The manuscript by Rosoldi et al. evaluates the calibration of the temperature sensors on an RS92 radiosonde and an RS41 radiosonde. The latter model is the successor of the former and plays an important role in the global radiosonde observing network. An understanding of the performance of the temperature sensors on these radiosonde models is important to evaluate their effect on numerical weather forecasting and climate records.

Brief summary of my main concerns

The authors have not prepared the ground check of the RS92 correctly and used an improper data table for the RS92 generated by the sounding software in their analysis. They missed the influence of the RS92 humidity sensor heater, which contributes to the temperature signal measured by the RS92 and which requires a different setup to reduce its effect. They have used only one sonde of each, making it difficult to transfer the results to an average production run. I am very doubtful that the current setup is suitable to address the temperature calibration of these two radiosondes to the level the authors try to achieve so that others may build upon.

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.
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.

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.

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.

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.

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.

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.

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.