Reply to the reviewer's comments on Manuscript No. AMT-2017-119
by K. Gierens, K. Eleftheratos, and R. Sausen

Remarks

Questions by the referees are reprinted in the following in *italic font* for convenience. Our replies are in upright font.

Changes to the revised manuscript are marked in red colour. Deleted text is simply marked as "text deleted".

Note also that we have added a figure to the manuscript and that the colour schemes of several figures has been changed.

1 Reply to reviewer No. 1

Before replying to the individual questions we want to stress that, contrarious to what the referee wrote, we do not "attempt to validate a method previously published by two of the authors". This is wrong. The question that the present paper tries to answer is clearly stated in the introduction (quotation on page 2), and its subquestion (1) (lines 19,20 on page 2) should make it sufficiently clear that the paper attempts to validate (or dis-validate) the statistical intercalibration method of Shi and Bates (2011).

1.1 Cloud and clear sky conditions

Cloud - the entire analysis is based on clear sky conditions, which may be appropriate for UTH analysis. However, as many UTH analyses will rely on the HIRS data themselves to set a cloud-free flag, these will themselves be sensitive to the HIRS calibration. The method described in the manuscript could easily be extended to a much more general application of all-sky radiances, by the addition of black cloud to layers in the radiosonde profiles layers close to saturation.

Indeed the analysis is based on clear sky conditions. The intercalibrated channel 12 data of Shi and Bates (2011) are cloud–cleared as stated in their paper (page 3, beginning of par. 9):

> "The HIRS data are first processed to remove cloudy pixels for the water vapor field. The cloud clearing procedure follows the method detailed by Jackson et al. [2003]. The process is accomplished using a simplified method based on the ISCCP cloud detection approach [Rossow and Garder, 1993]."

It seems the reviewer wants us to discuss the possibility that HIRS channel 12 data could be used to derive a cloud flag for other UTH studies using other platforms (if we understand the question correctly). This seems hardly possible if the HIRS channel 12 data are already cloud cleared as stated.

Of course, we could include cloud layers in the radiative transfer calculations, but to which end? It would not serve the purpose of the present paper and it would introduce additional degrees of freedom (optical thickness, crystal habits, 2D vs. 3D radiative transfer, ...) that it is difficult to see for us what gain of knowledge one could expect from such an effort.

1.2 Other platforms

What about the HIRS instruments on other NOAA and Metop platforms? These each have subtly different SRFs. A simple analysis should be conducted to show whether the results for the

HIRS on NOAA-14 and -15 can be generalised to all other HIRS/2 and HIRS/3 instruments, respectively, and by extension to HIRS/4.

The problem treated in the present paper is clearly the major change of SRF and the central channel wavelength between HIRS/2 on NOAA 14 and HIRS/3 on NOAA 15. This is clearly explained in the abstract and the Introduction. We show in the paper that the statistically based correction that Shi and Bates had performed can be supported by physical considerations. Otherwise there is nothing special with NOAA 14 and NOAA 15 that would exclude the application of the same arguments to other pairs of HIRS/2 carrying satellites on one hand and HIRS/3 or 4 carriers on the other. The only reasons to concentrate on NOAA 14 and 15 are a) that the change just happened between these two satellites and b) that therefore there is a certain period of time of common operation. Without common operation it is not possible to perform the necessary regression.

As argued in the paper, the wavelength change from 6.7 to 6.5 μm involves an upward shift of the weighting function maximum by about 1 km. Eventually this is the reason that specific features in the humidity profiles can be distinguished that lead one to look at channel 11 as well. Other modifications of the SRF do not involve a large jump in the position of the maximum of the weighting function. This renders it improbable that specific features in the humidity profiles could be found that serve a correction.

1.3 Temporally changing SRFs

Similarly, the SRF of several HIRS instruments has been suspected of changing during their lifetimes. It should be emphasised that this analysis represents a relatively short period in the lifetime of two HIRS instruments, and that their spectral and radiometric calibration was assumed to be constant over this period.

Certainly SRFs may change over their lifetime as a consequence of degradation of the instrument. The corresponding changes to the recorded brightness temperatures are probably small compared to the change due to the shift of the central wavelength, which is the problem we are concerned with. If it were not small, the satellite data could not be used in a meaningful way. The quoted NWPSAF web page from which we obtained the SRFs offer only one channel 12 SRF for each satellite; there is no information of temporal changes. The experts who provided the SRFs seem not to see a major problem with temporal changes. Thus, the working hypothesis that the SRFs are temporally constant is in fact necessary. We are not aware of intercalibration or other UTH related studies where temporal changes of SRFs of individual instruments ever have been taken into account or mentioned.

We have added the suggested sentence as the last sentence of the text.

1.4 Uncertainties

Uncertainties are generally neglected throughout the paper. While results are often expressed in terms of correlation coefficients, it would be more helpful to quantify the uncertainties - for example, with which HIRS/3 brightness temperatures can be predicted from HIRS/2 observations.

It is true that not all uncertainties are provided in the paper. Some are given now in this reply (cf. 1.6.11, 1.6.15, 1.6.18, 2.11), but we think they are too unimportant to cite them in the paper. Our conclusion will in no way be affected or modified when the uncertainties are considered.

1.5 Training and testing profiles

Two potentially serious deficiencies in the method related to the selection of profiles used for training and testing the regression should be addressed - see below.

To clarify the issue on page 3, line 24, we state clearly that the Lindenberg profiles are used for the regression ("our training data set"). In line 29 we state that the profiles from Sodankylä and Manus are used to test the regression.

1.6 Specific points

1.6.1 p3.14

What range of angles were covered?

DISORT uses Gauss-Legendre quadrature for integration along the polar angle. This has been described long ago by Chandrasekhar (1960, chapter II). 16 angles correspond to an 8-point quadrature in both hemispheres and the range of the abscissae, expressed as cosine of polar angle, is according to table 25.4 of Abramowitz and Stegun (1972), 0.09501... to 0.98940..., which corresponds to a set of abscissae for integration, in angular range of zenith angles, of 84.5° to 8.3°.

1.6.2 p3.17

It is not generally accurate to say that brightness temperatures from infrared sounders are only reported for cloud free situations .

This may be so and the statement has been made more precise: "brightness temperatures from the Shi and Bates data set are cloud cleared, see Shi and Bates, 2011".

1.6.3 p3.21

Why restrict the analysis to observations within 30° *of nadir?*

HIRS has a swath angle of almost 50°, but usually this angle is not used completely. In Gierens et al. (2014) we used only the inner 33°, as recommended then by our co-author from NOAA.

The analysis here shows clearly that the difference between the nadir and 30° off-nadir results is fairly constant. This is confirmed by the summary statistics for Fig. 1 that we have now provided, see sect. 1.6.11 of this reply. Thus, the gain in knowledge from other off-nadir angles is expected low.

1.6.4 p3.23

Are the set of profiles restricted to clear sky conditions? If so, how? Otherwise, the distribution could bias the regression results.

For the purpose of the present paper we could have simply invented test profiles but found it more convenient to take profiles from radiosonde data. Whether these profiles are from clear or cloudy situations is completely irrelevant. The radiative transfer calculations serve the purpose of finding features in the humidity profiles that are somehow related to the absolute size of the brightness temperature difference in clear sky. We stress that it is not the existence of clouds that cause the problem in the HIRS/2 to HIRS/3 transition.

1.6.5 p3.30

It is not clear exactly which profiles were used for training and testing the regression. It is good practice to subset the dataset used for training and testing.

See above: Section 1.5

1.6.6 p4.1

It would be worth pointing out that this corresponds approximately to the height of the tropical tropopause and that very little water vapour is expected above this level (or that it is poorly measured by radiosondes). However, it would be good to quantify its contribution to the Ch12 radiances.

The chosen cut-off pressure of 90 hPa for the radiosonde profiles means that the profiles above are extended using appropriate standard profiles that are provided within LibRadtran. 90 hPa corresponds roughly to 17 km altitude. From figure 7 it is possible to estimate that the weighting function values above 17 km are very small compared to their maxima. The contribution of these layers to measured brightness temperatures are thus very small.

The atmosphere is generally very dry above 17 km. The relative humidity in the lowermost stratosphere is typically around 5% as we know from MOZAIC-IAGOS data (Neis, 2017). Humidity features in this altitude range are thus too weak to have a significant impact on the measured brightness temperature differences. As the paper demonstrates, it is the middle troposphere whose humidity features have an impact. Taking standard humidity profiles above 90 hPa is justified.

We have made a slight rewording of the text to stress that the water vapour profiles are not cut-off at 90 hPa but replaced by the corresponding standard profiles.

1.6.7 p5.15

This contradicts the values given in par. 5.1. Please reconcile! It would also be helpful to compare the peak of the weighting functions of Ch11 and Ch12 here - especially bearing in mind how these depend on the water vapour column amount.

We don't understand! Here as in section 5.1 we assume a peak altitude of 5 km for channel 11. Where is the contradiction?

We agree that the peaks of the weighting functions depend on the column density of water vapour. A higher amount of water vapour lets these peaks rise and vice versa. Probably the vertical distance of these peaks is variable to a certain degree; this could be tested with radiative transfer calculations, but is not topic of the present paper. The derived regression covers such effects in a statistical manner.

1.6.8 p5.28

The choice of using brightness temperatures instead of radiances should be justified briefly.

There are a number of reasons for using brightness temperatures instead of radiances.
1) The HIRS data we have are given as brightness temperatures.
2) The formula for retrieval of UTH (upper tropospheric humidity) by Soden and Bretherton (1993) and Jackson and Bates (2001) uses brightness temperature as independent variables.
3) Brightness temperatures is a more natural measure for applications in atmospheric physics than radiances.

1.6.9 p5.31

Is it possible that the Ch12-Ch11 difference contains additional information in off-nadir directions?

It is improbable that there is additional information in the off-nadir radiances because. Old figures 1 and 2a (now 2 and 3a) show that the difference between nadir and 30° off-nadir radiances is quite constant and independent of the actual situation (i.e. the actual radiosonde profile).

1.6.10 p6.5

For completeness the covariance of the uncertainty of parameters should be included - to allow the evaluation of the uncertainty on the fitted brightness temperatures.

Unfortunately, the REGRESS function of IDL does not provide the covariance matrix, so we had to compute it from the design matrix and the variance of the residuals. The result is

$$C = \begin{pmatrix} 3.74562 & -0.00649644 & -0.00873263 \\ -0.00649365 & 0.000112047 & -7.66601 \cdot 10^{-5} \\ -0.00873516 & -7.66493 \cdot 10^{-5} & 0.000103991 \end{pmatrix}$$

This has been computed using relations from chapter 8.4 from von Storch and Zwiers (2001). The variance values of approximately 0.0001 for coefficients b and c can be read off the diagonal line, which leads to standard deviations of about 0.01 as given in the paper.

The corresponding correlation matrix is

$$\rho = \begin{pmatrix} 1 & -0.32 & -0.44 \\ -0.32 & 1 & -0.71 \\ -0.44 & -0.71 & 1 \end{pmatrix}$$

As these matrices do not modify any of our conclusions, we deem it unnecessary to show them in the paper.

1.6.11 p6.12

It is not sufficient to say Obviously, the superposition methods works well for these data... - quantitative statistical evidence is needed.

Here is the required evidence. The residual means ($T_{12/14} - \hat{T}_{12/15}$) are
-0.5 ± 1.1 K for Sodankylä and
-1.2 ± 0.4 K for Manus.

We have added this information in the paper (section 4.2).

Mean and standard deviations of the residuals for Lindenberg are -0.8 ± 0.5 K. Text added at the end of section 4.1.

1.6.12 p6.27

It is not sufficient to say Obviously, the superposition methods works well for these data... - quantitative statistical evidence is needed.

We have computed now two sets (NOAA 14 and NOAA 15) of actual weighting functions for three profiles from Lindenberg, one from each group of profiles given in the old figure 3 (now 4). The figure included at the end of this reply demonstrates what our discussion has stated, namely that HIRS 3 on NOAA 15 gets its photons from higher layers than HIRS 2 on NOAA 14. One of the NOAA 14 weighting functions has ground influence (simple black line). This is one of the profiles which give a large ΔT_{12} and which are characterised by a dry middle troposphere. The shown one is very dry down to low altitudes and thus it is possible that photons emitted at ground reach the satellite.

It is clear that channel 11 sees the ground whenever channel 12 sees it.

The text in section 5.1 has been slightly modified, the figure with the actual weighting functions is not included.

1.6.13 p7.12

Why is this range of latitudes selected? (70° N is outside most definitions of mid-latitude)

The reason to extend our study up to 70°N is that we are ultimately interested in ice supersaturation in the upper troposphere. Our experience, for instance from Gierens et al. (2004), is that nadir sounders find ice supersaturation mostly at latitudes higher than 55°N, although this phenomenon is not restricted to high latitudes. But nadir sounders can only detect ice supersaturation if the supersaturated layers are, say, 3 km thick or more. Such thick layers seem to be confined to high latitudes.

1.6.14 p7.17

Introducing the symbols x and y does not aid the understanding here. Please change to follow the same convention used in the rest of the paper ($T_{12/14}$, etc).

Done.

1.6.15 p7.20

Without uncertainties on the regression coefficients, their comparisons are meaningless.

As the number of data is large (more than 700000) the uncertainties are quite small, namely 0.3% for the intercept and 0.07% for the slope. We expect that the errors in GE17 regression parameters are similarly small because of the same number of data (we did not recalculate them).

There is no need to quote these numbers in the paper since we only say that the coefficients of the new paper are *similar* to those of GE17. This is a weak qualification that does not require to provide detailed error estimates.

1.6.16 p7.29

this problem - which problem?

"This problem" is the "considerable overestimation of the number of supersaturation events recorded with HIRS 3 and 4 instruments" (new text in the revised version).

1.6.17 p8.9

It is not at all obvious to me from Figure 9 what the problems in the low tail of channel 12 brightness temperatures are.

Yes, this is indeed confusing. The reason is that *low* channel 12 brightness temperatures lead to *high* UTHi values. Thus problems with the low tail of T_{12} lead to problems in the upper tail of the UTHi distribution. The text is reformulated in the revised version to avoid the confusion.

1.6.18 Fig. 1, summary statistics

Here follows the summary statistics that the reviewer likes to see.

Sodankylä:
mean difference at nadir: $-6.7 \pm 1.2\,\mathrm{K}$
mean difference at 30°: $-6.7 \pm 1.3\,\mathrm{K}$

Manus:
mean difference at nadir: $-7.1 \pm 0.9\,\mathrm{K}$
mean difference at 30°: $-7.1 \pm 0.9\,\mathrm{K}$

These numbers show, as stated in the paper, that the off-nadir direction does not present anything new that is not already found from the nadir direction.

These mean differences are now included in the paper (section 3, 1st par).

1.6.19 Old Fig.1 (now 2), scales

Done: same scales for both panels.

2 Reply to reviewer No. 2

2.1 General remarks

I find it problematic to call this a physics-based approach considering that the physical reasons for the differences in the measurement are mostly ignored. I know that modelling the instrument properly is a major undertaking and I do not expect the authors to do so for this paper, but I think the term physics-based promises too much. There are many other differences between HIRS/2 and HIRS/3, and even between HIRSes in the same series, such as calibration, footprint *size, and many technical details within the instrument, that should lead one to expect differences.*

For us, radiative transfer is physics. This is the physics that we mean. What the referee lists, "calibration, footprint size, and many technical details" are exactly this: technical details. Although technical details have physical effects, they are not physical details. So we think the critique at our use of "physical" is not justified. Furthermore, the paper text is clear enough that our choice of "physical" is intended to distinguish the approach from the "statistical" approach of Shi and Bates (2011).

We also agree that there are technical details that lead to differences in the response of various sensors to the same scene and that temporal changes (degradations) of these technical details even would lead to differences in the response of the same sensor when it could watch the same scene at different times. All this is correct. But it is very probable that response differences caused by such technical details are small compared to the response difference caused by the shift of the central wavelength. Otherwise the data could not be used in any meaningful way. We are not aware of any warnings to use the data because they might be spoiled by techical degradations. The question that we are interested in and that we try to solve by radiative transfer calculations is, whether the data can be used in a meaningful way considering the shift in the central wavelength. This is not so clear a priori since, as we wrote, this shift is equivalent to a shift of the most sensitive layer in the atmosphere by about 1 km or more. This can in principle break the desired 30+ years time series when the correlation of relative humidity in two altitudes one kilometre apart is low, which it probably is. We think that it is the width of the weighting functions of several kilometers that eventually saves the time series. But is was necessary to demonstrate this, and this is the purpose of the paper.

2.2 Page 3, line 6-7

Almost certainly the author mean the complete spectral response function for the entire optical path, not only for the filter. The overall spectral response function takes into account (in fact, is a multiplication of) all components in the optical path, including mirror, filter, detector, etc. In fact, some of those may degrade over the period of an orbit such that the overall SRF may change. In any case, I think the authors should use the term channel spectral response function (SRF) rather than filter response function, to avoid confusion about what is meant.

This is correct. The only information we have is the set of published channel spectral response functions which are those of the complete optical path. It is evident that we need the complete optical path since the published brightness temperatures are affected by the complete optical path. We agree to use the more specific term "channel spectral response function" instead of the misleading incorrect term "filter response function".

2.3 Page 3, line 16-17, Part 1

"channel 12 weighting functions do not reach the ground", this statement may be too general. It is true for the situation described by the authors, but is it also true in polar or high mountain conditions?

We agree. This too general statement has been deleted.

2.4 Page 3, line 16-17, Part 2

"brightness temperatures from infrared sounders are only reported for cloud-free situations". This statement is confusing. The authors must be talking about the simulations, not the measurements, which are all-sky top-of-atmosphere radiances. I suggest replacing "reported" by "calculated".

Yes, this statement is confusing. The intercalibrated channel 12 data of Shi and Bates (2011) are cloud cleared as stated in their paper (page 3, beginning of par. 9):

> "The HIRS data are first processed to remove cloudy pixels for the water vapor field. The cloud clearing procedure follows the method detailed by Jackson et al. [2003]. The process is accomplished using a simplified method based on the ISCCP cloud detection approach [Rossow and Garder, 1993]."

We state now, that these data are cloud cleared instead of that infrared data are reported as cloud-free data.

2.5 Page 3, line 30

I'm confused. Are you including the Sodankylä and Manus data for training, or merely for validation/comparing?

To clarify the issue on page 3, line 24, we state clearly that the Lindenberg profiles are used for the regression ("our training data set"). In line 29 we state that the profiles from Sodankylä and Manus are used to test the regression.

2.6 Page 4, line 12-15

If data are clearly wrong they shouldn't be shown in the first place. I don't think showing this tail in this figure adds any value to the paper. Just state in the text that a sanity check has discarded x/y, alternately retained $(y-x)/y$ of the radiosonde profiles.

Why not? To our view showing bad data along with the good data is a form of honesty. We agree that showing the tail may not add any independent information. But showing it does no harm to the paper.

2.7 Page 4, line 22-23

The authors find $\Delta = 6.5\,K$, Shi and Bates find 8 K. What causes the 1.5 K discrepency?

The authors do not find a temperature difference of 6.5 K. We state that in the upper troposphere the average temperature difference of two layers one kilometre apart in the vertical is 6.5 K. The average brightness temperature difference the authors found from the Lindenberg data is 7.2 K, while the quoted 8 K is approximately the average of the temperature-dependent corrections applied by Shi and Bates (2011). So, if the sensitivity maxima of HIRS 2 and 3 are apart in the vertical by a bit more than 1 km, the average temperature and brightness temperature differences agree very well and there is no discrepancy at all.

2.8 Page 5, line 11

The precision of this number is very high. How is it calculated?

We agree. This precision is not justified. We follow Soden and Bretherton (1996) and replace it by 7.3 μm.

2.9 Page 5, line 23-28

The authors conclude that "One additional independent variable is clearly insufficient to capture all this variability." As the authors show in Figure 5, the additional variable is not independent, but has a very strong correlation. They then proceed with the bilinear regression in equation (1) anyway, despite having concluded it is insufficient. This is somewhat surprising. Have the authors considered using a multilinear regression with additional HIRS channels, perhaps including those that add information on temperature (oxygen channels)?

We agree to replace the words "independent variable". We now use "piece of information". Indeed there is some dependence between what channel 11 and channel 12 measure because of the wide, thus overlapping, weighting functions.

Wo do not agree to use irrelevant information from oxygen channels etc. It is not clear to us in which way one could find information on mid-tropospheric relative humidity profiles in oxygen channels.

2.10 Page 5, line 27 and footnote

I don't think the specific software is relevant (except in the case of open-source software, where the license may require a mention+citation in the acknowledgements, but that is not applicable here)

We do not believe that readers will be confused by this information.

2.11 Page 6, line 9-12

I do not agree with the author's interpretation of the result. It would be easier if they would plot the difference as a function of reference, but even as shown it is evident there are large differences exceeding 4 K.

The residual means ($T_{12/14} - \hat{T}_{12/15}$) are
-0.5 ± 1.1 K for Sodankylä and
-1.2 ± 0.4 K for Manus.

These residuals are much smaller than the original differences between $T_{12/14}$ and $T_{12/15}$ shown in Figure 1. We interprete this as a good result. A few words are added in section 4.2.

Mean and standard deviations of the residuals are -0.8 ± 0.5 K. Text added as last sentence in section 4.1.

2.12 Page 6, line 15

Why are you showing such a generic weighting function? Please show actual weighting functions corresponding to typical radiosonde profiles for each station, both summer and winter, for all channels. That probably belongs in the methods section of the paper.

We show generic weighting functions because we want to present a *simple* argumentation of why the inclusion of channel 11, i.e. information on mid-tropospheric humidity, helps to bring $T_{12/14}$ and $T_{12/15}$ on a common level. For this purpose it is not at all necessary to use actual weighting functions, moreover use of individual weighting functions for special situations would impede argumentation and understanding ("Why this situation and not another one?", etc.).

2.13 Page 6, line 23

Although it will happen less frequently, channel 12 in both cases may also see the ground sometimes. What does a typical Sodankylä winter water vapour Jacobian look like for those channels?

Yes. The figure included at the end of this reply shows a couple of actual weighting functions (computed as the product of extinction coefficient and transmission function, as necessary for the calculation of brightness temperature), see above (section 1.6.12). One of these (simple black line) reaches the ground without getting zero. We cannot say, however, if this is a typical example. It happens in this case since the lower and middle troposphere is quite dry. The corresponding profile can be seen in fig. 3 (top) of the original manuscript.

The text in section 5.1 has been slightly modified, the figure with the actual weighting functions is not included.

2.14 Page 6, line 24-27

Certainly, in HIRS L1B data as obtained from NOAA CLASS, ground observations are /not/ flagged. Flagging in HIRS data is purely based on events on the spacecraft, such as bad calibration, mirror movements, etc. See the NOAA KLM C3 User's Guide and NOAA POD User's Guide. It should be easy to check whether the larger scatter is due to ground influence, by calculating corresponding Jacobians. You should also investigate what the winter vs. summer points look like for the Sodankylä method, and/or colour-code the points by IWV, which can be estimated from radiosonde profiles.

Please see sections 1.6.12 and 2.13 of this reply. If channel 12 reaches ground, channel 11 must do so as well.

And such a figure is more usefully plotted by Δ vs. reference, like was done for Figures 1 and 2 (this comment also applies to other scatter plots)

This is not generally true. We have discussed this in the reply to the reviewers of GE17, which you can find under https://www.atmos-meas-tech-discuss.net/amt-2016-289/amt-2016-289-AC1-supplement.pdf, section 1: "Review-independent major changes". The arguments given there will not be repeated here.

2.15 Page 7, line 1-9

I don't understand what point the authors are trying to make here. What is the relevance or implication of this observation?

Indeed there is no immediate (i.e. for the current paper) implication of this observation. However, who knows whether new theoretical radiative transfer derivations in the spirit of Soden and Bretherton (1993) or Stephens (1996) will not find this observation interesting or useful.

2.16 Page 7, line 19

For both the present paper and GE17, what are the uncertainties associated with the best estimate of those parameters?

As the number of data is large (more than 700000) the uncertainties are quite small, namely 0.3% for the intercept and 0.07% for the slope. We expect that the errors in GE17 regression parameters are similarly small because of the same number of data (we did not recalculate them).

There is no need to quote these numbers in the paper since we only say that the coefficients of the new paper are *similar* to those of GE17. This is a weak qualification that does not require to provide detailed error estimates.

2.17 Page 7, line 27

Why don't you show those data in a density plot as well, like for the other figures? Scatter plots have marked disadvantages (see Carr et al., 1987).

The data are displayed as a density plot in the revised version, as the reviewer suggests. We use again a single-hue display, hopefully with perceptually uniform colour scheme. The main text and figure captions are adjusted accordingly.

2.18 Page 8, line 6-7

I disagree with the conclusion. You have shown that two independent methods give /consistent/ results, but that does not imply that the data series can be considered as homogeneous. You have shown the weighting functions for both NOAA- 14 channel 12, NOAA-15 channel 12, and your estimated pseudo-channel 12, in Figure 7. To me, that figure shows it is /impossible/ to produce a homegeneous data series. The authors have reached this result but apparently not this conclusion in two independent ways. That is a useful (but modest and unsurprising) result. If a user wants to produce a climate data record (such as for UTH) spanning the entire HIRS series, s/he will in any case need to take note of all the inter-satellite differences (not only between HIRS/2 and HIRS/3, but between each and every copy of HIRS, none of which are identical) in their forward modelling.

We agree that our conclusion was too strong. We replace it with a more modest one. In fact, no time series is truly homogeneous and nobody can make this one homogeneous. If we do not want to stop analysing time series, we need a more practical view. The question is not, whether the time series is homogeneous, but whether it is justified to use it *as if it were* homogeneous. To our view, our results contribute to raise this justification.

We don't know, why this is rated a "modest and unsurprising" result. It seems the reviewer knew in advance that radiative transfer calculations and the superposition of channel 11 with channel 12 would lead to something similar (perhaps therefore modest) to the Shi and Bates results. This is astonishing!

2.19 Page 8, line 26-28

Again this conclusion is not justified. Just because there exists a correlation between A and B, *does not imply that B can be fully predicted from A. It is trivial to construct a counterexample: just take $B = 2 * A$, then add random noise to both A and B. They remain strongly correlated but it is impossible to predict B from A.*

We accept this comment. We think it is sufficient to add the words "close to" before "the brightness temperature" and to write that "close to" is meant as a mean residual difference of around 1 K instead of 7 K as before.

2.20 Page 9, line 4-6

For the reasons explained above, I consider this conclusion is not justified.

Although our wording here was weaker than before we accept this comment and replace the statement by a weaker one. Note that we have also modified the last sentence of the abstract in the same spirit.

2.21 Page 9, line 13-17

I think where you obtained code and data is important information in the methods/data section and should not be tucked away in a couple of single-line paragraph near the end. The methods section also needs an illustration of the SRFs for both NOAA/14-12 and NOAA/15-12.

It seems the referee is not familiar with the submission rules of Copernicus Journals. These single-line paragraphs are mandatory.

2.22 Old Figure 2 (now 3)

consider showing a hexplot instead, as Carr et al. (1987) have proposed as a superior alternative to the scatterplot and hexagons have less distortion than squares or rectangles. See figure 10 in Carr et al. (1987).

The information that a hexplot would provide, viz. the density of points in a certain region, is given already in the bottom part of the figure as a colour plot. However, the top panel is not superfluous. It has the same form as the panels of Fig. 1 and it has the information on both nadir and off-nadir viewing geometries which the bottom part has not. If we would produce a hexplot, we would repeat the information and we would lose the format coherence with Fig. 1. This is not desirable.

please use a perceptually uniform colourmap (see Borland and Taylor, 2007). The chosen colourmap goes from white to dark blue to light blue to dark red, which leads to misleading inferences. It was probably designed for showing diverging data, such as data that are either negative or positive. In recent years there has been considerable progress in the development of perceptually (nearly) uniform colourmaps that are now included with all popular data visualisation tooklits.

We have changed the colour scheme now to a single-hue one with increasing bluish intensity implying increasing number of events. This scheme is one of the sequential schemes from Data Graphics Research, Department of Geography, University of Oregon (http://geography.uoregon.edu/datagraphics/index.htm). Unfortunately, no information is given on their site whether these schemes are perceptually uniform, but we think they are.

A tiny cut-off of the colour bar label has been corrected as well.

2.23 Old Figure 3 (now 4)

Those figures are very hard to read. The colours are too light: green, cyan, and yellow are hard to read against a white background. It appears the authors may have used a plotting tool that defaults a black background, and have changed the background colour without changing the line colours. The bottom panel is completely impossible to read, because (1) there are too many lines, (2) the line colours, and (3) the reuse of colours. As the individual profiles are not important, I would suggest the authors plot the data as a large quantity of very thin black lines, which will give a visual effect of looking "dense" when there are many lines together, thus somewhat having the effect of a density plot. I have previously used lines of 0.01mm thickness to this end, with a visually pleasing result when plotting hundreds of lines. Although individual profiles will not be distinguishable in this case, that is not relevant for the message conveyed by this figure.

We follow the reviewer's advice and replace all curves by thin black lines. As the line key does then make no sense anymore, it has been removed from the panels. The relevant launch dates are instead listed in the figure caption.

There appear to be some data problems in these figures as well. I don't know if fast variations such as 2000 07 01 12 in the middle panel between 90 and 80 kPa are correct, but the blocky structure at low RH, low pressure at the bottom panel clearly isn't.

In the troposphere (from the Greek "tropein": to turn, to change) fast variations of relative humidity with altitude are not uncommon. In this respect the troposphere is sometimes like a "deck of cards", an expression Reginald Newell once used in a meeting that the first author attended (see for instance Newell et al., 1999). The blocky structure results from the form the data are given. RH is given as integer values which results in this blocky structure when the variation of RH with z is very small as it is above the local tropopause.

2.23.1 legend

The word "average" can be confusing because many people use it to mean "arithmetic mean". The authors may wish to choose a phrasing like "values near the median" or so.

Agreed. We replace "average values" by "values near to the mean".

2.24 Old Figure 6 (now 7)

The authors should show the independent data on the x-axis and the dependent on the y-axis; need to swap the axis on this figure.

The independent data is shown on the x-axis and the dependent on the y-axis. $\hat{T}_{12/15}$ is the independent data, $T_{12/14}$ is the dependent data.

2.25 Old Figure 7 (now 8)

Do you mean averaging kernels or weighting functions? In any case, they need be shown earlier as this is a fundamental instrument property, and thus belongs in the instrument description.

To our understanding, "averaging kernel" and "weighting function" are synonymous. By the way, the expression "averaging kernel" is not used in the manuscript, but we often use "weighting kernel", as we already did in an old paper (Gierens et al. 2004). The definition of our generic kernel is given in this old paper, and a derivation of it in Gierens and Eleftheratos (2016). The term "weighting" means that it weights contributions of water vapour in different altitudes to the measured intensity, the result is a "weighted average". The term kernel comes from its mathematical form and use:

$$K(z, z'; H) = H^{-1}\, e^{\,(z-z')/H} \exp(\, e^{\,(z-z')/H}).$$

It is a function of two variables, z and z' and it is used in an integral. See for instance the Encyclopedia of mathematics
(https://www.encyclopediaofmath.org/index.php/Kernel_of_an_integral_operator).

By checking the literature we find that the expression "kernel" is rarely used and that "weighting function" is more usual. We thus replace "kernel" by "function" or "weighting function".

The channel 12 spectral response functions are shown in the revised paper as Figure 1.

2.26 Old Figure 8 (now 9)

The bottom of the label "number of events" is off the figure.

We have changed as well to a single-hue colour bar and have corrected the size of the colour bar such that its label fits completely into the plot.

2.27 Textual

All changes done as suggested.

References

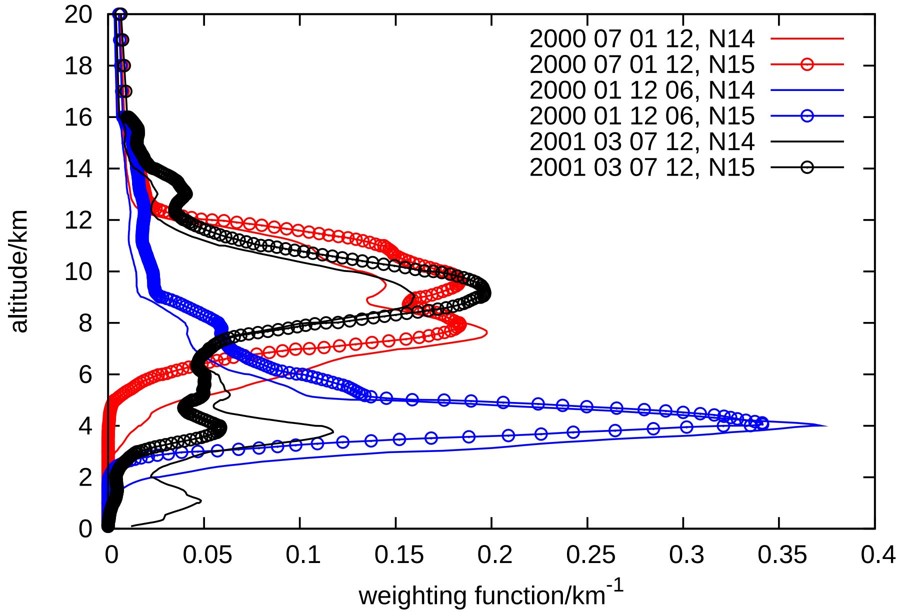
Abramowitz, M. and Stegun, I.: Handbook of mathematical functions, Dover, 9th edn., 1972.

Chandrasekhar, S.: Radiative Transfer, Dover Publications, Inc., 1960.

Neis, P.: Water vapour in the UTLS — climatologies and transport, Ph.D. thesis, Johannes Gutenberg-Universität Mainz, Germany, 2017.

Newell, R.E., Thouret, V., Cho, J. Y., Stoller, P.: Ubiquity of quasi-horizontal layers in the troposphere. Nature 398, 316-319, 1999.

von Storch, H. and Zwiers, F.: Statistical analysis in climate research, Cambridge University Press, Cambridge, UK, 1. edn., 2001.



Actual channel 12 weighting functions for NOAA 14 (simple lines) and NOAA 15 (lines with circles) for three radiosonde profiles from Lindenberg (dates given).

Intercalibration between HIRS/2 and HIRS/3 channel 12 based on physical considerations

Klaus Gierens[1], Kostas Eleftheratos[2], and Robert Sausen[1]

[1]Deutsches Zentrum für Luft- und Raumfahrt, Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany

[2]Department of Geology and Geoenvironment, National and Kapodistrian University of Athens, Athens, Greece

Correspondence to: Klaus Gierens (klaus.gierens@dlr.de)

Abstract. HIRS brightness temperatures at channel 12 (T_{12}) can be used to assess the water vapour content of the upper troposphere. The transition from HIRS/2 to HIRS/3 in 1999 involved a shift in the central wavelength of channel 12 from $6.7\,\mu\mathrm{m}$ to $6.5\,\mu\mathrm{m}$ causing a discontinuity in the time series of T_{12}. To understand the impact of this change in the measured brightness temperatures, we have performed radiative transfer calculations for channel 12 of HIRS/2 and HIRS/3 instruments, using a large set of radiosonde profiles of temperature and relative humidity from three different sites. For each radiosonde profile we performed two radiative transfer calculations, one using the HIRS/2 channel response function of NOAA 14 and one using the HIRS/3 channel response function of NOAA 15, resulting in negative differences of T_{12} (denoted as $\Delta T_{12} := T_{12/15} - T_{12/14}$) ranging between $-12\,\mathrm{K}$ and $-2\,\mathrm{K}$. Inspection of individual profiles for large, medium and small values of ΔT_{12} pointed to the role of the middle tropospheric humidity. This guided us to investigate the relation between ΔT_{12} and the channel 11 brightness temperatures which are typically used to detect signals from the middle troposphere. This allowed us to construct a correction for the HIRS/3 T_{12}, a correction that leads to a pseudochannel 12 brightness temperature such as if a HIRS/2 instrument had measured it. By applying this correction we find an excellent agreement between the original HIRS/2 T_{12} and the HIRS/3 data inferred from the correction method with $R = 0.986$. Upper tropospheric humidity (UTH) derived from the pseudoHIRS/2 T_{12} data compared well with that calculated from intersatellite-calibrated data, providing independent justification for using the two intercalibrated time series (HIRS/2 and HIRS/3) as a continuous HIRS time series for long term UTH analyses.

1 Introduction

Climate variability studies require the analysis of long homogeneous time series of climate data. E.g., a long time series to study the variability of upper–tropospheric water vapour can be derived from the brightness temperature measurements of the High–Resolution Infrared Radiation Sounder (HIRS) instrument aboard the National Oceanic and Atmospheric Administration (NOAA) polar orbiting satellites. The HIRS measurements started in mid–1979 and are still ongoing. They provide a unique long–term data set (covering nearly 4 decades) that can be exploited in climate research. When NOAA launched the weather satellite NOAA 15 in 1998, it was equipped like all its precursors with a HIRS instrument. This 20–channel instrument provides information of temperature and humidity in the troposphere, where channels 10 to 12 are sensitive to water vapour in different altitude bands (lower to upper troposphere, Soden and Bretherton, 1996). Unfortunately, with the launch of NOAA 15 the central frequency in channel 12 has moved from 6.7 μm to 6.5 μm. This is quite a large change, because it means that the channel has its maximum sensitivity about 1 km higher (and accordingly several degrees colder) than channel 12 of all previous satellites.

With that change, i.e. the transition from HIRS/2 on the older NOAA satellites to HIRS/3 on NOAA 15, the channel 12 time series became inhomogeneous. Shi and Bates (2011) performed an intercalibration, based on statistics of differences between the brightness temperatures measured by subsequent HIRS instruments (technically a regression of first kind). The intercalibration solved the problem of a broken time series for some of the statistics of the data, e.g. for the mean values. The intercalibrated time series was used for several studies (e.g. Gierens et al., 2014; Chung et al., 2016). Yet problems remained in the lower tail of the distribution of brightness temperatures, that is, at the lowest values of brightness temperatures, as has been detected by Gierens and Eleftheratos (2017, in the following cited as GE17).

However, the question arises whether it is sufficient to solve a physical problem (i.e. the different altitudes of peak sensitivity of the channel 12 on HIRS/2 and HIRS/3) with a purely statistical method. Hence, GE17 posed the following question:

> Is it justified at all to combine all HIRS T_{12} (the brightness temperature measured by channel 12) data into a single time series when it is a matter of fact that HIRS 2 and HIRS 3/4 sense different layers of the upper troposphere, layers that overlap heavily but whose centres are more than one kilometre apart vertically?

In fact, this question can be broken down into sub–questions: (1) Under which circumstances is the Shi and Bates (2011) intercalibration justified or not? (2) Which assumptions have to be made about the structure of temperature and moisture profiles? The present paper deals with such questions. Fortunately it turns out that it is possible and justified to combine the channel 12 time series on physical reasoning providing a homogeneous time series of 35+ years that can be used for climatological studies. In this paper we demonstrate that independent tests, based on results from radiative transfer calculations, lead to a comparison between NOAA 14 and NOAA 15 channel 12 brightness temperatures that is very similar to the same comparison performed with the intercalibrated data from Shi and Bates (2011). In other words, our new physics based procedure corroborates the statistically based procedure of Shi and Bates (2011) and this is good news.

The present paper is organised as follows. First, the radiative transfer model and its setup is introduced in section 2. Section 3 presents radiative transfer calculations for channel 12 on NOAA 14 and NOAA 15, using radiosonde profiles with high vertical resolution. From these calculations we find that certain profile characteristics in the middle troposphere yield either relatively small or relatively large differences between the computed channel 12 brightness temperatures. In section 4, HIRS channel 11 radiative transfer calculations are applied to get one piece of information more on these profile characteristics. It turns out that the channel 12 brightness temperature differences are linearly correlated with the channel 11 brightness temperatures. A bilinear regression is performed resulting in a superposition of HIRS/3 channel 11 and 12 brightness temperatures from NOAA 15 that produces a pseudo–channel 12 brightness temperature *as if* it was measured by the HIRS/2 instrument on

NOAA 14. A discussion of the method and an application to real HIRS data from NOAA 14 and NOAA 15 are presented in section 5, where we show that the comparison of the original NOAA 14 channel 12 brightness temperature with the pseudo–channel 12 brightness temperature from NOAA 15 is quite similar in its statistical properties to a corresponding comparison using the intercalibrated data. The concluding section 6 summarises the logic of the procedure and gives an outlook.

2 Radiative transfer simulations of channel 12 radiation for HIRS/2 and HIRS/3

In order to analyse the differences between channels 12 of HIRS/2 on NOAA 14 and of HIRS/3 on NOAA 15, respectively, we perform radiative transfer calculations using the channel 12 spectral response functions of the two instruments applied to a large set of atmospheric profiles of temperature and relative humidity. These functions are shown in figure 1. In particular, for each profile we perform two runs of the LibRadtran radiative transfer code (Emde et al., 2016), one for channel 12 on NOAA 14 and one for channel 12 on NOAA 15, i.e., we calculate the channel 12 brightness temperatures $T_{12/15}$ and $T_{12/14}$, which would have been measured by NOAA 15 and NOAA 14, respectively. We then calculate the brightness temperature differences $\Delta T_{12} := T_{12/15} - T_{12/14}$ and analyse how a given difference depends on the given profile characteristics. The channel spectral response functions have been obtained from EUMETSAT's NWP SAF [1].

LibRadtran is used with the following set-up: We use the DISORT radiative transfer solver (Stamnes et al., 1988) with 16 discrete angles and the representative wavelengths band parameterization (reptran, Gasteiger et al., 2014) with fine resolution ($1\,\mathrm{cm}^{-1}$). We assume a ground albedo of zero, **Text deleted** and cloud–free scenes (as brightness temperatures from the Shi and Bates data set are cloud cleared, see Shi and Bates, 2011). The background profiles of the absorbing gases are taken from implemented standard atmosphere profiles (Anderson et al., 1986), whereby the appropriate profile is automatically selected from the geographical position and the time to which the radiosonde profile refers. We calculate the channel–integrated brightness temperatures at top of the atmosphere for nadir and 30°off–nadir directions for these profiles.

The atmospheric profiles of temperature and relative humidity (with respect to liquid water) are taken from large sets of radiosonde data with high vertical resolution: (1) We use the set of profiles from the German weather observatory Lindenberg (52.21° N, 14.12° E, Spichtinger et al., 2003) similar to earlier satellite studies (Gierens et al., 2004; Gierens and Eleftheratos, 2016). We use this set of more than 1500 profiles to derive a regression based solution (our training data set). (2) To see whether there are systematic differences between latitude zones also radiosonde profiles from Sodankylä, Finland (67.37° N, 26.60° E), and Manus, Papua New–Guinea (2.06° S, 146.93° E) are used. These weather observatories belong to the Global Climate Observing System (GCOS) Reference Upper-Air Network (GRUAN). The data and products of the GRUAN network are quality-controlled as described by Immler et al. (2010); Dirksen et al. (2014). We use one year of profiles from both stations, 2013 for Manus and 2014 for Sodankylä. These profiles were used for testing the regression that we derived from the Lindenberg profiles. **Text deleted** The GRUAN profiles have a very high vertical resolution, too high for the radiative transfer

[1]Satellite Application Facility, SAF, for numerical weather prediction, NWP, https://nwpsaf.eu/site/software/rttov/download/coefficients/spectral-response-functions, last access in March 2017

calculation. Thus, only every 10th record has been used from the surface to 90 hPa. At higher altitudes (mainly in the dry stratosphere) we have replaced the radiosonde data by data from the standard atmospheres implemented in LibRadtran.

3 Discussion of radiative transfer results

Figure 2 displays the pseudo channel 12 brightness temperatures for $T_{12/15}$ against the corresponding brightness temperature differences (NOAA 15 minus NOAA 14), ΔT_{12}, computed with LibRadtran for the Sodankylä and the Manus profiles. $T_{12/15}$ and ΔT_{12} are presented for nadir and 30° off–nadir directions. As expected, the brightness temperatures for the two considered viewing directions differ, and their difference is rather constantly about 1 to 2 K. More precisely, the summary statistics for the two locations are: At Sodankylä the mean difference at nadir is -6.7 ± 1.2 K, and the mean difference at 30° is -6.7 ± 1.3 K. At Manus the mean difference at both nadir and at 30° are -7.1 ± 0.9 K. It is thus sufficient to only use the nadir radiances for further analyses. It can be noted that $T_{12/15}$ varies between 225 and 242 K for the Sodankylä profiles, while the corresponding ΔT_{12} ranges between -12 and -3 K. There is no obvious correlation between ΔT_{12} and $T_{12/15}$. For Manus, the data pairs show values of $T_{12/15}$ from roughly 229 to 241 K and brightness temperature differences ranging from -10 to -5 K. Again there is no obvious correlation between the brightness temperatures themselves and the corresponding differences.

Figure 3 (top) displays the corresponding results for the radiosonde profiles from Lindenberg. The data pairs form two groups: a large patch at low $T_{12/15}$ and a "tail" at small ΔT_{12}, but higher $T_{12/15}$. This tail has been discarded from further analysis since inspection of the corresponding profiles showed that the relative humidity sensor was obviously malfunctioning in the middle and upper troposphere (and in the stratosphere), indicating zero relative humidity. 1558 profiles out of the total 1660 profiles remain for the analysis. The bottom part of figure 3 shows the same data, without the mentioned tail, and represented as a 2-d histogram. The data at the maximum frequency (red colours) have a brightness temperature difference of about -7 K. Only a small set of the data pairs have $\Delta T_{12} > -5$ K and an even smaller set has $\Delta T_{12} < -11$ K.

At this point it is useful to recall that the weighting functions of the two considered channels peak in altitudes about 1 km apart because the water vapour optical thickness is larger at the central frequency of channel 12 on HIRS/3 than that on HIRS/2. The vertical distance of 1 km implies an air temperature difference of about 6.5 K on average in the upper troposphere, and this explains that an average ΔT_{12} of about the same value is found in the radiative transfer calculations. A similar (average) correction of 8 K has been derived by Shi and Bates (2011) and used by Chung et al. (2016).

Now the question arises how characteristics of humidity profiles are reflected in the brightness temperature differences. Figure 4 shows three sets of relative humidity profiles: 5 profiles with $\Delta T_{12} < -11$ K (left panel), 6 profiles with $-7.21\,\mathrm{K} < \Delta T_{12} < -7.19$ K (middle), and 20 profiles with a small difference, $\Delta T_{12} > -5$ K (right).

The first set of profiles with $\Delta T_{12} < -11$ K is characterised by high values of RH in the upper troposphere (200 to 400 hPa) and a very dry middle troposphere (450 to 650 hPa). Accordingly, channel 12 on NOAA 15 (ch. 12/15) gets more radiance from the upper levels than channel 12 on NOAA 14 (ch. 12/14) because it is more sensitive there. In turn, ch. 12/14 cannot balance this deficit in the middle tropospheric levels since it is too dry at this altitude. The result is a large negative difference of brightness temperatures. The profiles with $\Delta T_{12} > -5$ K are in turn characterised by a middle troposphere that has much higher relative humidity than the upper troposphere. Under such a circumstance the peak of the ch. 12/15 weighting function approaches the peak of the ch. 12/14 weighting function, that is, the brightness temperatures become more similar. Finally, an average brightness temperature difference is found for profiles without strong humidity contrast between the upper and the middle tropospheric levels, as shown in the middle panel of Figure 4.

This analysis shows that one can understand from consideration of the underlying radiation physics why the brightness temperature differences sometimes obtain large and sometimes relatively small values, and why the average difference is of the order $-7\,\mathrm{K}$. It is, however, clear that this additional knowledge is not available when satellite data analysis is confined to channel 12 only. To exploit this knowledge one need further pieces of information, in particular on the humidity in the mentioned middle tropospheric levels. Fortunately, this knowledge is available from the same HIRS instruments, from channel 11 (see, e.g., Soden and Bretherton, 1996).

4 Construction of a pseudo HIRS/2 channel 12

4.1 Regression using HIRS/3 channels 11 and 12

HIRS/3 channel 11 is centred at a wavelength of $7.3\,\mu\mathrm{m}$. While the strong water vapour ν_2 vibration–rotation band has its peak line strenghts at about the channel 12 wavelength ($\approx 6.5\,\mu\mathrm{m}$), channel 11 is centred on the longwave side of this band, off the peak with lower line strengths, and thus channel 11 is characteristic of the water vapour in lower levels than channel 12. In a standard midlatitude summer atmosphere channel 11 peaks at about $5\,\mathrm{km}$ altitude (see figure 2 of Gierens and Eleftheratos, 2016).

Using the channel spectral response function for channel 11 on NOAA 15, radiative transfer calculations have been performed for the radiosonde profiles used above. Figure 5 shows for the set of Lindenberg profiles the resulting brightness temperatures, $T_{11/15}$, plotted against the previously computed ΔT_{12}. As expected, $T_{11/15}$ is generally higher than the channel 12 brightness temperatures because it characterises the temperature in the middle troposphere where the channel 11 weighting function peaks. $T_{11/15}$ ranges from 248 to $268\,\mathrm{K}$ for the Lindenberg profiles. Figure 5 also shows a linear correlation between ΔT_{12} and $T_{11/15}$, although with a large scatter. The linear Pearson correlation coefficient is -0.68. Its square is 0.46, that is, variations of $T_{11/15}$ represent almost half of the variations in ΔT_{12}. The remaining scatter is not surprising given the tremendous variability of relative humidity profiles. One additional piece of information is clearly insufficient to capture all this variability. Nevertheless, the correlation is clearly visible. We have made use of it to construct a correction to the HIRS/3 measured channel 12 brightness temperatures, a correction that leads to a pseudo–channel 12 brightness temperature such *as if* a HIRS/2 instrument had measured it.

For that purpose we try a bilinear regression [2] of the following kind:

$$\hat{T}_{12/15} = a + bT_{12/15} + cT_{11/15}. \quad (1)$$

[2] The regression has been performed using IDL (Interactive Data Language) routine `REGRESS`.

Here, $\hat{T}_{12/15}$ is the desired pseudo–channel 12 brightness temperature that is equivalent to a HIRS/2 measurement, in other words the $T_{12/15}$ as it would have been $T_{12/14}$. For the calculation of $T_{12/15}$ only the nadir brightness temperatures have been retained as it seems that the off-nadir directions do not yield differing information. The two data vectors containing the brightness temperatures of channels 11 an 12 are linearly correlated with $R = 0.71$, but they point in different directions, that is, they are not co-linear. Regression thus yields a unique result, namely

$$a = -35.40\,\mathrm{K}, \quad b = 0.78, \quad c = 0.37. \tag{2}$$

The one σ uncertainty estimates of the parameters b, c are both ± 0.01. The corresponding data pairs are shown in figure 6. Slope and intercept of the regression (black line in the figure) are 1.000 and 2×10^{-4}, respectively, and the linear correlation between the linear superposition of channel 11 and 12 brightness temperatures and that of the pseudo–channel 12 brightness temperature is 0.986. Mean and standard deviations of the residuals are $-0.8 \pm 0.5\,\mathrm{K}$.

4.2 Test with independent radiosonde profiles

Using the linear superposition of channel 11 and 12 brightness temperatures for the considered atmospheric profiles from the two GRUAN stations, Sodankylä and Manus, leads to the data pairs shown in figure 7. The black diagonal in this figure is not the result of a best fit or a regression, it is $y = x$, plotted to guide the eye in checking the result. The residual means ($T_{12/14} - \hat{T}_{12/15}$) are $-0.5 \pm 1.1\,\mathrm{K}$ for Sodankylä and $-1.2 \pm 0.4\,\mathrm{K}$ for Manus. These residuals are much smaller than the original differences between $T_{12/14}$ and $T_{12/15}$ shown in Figure 2. Obviously, the superposition methods works well for these data representing a polar and an equatorial atmosphere, respectively.

5 Discussion

5.1 Superposition of weighting functions

The superposition of channels 11 and 12 is equivalent to a superposition of their weighting functions. Figure 8 gives an example. The weighting functions are generic functions as in Gierens and Eleftheratos (2016), assuming a water vapour scale height of $2\,\mathrm{km}$ and peak altitudes of $8.5\,\mathrm{km}$ for ch. 12/15 (red curve), $7.5\,\mathrm{km}$ for ch. 12/14 (black), and $5\,\mathrm{km}$ for ch. 11/15 (blue). The black curve with circles represents the superposition of channels 11 and 12 on NOAA 15 with the weights b and c derived above. The superposition curve (its upper tail, its peak, and about half of its lower tail) is between the corresponding channel 12 weighting functions. We note here that the superposition weighting function. has some weight at lower altitudes where both channel 12 weighting functions. are already very low. Overall, we see that the superposition method eventually brings the pseudochannel 12 brightness temperature of NOAA 15 closer to the level of the corresponding channel 12 brightness temperature of NOAA 14.

Figure 8 (and actual weighting functions calculated for the reply to the reviewers) shows that there is some possibility that channel 11 sees the ground when the atmosphere is quite dry. In such cases, which might occur at high latitudes, the superposition will not work. High brightness temperatures in both channels 10 and 11 could indicate such an event. **Text**

deleted Indeed, the (high-latitude) Sodankylä data show larger scatter in fig. 7 than the (equatorial) data from Manus, which might result from unwanted ground influence at the high-latitude station.

An interesting alternative interpretation of the coefficients resulting from the bilinear regression may derive from the following consideration: It is possible to rewrite Eq. 1 as a weighted mean of three temperatures:

$$\begin{aligned} \hat{T}_{12/15} &= a' T_0 + b T_{12/15} + c T_{11/15}, \qquad \text{with} \\ & a' + b + c = 1. \end{aligned} \tag{3}$$

From this interpretation and Eq. 2 follows $a' = -0.15$ and it turns out that $T_0 = 236\,\mathrm{K}$, which is remarkably close to 240 K, the T_0 used as a reference in the retrieval schemes developed by Soden and Bretherton (1993); Stephens et al. (1996); Jackson and Bates (2001). At the altitude where the channel 12 weighting function. peaks the temperature is, on average, close to T_0. The remarkable fact is that the regression results just in this T_0 for the constant part, not anything else, a finding that could not be expected *a priori*.

5.2 Application to real data

For the same set of 1004 days of common operation of NOAA 14 and NOAA 15 as used in GE17, we have compared the channel 12 brightness temperatures, daily averages on a $2.5° \times 2.5°$ grid in the northern midlatitudes, 30° N to 70° N. Differing from the previous paper, we use the original non–intercalibrated brightness temperatures. For NOAA 15 we compute the linear superposition derived above, that is, $\hat{T}_{12,15}$, while for NOAA 14 we use $T_{12,14}$. The 2-d histogram of these data pairs is shown in figure 9. It is remarkable how similar this histogram is to a corresponding one shown as figure 2 in GE17 which displays the intercalibrated data. The ordinary least squares linear fit through the new data pairs (solid line) has the equation:

$$(y/\mathrm{K}) = 47.72 + 0.8025\,(x/\mathrm{K}), \tag{4}$$

with a slightly smaller slope and a slightly larger intercept than in GE17 using the intercalibrated data pairs (0.8290 and 41.63, respectively). The bivariate regression straight line (dash-dotted) has the equation:

$$(y/\mathrm{K}) = 18.24 + 0.9256\,(x/\mathrm{K}), \tag{5}$$

and this has as well a slightly smaller slope and larger intercept than the corresponding fit through the intercalibrated data (which has 0.994 and 2.007, respectively). These regression lines and their coefficients just serve for comparison with the intercalibrated data pairs and this comparison shows remarkably that two essentially different methods to treat the HIRS 2 to HIRS 3 transition lead to very similar results.

In pursuit of the goal to study changes of upper tropospheric humidity with respect to ice (UTHi) we applied the retrieval formula of Jackson and Bates (2001) to $\hat{T}_{12,15}$ and to $T_{12,14}$ of the common 1004 days. A density plot of the corresponding data pairs of UTHi is displayed in figure 10. Obviously the result is not satisfying; the plot resembles closely the corresponding scatter of data pairs produced from the intercalibrated data (Shi and Bates, 2011) that has been shown in figure 1 of GE17. Unfortunately the superposition method does not solve the problem of a considerable overestimation of the number of supersaturation events recorded with HIRS 3 and 4 instruments and it seems that the pseudo–channel 12 data have to be treated with the cdf–matching technique developed by GE17 in the same way as the intercalibrated data. This is beyond the scope of the present paper.

The new method is an independent approach to an intercalibrated HIRS channel 12 data set, based on results of radiative transfer calculations, classification of profile characteristics and a superposition with information delivered by channel 11. The intercalibration of Shi and Bates (2011) is instead based on pixelwise direct corrections, where the brightness temperature dependent corrections are determined from regressions of the first kind between subsequent satellite pairs. As figures 9,10 show, both methods seem to produce very similar results. **Original text deleted and replaced by:** The statistically based method of Shi and Bates (2011) is thus supported by an independent method, and results obtained from data intercalibrated with either method should be more trustworthy. We thus consider our question from the beginning, whether combining HIRS 2 and HIRS 3 data into a single time series is justified, answered positively. Although both methods produce similar results as we see in fig. 10, neither way of intercalibration solves the problems with the discrepancy in the range of high UTHi values which results from a corresponding discrepancy at the low tail of channel 12 brightness temperatures (see GE17). It is probable that this problem does not originate from the intercalibration procedure, since for the radiative transfer calculation it makes no difference whether the humidity profile contains a very humid upper troposphere with supersaturated layers or not. In each case it provides the corresponding brightness temperature. It is more probable that the problem with the lower tail of the T_{12}-distribution comes from the retrieval method which is based on linearisations around certain "tangential points", thermodynamic properties typical of the upper troposphere (e.g. the $T_0 = 240\,\mathrm{K}$ mentioned above), and that this linear approach is not completely sufficient in cases where actual properties are too far away from the tangential points.

6 Conclusions

The procedure we have developed in the present paper proceeds along the following steps:

1. The difference, $T_{12/15} - T_{12/14} =: \Delta T_{12}$, calculated with LibRadtran for a set of radiosonde profiles, ranges from -12 to $-4\,\mathrm{K}$, with most cases around $-7\,\mathrm{K}$, which fits to the approximately $1\,\mathrm{km}$ altitude difference between the peaks of the channel 12 weighting functions of HIRS/2 and HIRS/3.

2. It turns out that the shape of the RH profile determines whether ΔT_{12} is close to one of the extremes or close to the average. It is particularly the shape of the humidity profile in the lower to middle troposphere that plays a role here.

3. Take channel 11 brightness temperatures as a proxy of that part of the profile, as that channel measures the humidity in the lower to middle troposphere.

4. Indeed, and fortunately, $T_{11/15}$ is correlated to ΔT_{12}; thus it can be used to identify in which cases ΔT_{12} is large, average, or small.

5. Thus it is possible to find a correction to $T_{12/15}$ such that the result is close to (with a mean residual difference of about 1 K) the brightness temperature that N14 would have measured if it had seen the same scene. This correction is a linear superposition of $T_{12/15}$ and $T_{11/15}$, measured by the same HIRS instrument.

Application of this superposition method to real data of 1004 common days of operation of NOAA 14 and NOAA 15, comparing $T_{12/14}$ with the pseudochannel 12 brightness temperature of NOAA 15, $\hat{T}_{12/15}$, yields a 2-d distribution that is very similar to the corresponding distribution obtained with the intercalibrated brightness temperatures by Shi and Bates (2011). Comparing the corresponding values of UTHi again yields a 2-d distribution very similar to that obtained from the intercalibrated data. From these findings we conclude that our method, which is based on radiative transfer calculations, i.e. physics, produces very similar results with Shi and Bates' statistical intercalibration method. The justification to use the intercalibrated channel 12 time series including its early HIRS 2 and later HIRS 3 and 4 phases is thus corroborated. **Old text deleted.**

Note, that this paper shows only the principle of method, how a pseudo HIRS/2 channel 12 brightness temperature can be computed from later HIRS versions, involving channels 11 and 12. As all HIRS instruments have slightly different channel spectral response functions, the regression parameters (a, b, c) will differ from one instrument pair to the other. They will also depend on which HIRS/2 instrument serves as reference. In this paper we used HIRS/2 on NOAA 14, but it certainly makes sense to additionally use HIRS/2 on NOAA 12 as Shi and Bates (2011) based their intercalibration on that satellite. This work is beyond the scope of the current paper and left for future exercise. We note also that this analysis represents a relatively short period in the lifetime of two HIRS instruments, and that their spectral and radiometric calibration was assumed to be constant over this period.

7 Code availability

The libRadtran radiative transfer software package is freely available under the GNU General Public License from http://www.libradtran.org/d

8 Data availability

The GRUAN radiosonde data are available from the GRUAN websites. The special Lindenberg radiosonde data set is available from the first author on request. The NOAA satellite data are available from NOAA public websites.

Author contributions. KG made the radiative transfer calculations and the analyses. KG and RS discussed the procedures and the statistical methods. KE prepared the satellite data in a useful form. All authors contributed to the text.

Competing interests. The authors declare no competing interests.

Acknowledgements. The authors thank the LibRadtran developer team for providing the radiative transfer code and Luca Bugliaro for checking the first author's setup of the radiative transfer job. We are grateful to all the people who provided the data used in this paper, which are colleagues from NOAA, the GRUAN network, and DWD. Christoph Kiemle read the pre-final version of the manuscript and made good suggestions for improvement and further discussion. Thanks for this!

References

Anderson, G., Clough, Kneizys, F., Chetwynd, J., and Shettle, E.: AFGL atmospheric constituent profiles (0–120 km), Tech. Rep. Tech. Rep. AFGL-TR-86-0110, Air Force Geophys. Lab., Hanscom Air Force Base, Bedford, Mass., 1986.

Chung, E.-S., Soden, B., Huang, X., Shi, L., and John, V.: An assessment of the consistency between satellite measurements of upper tropospheric water vapor, J. Geophys. Res., 121, 2874–2887, doi:10.1002/2015JD024 496, 2016.

Dirksen, R., Sommer, M., Immler, F., Hurst, D., Kivi, R., and Vömel, H.: Reference quality upper–air measurements: GRUAN data processing for the Vaisala RS92 radiosonde, Atmos. Meas. Tech., 7, 4463–4490, 2014.

Emde, C., Buras-Schnel, R., Kylling, A., Mayer, B., Gasteiger, J., Hamann, U., Kylling, J., Richter, B., Pause, C., Dowling, T., and Bugliaro, L.: The libRadtran software package for radiative transfer calculations (version 2.0.1), Geosci. Model Dev., 9, 1647–1672, 2016.

Gasteiger, J., Emde, C., Mayer, B., Buehler, S., and Lemke, O.: Representative wavelengths absorption parameterization applied to satellite channels and spectral bands, J. Quant. Spectrosc. Radiat. Transfer, 148, 99–115, 2014.

Gierens, K. and Eleftheratos, K.: Upper tropospheric humidity changes under constant relative humidity, Atmos. Chem. Phys., 16, 4159–4169, 2016.

Gierens, K. and Eleftheratos, K.: Technical Note: On the intercalibration of HIRS channel 12 brightness temperatures following the transition from HIRS 2 to HIRS 3/4 for ice saturation studies, Atmos. Meas. Tech., 10, 681–693, doi:10.5194/amt-2016-289, 2017.

Gierens, K., Kohlhepp, R., Spichtinger, P., and Schroedter-Homscheidt, M.: Ice supersaturation as seen from TOVS, Atmos. Chem. Phys., 4, 539–547, 2004.

Gierens, K., Eleftheratos, K., and Shi, L.: Technical Note: 30 years of HIRS data of upper tropospheric humidity, Atmos. Chem. Phys., 14, 7533–7541, 2014.

Immler, F., Dykema, J., Gardiner, T., D.N. Whiteman, P. T., and Vömel, H.: Reference quality upper-air measurements: guidance for developing GRUAN data products, Atmos. Meas. Tech., 3, 1217–1231, 2010.

Jackson, D. and Bates, J.: Upper tropospheric humidity algorithm assessment, JGR, 106, 32 259–32 270, 2001.

Shi, L. and Bates, J.: Three decades of intersatellite–calibrated High–Resolution Infrared Radiation Sounder upper tropospheric water vapor, J. Geophys. Res., 116, D04 108, doi:10.1029/2010JD014 847, 2011.

Soden, B. and Bretherton, F.: Upper tropospheric relative humidity from the GOES 6.7 μm channel: Method and climatology for July 1987, J. Geophys. Res., 98, 16 669–16 688, 1993.

Soden, B. and Bretherton, F.: Interpretation of TOVS water vapor radiances in terms of layer–averaged relative humidities: Method and climatology for the upper, middle, and lower troposphere, J. Geophys. Res., 101, 9333–9343, 1996.

Spichtinger, P., Gierens, K., Leiterer, U., and Dier, H.: Ice supersaturation in the tropopause region over Lindenberg, Germany, Meteorol. Z., 12, 143–156, 2003.

Stamnes, K., Tsay, S.-C., Wiscombe, W., and Jayaweera, K.: Numerically stable algorithm for discrete ordinate method radiative transfer in multiple scattering and emitting layered media, Appl. Optics, 27, 2502–2509, 1988.

Stephens, G., Jackson, D., and Wittmeyer, I.: Global observations of upper–tropospheric water vapor derived from TOVS radiance data, J. Climate, 9, 305–326, 1996.

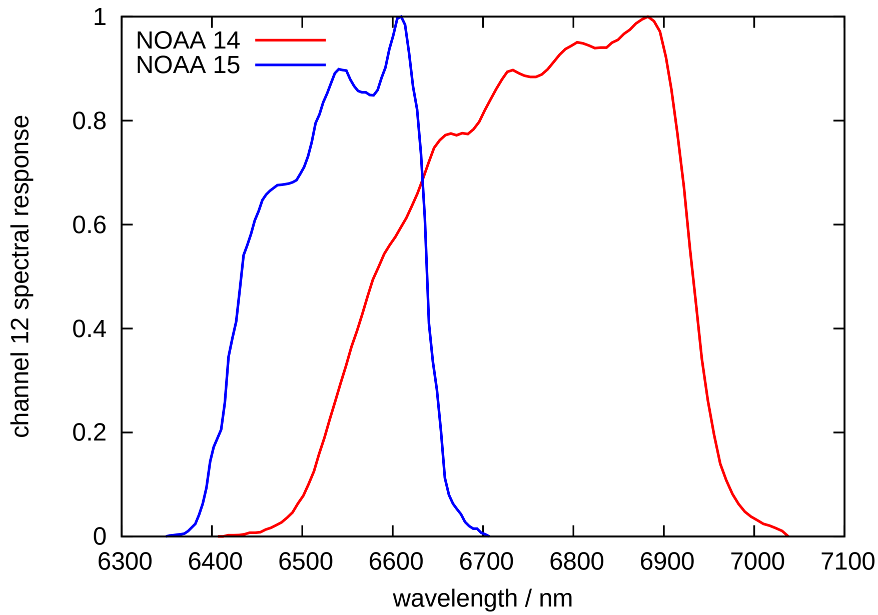


Figure 1. Channel 12 spectral response functions of the HIRS 2 instrument on NOAA 14 and of the HIRS 3 instrument on NOAA 15.

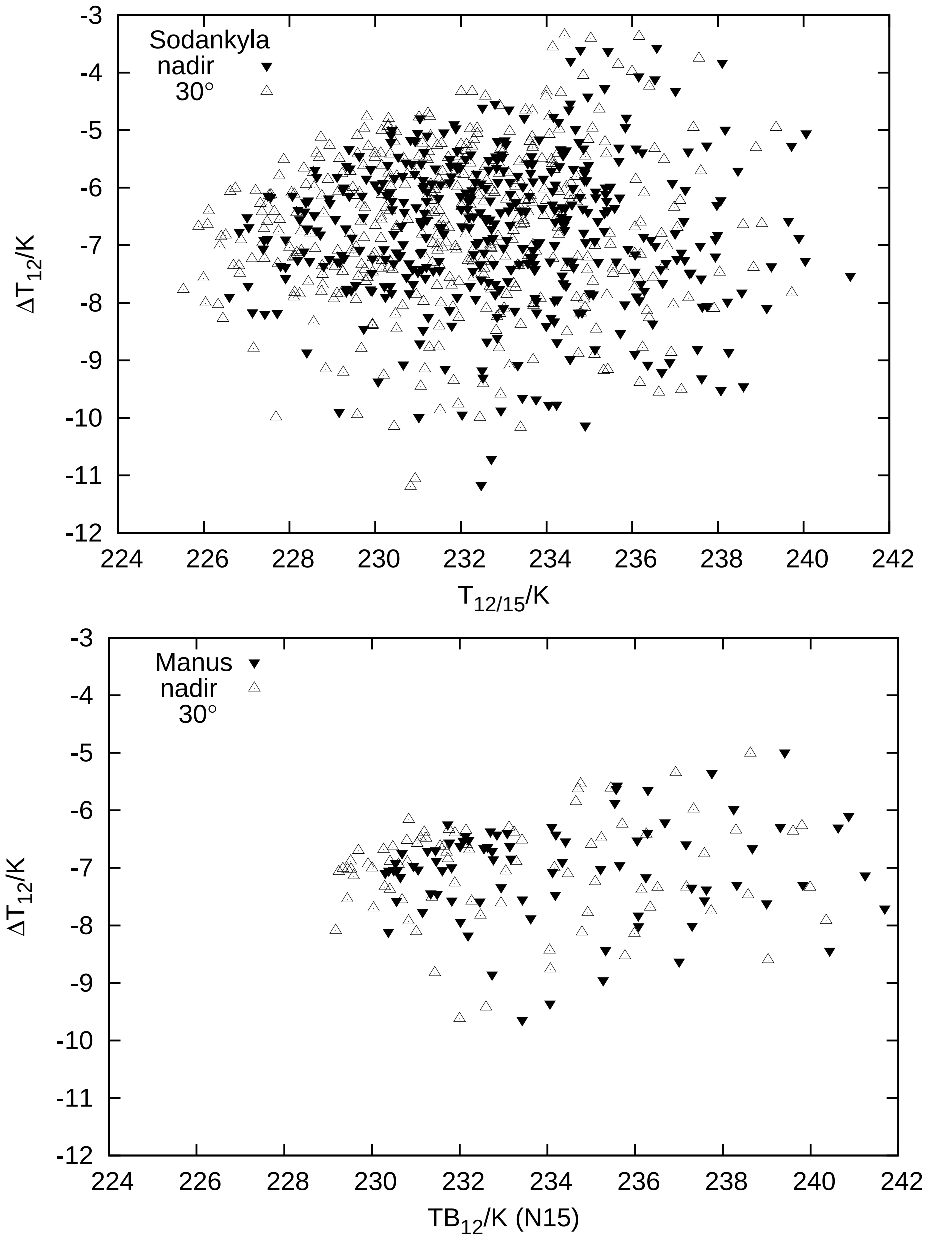


Figure 2. Scatter plot of brightness temperatures calculated with a radiative transfer model using radiosonde profiles from Sodankylä, Finland (top), and Manus, Papua-New Guinea (bottom). The abscissa represents the brightness temperature obtained with a channel 12 spectral response function for HIRS/3 on NOAA 15. The ordinate represents the difference between this brightness temperature and a corresponding one computed using the channel 12 spectral response function for HIRS/2 on NOAA 14. The calculations have been performed for both nadir and 30° off–nadir viewing directions.

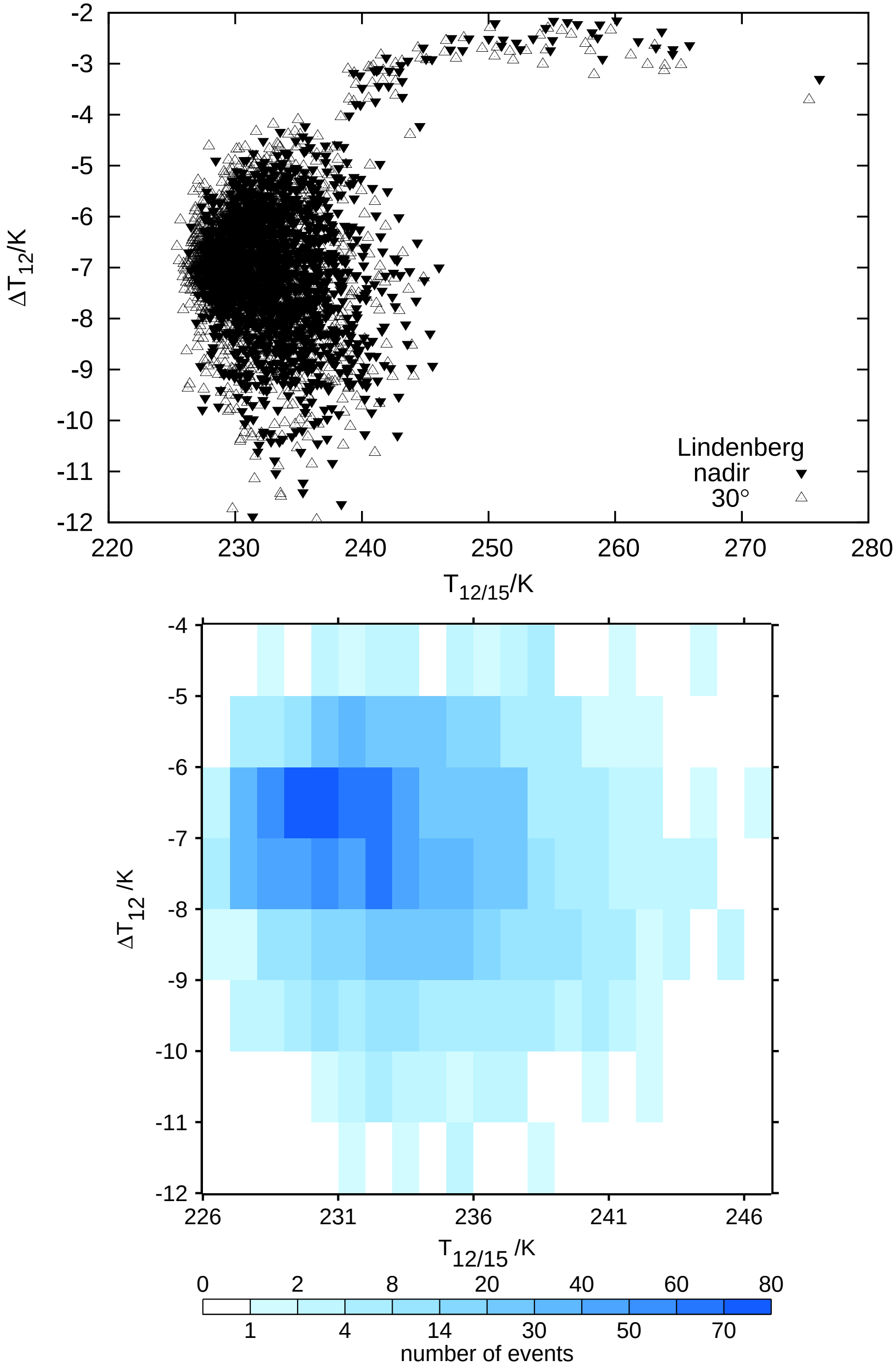


Figure 3. Scatter plot as in figure 2 (top) and a corresponding two–dimensional histogram for channel 12 brightness temperatures computed using radiosonde profiles from Lindenberg, Germany. Note the tail of high values in the scatter plot results from profiles with malfunctioning RH instrument. These 102 profiles have been discarded from further analysis. The 2-d frequency histogram does not contain them anymore. Calculations have been performed for nadir and 30° off-nadir directions, but the off–nadir results are only shown in the scatter plot.

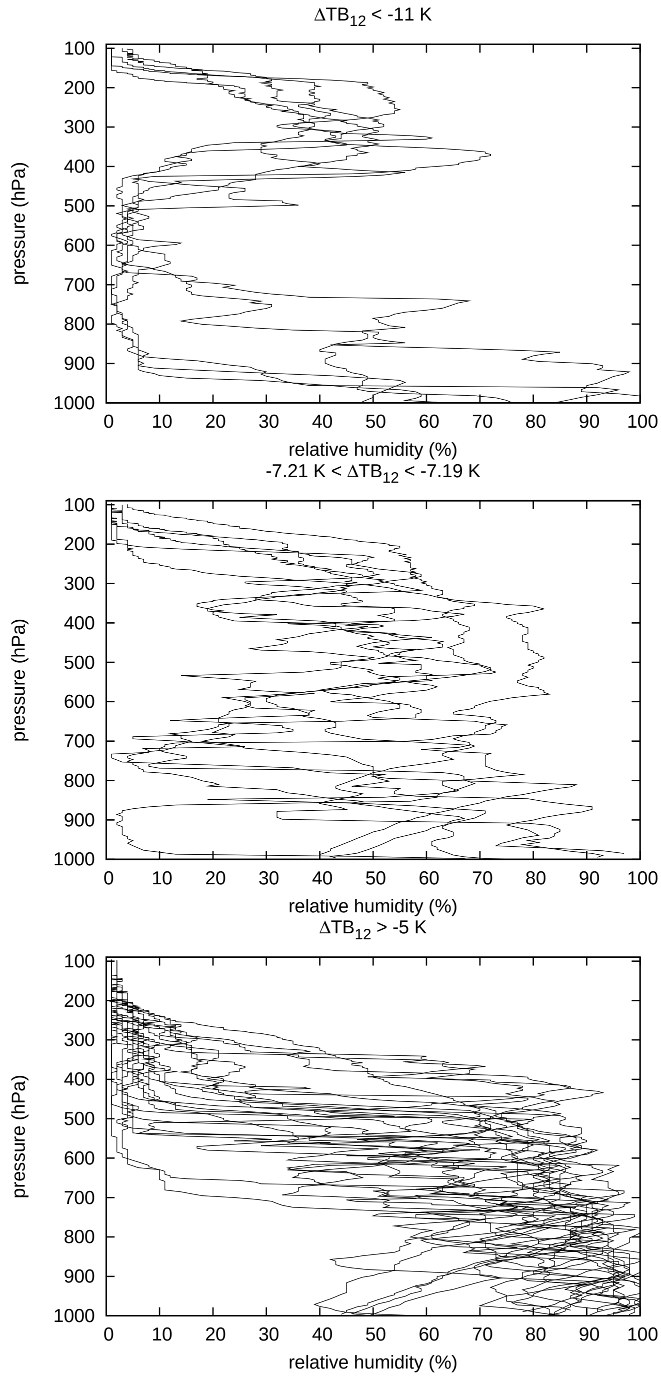


Figure 4. Lindenberg radiosonde profiles of relative humidity vs. pressure altitude that lead to brightness temperature differences in extreme ranges (top and bottom, values indicated in the figures) and to values near to the mean (middle panel). The profiles are obtained from the following launches (format yymmddhh): 00042806, 00122312, 01011506, 01021717, 01030712 (top); 00070112, 00111618, 00112318, 00123012, 01021612, 01040218 (middle); 00021306, 00021312, 00021406, 00022100, 00022106, 00052912, 00053018, 00060706, 00071506, 00080112, 00080200, 00111606, 00121606, 01020118, 01020218, 01022206, 01022218, 01022306, 01022312, 01022400 (bottom).

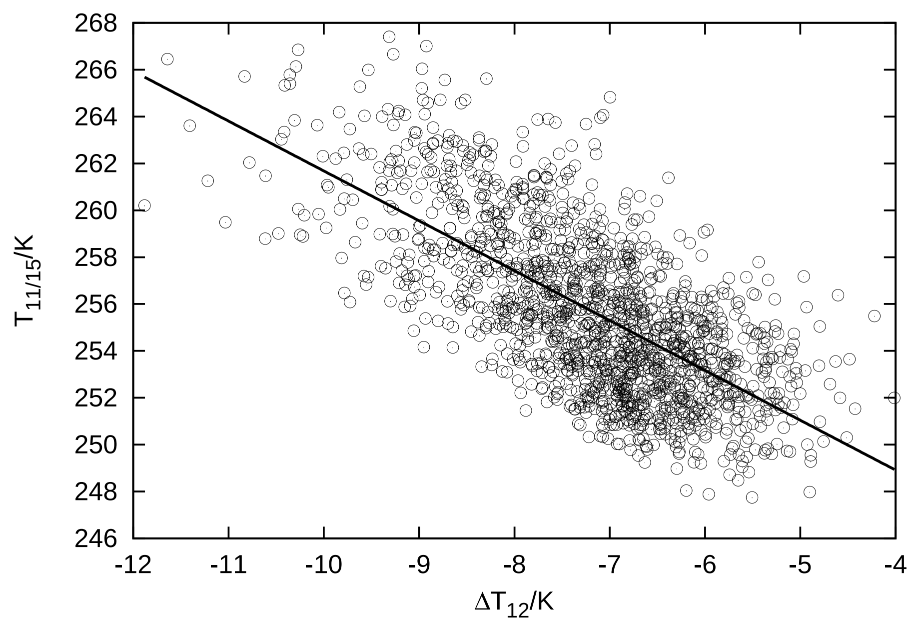


Figure 5. Scatter plot showing a linear correlation between the difference of channel 12 brightness temperatures (NOAA 15 minus NOAA 14) and the NOAA 15 channel 11 brightness temperature computed using the Lindenberg profiles. The linear Pearson correlation coefficient is -0.68.

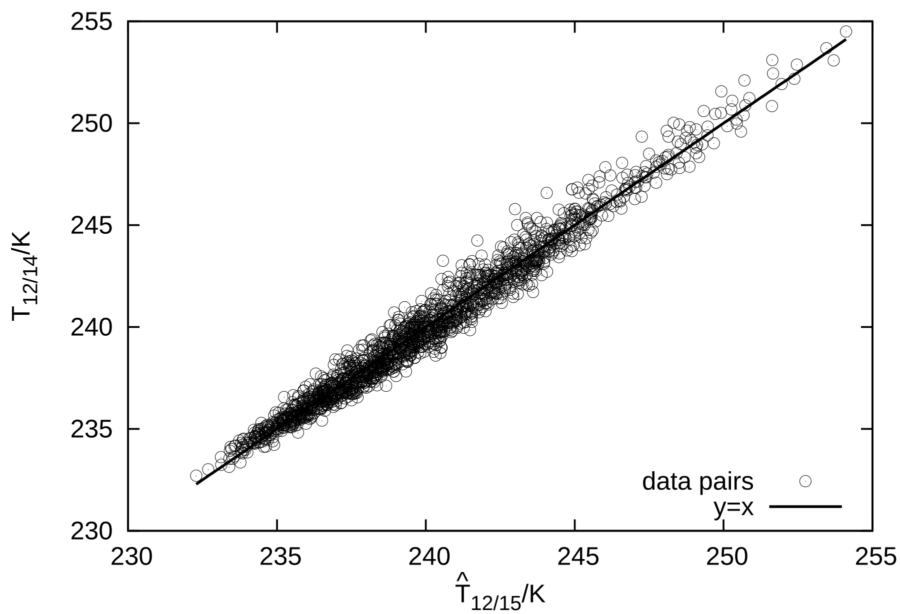


Figure 6. Scatter plot showing a linear correlation between a linear superposition of channel 11 and 12 brightness temperatures from NOAA 15 (abscissa, $\hat{T}_{12/15}$) with the corresponding channel 12 brightness temperature for the same profile but computed with the NOAA 14 channel response function. Note that the fit line has slope 1.000 and the intercept is close to zero (2×10^{-4}). The linear correlation is $R = 0.986$. All data computed using the Lindenberg profiles.

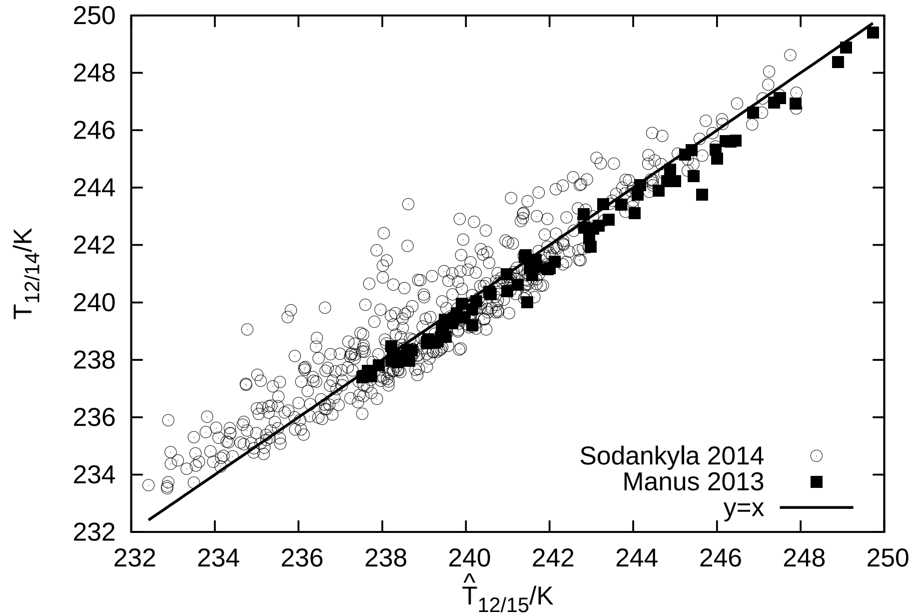


Figure 7. Test of the superposition method using radiosonde profiles from the two GRUAN stations Sodankylä, Finland, and Manus, Papua-New Guinea. The diagonal line ($y = x$) is included to check the result; it is not a fit.

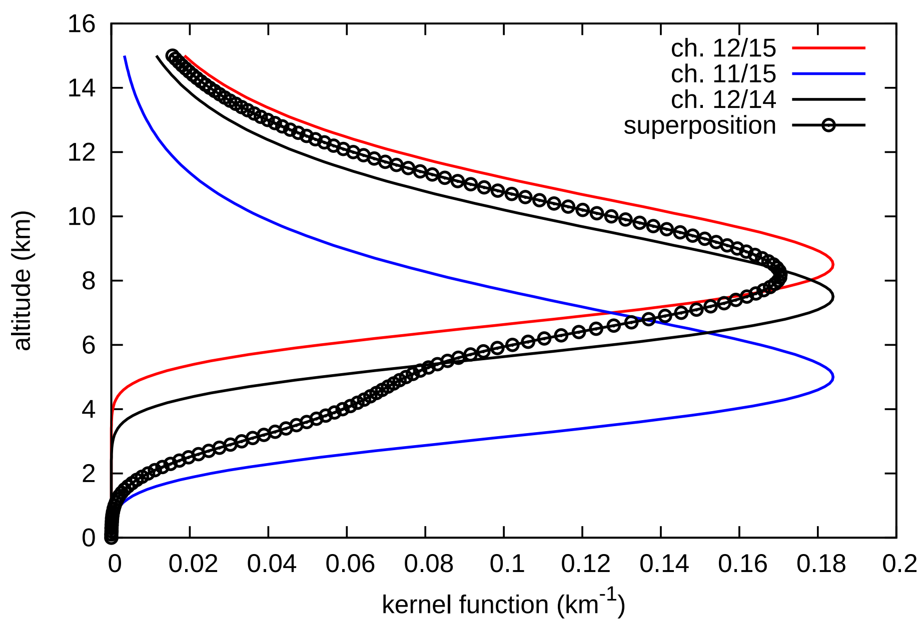


Figure 8. Examples of weighting functions for channels 11 and 12 on NOAA 15 (blue and red), their superposition (black with circles), and channel 12 on NOAA 14 (black).

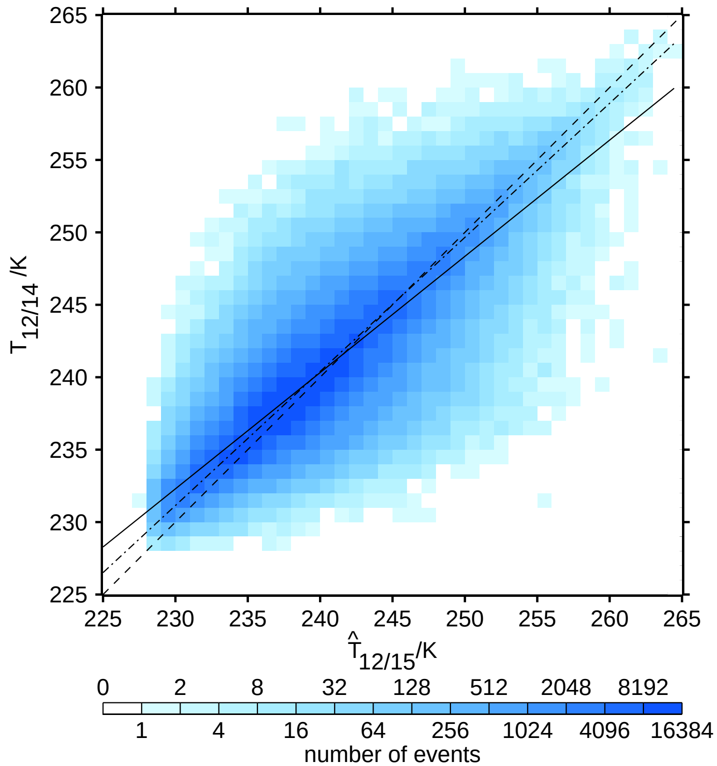


Figure 9. 2–d histogram of brightness temperatures, displaying $\hat{T}_{12/15}$ on the abscissa and $T_{12/14}$ on the ordinate axes, respectively. The data are from 1004 common days of operation of NOAA 14 and NOAA 15. The dashed diagonal line represents $x = y$, the solid line is the best fit according to an ordinary least squares regression and the dashed–dotted line is the bivariate regression line.

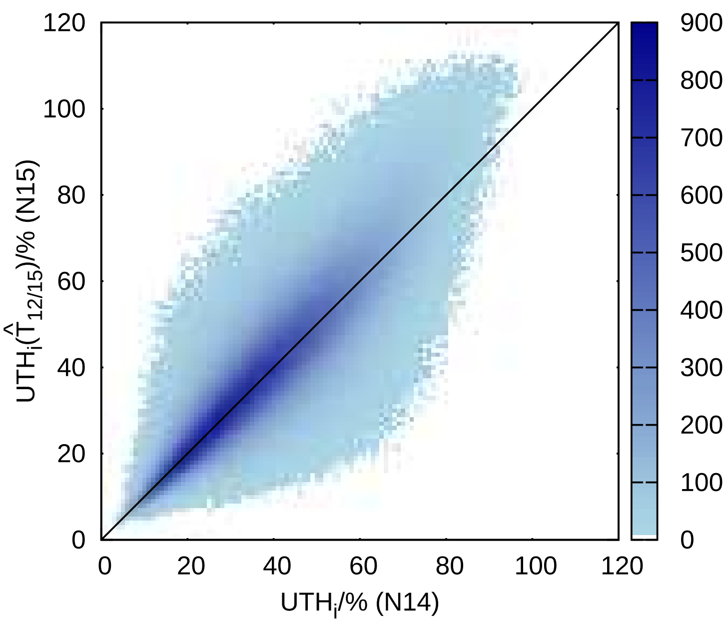


Figure 10. Heat map displaying UTHi computed using $\hat{T}_{12/15}$ on the ordinate against values computed from the original $T_{12/14}$ on the abscissa. Obviously the problem concerning the excess of supersaturation cases in the NOAA 15 data remains even with this new kind of data treatment. The colour scale shows the number of events in each $1\% \times 1\%$ pixel.