The authors thank the reviewer 1 for his time and comments. Please see below the authors' replies for each specific comment.

Reviewer 1 comment:

**Major comments**

The Vaisala RS41 radiosonde uses a platinum resistance thermometer as temperature sensor, which does not require re-calibration between production and use of the radiosonde in normal operations. A simple comparison with the temperature sensor of the humidity sensor is sufficient to justify the calibration stability at the time of launch and within preset limits.

Authors’ reply:

_What reported in the reviewer’s comment is in line with the manufacturer’s specifications, while the main goal of this work is to independently assess the reliability and the performance of the RS92 and RS41 temperature sensors. The calibration drift and uncertainty of the RS41 temperature sensor have been quantified independently of the manufacturer, showing a very good agreement with the manufacturer’s specifications and that this sensor does not require re-calibration before the radiosonde launch. Moreover, the fact that a platinum resistance thermometer does not require periodical recalibration is not supported by facts. The sensing platinum element is a delicate piece of equipment, and mechanical and thermal shocks can significantly alter their nominal resistance at a given temperature, see for instance Kowal et al 2020._

Reviewer 1 comment:

In contrast, the Vaisala RS92 radiosonde uses a capacitive sensing element, which has small inherent calibration drifts. To account for these drifts, Vaisala uses a ground check device (GC-25) to compare the measurement of the radiosonde against a platinum reference thermometer, built into the ground check unit. The processing software uses this measurement to correct the calibration drift between production and use of the radiosonde. A correction to the raw measurements is then applied to produce finalized data. Any study evaluating the calibration of the RS92 temperature sensor must use the processed data, not the raw data as the authors have done in their study. Although the
authors describe that this re-calibration is occurring in the Vaisala software, it is ignored.

Authors’ reply:

As for the RS41, the goal is the manufacturer (and GC25)-independent assessment of the calibration drift and uncertainty of the temperature sensor, in order to compare them, respectively, with the calibration drifts resulting from the GC25 and the calibration uncertainty declared by the manufacturer. Such an independent assessment is necessary to develop a transparent, reproducible and manufacturer-independent data processing starting from the same radiosonde’s raw data, following the approach of GCOS Upper-Air Reference Network (GRUAN) for providing reference measurements. Therefore, the RS92 raw temperature measurements instead of the manufacturer-processed measurements, corrected by the GC25 results, were compared with the measurements of the reference thermometer calibrated at INRIM. This independent comparison resulted in a cold bias in the calibration, with a correction factor ranging from 0.1 °C up to 0.3 °C, as well as a calibration uncertainty (k=1) less than 0.1 °C and 0.025 °C larger than that reported in the manufacturer specifications.

Reviewer 1 comment:

The authors point out that an improper ground check can make the observations worse. The conclusion should have been to do a proper ground check and to evaluate the calibration of the GC25 reference thermometer in order to evaluate the calibration of the processed RS92 temperature data. The accuracy of the RS92 temperature measurements depends on it. In operational use the GC25 reference temperature sensors should have been recalibrated in regular intervals of every one to two years, which high quality radiosonde stations typically did. Without an evaluation of the GC25 reference thermometer, an evaluation of the RS92 temperature sensor calibration is not very useful.

Authors’ reply:

The calibration drifts measured with the GC25 resulted in a warm bias compared to the platinum resistance thermometer inside the GC25 unit, with a correction factor ranging from 0.15 °C up to 0.27 °C. As a result, the application of this correction to RS92 raw measurements leads to an increase (up to 0.6 °C) of the cold bias compared to the reference thermometer calibrated at INRIM, worsening the accuracy of RS92 measurements rather than correcting them. This is clearly due to a not reliable correction of the GC25, presumably caused by not having recalibrated in the last two years the Pt100 thermometer inside the GC25 unit. Thus, the conclusion is that the RS92 temperature sensor requires both a pre-launch calibration correction with the GC25 and regular (at least every 2 years) quality assurance checks (recalibrations) of the Pt100 thermometer inside the GC25 unit to avoid significant biases (up to 0.6 °C) in radiosounding measurements. This clearly shows the usefulness of an evaluation of the RS92 temperature sensor calibration independent of the GC25, for example in order to estimate and possibly correct any biases that may affect the data records of not high quality (reference) radiosonde stations on a global scale, that may not always have performed the regular recalibrations mentioned above. Ultimately, our methodology and results confirm, independently of the manufacturer, the better performance of RS41 compared to RS92, in terms of both accuracy in pre-launch temperature measurements and less demanding procedures for the quality assurance of the ground check device.

Reviewer 1 comment:

The authors did not pick up on the fact, that the RS92 temperature shows a periodic signal of about 140 s or so. This disturbance is most likely caused by the heaters of the
two humidity sensors, which cycle at about that period. One of these two sensors is located closer to the temperature sensor than the other. The humidity sensors of the RS92 are heated much more strongly than on the RS41 and this heat source clearly affects the temperature reading of the RS92 temperature sensor. In normal sounding operations, this is not expected to be an issue due to the much stronger ventilation passing first over the temperature sensor. In the configuration shown here, multiple heat transfer paths are possible. The authors interpret this signal as additional noise, where in fact it is most likely due to artificial heating by one of the two humidity sensors. It is possible that the heating of the humidity sensor is also responsible for the temperature dependence that the authors observe in the calibration accuracy of the RS92, depending on the details of the heating of the humidity sensors.

Authors’ reply:

Actually, the periodic structure of the RS92 temperature signal immediately appeared to the authors, somehow related to the periodic switching on and off of the two humidity sensors and their heaters. However, although the temperature signal is certainty affected by the swapping cycle of humidity sensors and their heaters, this effect is challenging to be properly quantified, also considering the irregular duration and intensity of the signal maxima. Therefore, the RS92 temperature signal has been considered characteristic of the simultaneous operation of the radiosonde’s temperature and humidity sensors and appropriate to characterize the calibration accuracy. This is also in view to provide the results according to fairness criteria.

Certainly, it is very reasonable to assume that simulating conditions more similar to those of a real radiosounding with a stronger ventilation in the chamber can reduce the effects of the heating of the humidity sensors on the temperature signal. Therefore, in agreement with the reviewer's consideration, the text of the manuscript will be amended as follows:

- Mentioning the disturbance to the RS92 temperature signal due to the RS92 sensors’ architecture and periodic switching on and off of the two humidity sensors and their heaters, in particular those closer to the temperature sensor.
- Mentioning in the results and conclusions that the noise and the calibration uncertainty obtained for the RS92 are probably to some extent overestimated compared to the conditions of a real radiosounding due to the above disturbance, which, in real soundings, is mitigated by a stronger ventilation on the sensors.

It’s also useful to point out that, to our knowledge, it does not exist a publicly available documentation, from the manufacturer or independent, showing how and to what extent the signal and the calibration accuracy of the RS92 temperature sensor change with respect to those reported in this work under ventilation conditions similar to those of radiosoundings. Moreover, in the documentation provided by the manufacturer the ventilation and pressure conditions to which the calibration of RS92 temperature sensor refers are not reported. On the other hand, in the manuscript conclusions it will be added that further experiments in climatic chambers are recommended by using a measurement configuration suitable for simulating conditions more similar to those of a real radiosounding, with decreasing pressure levels and different ventilations on the sensors (although we are aware there are issues in controlling the air flow around the sensors and reproducing the real ventilation on the sensors in radiosoundings, which results from the complex combination of the balloon lifting vertical speed (typically 5m/s), the horizontal wind, and radiosonde rotations and pendulum motions).

Reviewer 1 comment:

Just using two radiosondes for this evaluation is not sufficient, since there is some production variability of this mass-produced radiosonde. To understand the calibration
stability of these sondes for the global network requires some statistical analysis of more than one sonde of each. Without that evaluation, the results are specific for the two tested radiosondes, but not applicable to any other. For the RS92, using sondes from different production batches would be useful, in particular, since the sensor did undergo some substantial changes during the lifetime of this sonde model.

Authors’ reply:

As clearly stated in the conclusions of the manuscript (see lines 668-669), the authors are aware that the results of their evaluation are specific for the two tested radiosondes and these results need to be consolidated by further tests with multiple pairs of radiosondes, in order to obtain results applicable to the different production batches. Nevertheless, in our opinion, the main contribution of this work consists in the introduction of a methodology to simultaneously and independently test the temperature sensors of two different radiosonde models within climatic chambers, in terms of noise, calibration accuracy and bias of sensors’ measurements. The latter have also been quantified, for the first time independently of the manufacturer, for the temperature sensors of Vaisala RS92 and RS41 radiosondes, although referred to a single pair of radiosondes. However, further tests with multiple pairs and production batches of radiosondes and under ventilation/pressure conditions more similar to those of radiosoundings are expected to be performed in the future.

Reviewer 1 comment:

Minor comments:

The dynamic tests are more or less meaningless. The time response of the sensors during a sounding profile play a very important role for the ability to resolve vertical structures. As the authors point out the balloon ascent provides a reasonably well defined ventilation speed of at least the balloon ascent velocity. The tests done have an undefined ventilation and are not representative for atmospheric observations. As the authors note, these dynamic tests may simulate taking a radiosonde from the preparation office to the outside. This transition is completely irrelevant for soundings.

Authors’ reply:

The time response of the sensors is not discussed in this work. Similarly to the tests performed at the first stage of the experiment, the subsequent tests with two climatic chambers were performed under the ventilation conditions (not well defined) generated by the chambers to homogenize the temperature field inside. However, the supposed invalidity or irrelevance for radiosoundings of the outcome of these tests needs to be demonstrated by means of similar tests performed under conditions more similar to those of a real radiosounding. To our knowledge, similar tests have never been carried out by the manufacturer or independently. More specifically, the dynamic tests of this work aim to investigate potential effects on the radiosondes’ temperature sensors of fast and steep thermal changes (in the order of about 20 °C) that radiosondes may meet when passing from indoor to outdoor environment before launch. These thermal changes are simulated by quickly moving the measurement frame equipped with the two radiosondes between the two climatic chambers. The test results reveal that such thermal changes may increase the noise and the calibration uncertainty of temperature sensors, at least during the first part of a radiosounding. In our opinion, this result can be of great interest for metrology and meteorology and climate communities, as it indicates a possible underestimation of the above uncertainty contributions in the algorithms currently used to process the raw measurements of both radiosonde models. Surely, this result needs to be confirmed by further tests with multiple pairs and production batches of radiosondes and under conditions more similar to those of a real radiosounding.
Reviewer 1 comment:

There is also some concern that placing two transmitting radiosondes in close proximity in an environmental chamber may cause some radio frequency interference effects that possibly do not occur in a normal sounding environment. Given the level of confidence the authors try to achieve (<0.1 K), evaluating whether RFI effects occur would be paramount. In particular the capacitive sensor of the RS92 may possibly be more susceptible to this effect under these conditions.

Authors’ reply:

As reported in the description of the experimental setup (see lines 161-168), the two Vaisala sounding systems used were configured to separately receive and process the signals transmitted by the two radiosonde models at two different frequencies, 402 MHz for the RS92 and 405 MHz for the RS41. The bandwidth of the telemetry signals (5 - 20 KHz) and the distance between the two selected frequencies ensures there is no interference between the signals received from the two radiosondes. More specifically, before the tests, each sounding processing subsystem (SPS311) was set to receive the signal from a single radiosonde model at the selected frequency and to communicate with a single ground-check device type. During the ground check procedure, the SPS311 recognizes the sonde, enables it to transmit at the selected frequency and then selectively receives and processes the signal transmitted at that frequency. On the other hand, no interference was reported for receiving systems similar to that of this experiment, used for comparing the same radiosonde models via dual or multiple soundings (e.g.: Nash et al., 2011; Jensen et al., Atmos. Meas. Tech, 2016), even with the two transmitting radiosondes in a closer proximity than in the present experiment (Kawai et al., Atmos. Meas. Tech, 2017). Finally, no interference from other sources was identified during the tests in the climatic chambers.

Reviewer 1 comment:

With all that criticism, I would like to point out that evaluating the calibration stability is only a minor factor for the measurement of atmospheric temperature. Much more significant is their behavior in real world conditions, i.e. evaluating the radiation correction, which is up to an order of magnitude larger than the calibration uncertainties discussed here. Additionally, the behavior in clouds under condensation conditions, is another essential challenge for in situ temperature measurements, where these two sondes may show significant differences. However, these factors were not addressed.

Authors’ reply:

The authors would like to stress that to provide reference measurements, all known systematic errors, as well as the uncertainty contributions related to these errors, should be properly quantified. Moreover, although the radiation uncertainty is dominant for daytime radiosoundings through the upper troposphere and stratosphere, the calibration uncertainty is a major contribution for night time radiosoundings (through the whole atmosphere) and for daytime radiosoundings through the lower troposphere. On the other hand, the radiation error and uncertainty for the temperature sensors of RS92 e RS41 have been characterized in other dedicated laboratory experiments (Dirksen et al., Atmos. Meas. Tech, 2014; von Rohden et al., Atmos. Meas. Tech, 2021). Finally, regarding the sensors’ behavior under condensation conditions, the tests performed in this work include the comparison of radiosondes’ temperature sensors under conditions very close to condensation, with RH values of 98% and 95% for temperature values of 20 °C and 40 °C respectively. Although further investigation is needed, by extending the tests above to lower temperatures and simulating ventilation and pressure conditions more similar to those of real radiosoundings, to our knowledge, no dedicated laboratory experiments have
so far been performed on this specific topic.