GRUAN site that launched RS92 sondes has transitioned to the RS41. It is important to stress that these decisions arose from the individual GRUAN sites and not from the network management. The consequence is that the network must transition between two instrument models from the same manufacturer. If some sites had chosen to switch to radiosondes from different manufacturers, GRUAN would have been required to develop a very different change management programme than the one described here.

Two of GRUAN's strengths are its ability to call on expertise from across the network to tackle such challenges and its ability to distribute required actions among the sites to share the burden. Currently, specialists from various fields of expertise within GRUAN are engaged in addressing the above mentioned points. This not only advantages the affected sites and GRUAN as a whole, but also helps other observational networks, such as e.g. GUAN, in managing the same transition.

3 Change management

Proper management of the change of a measurement system requires determining all relevant differences between both systems prior to the transition. Typically this means organizing a period of observational system overlap as well as laboratory-based characterization of the differences between the instruments. While the specialized facilities to perform extensive laboratory testing are not available at each site, there is no impediment to sites performing real-world intercomparisons like twin soundings, although the costs of extra receiving systems and sondes may pose limits on the number of flights that can be performed.

It is essential to quantify biases between the new and the old instruments as well as changes in calibration/measurement errors and uncertainties. These attributes may have complex interactions with covariates which complicate the quantification of the effects of the change. One example of such a measurement error, the solar radiation-induced temperature bias, ~~will vary~~varies with altitude, season and geographical location, because it depends on the ambient pressure, solar elevation angle and radiation intensity (Dirksen et al., 2014). The full range of sources of RS92 uncertainties

that may have complex spatio-temporal characteristics, including ventilation, sensor orientation, and pre-launch calibration, have been described in detail by Dirksen et al. (2014). These uncertainties will also vary with location due their dependence on solar elevation angle, cloudiness, and winds. Once the biases have been identified and corrected for, and the uncertainties have been determined, the data (m_1 and m_2) from both measurement systems should be consistent, meaning that the agreement criteria set out in Equation 1 are met with factor $k = 2$ (Immler et al., 2010).

Another way of saying this is that the measurement data consistently lie within each other's uncertainty coverage factors (u_1 and u_2) after accounting for any effects of non-coincidence. In other words, differences between coincident measurements by the two systems are consistently smaller than the 95 % confidence interval.

$$|m_1 - m_2| < k \cdot \sqrt{u_1^2 + u_2^2} \quad (1)$$

Verifying this consistency requires a sufficient population of coincident measurements to determine in a statistically robust manner that the data satisfy this condition. Neither the GRUAN Manual (WMO, 2013a) nor the GRUAN Guide (WMO, 2013b) provide a clear requirement for the duration and intensity (measurement frequency) of an intercomparison study because it is inherently dependent on instrument type and measurement principle. However, as mentioned in Section 1, the manual on the GOS suggests a one-year inter-comparison period. Different instruments measure different ECVs in very distinct manners. Some are episodic remote sensing, some are continuous remote sensing, others are periodic in-situ profiles. For each measurement type a different 'side-by-side' operation strategy would be required, which depends on (at a minimum): the instrument characteristics, variability in the target measurement, and cost and logistical considerations. Kobayashi et al. (2012) suggest that a total of 120 twin soundings spread across the seasonal cycle, at a given location, would be more than sufficient to characterize the effects of a change in radiosonde instrumentation, although the study did not consider the metrological quantification aspects that are necessary in the current case. Hence, for radiosondes, the GRUAN Lead Centre recommended that sites perform weekly, or with a bi-weekly interval, twin soundings for a period of two years. The two-year period ensures that seasonality is better probed, and a sounding interval of a week instead of a day mitigates the additional operational costs. The twin soundings should be equally distributed between day and night soundings. Above-mentioned metrological aspects refer to factors such as type B (instrumental) expanded uncertainties (absolute for each radiosonde), relative uncertainty in the comparison, traceability of involved instrumentation, and assessment of the comparability with independent ancillary results. This is the key aspect in adopting a metrological approach, where other reference instruments are used to make the evaluation traceable to standards and with a documented uncertainty budget.

Coordination with satellite overpass is highly encouraged, with particular emphasis on targeting times when a GNSS-RO and polar orbiter overpass may occur in the site's proximity. The

sonde-satellite data comparison would also serve as an additional station-to-station transfer-check of consistency of the results. This is further discussed in Section 7.

Laboratory testing is used to characterize the radiosonde's sensors, as is done to establish a
230 GRUAN data product for each radiosonde model. The results of the laboratory tests for the RS92-
RS41 transition are being employed in developing a GRUAN data product for the RS41. The pa-
rameters tested include: radiation error for the temperature and humidity sensors, sensor calibration
accuracy, and response time-lag and hysteresis effects of the humidity sensor. The Lead Centre facil-
ity, hosted by DWD at the Lindenberg Observatory, has access to a broad range of laboratory-based
235 facilities suitable for characterizing radiosonde performance. These facilities, shown in Figure 3,
were used to characterize the RS92 instrument as described in Dirksen et al. (2014) and include:

- standard humidity chambers (SHCs) validate RH sensor calibration,
- radiation test chambers to investigate the solar-induced temperature errors and dry biases,
- climate chamber operating between -75°C and $+20^{\circ}\text{C}$ for sensor response time-lag testing.

240 Section 5 gives a more detailed discussion of the application of these facilities to the characterization
of radiosondes, together with an overview of preliminary results from these laboratory tests.

4 Network coordination to share the burden

4.1 Introduction

From a GRUAN perspective, the collective transition from the RS92 to the RS41 radiosonde en-
245 courages approaches to burden sharing this change management. Here, burden sharing means that
selected sites, each representative of a different climate region (e.g., tropical, mid-latitude, polar)
perform prolonged intercomparison studies, while climatologically similar sites are not necessarily
required to perform such campaigns. This reduces the logistical burden on the site and the cost of
additional sondes and receiving systems. Other forms of burden sharing could be implemented. For
250 example, climatically equivalent sites can divide the workload such that one site conducts daytime
twin flights while another site performs the nighttime twin flights. A similar division of labor could
be agreed upon to balance sampling of different seasons.

This approach to burden sharing will promote quality over quantity, as it is envisaged that some
sites will opt only for a limited intercomparison period, or perhaps only a short campaign-like effort.
255 On their own, these limited intercomparison efforts would be insufficient to properly investigate
the seasonality of the differences between the old and the new sounding systems. But as a net-
work, the sum of these small contributions becomes substantial. Table 1 lists the GRUAN sites
that switched from RS92 to RS41 and, for those that decided to intercompare the two sondes, the
actual or proposed duration and frequency of their twin sounding programmes. Based on these

260 results, ~~the mid-latitudes climate regions in both hemispheres and the Arctic~~in the Arctic and the
mid-latitude climate regions in both hemispheres are covered by long-term intercomparison efforts.
Clearly lacking are intercomparison efforts in the tropics and the Antarctic, a consequence of there
being no GRUAN sites with an established operational RS92 measurement programme in those re-
gions. However, there is one (candidate) GRUAN site in the tropics that did employ the RS92 (the
265 BoM-operated site in Darwin, Australia). This site has also decided to switch to RS41 and has
performed a condensed intercomparison programme, which partially fills the gap in the tropics.

In addition to the efforts by GRUAN sites, several non-GRUAN sites have also performed inter-
comparison campaigns as a way to manage their RS92-RS41 transition. These include the UK Met
Office comparisons in St Helena and in Rothera in Antarctica. Some of these sites have agreed
270 to share their twin sounding data with GRUAN, providing critical data in regions that are not
~~well-served~~covered by GRUAN sites. These sites are listed in Table 2. Furthermore, the GRUAN
Lead Centre was involved in the StratoClim campaigns of 2016 and 2017 (India and Nepal, respec-
tively Brunamonti et al., 2018; Rex, 2014). During these campaigns, that took place in July-August
during the Indian Monsoon, RS92-RS41 twin soundings were performed to investigate the differ-
275 ences between both systems under these particular meteorological conditions.

All data from the intercomparisons listed in Tables 1 and 2 will be made available to the scientific
community via the GRUAN data servers, as discussed further in Section 8.3.

4.2 Support equipment

To support data telemetry for RS41 radiosonde measurements at GRUAN sites, the Lead Centre has
280 a spare, fully-equipped radiosounding receiving system that can be temporarily loaned to sites that
cannot afford to purchase a second receiving station. This fully functioning system consists of an
antenna, an MW41 receiving system, and a compact Vaisala WXT weather station for ~~collecting~~
~~metadata surface observations~~recording observations of surface meteorology (metadata) at the time
of the launch. In addition to this hardware, an SHC can be loaned for performing manufacturer-
285 independent pre-launch checks in a 100 %RH environment, as discussed by Dirksen et al. (2014).

Furthermore, Vaisala has several MW41 systems on hand that can be loaned to sites that wish to
conduct short to medium-term RS92-RS41 intercomparison campaigns. Various GRUAN sites have
used this option to manage their transition, and loaned systems have also been employed during
campaigns, such as the StratoClim campaigns, and the CONCIERTO campaign on Reunion Island in
290 2019.

5 Laboratory characterization

5.1 Introduction

The laboratory facilities at the Lead Centre/Lindenberg Observatory, photographically presented in Figure 3, are being used for extensive testing to characterize the measurement errors and uncertainties of the RS41 and the RS92, an activity which is essential for the development of a GRUAN data products for both radiosondes. The tests primarily focus on the error sources which are known to be dominant for radiosondes, i.e. the solar radiation heating of the temperature and humidity sensors, ~~the response time-lag, and the accuracy and reproducibility of the humidity sensor's calibration and for the sensor, its response time-lag and the accuracy and reproducibility of its calibration.~~

Radiation error tests are performed in an adapted SHC at pressures between ambient and 3 hPa (see Dirksen et al. (2014) for a description of the radiation tests and their configuration) as well as in a newly developed system that allows for improved ventilation and illumination. Preliminary radiation tests were performed on RS41 radiosondes from 2014 through 2019, and further tests are foreseen for 2020. The results indicate that the temperature sensor of the RS41 radiosonde is less susceptible to heating by solar radiation than that of the RS92. However, these results apply to raw (uncorrected) measurement data and it is not possible to draw direct conclusions on the resulting temperature bias between the Vaisala-processed data products of RS92 and RS41 ~~since in the data processing of both sondes different corrections are applied for the radiative heating, since different corrections for radiative heating are applied to the two sonde models by the Vaisala processing software.~~

The calibration accuracy and hysteresis effects of the humidity sensor have been investigated by placing the radiosonde's sensor boom in an SHC ~~having with~~ a stable, well-defined RH between 0 % and 100 %. The stable RH environment inside each SHC is achieved using one of the saline mixtures listed in Table 3. In addition, each SHC is equipped with a Pt100 reference thermometer which tests the calibration accuracy of the radiosonde temperature sensor. The reference Pt100 temperature sensors were certified by DAkkS (the German national accreditation agency), showing an uncertainty of 0.04 unitK (1σ). The RH values over reference salt mixtures are taken from Greenspan (1977), who specifies the upper limit of the uncertainty of the employed RH values at 25°C as 0.5 %RH.

The response time-lag of the humidity sensor is determined by measuring its reaction to a stepwise change in the humidity of the airflow, while keeping the temperature of the airflow constant. The time-lag effect becomes significant at temperatures below -40°C , the point at which the response time of the humidity sensor starts to exceed 10 s (Miloshevich et al., 2004; Dirksen et al., 2014). The time-lag tests are performed in a climate chamber which can reach temperatures as low as -75°C . Table 4 summarizes the laboratory experiments that have been performed to date to characterize the RS41 radiosonde.

The lab-based characterization results will be included in the scientific database holding all data

that are relevant to the RS92 to RS41 transition (discussed in Section 8.3).

5.2 Results of the laboratory characterization

To assess the calibration accuracy of the humidity and temperature sensors, more than 150 RS41
330 radiosondes from various production batches were tested in SHCs at the relative humidities listed
in Table 3. In a typical experiment, each RS41 was sequentially placed in a series of six SHCs
with increasing relative humidity, from 0 to 100 %, then back through the sequence of drier SHCs
to 0%RH. This sequence also allows assessment of hysteresis of the humidity sensor. At each
humidity level, the ~~radiosonde was immersed~~radiosonde's sensor boom was inserted in the SHC for
335 approximately 4 minutes while its readings were recorded. The air inside each SHC was circulated
at 5 m/s by a fan, and the temperatures of the air and saline solution inside the SHCs were measured
by Pt100 reference thermometers. For some salts both air and solution temperatures are needed to
accurately determine the relative humidity in the SHC, since the humidity over the reference salt
depends on both quantities.

340 Figure 4 shows that the majority of the temperature measurements by the RS41 are within ± 0.5 K
of the reference temperature. Although the tail of the distribution extends to 1 K (not shown), ~~only~~
a fewer than 2 % of the measurements show differences beyond 0.5 K. The mode of the distribution
indicates a bias of -0.025 K, while ~~from its width it can be inferred that the calibration uncertainty~~
is smaller than 0.1 K, a calibration uncertainty of < 0.2 K (1σ) can be inferred from the sensor's
345 uncertainty ± 0.04 K and the 16th and 84th percentiles of the differences. The ~~As is apparent from~~
the percentile marks, the histogram of temperature differences between the RS41 and the Pt100
reference is not Gaussian, something that cannot yet be explained but will be the subject of further
investigation.

Figure 5 shows that all humidity measurements by the RS41 are within 2 %RH of the SHC refer-
350 ence RH value ~~with, and~~ only a small fraction of the measurements show ~~ing~~ differences larger than
1 %RH ~~, which means that the uncertainty of the humidity calibration is smaller than 1%RH.~~ The
uncertainty of the humidity calibration is < 1.5 %RH given the < 0.5 %RH uncertainties of the SHC
RH references listed in Table 3.

~~The histogram of RH differences between the RS41 and the reference RH value has a pronounced~~
355 ~~peak around 0%RH, due to the fact that the calibration errors are dependent on RH. At low RH~~
~~the differences, and their spread, between the reference and the RS41 values are small, whereas~~
~~at high RH the differences and spread are larger. As a result, all observed calibration errors at~~
~~0%RH humidity cluster around 0%, thereby over-representing these values in the distribution.~~ The
histograms of RH differences at the various RH levels, represented by the colored traces in Fig 5,
360 show that the calibration errors are dependent on RH. At low RH the differences and their spread are
small, whereas at high RH the differences and spread are larger. As a result, the observed calibration
errors at 0 %RH humidity cluster around 0 %, (blue trace in Fig 5), thereby over-representing these

values in the overall distribution (black trace in Fig 5). The larger differences and spread at high RH indicate that there are discrepancies between the reference salt mixtures and the calibration by Vaisala.

365

Radiation tests were conducted in the modified SHC as described by Dirksen et al. (2014). Measurements were performed at various settings within the ranges:

– pressure between 3 hPa and ambient,

– irradiance from 200 to 1000 W/m²,

370

– ventilation speed of either 2.5 or 5 m/s,

– illumination times of 1-4 minutes.

The illumination times depended on pressure, with longer exposures required for lower pressure environments to reach equilibrium. Figure 6 shows similar heating of the RS92 and RS41 temperature sensors at 300 hPa (0.15 K), but at 3 hPa (Figure 7) the RS41 (1.4 K) heats only half as much as the RS92 (2.8 K). Furthermore, at 3 hPa, the RS41 sensor has reached equilibrium after approximately 30 s of illumination, whereas after 4 minutes of illumination the RS92 still hasn't reached equilibrium. The investigation of the RS41's radiation errors continues, including experiments with the newly developed laboratory radiation system, and the full analysis of these tests will be reported in a separate paper.

375

380 6 Metrology

A fundamental metrological principle stipulates that replacing one operational instrument with another should pose no problem provided that the results from both instruments are fully traceable to SI standards. Consequently, the new instrument could (almost) instantaneously be included in the traceability chain without the need for parallel testing or comparison with the replaced device. In practice, this idealized concept can rarely be adopted, even in primary metrology laboratories or in National Metrology Institutes. The problem is that different instruments or sensors may show different responses to external environmental factors.

385

Concerning radiosondes, the sensors, especially for humidity and temperature, may ~~during a sounding~~ be exposed to unavoidable atmospheric or ascent-related effects during a sounding. Some of these effects cannot ~~fully be included in the realization of~~ be completely duplicated during the controlled laboratory conditions which are provided during the preceding metrological instrument characterization and calibration procedures. Consequently, differences in the responses of sensors from different sonde models may still exist during ascents. An example is the warm bias of the radiosonde's temperature sensor caused by solar radiation.

390

395

For this reason it is essential to identify and quantify these differences between the old and the new measurement system at each time of a sonde replacement, not only by laboratory work, but also

by comparison flights where the instruments ~~to be compared~~ are ~~mounted under~~carried by the same balloon. The same should also be done when there have been significant changes in the design, sensor technology, or operation of an actual instrument.

400 An adequate characterization of the measurement uncertainties for entire radiosonde profiles includes the following components:

- Calibration uncertainty, given by the manufacturer (see product data sheets for Vaisala RS92 and RS41 available the manufacturer’s web site, <https://www.vaisala.com>). The sensors are calibrated in Vaisala’s CAL4 calibration facility (Vaisala, 2002) that contains PTU (pressure, 405 temperature, and humidity) reference sensors, routinely re-calibrated against National Institute of Science and Technology of the United States of America (NIST)-traceable standards (for pressure and temperature) and the Finnish Centre for Metrology and Accreditation (Mittatekniiikan keskus – MIKES, which has become part of VTT Technical Research Centre of Finland Ltd and is now officially known as VTT-MIKES) for humidity. Note that for GRUAN 410 the uncertainties of the calibration curves are important, not so much the information about the overall uncertainties in soundings which are related to the manufacturer-provided data product,
- Uncertainties from GRUAN-prescribed pre-flight ground check procedures,
- Uncertainties estimated during GRUAN post-flight processing, including uncertainties of corrections to the measured raw data for systematic effects which are mainly derived from laboratory tests. 415

~~Disclosed differences found in~~Differences revealed by instrument comparison flights indicate that some systematic effects have not yet been correctly assessed or even identified. To assure consistency of the measurement results of different instruments, the differences are to be evaluated in terms 420 of possible corrections, or, if this is not possible, the uncertainty budget is to be ~~extended~~widened properly to account for them. It is important to arrange the comparisons network-wide in a coordinated manner, covering different ~~time scales~~seasons and locations. This is to ensure that ~~as much aspects as possible are covered which may introduce systematics~~This is to ensure that all possible sources of systematic errors such as latitude, climate, environmental and technical conditions before and during launch, local specifics in the sounding procedures or setups are tested during twin 425 soundings.

An advantage of performing dual launches of RS92 and RS41 radiosondes is that some of the uncertainties related to the change process are moved from absolute evaluations to relative ones, making second order contributions to the uncertainty budget limited, if not negligible. Second order 430 contributions are uncertainties in the quantification of any environmental factor that affects the measurement of the primary quantity, like the intensity of solar radiation impinging on the temperature sensor.

The investigation of external environmental factors on the measurement uncertainty of meteorological instruments is one of the key elements of the MeteoMet project (Merlone et al., 2015, 2018)
435 , which also included dedicated campaigns for the validation of radiosonde data, for example in a polar environment (Musacchio et al., 2015).

~~MeteoMet aimed to link the metrological and meteorological communities and to apply robust metrological methods for retrieving traceable reference-quality meteorological observations that can be used for climate monitoring. For this purpose, MeteoMet advocates the application of climatic chambers which create environments of controlled temperature, pressure, and humidity settings, similar to the one at Lindenberg Observatory (Section 5). Besides laboratory facility, transportable calibration chambers have been developed and involved in a field campaign at the Ny-Ålesund GRUAN site, for the calibration of ground-based pre-launch check sensors for pressure and temperature (Musacchio et al., 2015). The novelty of the system was its capability to generate~~
440 ~~pressure and temperature independently, thus making the calibration curves inclusive of mutual factors of influence (mainly temperature on barometers). Measurement results become more accurate, being the calibration curves (or planes) more representative of the field conditions met also in such extreme polar environment.~~

7 Ancillary measurements

450 7.1 Introduction

Many GRUAN sites employ the principle of deliberate redundancy by simultaneously measuring a specific atmospheric parameter using different techniques. Examples include the use of remote sensing techniques such as Global Navigation Satellite System (GNSS), lidar, microwave radiometer (MWR), or Fourier-Transform Infrared Spectrometry (FTIR) in addition to the regular radiosonde
455 flights. Ground-based observations, together with collocated satellite-based measurements, are referred to as ancillary data. These ancillary data add significantly to the analysis of the twin-launch data, because they form provide an independent source of data ~~that is not affected in the same way by the error sources that are typical for radiosoundings and therefore can be used to validate the radiosonde intercomparison data.~~
460 unaffected by the same error sources as the radiosonde sensors, therefore allowing independent validation of the radiosonde intercomparison data.

Furthermore, within one orbit and among a limited number of consecutive orbits, satellites provide a consistent background on a global scale. However, the long-term calibration drifts and retrieval errors of space-borne instruments must be taken into account when comparing long-term data sets of coincident radiosoundings and satellite overpasses. See e.g. (Hurst et al., 2016)

465 The GRUAN Lead Centre, in cooperation with the GRUAN Task Team on Ancillary Measurements, is working with each GRUAN site to establish respective ancillary measurement data streams and ascertain which of these streams contain relevant data that could be used to support the RS92

to RS41 transition. A key goal is to establish scheduling and sampling protocols to provide ancillary information that is spatially and temporally synchronized and internally consistent (Equation 1).

470 Protocols are being developed and deployed to ensure that such data are submitted to the scientific database (see Section 8.3) and tagged as ancillary information to facilitate future analyses.

The use of ancillary data from GRUAN sites in such a manner is currently in an early stage of development. Note that there are not yet ~~officially~~certified GRUAN data products for the ancillary measurements, although some like the GNSS-based total water vapor column product is likely to
475 be certified in the near future (Ning et al., 2016). Protocols for the development and the delivery of geophysical profiles from the various remote sensing techniques are being finalized.

The National Oceanic and Atmospheric Administration (NOAA) Products Validation System (NPROVS) (Reale et al., 2012) facility routinely associates satellite measurements with any tagged GRUAN and operational observation data within a given spatiotemporal collocation window. This system
480 can also process associated output from NWP models and profile data from coincident ancillary measurements, including their associated uncertainties. This includes facilitates redundancy testing (Immler et al., 2010) and integration of consistent ancillary profiles into Site Atmospheric State Best Estimates (SASBE) (Tobin et al., 2006) as recommended by the Lead Centre, which is another objective of GRUAN. However, the processing, packaging and use of the integration of satellite,
485 radiosonde, and ancillary data in a spatially and temporally coherent manner is a complicated task that remains under discussion.

~~The currently proposed plan is that the Lead Centre will identify each twin launch for the NOAA NPROVS team, including site-specific ancillary data, once coincident measurement strategies and data streams are established. The NPROVS team will append level-2 (geophysical profile) data
490 from satellites with special emphasis on those targeted for satellite overpasses and provide routine monitoring, analysis and distribution.~~

7.2 Scheduling

In addition to ground-based ancillary measurements such as GNSS, lidar, MWR, and FTIR, ~~satellite-based~~
space borne observations ~~also from polar orbiting radiometer sounders and satellite-based GNSS~~
495 radio occultation (GNSS-RO) present a valuable and abundant source of additional redundant observations for comparisons with radiosounding data. To maximize their potential for scientific exploitation, the radiosonde launches should be scheduled to be coincident with satellite overpasses and/or the occurrence of ~~GNSS-radio occultations (GNSS-RO)~~ over the site. ~~This would maximize the informational content available from instrumentation both on-site and arising from satellite-based~~
500 ~~capabilities.~~

Since the early 2000's, NOAA cooperates with the U.S. Department of Energy Atmospheric Radiation Measurement (ARM) program to perform radiosoundings that are collocated with satellite observations. A joint service between GRUAN and European Organisation for the Exploitation of

Meteorological Satellites (EUMETSAT) provides predictions of these overpasses, including so-called "golden overpasses" where the measurements of a polar orbiter, such as MetOp-A or MetOp-B, are spatially (distance < 200 km) and temporally (within 30 minutes) ~~coincident~~collocated with a GNSS-RO measurement. In addition, at GRUAN sites where ground-based ancillary measurements are also being made, there could be multiple redundant observations of target ECVs to enable better exploitation; see Section 7.1 for further discussion.

510 In the most basic context, satellites provide a consistent background or traveling calibration standard on a global scale to further interpret twin radiosonde results at a given site. Potential benefits include the identification of possible site bias and ~~whether a given set of twin results indicate an improved (or degraded) representation of the actual atmospheric state. in case of twin soundings, a measure of the performance of the radiosondes in various atmospheric environments.~~ The additional information from satellite-synchronized launches can potentially minimize the number of twin
515 launches required, reducing costs to the sites and maximizing opportunities for subsequent exploitation by the broader science community for myriad applications on a global scale.

Figure 8 presents examples ~~Examples~~ of the potential benefits ~~gained~~ from redundant observations ~~(Figure 8) were~~compiled and processed using the NPROVS ~~(Reale et al., 2012)~~. Between
520 September 2015 and November 2016, 58 RS92-RS41 twin soundings were performed at the GRUAN site in Lauder, New Zealand. Among these, 15 were scheduled to coincide within one hour with overpasses of EUMETSAT's MetOp-B satellite that hosts the Infrared Atmospheric Sounding Interferometer (IASI) and the Advanced Microwave Sounding Unit (AMSU). The comparisons (Figure 8) show that the wet bias in the upper troposphere between satellite and European Centre for
525 Medium Range Weather Forecast (ECMWF) profiles and the Vaisala-processed radiosonde data is smaller for the RS41 radiosonde than for the RS92 radiosonde, meaning that the RS41 consistently reports higher RH values in the upper troposphere than the RS92. Both over Lauder and Europe, the water vapor levels measured by the RS41 are up to 10 % higher than for the RS92. These results are consistent with Sun et al. (2016) who found that the bias of water vapor mixing ratio in the
530 upper troposphere (UT) from satellite observations and radiosonde measurements (from GRUAN and non-GRUAN sites) was at least 10 % smaller for the Vaisala RS41 compared to RS92. This study was based on 6 month global samples of conventional RS41 and RS92 radiosondes collocated with spaceborne infrared and microwave soundings from NOAA's Suomi National Polar-orbiting Partnership (Suomi-NPP) satellite.

535 The decreasing differences between the radiosonde and both the satellite and the model data is tentatively interpreted as the RS41 providing better RH measurements in the upper troposphere than the RS92. However, validation by additional, independent measurements is needed to substantiate this.

Figure 8 also shows that the bias between radiosonde and satellite/model data in the UT is up to
540 a factor two smaller for Europe than for Lauder. In addition, the shape of the difference profiles for

Lauder and Europe are different and analysis of this discrepancy is ongoing.

The finding that the RS41 measures higher humidity values in the UT than the RS92 is also consistent with results for RS92-RS41 twin soundings that were performed at the GRUAN site in Lamont/Southern Great Plains (SGP) in Oklahoma, USA, under management of ARM (Jensen et al., 545 2016). However, these soundings were not ~~performed in coincidence~~ collocated with satellite overpasses, so comparisons with satellite data are not possible.

The collocated observations also permit assessment of calculated radiances, derived from radiosonde profiles using radiative transfer models, versus observed satellite radiances. Calbet et al. (2017) used this method to evaluate RS92 data, and when applying the same method to RS41 data it 550 can provide additional information on the differences between RS92 and RS41.

In summary, scheduling and targeting RS41-RS92 twin soundings with satellite overpasses brings more systems into the comparison and can enhance the transition analysis and provides more robust physical interpretations of results under a wider variety of atmospheric conditions and locations. The NPROVS programme provides routine coincidences of global conventional (including GUAN) 555 and GRUAN reference radiosondes with environmental satellite observations (including GNSS-RO) within 6 hours and 250 km. ~~This spatiotemporal criterion~~ These spatiotemporal criteria, taken from Reale et al. (2012), is a compromise between representativeness and sample size. This establishes a powerful baseline data set which can then be sub-sampled for detailed analysis depending on specific requirements for synchronicity, cloudiness, etc.

560 **8 Scientific analysis and publication**

8.1 Introduction

The wide range of research activities investigating the RS92-RS41 transition that are outlined in this paper will result in a substantive, and valuable, data archive.

The analysis of this data archive will be done from various perspectives by scientists with different 565 areas of expertise. Their results will need to be shared with the atmospheric science community. Table 5 lists a preliminary allocation of research analysis tasks and their principal investigators. With this multi-disciplinary approach to the analysis of the data it is anticipated that the differences between the RS41 and RS92 radiosondes will be well understood and that inhomogeneities in their combined long-term data records can be minimized.

570 The list in Table 5 is incomplete and will be expanded as needs and requirements become clearer. The aim is to publish several distinct papers that describe the results. As further outlined in Section 9, we strongly welcome engagement by the broader atmospheric science community to analyze and publish the results.

8.2 Preliminary results

575 Up to September 2019, approximately 1500 RS92-RS41 twin soundings have been performed within GRUAN. A comprehensive analysis of this extensive data set is still ongoing, but as an example of this larger effort we present the preliminary results of the twin soundings performed at Lindenberg. According to Table 1, more than 400 twin soundings were performed in Lindenberg between December 2014 and July 2019. A subset of 224 twin soundings was available for the current analysis.

580 The majority of these were performed with a payload consisting of an RS92 and RS41 radiosonde only, but a substantial number were performed with the RS92 and RS41 as part of an expanded scientific payload ~~, consisting of additional instruments such as the~~ that included a Cryogenic Frost-point Hygrometer (CFH) and Electrochemical Concentration Cell (ECC) ozonesonde. For the twin soundings, the payload rig was configured ~~consistent with~~ by conforming to the recommendations

585 given in GRUAN-TN-7 (von Rohden et al., 2016), such that the radiosondes were attached with a 80 cm long string to each end of a 1.5 m long rod, ensuring free rotational movement of the radiosondes and minimizing potential contamination by water ~~evaporating/outgassing~~ from the rod. In the analysis, GRUAN-processed RS92 profiles (RS92-GDP.2 Dirksen et al., 2014) are used whereas the RS41 data are processed by the Vaisala MW41 system.

590 Figure 9 shows the profiles for such a twin sounding performed during daytime. The difference plot (right-hand panel) shows that up to the tropopause (in this particular case at approximately 14 km) both sondes capture the variations and structures in the humidity profile equally well, and the RS41 reports slightly higher (up to 2%RH) humidity values than the RS92. Above the tropopause, the RS41 reports lower RH values than RS92, which is attributed to the shorter time-lag of the

595 RS41 humidity sensor that ~~is better able to capture the steep negative gradient in water vapor at the tropopause~~ can more accurately measure the steep drop-off in water vapor as a function of altitude near the tropopause.

The plots in Figure 10 (left-hand panel) show that for nighttime measurements the absolute temperature differences between the two sonde models are generally smaller than 0.05 K up to 30 km

600 altitude with the RS92 (GRUAN-processed data) reporting slightly higher temperatures than RS41 (Vaisala-processed data). Above 30 km, T_{RS41} increasingly exceeds $T_{RS92-GDP.2}$, by 0.1 K at 35 km. This indicates differences in the corrections for radiative cooling at the top of the profile for each radiosonde type. The right-hand panel of Figure 10 shows that the temperature differences for daytime measurements in the troposphere are smaller than 0.1 K, with $T_{RS92-GDP.2}$ larger than T_{RS41} . Above

605 the tropopause ~~thisese~~ temperature differences gradually increases with altitude to approximately 0.6 K at 35 km.

Figure 11 shows that the tropospheric humidity values in the Vaisala-processed RS41 data are on average up to 5% higher at night and up to 10 % higher during daytime. For the daytime measurements (right-hand panel) the relative differences increase with altitude, starting with a mean

610 difference of 2.5 % at the surface and reaching approximately 8% at 10 km. This observed higher

RH_{RS41} in the Lindenberg twin soundings is consistent with the results presented in Section 7.2 and in Figure 8.

615 More detailed and elaborate analyses of the differences between RS92 and RS41, including twin soundings from other (GRUAN) sites, will be performed in subsequent studies. There is already a considerable amount of data available, as is summarized in Tables 1 and 2. These studies will investigate in detail the influence of geographical and climatological effects, such as solar elevation angle, clouds, and winds, on the RS92-RS41 differences.

8.3 Scientific database

620 A dedicated database, containing all data pertaining to the RS92-RS41 transition has been created and will be maintained. This database will be given its own digital object identifier (doi) and be preserved over the long-term. The purpose of the database is to have all relevant data for the transition available at a centralized location, thereby serving as a central point of access for users. This will facilitate a multi-faceted analysis of the effects of the transition by experts from various fields of expertise.

625 The database will include the data from the laboratory measurements and the intercomparison launches made at the sites. Furthermore, it will include coincident ancillary measurements from e.g. satellite overpasses and/or ground-based remote sensors such as those identified in Section 7. Making ancillary measurements available together with the radiosonde intercomparison data allows for in-depth analysis and understanding of the differences between the RS92 and RS41 radiosondes, and is commensurate consistent with one of the key principles of GRUAN: to have measurement
630 redundancy.

The data format of the files in the ancillary database will be CF-compliant NetCDF for ease of access and the database will be built for easy web-based data discovery and access. It will be available as it is being populated with data to enable scientific analysis from the outset.

635 Although it may not be possible to analyze all aspects of the data immediately, building a long-term database will enable exploitation by the expert community well into the future and represent a substantial value-added outcome. The database availability will be advertised via the GRUAN website at <https://www.gruan.org> and readers should check this source for the latest status.

8.4 Technical documentation

640 Documentation is a ~~foundation~~ cornerstone of a reference network such as GRUAN. It is essential for the transfer of knowledge, ranging from describing operational procedures and best practices to perform measurements, via a detailed description of correction algorithms, to documenting changes to measurement systems. In a broader sense, robust documentation ensures the traceability of the data products, a requirement for reference data. Only through the existence of proper documentation
645 is it possible to assure the quality of the measurement data within GRUAN. All GRUAN technical

documentation is available on the GRUAN website under <https://www.gruan.org/documentation/gruan/>.

650 In the specific case of the RS92 to RS41 transition, comprehensive documentation will provide the required transparency on how the change was managed, and this will make it possible to reconstruct and scrutinize the reported differences between the RS92 and the RS41 radiosondes, even after many years. Furthermore, this documentation will serve as a template for managing any future transitions of measurement systems within GRUAN. Finally, with GRUAN documentation available to the wider scientific community, other networks, such as GUAN or the Global Observing System (GOS) might be assisted in managing changes in their observational radiosonde systems.

655 This paper serves as an overarching document, outlining the strategy of managing the RS92-RS41 transition within GRUAN. Other GRUAN technical documents and publications will cover various aspects in more detail. For example, a technical note was released which outlines the GRUAN recommendations for the rig configuration for performing twin soundings (von Rohden et al., 2016), whereas rig configurations for extended payloads consisting of multiple instruments and radiosondes are discussed in Jauhiainen et al. (2016). It is foreseen that separate papers will be written that report
660 on:

- The results of the laboratory characterization of RS41 sensors described in Section 5
- Synthesis of the RS92-RS41 intercomparison studies
- Comparison against non-radiosonde measurements (e.g. ancillary data)

665 In addition, a final paper will be drafted that collates [and summarizes](#) the results of the separate reports and summarizes and evaluates the outcomes of the RS92-RS41 transition for GRUAN.

9 How to get involved

The GRUAN change management programme envisaged in this paper could, in principal, be completed solely by current GRUAN members (sites, Lead Centre, scientists in the working group on
670 GRUAN, and task teams). However, we explicitly recognize that there are substantial resources and expertise beyond the immediate GRUAN community which could increase the robustness of all aspects of the envisaged programme. Some specific potential suggestions are given below but there are undoubtedly many more ways to get involved.

675 Participation of non-GRUAN sites, who plan to undertake an intercomparison of their own, is strongly encouraged. Sites need not undertake the full multi-season campaign to contribute substantive value. Any additional intercomparison data will provide either additional training data sets or a means to independently validate results and ensure that any geographical effects have been adequately accounted for. Sites should contact the Lead Centre staff (lead author) to initiate a discussion around data submission requirements.

680 Participation of experts in the analysis of the results is strongly encouraged. Research results
are likely to be more robust and comprehensive after accounting for a broad range of user inputs.
The GRUAN community, although broadly diverse, likely misses some important types of expertise.
The GRUAN Lead Centre and Working Group chairs can provide letters of support and further
information to investigators wishing to apply for grant support to aid their involvement in the analysis
685 of the transition from the RS92 to other models of radiosonde.

Dissemination and outreach of results leading to impact for both NRT and long-term applications
will require sustained community engagement. It is important that the research results translate to
real-world applications and that, ultimately, will require user uptake.

10 Summary and outlook

690 In this paper we have described the ongoing GRUAN-wide coordinated approach to managing the
change from the Vaisala RS92 to the RS41 as an operational radiosonde system within GRUAN.
Since the network's goal is to provide long-term reference-quality observations of ECVs such as
temperature and water vapor for the purpose of e.g. climate monitoring, it is vital that this change
of measurement system does not introduce discontinuities or inhomogeneities in the GRUAN data
695 records. The majority of the 27 GRUAN sites were launching the RS92 as their operational ra-
diosonde ~~, which was phased out in the last quarter of 2017, and until its production ceased in late~~
~~2017~~; most of these sites have ~~since now~~ switched to the RS41.

Such a large-scale change in instrumentation is unprecedented in the history of GRUAN and
poses a challenge for the network. To ensure the integrity of the data record before and after the
700 transition, it is necessary to fully understand and characterize the differences, the bias adjustment,
and measurement uncertainties between the RS92 and RS41 radiosondes. Within GRUAN several
different, but aligned, programmes are generating a body of knowledge to underpin the RS92-RS41
transition, involving laboratory characterization of measurement errors due to external factors (e.g.
solar radiation, time-lag), extensive twin sounding studies with RS92 and RS41 on the same balloon,
705 and comparison with ancillary measurements.

Regular twin soundings for a period of two years is considered sufficient to capture seasonal-
ity in the differences. Since not all sites are able to implement such an extended intercomparison
programme, burden-sharing is employed, whereby at sites with similar climatological conditions,
only one site needs to perform the intercomparisons. All data relevant to the RS92-RS41 transition
710 are archived in a scientific database that will be accessible to the scientific community for exter-
nal scrutiny, enabling transparency and traceability. These data include radiosoundings, collocated
satellite observations and other ancillary measurements as well as the data from the laboratory mea-
surements. Furthermore, data from intercomparison studies performed at several non-GRUAN sites
have been shared with GRUAN by cooperating meteorological services and institutes and these are

715 also included in the database.

Preliminary analysis of the laboratory experiments indicates that the [manufacturer](#) calibrations of the RS41's temperature and humidity sensors are more accurate than for the RS92. Comparison with external references show calibration uncertainties of < 0.42 K for temperature and < 1.5 %RH for the humidity sensor. Preliminary analysis of 224 RS92-RS41 twin soundings performed at Lindenberg Observatory show that differences between RS92-GDP.2 and Vaisala-processed RS41 nighttime temperature measurements are smaller than 0.1 K over the entire profile. However, daytime temperature differences in the stratosphere increase steadily with altitude, with $T_{\text{RS92-GDP.2}}$ 0.6 K higher than T_{RS41} at 35 km. RH values measured by the RS41 in the troposphere are up to 8% higher. These higher RH_{RS41} values are consistent with the analysis of satellite-radiosonde collocations. A comprehensive analysis of all twin soundings performed within GRUAN, that will also evaluate the effect of e.g. climatological factors and will also include ancillary data, is ongoing.

720
725
730 Commensurate with the importance of detailed and comprehensive documentation to GRUAN's operations, the RS92-RS41 transition will be extensively documented to ensure traceability of the process. Furthermore, the documentation will help to convey the experience and knowledge gathered to other networks to aid them in managing any changes in their operational radiosonde systems. Future publications will report the results of the characterization of the RS92-RS41 differences from various perspectives, and a final publication will evaluate the impact of the RS92-RS41 transition on GRUAN data records.

Acknowledgements. The authors wish to thank David Smyth of Maynooth university, and other members of the GRUAN community, for their contribution to the discussion of the GRUAN change management, which greatly helped to improve this paper.

References

- Bodeker, G. E., Bojinski, S., Cimini, D., Dirksen, R. J., Haeffelin, M., Hannigan, J. W., Hurst, D. F., Leblanc, T., Madonna, F., Maturilli, M., Mikalsen, A. C., Philipona, R., Reale, T., Seidel, D. J., Tan, D. G. H., Thorne, P. W., Vömel, H., and Wang, J.: Reference upper-air observations for climate: From concept to reality, *Bull. Amer. Meteor. Soc.*, 97, 123–135, 2016.
- Bojinski, S., Verstraete, M., Peterson, T. C., Richter, C., Simmons, A., and Zemp, M.: The concept of Essential Climate Variables in support of climate research, applications, and policy, *Bull. Amer. Meteor. Soc.*, 95, 1431–1443, 2014.
- Brunamonti, S., Jorge, T., Oelsner, P., Hanumanthu, S., Singh, B. B., Kumar, K. R., Sonbawne, S., Meier, S., Singh, D., Wienhold, F. G., Luo, B. P., Böttcher, M., Poltera, Y., Jauhiainen, H., Kayastha, R., Dirksen, R., Naja, M., Rex, M., Fadnavis, S., and Peter, T.: Balloon-borne measurements of temperature, water vapor, ozone and aerosol backscatter at the southern slopes of the Himalayas during StratoClim 2016–2017, *Atmos. Chem. Phys.*, 18, 15937–15957, <https://www.atmos-chem-phys.net/18/15937/2018/>, 2018.
- Calbet, X., Peinado-Galan, N., Ripodas, P., Trent, T., Dirksen, R., and Sommer, M.: Consistency between GRUAN sondes, LBLRTM and IASI, *Atmos. Meas. Tech.*, 10, 2323–2335, <https://www.atmos-meas-tech.net/10/2323/2017/>, 2017.
- Dirksen, R. and Sommer, M.: Data package related to article 'Progress in managing the transition from the RS92 to the Vaisala RS41 as the operational radiosonde within the GCOS Reference Upper-Air Network', <https://www.gruan.org/data/data-packages/dpkg-2019-1/>, [Accessed 22 November 2019], 2019.
- Dirksen, R. J., Sommer, M., Immler, F. J., Hurst, D. F., Kivi, R., and Vömel, H.: Reference quality upper-air measurements: GRUAN data processing for the Vaisala RS92 radiosonde, *Atmos. Meas. Tech.*, 7, 4463–4490, <http://www.atmos-meas-tech.net/7/4463/2014/>, 2014.
- Edwards, D., Anderson, G., Oakley, T., and Gault, P.: Met Office Intercomparison of Vaisala RS92 and RS41 Radiosondes, Tech. rep., UKMO, go.vaisala.com/gen4/downloads/Met_Office_Intercomparison_of_Vaisala_RS41_and_RS92_Radiosondes.pdf, 2014.
- Greenspan, L.: Humidity Fixed Points of Binary Saturated Aqueous Solutions, *Journal of Research of the National Bureau of Standards Section A: Physics and Chemistry*, 81A, 89–96, 1977.
- Hurst, D. F., Read, W. G., Vömel, H., Selkirk, H. B., Rosenlof, K. H., Davis, S. M., Hall, E. G., Jordan, A. F., and Oltmans, S. J.: Recent divergences in stratospheric water vapor measurements by frost point hygrometers and the Aura Microwave Limb Sounder, *Atmos. Meas. Tech.*, 9, 4447–4457, <http://www.atmos-meas-tech.net/9/4447/2016/>, 2016.
- Immler, F. J., Dykema, J., Gardiner, T., Whiteman, D. N., Thorne, P. W., and Vömel, H.: Reference Quality Upper-Air Measurements: guidance for developing GRUAN data products, *Atmos. Meas. Tech.*, 3, 1217–1231, <http://www.atmos-meas-tech.net/3/1217/2010/>, 2010.
- Jauhiainen, H., Fujiwara, M., Philipona, R., Dirksen, R., Hurst, D. F., Kivi, R., Vömel, H., Demoz, B., Kizu, N., Oakley, T., Shimizu, K., Maturilli, M., Leblanc, T., Madonna, F., and Querel, R.: GRUAN Technical document 7 - Review of Multiple-payload Radiosonde Sounding Configurations for Determining Best-Practice Guidance for GRUAN Sites, Tech. rep., GRUAN Lead Centre, Lindenberg, <https://www.gruan.org/gruan/editor/documentation/gruan/td/gruan-td-7/>, version 1.0, 2016.
- Jensen, M. P., Holdridge, D. J., Survo, P., Lehtinen, R., Baxter, S., Toto, T., and Johnson, K. L.: Comparison

- of Vaisala radiosondes RS41 and RS92 at the ARM Southern Great Plains site, *Atmos. Meas. Tech.*, 9, 3115–3129, <http://www.atmos-meas-tech.net/9/3115/2016/>, 2016.
- 780 Kawai, Y., Katsumata, M., Oshima, K., Hori, M. E., and Inoue, J.: Comparison of Vaisala radiosondes RS41 and RS92 launched over the oceans from the Arctic to the tropics, *Atmos. Meas. Tech.*, 10, 2485–2498, <https://www.atmos-meas-tech.net/10/2485/2017/>, 2017.
- Kizu, N., Sugidachi, T., Kobayashi, E., Hoshino, S., Shimizu, K., Maeda, R., and Fujiwara, M.: GRUAN Technical Document 5 - Technical characteristics and GRUAN data processing for the Meisei RS-11G and iMS-100 radiosondes, Tech. rep., GRUAN Lead Centre, Lindenberg, <https://www.gruan.org/documentation/gruan/td/gruan-td-5>, version 1.0, 2018.
- 785 Kobayashi, E., Noto, Y., Wakino, S., Yoshii, H., Ohyoshi, T., Saito, S., and Baba, Y.: Comparison of Meisei RS2-91 Rawinsondes and Vaisala RS92-SGP Radiosondes at Tateno for the Data Continuity for Climatic Data Analysis, *Journal of the Meteorological Society of Japan. Ser. II*, 90, 923–945, 2012.
- 790 Kobayashi, E., Hoshino, S., Iwabuchi, M., Sugidachi, T., Shimizu, K., and Fujiwara, M.: Comparison of the GRUAN data products for Meisei RS-11G and Vaisala RS92-SGP radiosondes at Tateno (36.06° N, 140.13° E), Japan, *Atmos. Meas. Tech.*, 12, 3039–3065, <https://www.atmos-meas-tech.net/12/3039/2019/>, 2019.
- Merlone, A., Lopardo, G., Sanna, F., Bell, S., Benyon, R., Bergerud, R. A., Bertiglia, F., Bojkovski, J., Böse, N., Brunet, M., Cappella, A., Coppa, G., del Campo, D., Dobre, M., Drnovsek, J., Ebert, V., Emardson, R., 795 Fernicola, V., Flakiewicz, K., Gardiner, T., Garcia-Izquierdo, C., Georgin, E., Gilabert, A., Grykałowska, A., Grudniewicz, E., Heinonen, M., Holmsten, M., Hudoklin, D., Johansson, J., Kajastie, H., Kaykısızlı, H., Klason, P., Kňazovická, L., Lakka, A., Kowal, A., Müller, H., Musacchio, C., Nwaboh, J., Pavlasek, P., Piccato, A., Pitre, L., de Podesta, M., Rasmussen, M. K., Sairanen, H., Smorgon, D., Sparasci, F., Strnad, R., Szymrka-Grzebyk, A., and Underwood, R.: The MeteoMet project - metrology for meteorology: challenges and results, *Meteorological Applications*, 22, 820–829, <http://dx.doi.org/10.1002/met.1528>, 2015.
- 800 Merlone, A., Sanna, F., Beges, G., Bell, S., Beltramino, G., Bojkovski, J., Brunet, M., Del Campo, D., Castriello, A., Chiodo, N., Colli, M., Coppa, G., Cuccaro, R., Dobre, M., Drnovsek, J., Ebert, V., Fernicola, V., Garcia-Benadí, A., Garcia-Izquierdo, C., Gardiner, T., Georgin, E., Gonzalez, A., Groselj, D., Heinonen, M., Hernandez, S., Högström, R., Hudoklin, D., Kalemci, M., Kowal, A., Lanza, L., Miao, P., Musacchio, C., 805 Nielsen, J., Noguera-Cervera, M., Oguz Aytakin, S., Pavlasek, P., De Podesta, M., Rasmussen, M., Del-Río-Fernández, J., Rosso, L., Sairanen, H., Salminen, J., Sestan, D., Šindelářová, L., Smorgon, D., Sparasci, F., Strnad, R., Underwood, R., Uytun, A., and Voldan, M.: The MeteoMet2 project – Highlights and results, *Meas. Sci. Technol.*, 29, <https://doi.org/10.1088/1361-6501/aa99fc>, 2018.
- Miloshevich, L. M., Paukkunen, A., Vömel, H., and Oltmans, S. J.: Development and Validation of a Time-Lag Correction for Vaisala Radiosonde Humidity Measurements, *J. Atmos. Ocean. Technol.*, 21, 1305–1327, 2004.
- 810 Musacchio, C., Bellagarda, S., Maturilli, M., Graeser, J., Vitale, V., and Merlone, A.: Arctic metrology: calibration of radiosondes ground check sensors in Ny-Ålesund, *Meteorological Applications*, 22, 854–860, <http://dx.doi.org/10.1002/met.1506>, 2015.
- 815 Nash, J., Oakley, T., Vömel, H., and Wei, L.: WMO Intercomparison of High Quality Radiosonde Systems Yangjiang, China, 12 July – 3 August 2010, Tech. rep., WMO, <https://www.wmo.int/pages/prog/>

- www/IMOP/publications/IOM-107_Yangjiang/IOM-107_Yangjiang.zip, WMO/TD-No. 1580, Instruments And Observing Methods Report No. 107, 2011.
- 820 Ning, T., Wang, J., Elgered, G., Dick, G., Wickert, J., Bradke, M., Sommer, M., Querel, R., and Smale, D.: The uncertainty of the atmospheric integrated water vapour estimated from GNSS observations, *Atmos. Meas. Tech.*, 9, 79–92, <http://www.atmos-meas-tech.net/9/79/2016/>, 2016.
- Peterson, T. C. and Durre, I.: A Climate Continuity Strategy for the Radiosonde Replacement System Transition, https://ams.confex.com/ams/84Annual/techprogram/paper_71705.htm, [Accessed 2 September 2019], 2004.
- 825 Reale, T., Sun, B., Tilley, F. H., and Pettey, M.: The NOAA Products Validation System (NPROVS), *J. Atmos. Ocean. Technol.*, 29, 629–645, <http://dx.doi.org/10.1175/JTECH-D-11-00072.1>, 2012.
- Rex, M.: StratoClim, Stratospheric and upper tropospheric processes for better climate predictions, <http://www.stratoclim.org>, [Accessed 15 January 2019], 2014.
- 830 Seidel, D. J., Berger, F. H., Immler, F., Sommer, M., Vömel, H., Diamond, H. J., Dykema, J., Goodrich, D., Murray, W., Peterson, T., Sisterson, D., Thorne, P., and Wang, J.: Reference Upper-Air Observations for Climate: Rationale, Progress, and Plans, *Bull. Amer. Meteor. Soc.*, 90, 361–369, 2009.
- Sun, B., Reale, A., Tilley, F., Pettey, M., Nalli, N., and Barnett, C.: Assessment of NUCAPS S-NPP CrIS/ATMS Sounding Products Using Reference and Conventional Radiosonde Observations, abstract AMS, <https://ams.confex.com/ams/97Annual/webprogram/Paper313881.html>, 2016.
- 835 Tobin, D. C., Revercomb, H. E., Knuteson, R. O., Lesht, B. M., Strow, L. L., Hannon, S. E., Feltz, W. F., Moy, L. A., Fetzer, E. J., and Cress, T. S.: Atmospheric Radiation Measurement site atmospheric state best estimates for Atmospheric Infrared Sounder temperature and water vapor retrieval validation, *J. Geophys. Res.*, 111, <http://dx.doi.org/10.1029/2005JD006103>, d09S14, 2006.
- Vaisala: CAL4 Calibration machine Traceability and Uncertainty, Technical Document DOC210645, Vaisala
840 DCO210645, 2002.
- von Rohden, C., Sommer, M., and Dirksen, R.: GRUAN Technical Note 7 - Rigging recommendations for dual radiosonde soundings, Tech. rep., GRUAN Lead Centre, Lindenberg, https://www.gruan.org/gruan/editor/documents/gruan/GRUAN-TN-7_Comparison_setup_v1.0_final.pdf, version 1.0, 2016.
- WMO: The GCOS reference upper air network (GRUAN) Manual, Tech. Rep. 170, WMO/WIGOS Technical
845 Report 2013-02, <http://www.wmo.int/pages/prog/gcos/Publications/gcos-170.pdf>, [Accessed 8 April 2016], 2013a.
- WMO: The GCOS reference upper air network (GRUAN) Guide, Tech. Rep. 171, WMO/WIGOS Technical Report 2013-03, <http://www.wmo.int/pages/prog/gcos/Publications/gcos-171.pdf>, [Accessed 8 April 2016], 2013b.
- 850 WMO: Manual on the Global Observing System, World Meteorological Organization, Geneva, 2015 edn., https://library.wmo.int/doc_num.php?explnum_id=3960, [Accessed 30 September 2019], 2017.



Fig. 1. Locations of GRUAN sites and other transition-related sites (September 2019). Red dots represent sites that switched from RS92 to RS41 as their operational radiosonde. Blue dots represent sites that employ another type of radiosonde or that transitioned from another type to the RS41. Green dots represent sites which performed additional comparison launches and provided these data to GRUAN.

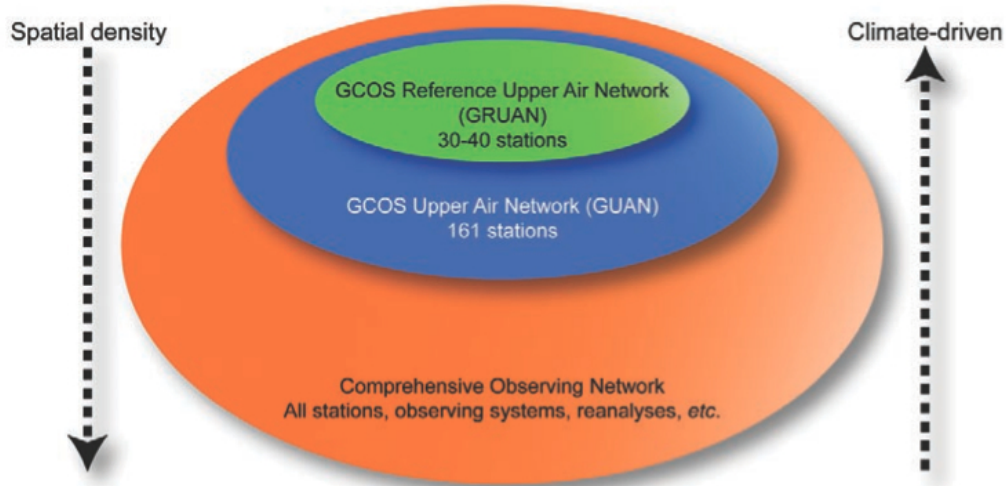


Fig. 2. Schematic diagram of the GCOS tiered-network concept of upper-air observations for climate. GRUAN offers less spatial density than either GUAN (a subset of the global radiosonde network stations that have made commitments of longevity and metadata collection) or the comprehensive network (which includes all ground-based and satellite upper-air observations and reanalyses). From Seidel et al. (2009).

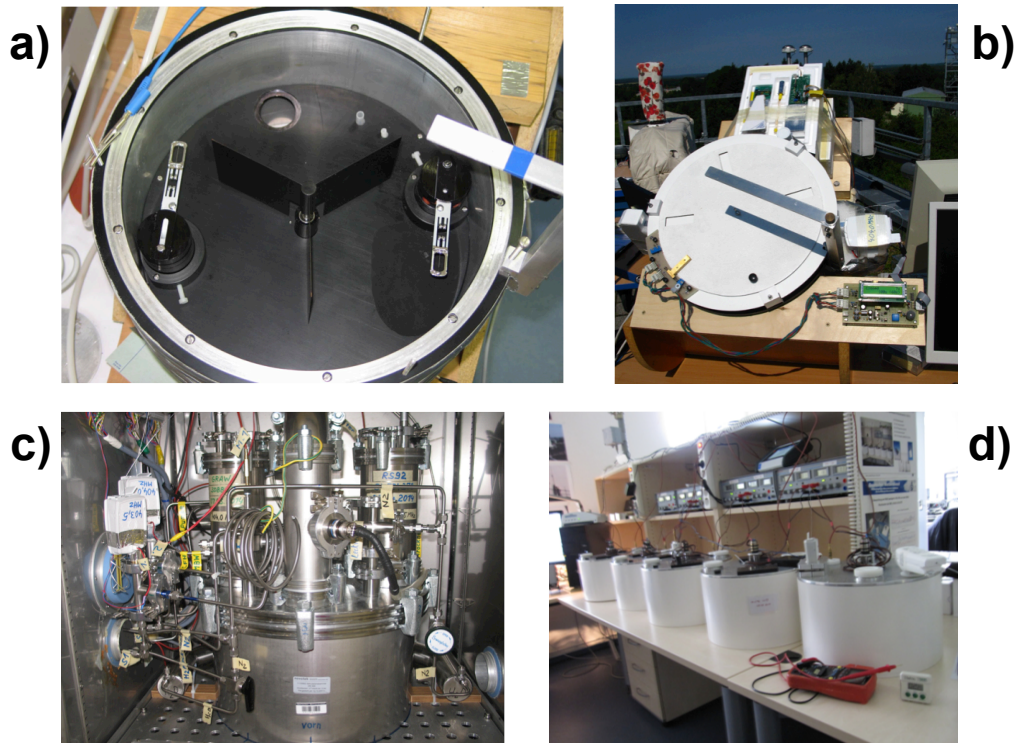


Fig. 3. Overview of the experimental equipment used to characterize the measurement errors and uncertainties of the RS92 and RS41 sensors. a) Radiation test chamber (top view) b) Radiation test chamber in operation, oriented perpendicular to the Sun (light source). Visible on top of the quartz plate is the shutter, with 3 opening slits, allowing for simultaneous testing of 3 radiosondes. c) Climate chamber that is used to test the performance of radiosondes under various temperature, pressure and humidity conditions. In the configuration shown here, the response time-lag of the humidity sensor is investigated. d) Standard humidity chambers that contain reference saline solutions that generate RH environments between 0 and 100 %RH.

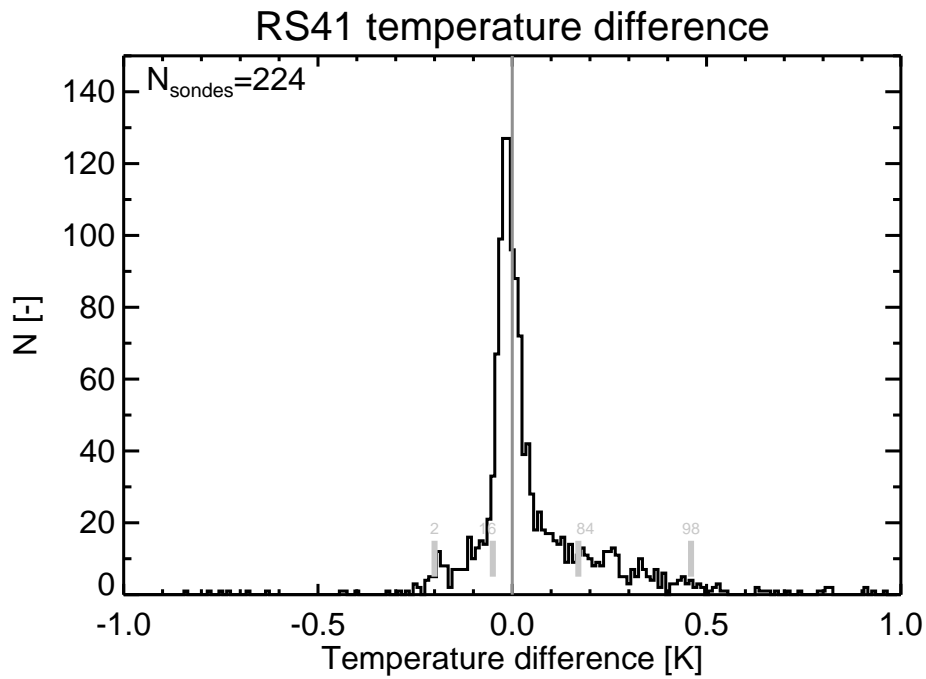


Fig. 4. Histogram of the differences between the temperature recorded by the RS41 and the reference Pt100 probe at room temperature under various humidity conditions inside a standard humidity chamber. The plot displays the collated results for tests in RH environments of 0, 11, 33, 75, and 100 %RH (Table 3) that were performed between 2014 and 2018. Data are collected in 0.01 K wide bins. [The grey bars indicate the percentiles that correspond to the boundaries of the \$1\sigma\$ \(16 & 84\) and \$2\sigma\$ \(2 & 98\) ranges.](#)

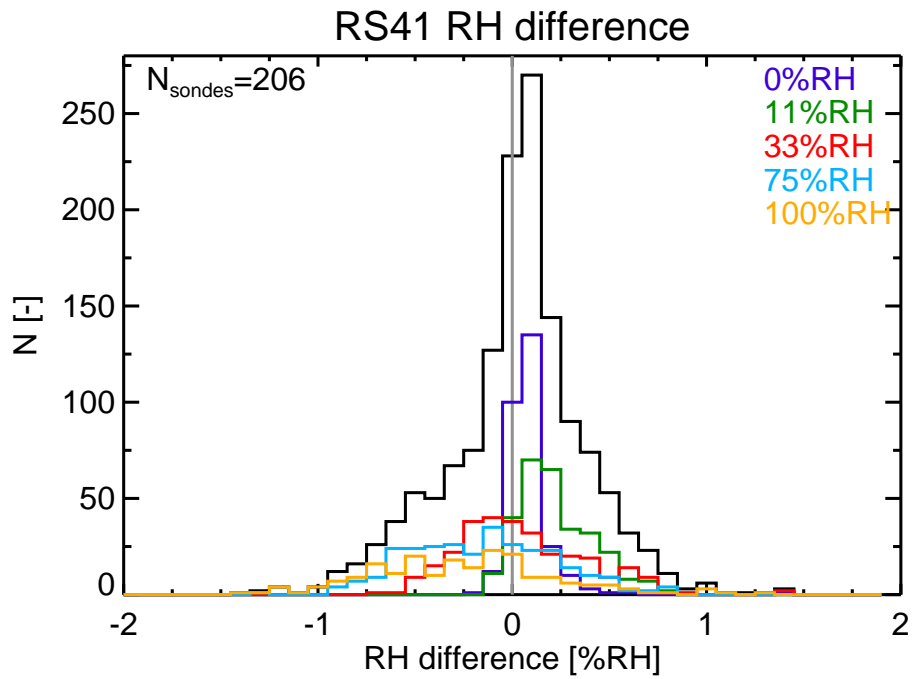


Fig. 5. Histograms of the differences between the RH value recorded by the RS41 and the reference value at room temperature for reference RH values of 0, 11, 33, 75, and 100 %RH in the standard humidity chamber (Table 3) that were performed between 2014 and 2018. The black trace represents the histogram of all measurements, the colored traces represent the histograms for the reference RH values of 0 (blue), 11 (green), 33 (red), 75 (light blue), and 100 (orange) %RH. The plot displays the collated differences for all 5 reference RH values. Data are collected in 0.1 %RH wide bins.

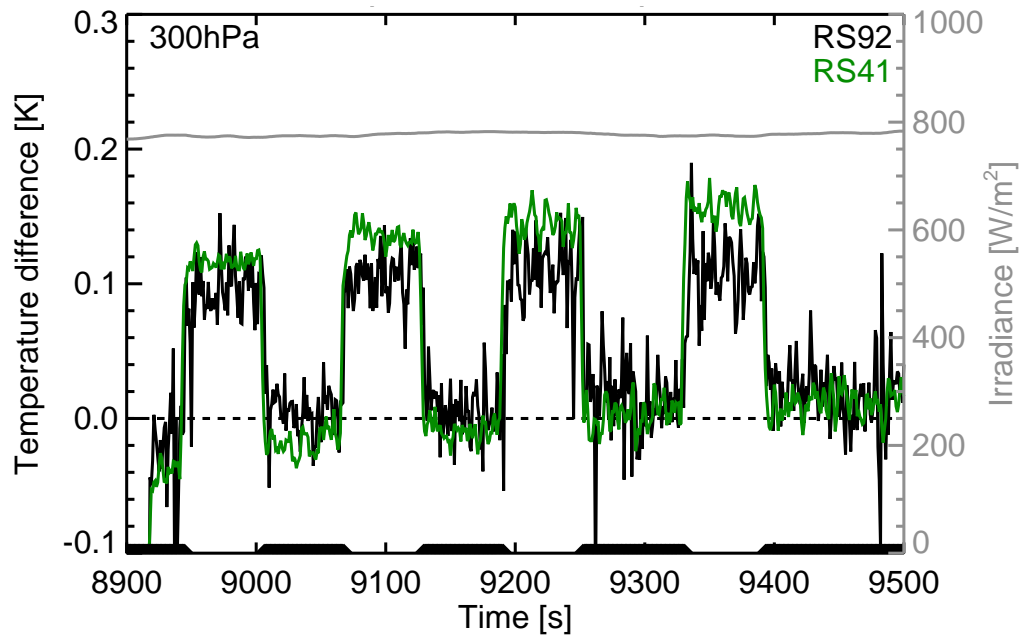


Fig. 6. Results of radiation tests with the RS41 (green) and RS92 (black) radiosondes at 300 hPa with 5 m/s ventilation flow. The solar irradiance during the test (thin grey trace) was approximately 800 W/m^2 . The thick black segments at zero temperature difference indicate when the shutter was closed. During shutter open periods the sensor was illuminated for 60 s.

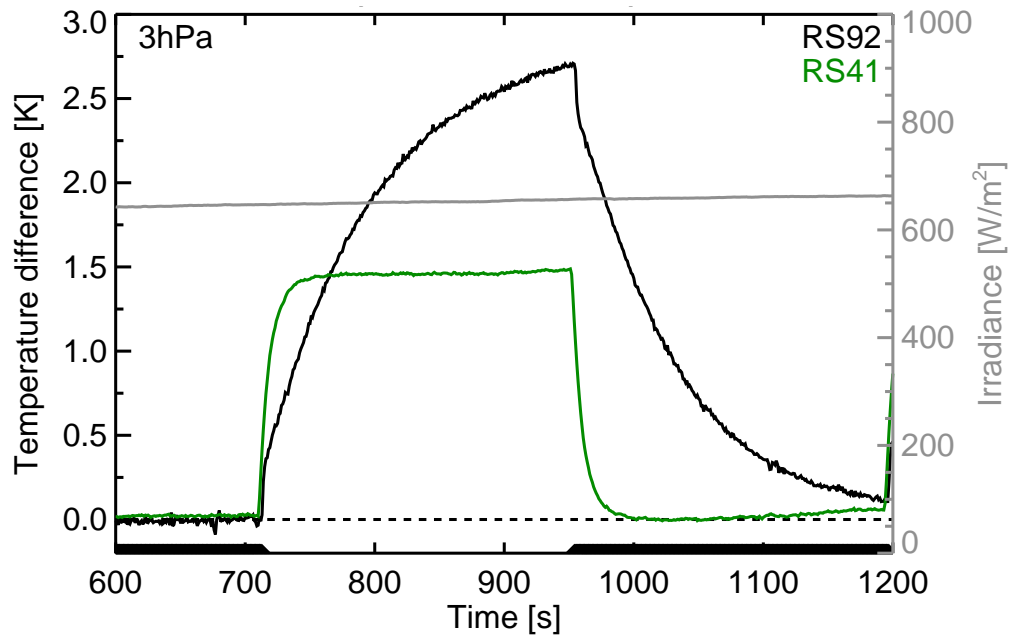


Fig. 7. Results of radiation tests with the RS41 (green) and RS92 (black) radiosondes at 3 hPa with 5 m/s ventilation flow. The solar irradiance during the test (thin grey trace) was approximately 650 W/m^2 . The thick black segments along the x-axis indicate when the shutter was closed. During shutter open periods the sensor was illuminated for 240 s.

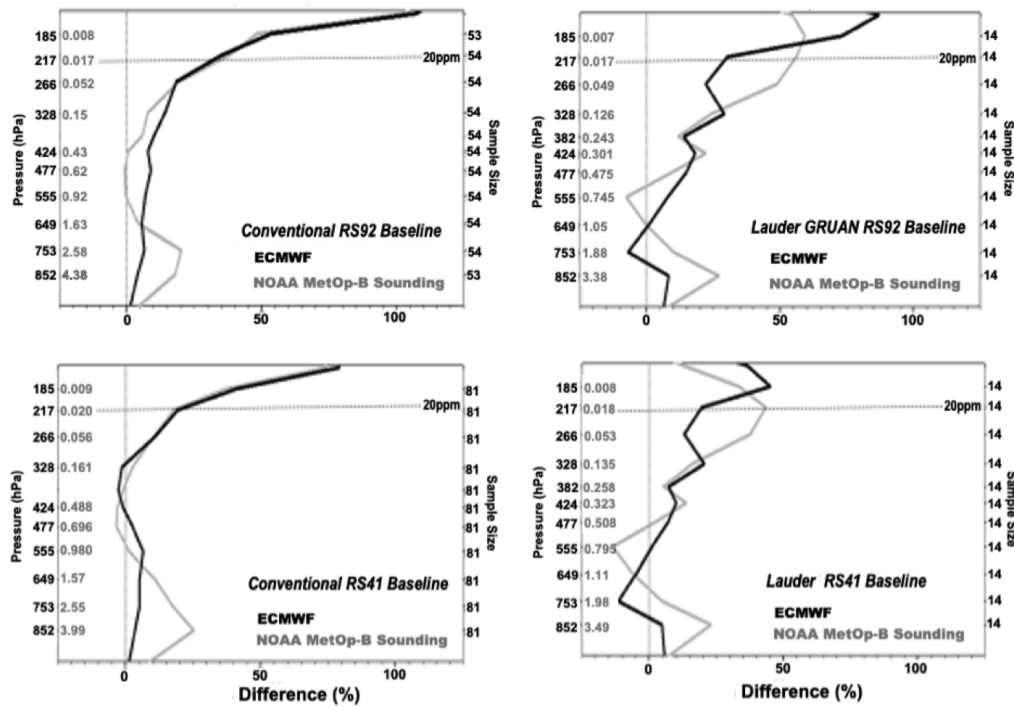


Fig. 8. Mean biases between water vapor profiles from radiosonde and satellite observations from IASI MetOp-B (red). The relative differences are expressed as $100 \times (WVMR_{sat} - WVMR_{radiosonde}) / WVMR_{radiosonde}$. The differences with associated ECMWF analysis data are represented by the purple trace. Mean biases between water vapor profiles from ancillary and radiosonde data from the RS92 (top) and RS41 (bottom). The relative differences are expressed as $100 \times (WVMR_{anc} - WVMR_{radiosonde}) / WVMR_{radiosonde}$. Grey trace: differences with satellite observations from IASI MetOp-B; solid black trace: differences with associated ECMWF analysis data. The relative differences with the RS92 are shown in the left-hand column, differences with RS41 in the right-hand column. The upper panels (The right-hand column) presents the results for RS92-RS41 twin soundings performed at the GRUAN site in Lauder, New Zealand; only the RS92 data are GRUAN processed, the GRUAN site in Lauder, New Zealand, whereas the lower panels (The left-hand column) presents the results for conventional RS92 and RS41 radiosoundings performed during a 10-day period in Oct/Nov 2017 at conventional radiosonde selected WMO sites in Europe (35–65°N, 10°W – 40°E). Coincidence/Collocation criteria: less than 1 hour (2 hours for ECMWF data) at approximately 500 hPa and within 50 km at the surface. The labels at the y-axis depict the pressure level (black); the average water vapor mixing ratio (g/kg) is given in grey. The numbers at the right-hand y-axis represent the number of coincident observations at each pressure level.

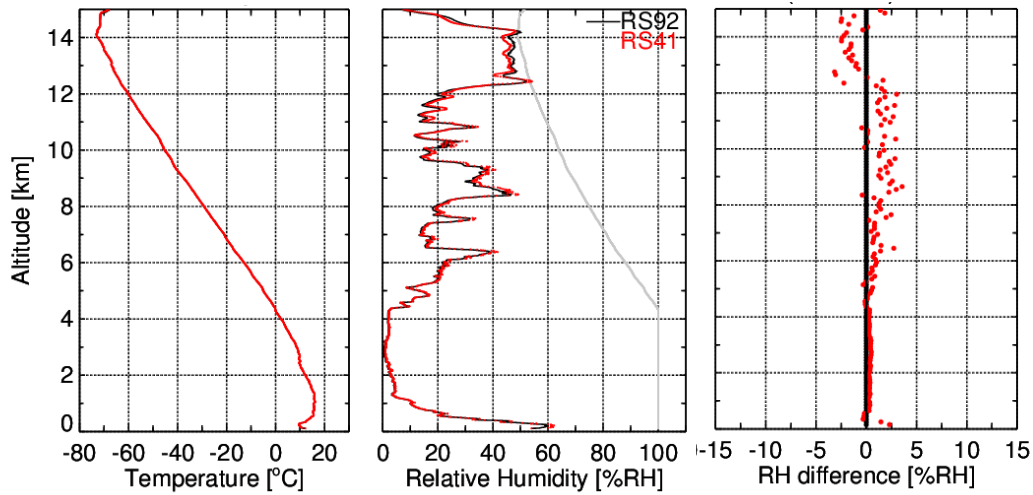


Fig. 9. Twin sounding with RS92 and RS41 on the same balloon performed on 2 November 2015, 12 UTC, at Lindenberg meteorological Observatory. Left-hand panel shows the observed temperature profile of RS92 (black) and RS41 (red). The middle panel shows the observed relative humidity profile over water, the grey trace represents ice saturation. The right-hand panel shows the differences between the humidity of both sondes ($RH_{41} - RH_{92}$), the red dots represent data collected in 500 m wide altitude bins.

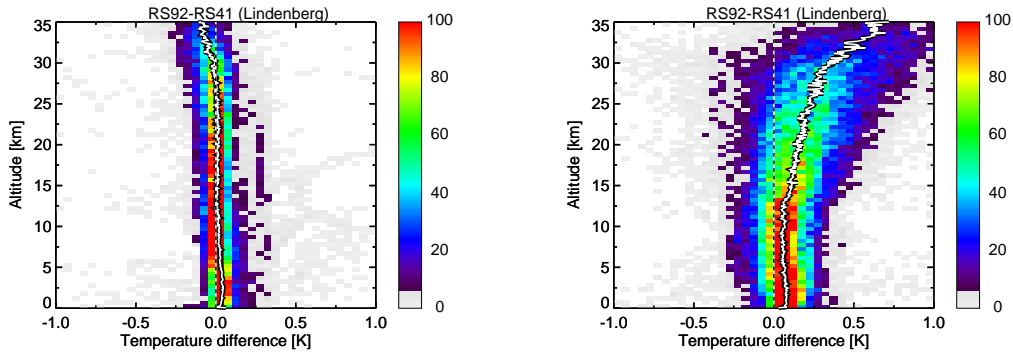


Fig. 10. Temperature differences between RS92-GRUAN data product (RS92-GDP.2) and Vaisala-processed RS41 data (RS92_{GDP.2}-RS41_{EDT}) for comparison soundings performed at the Lindenberg GRUAN site. For each sounding profile the mean $T_{RS92-GDP.2} - T_{RS41}$ was calculated in a 500 m-wide altitude bin. Data are presented as a scatter density plot in altitude-temperature bins of 0.5 km x 0.05 K. The color represents the number of data points in each bin. The thick white trace represents the averages of the temperature differences for each 500 m-wide altitude bin. Left-hand panel displays the results for 77 nighttime soundings, whereas the right-hand panel shows the results for 147 daytime soundings.

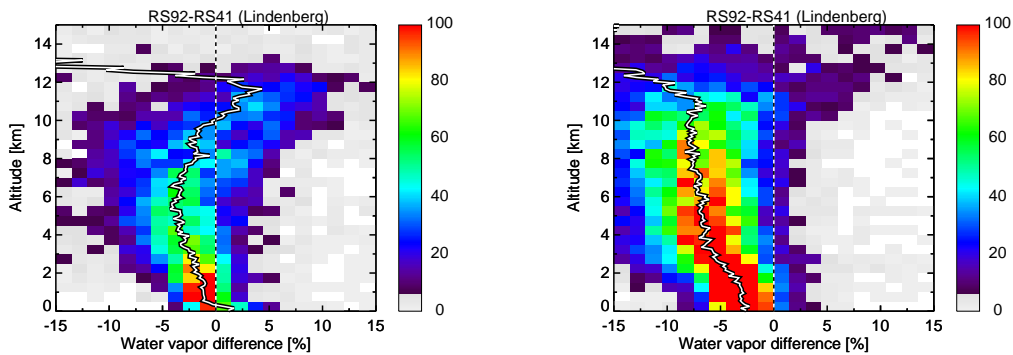


Fig. 11. Same as Figure 10 but for relative humidity. Shown is the relative difference $100 \times (RH_{RS92-GDP.2} - RH_{RS41}) / RH_{RS92-GDP.2}$. Bin-size for the scatter density plot is 0.5 km x 2.5 %RH. The average tropopause height over Lindenberg is approximately 10-12 km.

Table 1: List of GRUAN sites that are involved in the RS92-RS41 transition. Listed are: the dates at which a site switched from RS92 to RS41 as operational radiosonde for the sites, the length and frequency of the twin sounding programs, and the number of twin soundings performed (y - year, m - month, w - week, d - day, h - hour).

Site	Position (Lat °N/Lon °E)	Change Date	Duration, Interval	Amount (+ ongo.)	Remark
Alice Springs	-23.79/133.89	2018-08-22	-	-	-
Barrow	71.32/-156.61	2017-12-15	-	-	ARM RIVAL campaign canceled
Beltsville	39.05/-76.88	-	3 y, ≥ 1 w	126+	regular twin soundings (2016-12 to 2019-09, ongoing), no daily operational soundings
Boulder	39.95/-105.20	2017-01	-	-	no daily operational soundings
Cabauw	51.97/4.92	2017-05-11	-	-	-
Darwin	-12.43/130.89	2018-07-01	1 m, 12 h	57	campaign (2018-06)
Davis	-68.57/77.97	2019-02-23	-	-	-
Graciosa	39.09/-28.03	2019-01-12	1 y, ≥ 1 w	30	ARM RIVAL campaign (2018-04 to 2019-03)
La Reunion	-21.08/55.38	-	1 w + 1 w, 1 d	38	two short campaigns (2015-05 and 2019-01), no operational use of RS92 and RS41
Lamont	36.6/-97.49	2017-12-15	1 w, ≤ 5 /d 1 y, ≥ 1 w	20 49	one short campaign (2014-06) (see Jensen et al., 2016). ARM RIVAL campaign (2018-02 to 2019-04)
Lauder	-45.05/169.68	2016-07, 2018-03-28	1 y, 1 w	57	regular twin soundings (2015-10 to 2016-11)
Lindenberg	52.21/14.12	2017-03-22	5 y, 1 w to 12 h	410+	intensive campaigns in addition to weekly twin soundings (2014-12 to 2019-07, ongoing)
Macquarie Island	-54.50/158.94	2019	-	-	transition planned for 2019
Melbourne	-37.67/144.83	2018-04-01	-	-	-

This table is continued on the next page.

Table 1 – *Continued from previous page*

Site	Position (Lat °N/Lon °E)	Change Date	Duration, Interval	Amount (+ ongo.)	Remark
Ny-Alesund	78.92/11.93	2017-04-01	3 y, 1 w	125	regular twin soundings (2015-03 to 2018-03)
Payerne	46.81/6.95	2018-03-29	4.5 y, ~1 w	143	regular twin soundings (2014-08 to 2018-12), switch of operational radiosonde: Meteolabor SRS-C50 to RS41
Potenza	40.60/15.72	2016-11	–	–	no daily operational soundings
Ross Island	–77.85/166.66	2016-02-12	–	–	–
Sodankyla	67.37/26.63	2017-03-28	–	8	sporadic twin soundings (2015-10 to 2018-07)
Tateno	36.06/140.13	–	1 w, 12 h	10	campaign, switch of operational radiosonde: RS92 to Meisei RS-11G
Tenerife	28.32/–16.38	2017-12-13	–	–	–

Table 2. Non-GRUAN sites that provided data of RS92-RS41 intercomparison flights to GRUAN. (y - year, m - month, w - week, d - day, h - hour).

Site	Position (Lat °N/Lon °E)	Duration, Interval	Amount	Remark
Camborne	50.22/-5.33	2 w, $\leq 4/d$	120	campaign organized by Vaisala in collaboration with UK MetOffice (2013-11), 30 flights with 2xRS92 and 2xRS41 on one rig (see Edwards et al., 2014)
		7 m, ≥ 1 w	25	regular twin soundings (2018-04 to 2018-10)
Kathmandu	27.62/85.54	2 w, $\leq 4/d$	28	StratoClim campaign (2017-08)
Nainital	29.35/79.46	1 m + 1 w, $\leq 3/d$	35	StratoClim campaign (2016-08, 2016-11)
Palau	7.34/134.47	3 w, 48 h	20	StratoClim campaign (2018-03, 2018-10, 2019-05)
Rothera	-67.57/-68.13	1 y, ≥ 1 w	48	regular twin soundings (2017-05 to 2018-08)
St. Helena	-15.94/-5.67	1 y, ≥ 1 w	51	regular twin soundings (2017-12 to 2018-12)
Table Mountain	34.39/-117.7	5 y, sporadic	53	sporadic twin soundings (2014 to 2018)
Ship R/V Mirai	(cruise)	5 m, sporadic	36	sporadic twin soundings (2015-08 to 2015-12), (see Kawai et al., 2017)

Table 3. Humidity values achieved in the SHC using desiccant, pure water and 4 different saturated salt solutions at 25°C. The uncertainties of the RH values are given between parentheses. Data taken from Greenspan (1977). All radiosondes were tested twice at each RH level, except at 100 %RH.

Relative Humidity (%RH)	0	<u>11.30 (0.27)</u>	<u>33.78 (0.18)</u>	<u>75.29 (0.12)</u>	<u>97.30 (0.45)</u>	100
SHC contents	desiccant	LiCl	MgCl	NaCl	K ₂ SO ₄	distilled H ₂ O

Table 4. Overview of laboratory experiments performed at the GRUAN Lead Centre to characterize the RS41 sensors.

Experiment	Sensor	Range	Set-up	Time frame
Radiation	T, RH	3-1013 hPa	radiation set up	2014-2020
Time-lag	RH	-75°C - -40°C	climate chamber	2015-2020
Calibration	RH	0-100 %	SHC, reference salts	since 2014

Table 5. List of scientific analysis tasks with their corresponding lead investigator(s).

Task	Lead
Laboratory experiments	
Radiation tests	Lead Centre
Time-lag	Lead Centre
Development GRUAN data product for RS41	Lead Centre
Twin soundings	Lead Centre + Task Team (TT) radiosonde
Ancillary measurements	TT ancillary measurements
NPROVS	Tony Reale
Metrology	Andrea Merlone
Statistical analysis	Alessandro Fasso
Evaluation of RS92-RS41 transition	Lead Centre/Working Group