

**Reply to the reviewer’s comments on Manuscript No. AMT-2016-289**  
**by K. Gierens and K. Eleftheratos**

## Remarks

We thank all reviewers for their very constructive comments which helped improving the paper. Below you find a detailed response to all questions raised. Reviewer comments are reprinted in italics font, our replies in times roman.

According to the replies presented below we have revised our manuscript. You will find passages with substantial changes marked in red in the revised version.

Here we would like to start with a comment on the significance of this research which was rated fair by two reviewers. We believe that this assessment is wrong. In the revised paper we have added arguments that show why such studies as this one are important.

As we state in the paper, the original intercalibration by Shi and Bates (2011) is quite successful for the bulk of the channel 12 data. That is, if the mean of  $T_{12}$  or of UTHi is relevant, the intercalibration is sufficient. However, the mean means nothing in non-linear processes like radiation and cloud formation. This implies that more than just the first (and perhaps the second) moment of the UTHi distribution is needed, in particular characteristics of the tails of its distribution. It is clear that for cloud research and how cloudiness will change with climate change, information on the upper tail of UTHi is needed. For questions of the radiative balance of the Earth it is important how the very dry regions of the subsidence zones (termed “radiator fins” by Pierrehumbert, 1995; see also Schröder et al. 2014 and Roca et al. 2011) behave with ongoing climate change; thus the dry end of the UTHi distribution is of immense interest as well. These arguments show that homogeneous time-series of the whole UTHi distribution are needed, it is not sufficient that just the time-series of the mean is smooth. From these considerations we believe that technical studies like ours are not only of marginal interest. They are important to improve the quality of the time-series for the whole of the UTHi distribution.

These arguments are now included as the last paragraph in the discussion section 3.3 of the revised manuscript.

## 1 Review-independent major changes

In the course of the revision process we noticed that it makes no sense to compute a regression line for a plot like in our old figure 1. The reason is simple, but the problem has not been noted by anyone so far. As UTHi cannot be negative, the difference  $UTHi(N15) - UTHi(N14)$  strongly tends to negative values (i.e.  $UTHi(N14) > UTHi(N15)$ ) when  $UTHi(N15)$  is small. Given that  $UTHi(N15)$  is small, it is quite improbable that  $UTHi(N14)$  is even smaller. At the other end of the distribution we have a similar phenomenon, as values exceeding 115% do not occur in our data sets. Thus, given that  $UTHi(N15)$  reaches the upper extreme, it is much more probable that  $UTHi(N14)$  remains smaller than that it would be even larger. This means that, unless all data pairs agree perfectly, a scatter plot like that in figure 1 *must* have a rhombic shape with an surplus of negative ordinate values at small abscissa values and a surplus of positive ordinate values at large ordinate values. It is clear that a linear fit through a such shaped cloud of data points must have a positive slope. It might be that the slope depends on the width (standard deviation) of the individual distributions but our statement that it “differs quite substantially from the ideal value of zero” is, albeit true, meaningless for the problem at hand.

Therefore we replace such plots by simple  $y$  vs  $x$  plots, see the new figure 1 in the revised version. In such a plot the problem becomes evident through an unequal number of points above

and below the  $y = x$  diagonal line.

Similar changes apply to figures 3 and 7.

The old scatterplots were not wrong but their regression lines were misleading. They are appropriate to perform a bin-wise regression (regression of the 1st kind) but not for a regression of the 2nd kind. We have redrawn all figures in typical x-y form, which is typically used to study the linear correlation between two parameters.

Figure 5 has been completely revised. It has been reduced to show the corrections and their actual values more clearly, but not the data points already shown before which is not necessary.

## 2 Reply to reviewer No. 2

### 2.1 Cloud contamination

*Cloud contamination of infrared measurements can cause a positive bias in the upper tropospheric humidity estimated from T12 observations. This means that differences in the cloud clearance between HIRS 2 and HIRS 3/4 may contribute to the discontinuity in the time series. As a result, it is important to make sure that the cloud clearance at the low end of T12 is consistent between HIRS 2 and HIRS 3.*

Note that we have used the intercalibrated data of Shi and Bates (2011). These data 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 add a statement on this in section 3.2.

### 2.2 Diurnal variations

*Due to the diurnal variations of cloud ice mass and humidity in the upper troposphere, difference in the local observation time between satellites may lead to discrepancy in the observed brightness temperatures. In addition, the orbit of NOAA-14 has substantially drifted during the transition period. Have potential biases arising from these factors been taken into account here?*

Indeed, diurnal variations of relative humidity in the upper troposphere could lead to spurious variations in the NOAA channel 12 time series. But such effects would occur not just related to the HIRS/2 to HIRS/3 transition. The NOAA series is a succession of morning and afternoon satellites. Thus, if systematic diurnal humidity variations have indeed an effect on the time series, this effect must be present since N7 started. We have looked at the cumulative probability distributions of UTHi (as well as T12 and T6) for all satellites and all years. We find essentially two groups of curves: (i) N6 to N14, (ii) N15 and all later satellites, see Fig. 1 at the end of this reply. Please note the gap between the dark brown (N14) and the violet (N17) curves. The cdf curves of all satellites before N14 are left of the N14 curves while all satellites of the HIRS3/4 era are to the right. There is nothing in these curves that suggests effects of diurnal variations, since then we would expect two other groups, namely one containing morning satellites and the other containing afternoon satellites. A strong orbital shift would then smear out the distinction between these hypothetical groups.

It is however remarkable that the N14 cdf curves are at the right edge of the HIRS2 group. The reason for this and whether it results from a fact other than typical interannual variability is unknown and a topic for future research.

### 2.3 Assumption that supersaturation did not change

*The proposed intercalibration method is based on the assumption that the probability of supersaturation did not change during the transition period from the HIRS 2 to the HIRS 3 instrument. The validity of this assumption can be assessed using microwave observations (e.g., Buehler et al., 2008, JGR) or the free-tropospheric humidity data set constructed from the Meteosat MVIRI and SEVIRI observations (e.g., Schroder et al., 2014, ACP).*

This is indeed a working hypothesis that was necessary to do the correction. Of course the frequency of supersaturation might have changed over time, which is not known and which is a reason for our studies. It is however very implausible that it has changed such dramatically just at the transition to HIRS/3. The increase of the frequency of threshold exceedances is not small, it is more than a 3-sigma increase when we compute the sigma from the first ten years of the time series. It is hardly conceivable that such a dramatic change could have happened unnoticed in other variables (for instance frequency and coverage of persistent contrails). Such changes have, at least to the author’s knowledge, never been reported. Gierens, Eleftheratos and Shi (2014) indeed found a small decadal increase of UTHi in large regions of the northern mid-latitudes using the intercalibrated HIRS data. These decadal changes refer to the whole range of UTHi, not just the high humidity cases. It might be that the probability density function of UTHi has changed such that high humidities had experienced a stronger increase than the bulk of the distribution. These questions are not yet solved and their solution needs much more research (including analyses of the microwave data mentioned by the reviewer). This research is far beyond the topic of the current paper.

### 2.4 Minor comments

#### 2.4.1 L400-410, Fig. 8

*This portion is confusing because the data pairs of the unmodified values show better agreement than the pairs of the modified values.*

We agree with the referee’s observation. It is indeed the case that for these 256 data pairs the statistics is worse after the correction than before. Insofar, this selection led to a somewhat unfortunate example. However, there are much more cases, as stated, where N15 shows supersaturation while N14 does not. For these cases, the statistics is better after the correction. In the revised version we will add a comment similar to what we write here, but we will retain the somewhat “bad” example, since it shows honestly that every “bulk” correction (i.e., a correction that is not point by point) inevitably has its pros and cons.

#### 2.4.2 L515-519, Table 1

*The mean fraction of exceedances is very small for the examined UTHi thresholds. This gives an impression that it might be okay not to correct for the discontinuity at the low end of T12.*

Indeed, the mean fraction of exceedances is small for the examined UTHi thresholds. However, we cannot ignore the fact these small values changed artificially following the transition

from HIRS 2 to 3. As we focus on ice saturation and supersaturation cases and since we know what caused this unnatural discontinuity in the time series, it is important for us to find and apply methods that take care of this problem. Our method (cdf-based intercalibration) indicates that it is necessary to correct for the discontinuity at the low end of T12, when it comes to assess extreme UTHi values as in our case, and appears to solve the problem satisfactorily.

### 2.4.3 Fig. 1

*To further demonstrate that the discontinuity during the transition period is caused by the shift of the central wavelength, the authors may also present scatter plots for matching pairs between NOAA-12 and NOAA-14 and between NOAA-15 and NOAA-16.*

We refer back here to Fig. 1. The grouping of the cdfs is evidently according to the version of the HIRS instrument used on the various satellites. It is also clear from the plot that there are interannual variations and perhaps remaining sensor differences. Scatter plots of N12 vs. N14 and N15 vs N16 would lay the focus on these other differences which would be inappropriate as these are not the focus of the current paper.

### 2.4.4 Figs. 3, 5

The figures have been replaced and the problems don't not exist anymore.

### 2.4.5 Figs. 4, 9

All corrections done.

## 3 Reply to reviewer No. 3

### 3.1 Consideration of regression dilution

*I have no problems with the CDF approach the authors chose. But I would like to see it compared to something that doesn't fail as terribly as the OLS regression line. Glancing at figure 2, it can be seen that the regression curve is flat, thanks to regression dilution. (See Pitkänen et al., 2016; doi:10.1002/2016GL070852). The elimination of super-saturation upon its application is a direct result. Rather than continuing with their critique of the linear regression method - with fails almost trivially - some other standard technique ought to have been applied. May I suggest calculating instead a bivariate regression (see York reference in in Pitkänen)? Practically, this can be done by choosing a line that goes through the center of mass of the scatter plot, with the same slope as the first eigenvector of the 2x2 covariance matrix of the 2xN time series of pairs. The first eigenvector points in the direction of maximum variance, thus minimizing the residuals perpendicular to the line, rather than in an arbitrarily chosen y-direction.*

#### 3.1.1 Bivariate regression

We thank the reviewer to make us aware of the problems with the ordinary least squares regression that seems to be largely unknown in the atmospheric sciences. The reviewer asked in particular to compute a bivariate regression fit for the bivariate  $T_{12}$  distribution shown in figure 2 and provided a recipe for the procedure. We followed the suggestion and got the following results:

The covariance matrix for the original intercalibrated data pairs is

$$C = \begin{pmatrix} 23.0041 & 19.0753 \\ 19.0753 & 22.7789 \end{pmatrix}$$

The eigenvalues are 41.9671 and 3.81587, reflecting that the data cloud is much more elongated along the diagonal than perpendicular to it. The first eigenvector is proportional to  $(1, 0.994114)^T$ , that is, the slope of the regression is very close to unity when the errors in the data on the x-axis are taken into account. (Of course, the second eigenvector is perpendicular to the first). According to York et al. (2004, eq. 13a), the regression line crosses the bivariate mean, that is in our case (240.029, 240.663) and thus the intercept of the bivariate regression line is 2.04681. The bivariate regression coefficients indicate that indeed, a bivariate least squares fit as suggested by the reviewer fits better to the data than the ordinary least squares fit. These results show again how good the original intercalibration by Shi and Bates (2011) was.

Figure 2 shows the new figure 2 in the revised manuscript, displaying again the ideal fit line (the diagonal  $y = x$ , dashed), and the ordinary least squares fit (solid). Additionally we have plotted the bivariate least squares fit (dashed-dotted) that represents the best fit when the errors in the independent values (here  $T_{12}$  (N15)) are taken into account.

Note that the marginal means of  $T_{12}$ (N14) follow rather the ordinary least squares than the bivariate regression line (shown in the new figures 2 and 6 in the revised paper). Thus a regression of the 1st kind (which uses the marginal means for correction) is closer to the traditional regression line than to a bivariate regression line.

### 3.1.2 Using bivariate regression for correction is inconsistent

Finally we want to point out that while the bivariate regression line provides in some sense the best fit through a bivariate distribution of data with uncertainties in both dimensions, it seems just therefore inappropriate to derive from it corrections to the quantity on the  $x$ -axis. To correct  $x$  needs a fixed value of  $x$ , as the OLS regression and regression of the first kind assume. If, however, uncertainties in the  $x$ -dimension of the data are explicitly considered, it is not immediately clear to which value the correction should be applied or how it may be derived and formulated.

Look at the sketch given in fig. 3 of this reply. Regression works on data pairs  $(x_i, y_i)$  and determines from these a best straight line that minimises a certain distance norm. In the case of OLS regression the sum of the lengths of the solid red lines is minimised and the result is a straight line  $a + bx$ . In the case of bivariate regression the sum of the lengths of the solid blue lines is minimised, resulting in a straight line  $A + Bx$ . In the sketch we have plotted only one fit line for convenience, but usually  $|B| > |b|$ . Note that the OLS regression yields the prediction  $y_{OLS}$  as a function of  $x_i$  only, once the fit coefficients are given, while the bivariate regression predicts  $y_a$  as a function of  $x_i$  and  $y_i$  (once the coefficients  $A, B$  are known). Independent data which are to be corrected using the regression fits *do not come as pairs*, there is only data  $x_o$ . An OLS-type correction is possible here as  $x_{o,corr} = a + bx_o$ . Of course, a similar correction can be performed using the coefficients  $(A, B)$  of the bivariate fit. However, this correction is inconsistent with the bivariate regression, which should properly involve  $x_{o,a}$ , the adjusted  $x$ -coordinate that belongs to the actually observed value,  $x_o$ . However,  $x_{o,a}$  cannot be determined, since a corresponding datum  $y_o$  is not given.

Thus, in regression-type correction one has two non-ideal possibilities:

- a) to use the non-ideal OLS fit and make the corrections consistently
- b) to use the ideal bivariate fit and make the corrections inconsistently.

To illustrate these arguments we apply the correction here to the “training” data, that is, the data pairs from which the regression coefficients have been determined.

The correction that is consistent with the bivariate regression is in our case:

$$\hat{T}_{12/15} = A + \frac{B}{1+B^2} (T_{12/15} + B T_{12/14} - AB). \quad (1)$$

(The formula can be derived from the equations given by York et al., 2004, assuming independent data and equal weights). It contains  $T_{12/14}$  explicitly and this shows that the formula cannot be applied for correction of independent data.

The inconsistent correction is simply

$$\hat{T}_{12/15} = A + B T_{12/15}. \quad (2)$$

The result of both exercises is plotted in fig. 4 of this reply. Inconsistent use of the regression coefficients leads to the blue points that are arranged on the bivariate fit line. They are all above the black diagonal, that is, the brightness temperatures are corrected upward as they should, at least in the lower tail of the data. The correction above 240 K is not really necessary as we have demonstrated in the manuscript. The red dots represent the result of consistent use of the regression coefficients. As desired, in the low tail this leads to an upward correction. But at higher temperatures the corrections go up and down, perhaps with an average correction close to zero. But this leads to noise that is unwanted, and — as we have seen — it is unnecessary. According to these findings, we cannot derive corrections for satellites beyond N15 using the consistent formula (1), simply because there are no  $T_{12/14}$  data beyond 2005. Corrections using the inconsistent formula (2) would not be appropriate to be derived.

### 3.2 Influence of close data pairs

*I think the authors should be more clear how they choose pairs of data points. For example 2 HIRS/2 points "A" and "B", and 2 HIRS/3 points "1" and "2", if all close together, can produce 4 pairs for comparison: A-vs-1, A-vs-2, B-vs-1, and B-vs-2. Do the authors avoid this sort of inflation?*

We do not construct data pairs from data in adjacent grid points. A data pair consists of two daily averages, one from N14 measurements and one from N15 measurements, in the same grid box. We have added a few sentences in section 2.1 and hope that the new text is clear now.

### 3.3 Minor corrections and comments

#### 3.3.1 Figures and colour tables

We have remade the scatter plots with slightly larger symbols (for the expert: In Gnuplot we use “points” now with pointsize 0.3 instead of “dots”). The rainbow colour scheme has been replaced by a blue-yellow-red colour series which seems to be distinguishable by colour-blind readers. Without knowing what is meant with “perceptually uniform”, we hope that the new colour table meets that category as well.

#### 3.3.2 Others

P1 7: Ok

P2 1-3: ok, references added.

P3 27-28: ok, in section 4. Added.

P4 29-30: This was meant as a mathematical bracket because both numbers have the unit K. We have removed all brackets in the revised version.

P7 26-27: We have modified the sentence into “Although it looks like a manifestation of climate change it is rather a manifestation of the change from HIRS 2 to HIRS 3”.

Fig. 9: yes, that was wrong. Corrected.

## References

Please look in the revised manuscript’s reference section for the literature quoted in this reply.

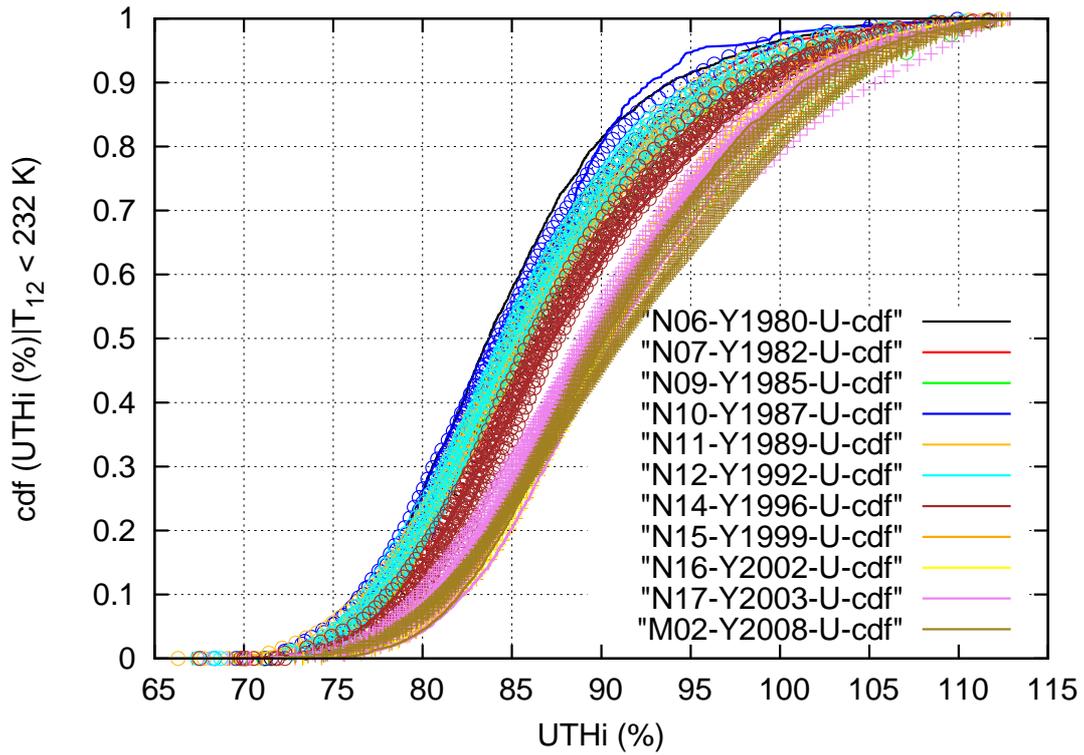


Figure 1: Cumulative distribution functions of UTHi for all satellites of the NOAA series conditioned on  $T_{12} < 232 \text{ K}$ , that is for the low tail of channel 12 brightness temperatures. It is evident that the cdfs form two groups that are distinguished by the HIRS version that the respective satellite carries. Different colours refer to different satellites, different symbols along the curves refer to different years of operation.

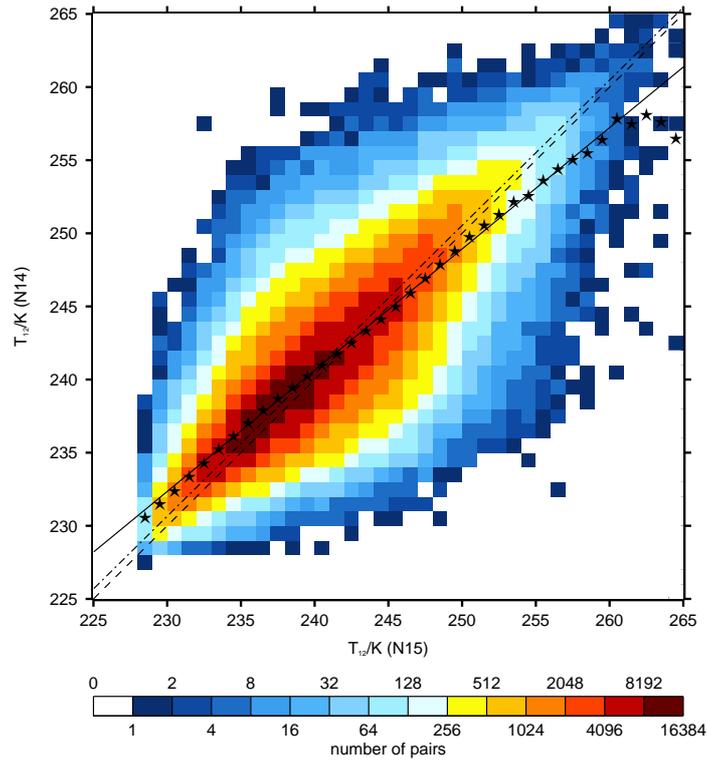


Figure 2: Bivariate distribution of channel 12 brightness temperatures intercalibrated by Shi and Bates (2011), measured by NOAA 15 (x-axis) and NOAA 14 (y-axis), on 1004 common days in the same  $2.5^\circ \times 2.5^\circ$  gridboxes between  $30^\circ\text{N}$  and  $60^\circ\text{N}$ . The three straight lines are: the diagonal  $y = x$  (dashed), the ordinary least squares fit (solid), and the bivariate least square fit (dashed–dotted). The stars are the marginal means of  $T_{12}$  (N14) for each 1 K interval of  $T_{12}$  (N15).

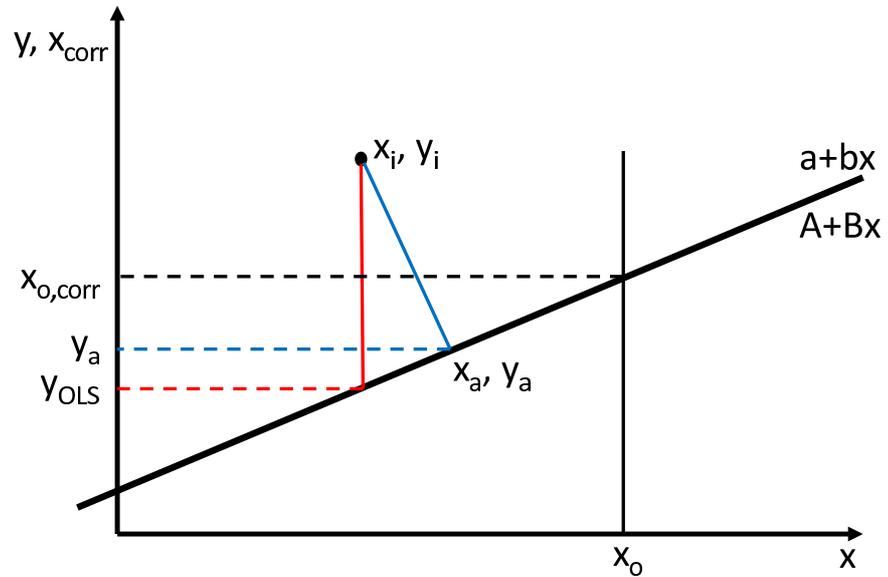


Figure 3: Sketch explaining the subtle difference between regression for finding the best line through a collection of data pairs and regression for formulating a correction of  $x$ -values.

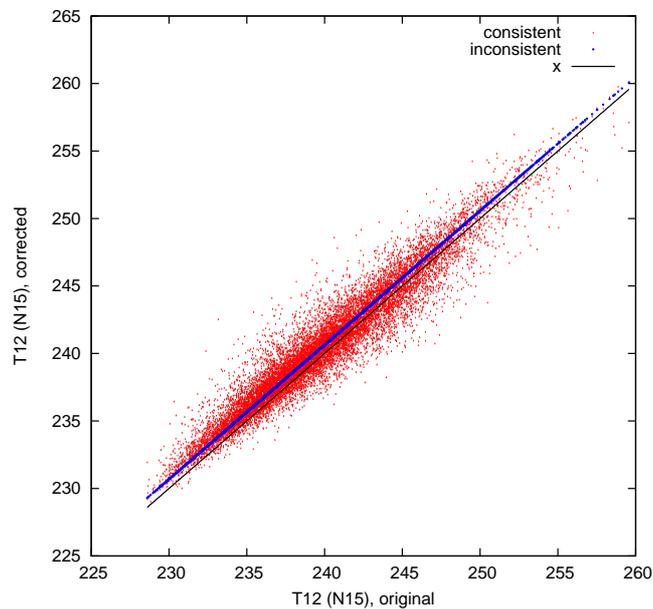


Figure 4: Consistent (red dots) and inconsistent (blue points) use of the bivariate regression coefficients for correction of channel 12 brightness temperatures from NOAA 15. The black line is the diagonal  $y = x$ , included to guide the eye.

# 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

Klaus Gierens<sup>1</sup> and Kostas Eleftheratos<sup>2</sup>

<sup>1</sup>Deutsches Zentrum für Luft- und Raumfahrt, Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany

<sup>2</sup>Faculty of Geology and Geoenvironment, University of Athens, Athens, Greece

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

**Abstract.** In the present study we explore the capability of the intercalibrated HIRS brightness temperature data at channel 12 (the HIRS water vapour channel;  $T_{12}$ ) to reproduce ice supersaturation in the upper troposphere during the period 1979–2014. Focus is given on the transition from the HIRS 2 to the HIRS 3 instrument in the year 1999, which involved a shift of the central wavelength in channel 12 from 6.7  $\mu\text{m}$  to 6.5  $\mu\text{m}$ . It is shown that this shift produced a discontinuity in the time series of low  $T_{12}$  values ( $< 235$  K) and associated cases of high upper-tropospheric humidity with respect to ice ( $\text{UTHi} > 70\%$ ) in the year 1999 which prevented us from maintaining a continuous, long term time series of ice saturation throughout the whole record (1979–2014). We **show** that additional corrections are required to the low  $T_{12}$  values in order to bring HIRS 3 levels down to HIRS 2 levels. The new corrections are based on the cumulative distribution functions of  $T_{12}$  from NOAA 14 and 15 satellites (that is, when the transition from HIRS 2 to HIRS 3 occurred). By applying these corrections to the low  $T_{12}$  values we show that the discontinuity in the time series caused by the transition of HIRS 2 to HIRS 3 is not apparent anymore when it comes to calculate extreme UTHi cases. We come up with a new time series for values found at the low tail of the  $T_{12}$  distribution, which can be further exploited for analyses of ice saturation and supersaturation cases. The validity of the new method with respect to typical intercalibration methods such as regression-based methods is presented and discussed.

## 1 Introduction

Ice supersaturation is a frequent phenomenon in cold regions of the troposphere (below 0°C, in particular in the upper troposphere), important for the weather state, cirrus cloud formation and climate (Gierens et al., 2012). The probability density function of the degree of ice supersaturation is approximately an exponential distribution with a mean supersaturation value of about 15%. A slight change of the mean value implies a large change in the tail of the exponential distribution, thus conditions for in-situ cirrus formation can occur much more frequently or much more seldom than today after a slight change of the mean supersaturation. Such subtle changes cannot reliably be predicted with climate models; hence the prediction of future cirrus coverage is **challenging**. Moreover, cirrus clouds are a component of the climate system and their feedback on climate change is one of the most uncertain issues in climate research (e.g. Ou and Liou, 1995; Stephens, 2005). Any short or long term change in the frequency of occurrence of ice supersaturation and in its probability density function is expected to have an

influence on the cirrus cloud field and therefore on climate change (e.g. Irvine and Shine, 2015). Relatively few papers (e.g. Bates and Jackson, 2001; Soden et al., 2005; Chung et al., 2014; Gierens et al., 2014) appear in the literature describing the large and small scale distribution and seasonal, annual and longer time scale changes of relative and absolute humidity of the upper troposphere. A lack of observations, especially those at regional and global scales, has hampered our ability to study the changes in this important climate variable.

An ideal data set with which to study long-term changes of upper-tropospheric humidity (UTH) is provided by the series of polar orbiting satellites of the National Oceanic and Atmospheric Administration (NOAA), which has started in the late 1970s and is still ongoing, meanwhile in co-operation with the European Organisation for the Exploitation of Meteorological Satellites (Eumetsat). The satellites all carry the High-Resolution Infrared Radiation Sounder (HIRS). Channel 12 of this instrument can be used to retrieve UTH. It is a radiance based quantity that represents a weighted mean over a vertical profile of relative humidity with a peak of weighting function in the upper troposphere. The retrieval method has been developed by Soden and Bretherton (1993) and improved by Jackson and Bates (2001).

All the NOAA satellites from N06 (launched 1979) to N14 (launched 1994) carried version 2 of the HIRS instrument while from N15 on (launched 1998) version 3 and later version 4 of HIRS was installed. The transition from HIRS 2 to HIRS 3 involved a shift of the central wavelength in channel 12, from  $6.7\ \mu\text{m}$  to  $6.5\ \mu\text{m}$ . Unfortunately, this is not a small change as it may appear. The atmosphere is nearly 1.5 times as opaque at  $6.5\ \mu\text{m}$  than at  $6.7\ \mu\text{m}$  (see, for instance, the black curve in fig. 1 of Shi and Bates, 2011). Thus the kernel function for the retrieval of UTH peaks about one kilometre higher in the atmosphere for HIRS 3 and 4 than for HIRS 2 (cf. fig. 2 of Gierens and Eleftheratos, 2016), or in other words, channel 12 of N15 and the later satellites is sensitive to a more than 1 km higher layer in the atmosphere than channel 12 of the older satellites of the NOAA series; yet the layers strongly overlap due to large half widths of the corresponding weighting kernels of, say, 4 to 5 kilometres. As temperature decreases on average by  $6.5\ \text{K km}^{-1}$  in the troposphere, the change of the wavelength and the corresponding increase of the weighting function peak altitude led to a discontinuous shift in the corresponding brightness temperatures of about 8 K (Shi and Bates, 2011; Chung et al., 2016).

Such a strong discontinuity would break the desired long-term time series, but Shi and Bates (2011) were successful in solving the problem. They performed an intercalibration of the channel 12 brightness temperature,  $T_{12}$ , of all NOAA satellites, using N12 as a reference. They compute for each satellite monthly and zonal averages, with  $10^\circ$  latitude belts centred on  $85^\circ\ \text{S}$  to  $85^\circ\ \text{N}$ . Thus, they obtain a set of mean brightness temperature values  $T_{L,YM}^N$ , where the upper index  $N$  is satellite number, and the lower indices are latitude belt and year/month combination. Biases are then computed as individual differences  $T_{L,YM}^N - T_{L,YM}^{N+1}$ , that is for pairs of subsequent satellites operating in the same months and years. The individual bias values are then put into 5 K wide classes of brightness temperatures. The result of this is a data set providing temperature dependent corrections for each satellite pair. These corrections are applied pixel-wise (i.e. not simply by adjustment of the time-series means), with N12 taken as reference. The intercalibration procedure solves not only the problem with the wavelength change, minor changes due to variations in filter functions and calibration loads are covered automatically as well.

The inter-calibrated HIRS brightness temperature (BT) data for the past 35 years (1979–2014) have been used to study long term changes in the upper tropospheric water vapour (Chung et al., 2016). With this long term data set we can also study the upper tropospheric humidity with respect to ice (UTHi, Gierens et al., 2014).

In the present paper this radiance based quantity is used for the first time to study ice supersaturation cases in the upper 5 troposphere with such a long time series. As ice-supersaturated layers are typically much shallower than the layer where channel 12 of HIRS is sensitive to, only a very small fraction of UTHi values exceeds 100% (Gierens et al., 2004). Yet one can argue that there is sometimes ice supersaturation in the upper troposphere when UTHi is of the order 70% and that the probability of occurrence of ice supersaturation increases with the measured value of UTHi in a certain fashion (Lamquin et al., 2009; Dickson et al., 2010). The research focuses on UTHi values exceeding 70% and higher thresholds. Preliminary findings 10 show that the extreme UTHi situations might have increased in the past decade, whereas the zonal mean UTHi remained almost unchanged. These results are very interesting; they contribute to an ongoing debate whether the free troposphere is moistened as a consequence of global warming (e.g. Paltridge et al., 2009; Dessler and Davis, 2010).

Chung et al. (2016) stated that the discontinuity in the time series caused by the transition of HIRS 2 to HIRS 3 has been almost completely removed by the calibration process conducted by Shi and Bates (2011), in which the influence of the filter 15 change was adequately taken into account by a scene radiance-dependent bias correction. Indeed, there is no evidence for a discontinuity in their time series of  $T_{12}$  anomalies in the period 1979 to 2015. Although this is true for the mean  $T_{12}$ , two interesting questions raised here are a) whether Shi and Bates's inter-calibration process is also valid for values found at the low tail of the distribution of  $T_{12}$  when it comes to calculate extreme UTHi cases as in our case, and b) whether it is actually correct to combine the two HIRS time series (HIRS 2; 1979–2005 and HIRS 3/4; 1999–2014) into a single one for the case of 20 low  $T_{12}$  values, given that HIRS 2 and HIRS 3 actually sense different layers in the upper troposphere. Assuming that we can physically combine the two time series into one, like Chung et al. have done, our findings indicate that the discontinuity caused by the transition of HIRS 2 to HIRS 3 is not completely removed when looking at the low  $T_{12}$  values, so that further corrections are needed in order to bring HIRS 3 levels down to HIRS 2 levels. By applying additional corrections to the low  $T_{12}$  values, we come up with a more consistent intersatellite-calibrated  $T_{12}$  time series with reduced errors at the low  $T_{12}$  values due to 25 the transition from HIRS 2 to HIRS 3, which can be further used for analyses of extreme UTHi cases.

In the following we first show how high values of UTHi and ice-supersaturation behave when the transition between the two HIRS instruments occurs. Then we discuss several refinements to the intercalibration (that is, we work on the data that are already intercalibrated by Shi and Bates, 2011). A new procedure is devised and will be explained (section 2). A couple of simple results from the new method are presented, and the new method is discussed in comparison to more traditional 30 methods (section 3). Finally our results are summarised, conclusions are drawn and an outlook on future research necessities and possibilities is given in section 4.

## 2 The intercalibration problem

### 2.1 Retrieval of upper-tropospheric humidity and ice supersaturation

When we used these intercalibrated data to set up a time series of the number of occurrences of cases with ice supersaturation we found a strong increase, seemingly coincident with the transition from HIRS 2 to HIRS 3 and this unwanted surprise led us to check the intercalibration especially for the transition again. The check disclosed problems especially at the low end of channel 12 brightness temperatures, i.e. at those data that are characteristic for the supersaturation cases.

We believe that the intercalibration of Shi and Bates (2011) works well for the bulk of the data but not so well in the tails of the  $T_{12}$  distribution. Recall that the intercalibration was based on monthly and zonal averages of  $T_{12}$ , in other words, on a distribution with clipped tails (as averaging eliminates extremes). It is appropriate to consider intercalibration as an exercise in linear regression. With clipped tails, the regression sees only the central part of a distribution, however the tails could in principle change the regression coefficients quite substantially because of a leverage effect (the distance of tail values to the pivot at the mean value is evidently particularly large, that is, they have a large lever, see von Storch and Zwiers, 2001, sect. 8.3.18).

In order to make progress and avoid excessive averaging we consider daily averages of  $T_{12}$  in  $2.5^\circ \times 2.5^\circ$  grid boxes of the 30 to  $70^\circ\text{N}$  zone, similar to the data we have produced for the study in Gierens et al. (2014). We use all days with common operation between N14 (HIRS 2) and N15 (HIRS 3). In total we have 1004 common days (between 1 January 1999 and 7 April 2005). For each of these days we select those grid boxes where both satellites gave valid data, usually overpassing at different time of day. Grid boxes with data from only one satellite are not considered. Two such averages for a certain grid box and a certain day, one from N14 and one from N15, form a data pair. In total we have 730473 data pairs.

Figure 1 shows a scatter plot of randomly selected 2% of the data pairs for the upper-tropospheric humidity with respect to ice, UTHi. Note that calculations have been done with all data. The abscissa shows values measured by N14, while the ordinate shows corresponding values measured by N15. Ideally the data pairs should lie on the diagonal (the dashed red  $y = x$  line) or at least they should be dispersed symmetrically around it. However, one can notice a tendency of UTHi (N15) values measured by HIRS 3 to be higher than their N14 counterparts measured with HIRS 2. While the considered N14 data contain 636 records with  $\text{UTHi} > 100\%$ , 2739 records of N15 have  $\text{UTHi} > 100\%$ . There are only 256 cases where both N14 and N15 show supersaturation in the same grid box and on the same day. In spite of the apparent tendency of N15 to show more supersaturation, the maximum values are equal, 113% for both instruments. These results suggest that the intercalibration of channel 12 must be improved if one is interested in high humidity cases and, in particular, in ice supersaturation.

### 2.2 Regression-based intercalibration

Let us make a step back and consider the brightness temperatures  $T_{12}$  measured with the HIRS instruments. Figure 2 shows the two-dimensional histogram of  $T_{12}$  pairs (i.e.  $\{T_{12}(N15), T_{12}(N14)\}$ ) in 1 K resolution. Recall that these data are both intercalibrated by Shi and Bates (2011) to N12, and indeed the data pairs cluster nicely and symmetrically around the  $y = x$  line (black dashed line, hereinafter simply referred to as “diagonal”), which demonstrates that the intercalibration was quite

successful. This statement can be corroborated quantitatively, as both data sets have similar measures: mean values and standard deviations of about  $240 \pm 5$  K, a total range from 228 to 265 K, similar quartiles and medians. In spite of this, the maximum of the joint distribution (dark red pixels) is not centered on the diagonal axis. In particular at low  $T_{12}(N15)$  there appears a tendency of N15 to display lower values than N14, leading to the observed surplus of supersaturation cases relative to N14.

5 The diagonal does not, therefore, represent the best (least squares) fit, that is, the intercalibration can be improved. Ordinary least squares (OLS) linear regression (black solid line) yields the following fit:

$$(y/K) = 41.63 + 0.8292(x/K), \quad (1)$$

with a slope that is not very close to unity and an intercept that differs quite substantially from zero. (Note that a quadratic fit is not required; it does hardly give an improvement). The relatively small value of the slope is the result of an effect

10 termed “regression dilution” or “regression attenuation” (Cantrell, 2008; Pitkäinen et al., 2016) which results from neglecting the measurement error in the  $x$ -component of the regression pair. This problem can be overcome using a bivariate regression which accounts for errors in both components of the data pairs (see Appendix). The bivariate regression straight line fit is shown in the figure as a dash-dotted line. Its equation is

$$(y/K) = 2.05 + 0.994(x/K), \quad (2)$$

15 with a slope that is indeed very close to unity. However, similar to the diagonal it does not really represent an optimum fit for the lower range of brightness temperatures because the majority of data pairs lie above the bivariate fit line in this range of brightness temperatures. As we argue in the appendix, a correction using the bivariate fit can only be done in an inconsistent fashion, using the bivariate fit coefficients in a way as if the fit was an OLS fit. Thus we do not use the bivariate regression for correction of the N15 brightness temperatures.

20 The stars in figure 2 represent the marginal means of  $T_{12}(N14)$  per 1 K interval of  $T_{12}(N15)$ . These represent bin-wise mean differences which could be used for intercalibration as well (regression of the first kind). The difference between the OLS regression and bin-wise correction is yet small.

For the moment this demonstrates that the intercalibration of the channel 12 brightness temperatures can be improved using common daily data for single grid cells instead of zonal/monthly averages. Whether this improvement is useful as well for the  
25 retrieval of upper-tropospheric humidity values has still to be shown.

If we perform the retrieval of UTHi (Jackson and Bates, 2001) for N15 using the OLS regression-corrected values of the channel 12 brightness temperature,  $\widehat{T}_{12}$ , where

$$\left(\frac{\widehat{T}_{12}}{K}\right) = 41.63 + 0.8292\left(\frac{T_{12}}{K}\right). \quad (3)$$

then the resulting scatter plot of the corresponding values of UTHi is shown in figure 3 in the same format as in figure 1. It  
30 is obvious that the N15-retrieved values are lower than before and that the excess of data points above the diagonal line is no longer present.

Yet unfortunately we must note that the range of UTHi (N15) is dramatically decreased at the high end and that all cases of supersaturation are eliminated when this kind of intercalibration is indeed applied. So ironically, instead of reducing the number of supersaturation cases in N15 data to a level given by the corresponding number of such events in N14 data, the new regression-based intercalibration eliminates all supersaturation. The comparison of this feature between N14 and N15 has in  
5 no way been improved, it has merely been turned upside-down. We note that similar procedures like bin-wise intercalibration with and without outlying data pairs (more than  $\pm 3\sigma$  distance from the regression line) does only lead to minor modifications. The basic problem, that is, the strong elimination of high UTHi values and the complete loss of supersaturation, remains. Thus, the **OLS** regression method, however a natural choice it might appear for the purpose of intercalibration, does not lead to plausible results. We need another procedure.

### 10 2.3 Intercalibration via the distribution function

The goal of the new intercalibration exercise is to have similar number of supersaturation cases for the data overlap period of N14 and N15, because the strong jump detected in the original data seems implausible even when one acknowledges that the two satellites see the same grid cell at different times during a day. Looking at the cumulative distribution functions (cdf) of the corresponding channel 12 brightness temperatures (fig. 4) discloses the origin of the difference in supersaturation cases: there  
15 are much more (exceeding a factor of three) cases of very low  $T_{12}$  values measured by HIRS 3 than by HIRS 2, a tendency that could already be observed in the 2-D histogram of fig. 2. As low  $T_{12}$  produces high UTHi in the retrieval, this difference at the low  $T_{12}$  tail produces the corresponding difference in the high UTHi tail.

We devised an alternative intercalibration procedure that yields similar distribution functions (with the N14 cdf as reference) as follows: The data sets are grouped in  $T_{12}$ -bins first, the data in each bin are counted, resulting in numbers  $n_t^s$  (where the  
20 upper index  $s$  labels the satellite and the lower index  $t$  the  $T_{12}$  interval). We start with the lowest bin and compare  $n_1^{15}$  with  $n_1^{14}$ . As there are more cases with low brightness temperature measured by N15,  $n_1^{15} - n_1^{14} = \delta n_1 > 0$ . Now we determine a minimal temperature correction  $\Delta T_1$  such that if all  $T_{12}$  in the first bin of the N15 data set are incremented by this value, the surplus  $\delta n_1$  of them get shifted to the next bin, and as a result, the first bin contains an equal number of data from N14 and N15, as desired. For the next bin we use the same procedure where we take into account the  $\delta n_1$  additional values that have been  
25 shifted from the foregoing bin. The process is stopped when a bin is reached where either the ratio of the two cdfs is already close to unity or where this happens after the data from the bin below are shifted up. Note that we take the ratio between the cdfs, not their difference. This has the consequence that the corrections approach zero as the cdfs both approach unity, that is, the corrections are applied just at the low  $T_{12}$  tail where we want to apply it; unnecessary corrections in the upper bins are avoided.

30 What is the best bin width  $\Delta$  for such a procedure? We could use Sturges' rule (or similar ones) to determine it:

$$\Delta \approx \frac{\max(T_{12}) - \min(T_{12})}{1 + \log_2 n} \quad (4)$$

which gives a  $\Delta$  of approximately 1 K. Indeed the maximum correction  $\Delta T_t$  is smaller than 0.8 K when a bin width of 1 K is chosen. If the bin width is smaller the necessary shifts get smaller as well, but at a low rate such that the maximum correction

can exceed  $\Delta$ , which means that some data would have to be shifted by more than one bin. This happens for  $\Delta = 0.5$  K where the maximum shift computed exceeds 0.6 K. Shifting data by more than one bin would render the bookkeeping of shifted data unnecessarily complicated; thus we avoid it. **The corrections for 1 K bins are shown in fig. 5 together with their respective values for convenience. The new corrections for  $T_{12}$  (N15) are smaller than those determined by the OLS regression fit of Eq.**

5 (1). The corrections are even zero above  $T_{12} > 240$  K, due to the termination criteria of our algorithm.

The result of this kind of intercalibration for the intercomparison of the two brightness temperature data sets is shown in Fig. 6. Although the 2-D histogram is very similar to the one shown in Fig. 2, there are notable differences. The gravity centre of the joint distribution (dark red pixels) is now following the diagonal axis (dashed black line), a desired feature. **Also the bivariate fit (dash-dotted line) crosses the middle of the distribution's gravity centre.** The best OLS fit (solid black line) is still  
 10 tilted against the diagonal; its equation is

$$(y/\text{K}) = 29.89 + 0.8771 (x/\text{K}). \quad (5)$$

The intercept is much smaller than for the original data, and the slope is a bit closer to unity than before. Marginal means of  $T_{12}$ (N14) (stars) again closely resemble the OLS linear regression. The marginal means and the OLS regression are very close to the  $y = x$  diagonal **and the bivariate fit in the gravity centre of the distribution. In spite of this, the bivariate fit line**  
 15 **has worse parameters than before the correction (see the appendix), although it fits the data better in the central region. How is this possible? In fig. 2 (original data) the bivariate fit is nearly parallel to the diagonal, but lies above it, clearly reflecting the problem: the data pairs concentrate above the diagonal. As these lines are nearly parallel, the fit's slope is nearly unity (0.994) and the intercept is very small (2.05). In fig. 6, the cdf-correction shifts the highest concentration of data pairs onto the diagonal, hence the bivariate fit and diagonal almost coincide there. However, only  $T_{12}$  (N15) has been corrected, not  $T_{12}$**   
 20 **(N14), which, so to speak, rotates the data in the dark red patch and the lower values clockwise. Accordingly the bivariate fit got a bit steeper than before (slope 1.06) its intercept got further away from zero (-14.8).**

The result of the cdf-based intercalibration is shown for UTHi in Fig. 7. It is seen that high and supersaturation values of UTHi are retained, as desired. **The scatter of the data points around the diagonal is more symmetric than with both the original and the OLS regression-intercalibrated data (cf. figs. 1 and 3).**

25 It is not necessary to show the  $T_{12}$  cumulative distribution functions after the correction; these are almost equal *qua* construction.

### 3 Results and discussion

#### 3.1 Overall improvement

Simple statistical measures, computed with the set of the common daily and grid based data, may show that indeed an improve-  
 30 ment results from the cdf-based intercalibration. The indicators are the following:

- The mean difference of channel 12 brightness temperature (N15 minus N14) is  $(-0.63 \pm 2.76)$  K in the original data (mean and one standard deviation). With the correction applied to N15 it reduces to  $(-0.35 \pm 2.70)$  K.

- The mean difference of the corresponding UTHi is  $(3.24 \pm 12.41)\%$  in the original data. With the correction it reduces to  $(0.54 \pm 11.50)\%$ .

Thus the mean temperature difference is almost halved, the mean UTHi difference is even reduced by a factor of six.

### 3.2 Simple applications

5 For testing the procedure further we consider the 256 data records indicating ice supersaturation in both measurements (N14 and N15). These pairs of brightness temperature and UTHi are shown in Fig. 8 with black points showing the original values and red points the modified ones, after application of the cdf-based intercalibration. All N15-brightness temperatures of these cases are shifted to slightly higher values and thus all corresponding UTHi values are decreased. 176 of the cases (more than two thirds of them) change from supersaturated to subsaturated in the N15 data, but all remain at above 90%, that is, they  
10 still indicate quite moist conditions. **This example shows that the correction can worsen the relation between the brightness temperatures in certain cases. For the majority of data pairs, however, it improves the relation, as for instance for the more than 2000 cases where N15 indicates supersaturation ( $UTHi > 100\%$ ) while N14 does not.**

Figure 9 shows 35-year time series of UTHi threshold exceedances. This is the fraction of data with  $UTHi \geq X\%$ , where  $X$  is 70, 80, 90, and 100. This counting exercise has been performed with the original data (shown in the upper panel) where  
15 a strong increase in high UTHi cases can be observed from about 1999 onwards for all selected thresholds. **Although it looks like a manifestation of climate change it is rather a manifestation of the change from HIRS 2 to HIRS 3. We note again that the plot shows data intercalibrated by Shi and Bates (2011). These data have been cloud-cleared in a consistent fashion. The strong increase that we see is not an artefact of missing or inconsistent cloud-clearance.** A similar analysis with the modified data shows no obvious signs of a trend and it will need sophisticated time-series-analytical methods to find out if there are any  
20 trends in the data at all. A deeper analysis of the four time series will be reported in a forthcoming paper.

### 3.3 Discussion

Actually there are two questions to be discussed:

- Is it justified at all to combine all HIRS  $T_{12}$  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  
25 one kilometre apart vertically?
- Is it justified to use a cdf-based intercalibration procedure?

The first of these questions is a difficult one; and it is just the basic question of a number of subsequent problems as, for instance, under which circumstances is it justified or not? Which assumptions have to be made about the structure of temperature and moisture profiles, etc. This technical note is not the place to answer these questions; but it certainly deserves  
30 much more research in order to be sure that results obtained so far (Gierens et al., 2014; Chung et al., 2016) are reliable. This should be a topic for the near future.

To discuss the second question needs an analysis and comparison of what is effectively done in the cdf-based and the regression-like methods. It should be noted that the only subjective element in the intercalibration problem is the choice of the method. Once the method has been chosen, everything else is based on fixed rules and is therefore objective. The difference in the methods is the different set of rules and the reasoning from which these rules are derived. In the end, the procedures are similar again: All methods are used to determine a  $T_{12}$ -dependent correction which is then applied.

- The OLS regression method is based on the postulate that the mean squared difference between *all* data pairs is a minimum (regression of the second kind).
- The method of Shi and Bates (2011) is based on the postulate that the mean squared difference between data pairs in given intervals (bins) of  $T_{12}$  is a minimum (regression of the first kind). This method is more flexible than the OLS regression-based method since it does not assume a linear relation between the two data sets. As one can see in fig. 2 (black line and stars), both methods give very similar results.
- The cdf-based method is based on the postulate that  $P\{\widehat{T}_{12}(N15) \leq T\}/P\{T_{12}(N14) \leq T\} \approx 1$  ( $P\{\cdot\}$  is the probability of the event stated in the brackets), i.e. that both cumulative distributions are similar.

There might be further possibilities which can be based on still other postulates. For instance, instead of considering the relative differences between the two cdfs one could as well use the absolute differences and postulate that these are close to zero. To our knowledge there is no principle argument favouring one or another of these. **The bivariate regression cannot be used in a consistent way for the desired correction, but the inconsistent way may produce good results as well. Nevertheless, we did not consider it appropriate here to apply inconsistent corrections to the data of  $T_{12}$  from HIRS 3 and HIRS 4.**

One essential difference between regression-based and cdf-based methods is that the first consider the data as pairs while this connection is given up in the latter method. The latter instead considers the statistical properties of the data as two independent populations. The reasoning for that has pros and cons. Considering the data as pairs is justified to a certain degree since they are taken on the same day in the same grid cell. But they are also taken at different times of the day which loosens the connection. **In addition, statistical errors arising from the use of OLS regression (i.e. regression dilution) may cause difficulties in determining a correct connection between data pairs.**

Since the truth is unknown nobody can decide which method gives results closer to reality. However, it would be very implausible that supersaturation would suddenly occur much more frequently than before (original data), or not anymore (OLS regression-based method). If the NOAA HIRS channel 12 time series can be combined at all (a question not to be solved here) we need an intercalibration that keeps a certain level of supersaturation frequency and the most conservative choice is then that a change of the UTH distribution functions during the 1004-day transition period from one to the next satellite should be small. So for us it was simply a practical decision guided by this conservative assumption to choose the cdf-based method.

Further evidence for choosing the cdf-based method, as a plausible intercalibration method to account for values found at the low tail of  $T_{12}$  distribution when it comes to analyse high UTH<sub>i</sub> values, is provided in Table 1, which shows the average fraction of UTH<sub>i</sub> exceedances from 70% to 100% during three periods of interest: the period before the transition from HIRS 2

to HIRS 3/4 (1980–1999), the period during the transition (1999–2005) and the period after the transition to HIRS 3/4 (2006–2014). The table also shows the mean fraction of exceedances before (a) and after (b) the corrections applied based on the cdf-method, together with the differences of the means between (b) and (a), i.e. cdf-corrected data minus original data, indicative of improvements performed in the original data. All averages and corresponding differences are expressed in percent.

5 For the case of 70% UTHi threshold, the original data suggest that the mean fraction of exceedances increased from about 1.6% in the period 1980–1999 to about 3.8% in 2006–2014, corresponding to an overall increase of about 138% within about two decades or so. The respective changes for the cases of 80%, 90% and 100% UTHi thresholds by the original data were even larger. Although the mean fraction of exceedances is generally small for the examined UTHi thresholds, such large changes from one period to another do not sound reasonable and are indicative that something may be wrong in the data. Application of  
10 the cdf-based correction to the UTHi threshold data of 70% reduces the change from 138% to 9%. Significant improvements are also found at the other UTHi thresholds. The differences between the cdf-corrected data and the original data in the periods examined are obvious (Table 1c). Our findings suggest that extreme UTHi cases might have increased in the past 35 years. However, given that the zonal mean UTHi remained almost unchanged during the period 1979–2014 (Chung et al., 2016) it is doubtful whether the observed changes estimated with the original data are real. The observed changes estimated with the  
15 cdf-based method (Table 1b) look more reasonable than those calculated with the original data (Table 1a).

Our proposed intercalibration method is based on the assumption that the probability of supersaturation did not change during the transition period from the HIRS 2 to the HIRS 3 instrument. This is indeed a working hypothesis that is necessary to do the correction. Of course the frequency of supersaturation might have changed over time, which is not known and which is a reason for studying high UTHi values. It is however very implausible that it has changed such dramatically just at the  
20 transition to HIRS 3. The increase of the frequency of threshold exceedances is not small; it is more than a  $3\sigma$  increase when we compute the  $\sigma$  from the first ten years of the time series. It is hardly conceivable that such a dramatic change could have happened unnoticed in other variables (for instance frequency and coverage of persistent contrails). Such changes have, at least to the authors knowledge, never been reported. Gierens et al. (2014) found a small decadal increase of UTHi in large regions of the northern midlatitudes using the intercalibrated HIRS data. These decadal changes refer to the whole range of UTHi, not  
25 just the high humidity cases. It might be that high humidity cases have experienced a much stronger increase than the bulk of the distribution. These questions are not yet solved and their solution needs more research, including analyses of microwave data (Buehler et al., 2008, e.g.) or of free-tropospheric humidity data from geostationary satellites (e.g. Schröder et al., 2014). This research however is beyond the topic of the current paper. Another issue worth noting is the small fraction of exceedances for the examined UTHi thresholds, which may give an impression that it might be okay not to correct for the discontinuity at  
30 the low end of  $T_{12}$ . Indeed the values are small, however we cannot ignore the fact that these small values changed artificially during the transition period from HIRS 2 to 3. As we are interested in near and supersaturated relative humidity with respect to ice, and since we know what caused this unnatural discontinuity in the time series, it is important for us to find and apply methods that take care of this problem. Our method (cdf-based intercalibration) indicates that it is necessary to correct for the discontinuity at the low end of  $T_{12}$ , when it comes to assess extreme UTHi values as in our case, and appears to solve the

problem satisfactorily. Indeed, the corrections performed at the tail end of the distribution of brightness temperatures render a time series for the UTHi threshold exceedances without evident strange jumps during the transition period from HIRS 2 to 3.

5 Finally we want to stress that it is really important to have homogeneous time-series over the whole range of brightness temperature and UTHi values. For applications where the mean of  $T_{12}$  or of UTHi is relevant, the intercalibration of Shi and Bates (2011) is sufficient. However, the mean means nothing in non-linear processes like radiation and cloud formation. This implies that more than just the first (and perhaps the second) moment of the UTHi distribution is needed, in particular characteristics of the tails of its distribution. It is clear that for cloud research and how cloudiness will change with climate change, information on the upper tail of UTHi is needed. For questions of the radiative balance of the Earth it is important how the very dry regions of the subsidence zones (termed “radiator fins” by Pierrehumbert, 1995) behave with ongoing climate change; thus the dry end  
10 of the UTHi distribution is of immense interest as well (see also Schröder et al., 2014; Roca et al., 2011). These arguments show that homogeneous time-series of the whole UTHi distribution are needed, it is not sufficient that just the time-series of the mean is smooth. This paper is intended as a step in this direction.

#### 4 Conclusions

We developed a new method for intercalibration of satellite data that is based on a comparison of distribution functions of  
15 brightness temperatures instead of regression methods. We applied this intercalibration to channel 12 brightness temperatures measured with the HIRS 2 instrument on NOAA 14 and the HIRS 3 instrument on NOAA 15. These data had already been intercalibrated by Shi and Bates (2011) but there were still discrepancies at the low end of the distribution, perhaps a consequence of basing their intercalibration on monthly and zonal means which can smooth extremes away. Here we based our additional intercalibration on daily data in  $2.5^\circ \times 2.5^\circ$  grid boxes. The originally intercalibrated data show a very strong increase of very  
20 low brightness temperatures with the transition from HIRS 2 to HIRS 3, and this translates into a correspondingly strong increase in the frequency of occurrence of ice supersaturation in upper-tropospheric humidity with respect to ice retrieved from the brightness temperatures. This seemed to us unphysical and implausible.

We tried regression-based intercalibration procedures first but without success. Instead of less ice supersaturation in HIRS 3 data, all supersaturation cases were eliminated because the corrections were too large. This again seemed to us unphysical and  
25 implausible.

The new intercalibration method is constructed in a way that the probability of supersaturation does not change in the transition from HIRS 2 to HIRS 3. Of course, we do not know whether this assumption is correct; it is simply the most conservative assumption. Other data sets for the transition period (1999–2005) are needed to check the validity of this assumption. This is beyond the scope of the present paper.

30 The overall discrepancies between the  $T_{12}$  data pairs of HIRS 2 on NOAA 14 and HIRS 3 on NOAA 15 are reduced when the new intercalibration is applied. The mean difference in terms of brightness temperature is almost halved, and the mean difference of the retrieved UTHi is even reduced by a factor of six.

A fundamental question is whether and under which conditions HIRS 2 and HIRS 3 data can be combined into a single time-series at all, since they sense different layers in the upper troposphere. For the present investigation we have assumed, as a working hypothesis, that such a combination is admissible. It is not in the scope of this paper to begin an investigation of this difficult problem, but it is certainly a topic for the next future.

## 5 5 Code availability

IDL code for the cdf nudging can be obtained from the first author on request.

## 6 Data availability

HIRS data in general are available from NOAA. The data used for the present paper can be obtained from the authors on request.

## 10 Appendix A: Bivariate regression

Following the suggestion of one reviewer we added bivariate regression lines to figures 2 and 6. These are computed as follows. First, we determine the covariance matrix,  $C$ , of the respective bivariate distribution. We then determine the eigenvalues and eigenvectors of these matrices. The eigenvectors are perpendicular to each other, as they must for real symmetric matrices. We multiply the larger eigenvalue with a scalar such that its  $x$ -component becomes unity; then the resulting  $y$ -component is  
 15 identical to the slope of the bivariate regression line. The latter crosses the bivariate mean,  $(\bar{x}, \bar{y})$ ; thus we have a direction and one point which determines the whole line, including its intercept at  $x = 0$ .

Specifically, we have for the original data:

$$C = \begin{pmatrix} 23.0041 & 19.0753 \\ 19.0753 & 22.7789 \end{pmatrix}$$

The eigenvalues are 41.9671 and 3.81587, reflecting that the data cloud is much more elongated along the diagonal than perpendicular to it. The first eigenvector is proportional to  $(1, 0.994114)^T$ , that is, the slope of the regression is very close to unity. The regression line crosses the bivariate mean, that is in our case  $(240.029, 240.663)$  and thus the intercept of the  
 20 bivariate regression line is 2.04681.

For the cdf-corrected data we have:

$$C = \begin{pmatrix} 20.5694 & 18.0412 \\ 18.0412 & 22.7789 \end{pmatrix}$$

The eigenvalues are 39.7491 and 3.59916. The first eigenvector is proportional to  $(1, 1.06311)^T$ . The regression line crosses the bivariate mean at  $(240.309, 240.663)$  and thus the intercept of the bivariate regression line is  $-14.81190$ .

Although the bivariate regression line provides in some sense the best fit through a bivariate distribution of data with uncertainties in both dimensions, it seems just therefore inappropriate to derive from it corrections to the quantity on the  $x$ -axis. To correct  $x$  needs a fixed value of  $x$ , as the OLS regression and regression of the first kind assume. If, however, uncertainties in the  $x$ -dimension of the data are explicitly considered, it is not immediately clear to which value the correction should be applied or how it may be derived and formulated. For instance, when uncertainties are equal in the  $x$  and  $y$  dimensions, the bivariate regression minimises the sum of squares of the distances between the data points and the fit line in a direction perpendicular to the fit line (instead of parallel to the  $y$ -axis, as does OLS regression). The point on the fit line that is in this sense closest to the original point represents in the same sense the best (or expected) value of the  $x$ -coordinate of the original data point. York et al. (2004) call this point the adjusted point. In order to correct consistently with the bivariate regression one would have to interpret the adjusted  $y$ -coordinate as the corrected value. In our case the formula is

$$\hat{T}_{12/15} = A + \frac{B}{1+B^2} (T_{12/15} + BT_{12/14} - AB), \quad (\text{A1})$$

where  $A$  and  $B$  are the coefficients of the bivariate regression. Note that this formula explicitly needs  $T_{12}$  from N14 (here written as  $T_{12/14}$ ), a quantity that is not available for correction of independent data. Alternatively, one can formulate the regression similarly to the OLS case, but with the bivariate regression coefficients, as

$$\hat{T}_{12/15} = A + BT_{12/15}. \quad (\text{A2})$$

This use of the coefficients is, however, inconsistent.

The result of both corrections is plotted in fig. A1. Inconsistent use of the regression coefficients leads to the blue points that are arranged on the bivariate fit line. They are all above the black diagonal, that is, the brightness temperatures are corrected upward as they should, at least in the lower tail of the data. The correction above 240 K is not really necessary as we have demonstrated. The red dots represent the result of consistent use of the regression coefficients. As desired, in the low tail this leads to an upward correction. But at higher temperatures the corrections go up and down, perhaps with an average correction close to zero. But this leads to noise that is unwanted, and — as we have seen — it is unnecessary.

Bivariate regression is a method to avoid regression dilution (Cantrell, 2008; Pitkäinen et al., 2016), but not for the derivation of corrections. Further information on bivariate regression for general cases can be found in York et al. (2004) who provide unified equations for general least square regression methods.

*Author contributions.* KE provided the raw UTHi and BT data, KG wrote and ran the IDL codes. Both analysed the results and wrote the text.

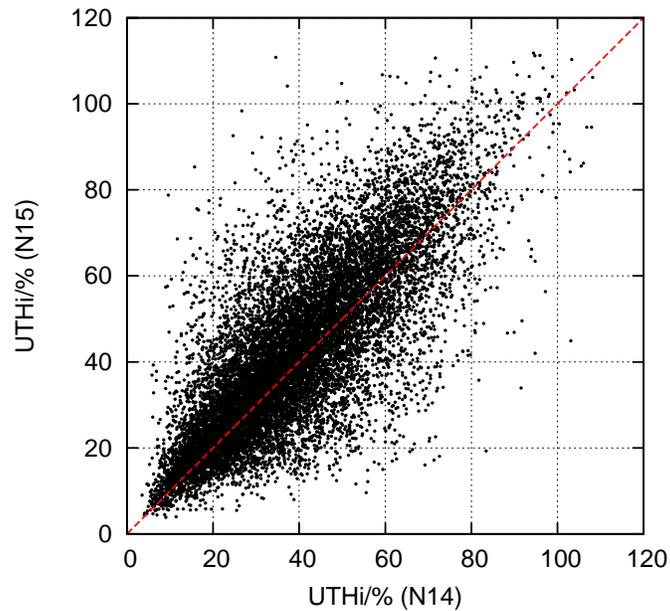
*Competing interests.* The authors declare that they have no conflict of interest.

*Acknowledgements.* We thank Dr. Shi (NOAA) for providing 35 years worth of HIRS brightness temperature data and for her help and advice when problems occurred. Christoph Kiemle and Robert Sausen read a manuscript version of the paper and helped to improve it.

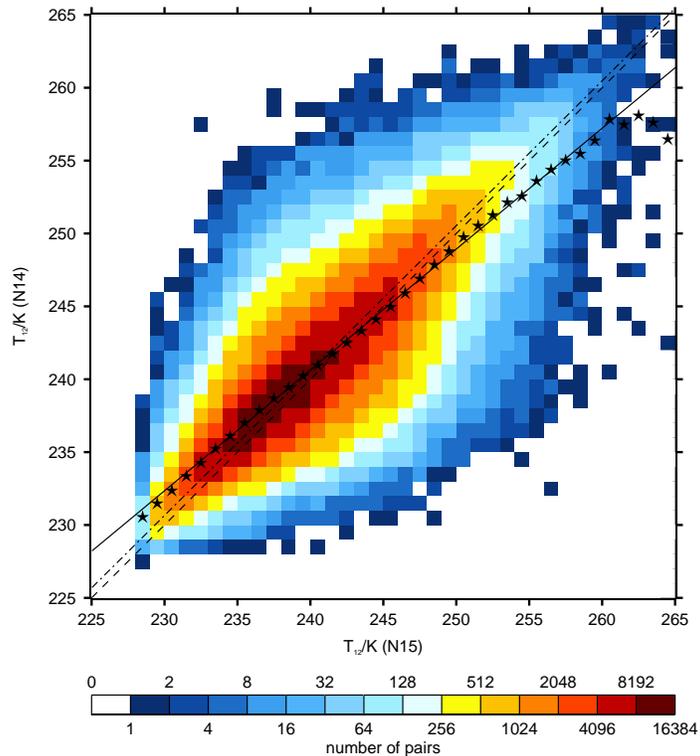
## References

- Bates, J. and Jackson, D.: Trends in upper-tropospheric humidity, *Geophys. Res. Lett.*, 28, 1695–1698, 2001.
- Buehler, S., Kuvatov, M., John, V., Milz, M., Soden, B., Jackson, D., and Notholt, J.: An upper tropospheric humidity data set from operational satellite microwave data, *J. Geophys. Res.*, 113, D14 110, doi: 10.1029/2007JD009 314, 2008.
- 5 Cantrell, C.: Technical Note: Review of methods for linear least-squares fitting of data and application to atmospheric chemistry problems, *Atmos. Chem. Phys.*, 8, 5477–5487, 2008.
- Chung, E.-S., Soden, B., Sohn, B., and Shi, L.: Upper-tropospheric moistening in response to anthropogenic warming, *Proc. Natl. Acad. Sci.*, 111, 11 636–11 641, 2014.
- Chung, E.-S., Soden, B., Huang, X., Shi, L., and John, V.: An assessment of the consistency between satellite measurements of upper  
10 tropospheric water vapor, *J. Geophys. Res.*, 121, 2874–2887, doi:10.1002/2015JD024 496, 2016.
- Dessler, A. and Davis, S.: Trends in tropospheric humidity from reanalysis systems, *J. Geophys. Res.*, 115, D19 127, doi:10.1029/2010JD014 192, 2010.
- Dickson, N., Gierens, K., Rogers, H., and Jones, R.: Probabilistic description of ice-supersaturated layers in low resolution profiles of relative humidity, *Atmos. Chem. Phys.*, 10, 6749–6763, 2010.
- 15 Gierens, K. and Eleftheratos, K.: Upper tropospheric humidity changes under constant relative humidity, *Atmos. Chem. Phys.*, 16, 4159–4169, 2016.
- 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., Spichtinger, P., and Schumann, U.: Ice supersaturation, in: *Atmospheric Physics. Background — Methods — Trends*, edited by  
20 Schumann, U., chap. 9, pp. 135–150, Springer, Heidelberg, Germany, 2012.
- 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.
- Irvine, E. and Shine, K.: Ice-supersaturation and the potential for contrail formation in a changing climate, *Earth Syst. Dynam. Discuss.*, 6, 317–349, 2015.
- 25 Jackson, D. and Bates, J.: Upper tropospheric humidity algorithm assessment, *JGR*, 106, 32 259–32 270, 2001.
- Lamquin, N., Gierens, K., Stubenrauch, C., and Chatterjee, R.: Evaluation of Upper Tropospheric Humidity forecasts from ECMWF using AIRS and CALIPSO data, *Atmos. Chem. Phys.*, 9, 1779–1793, 2009.
- Ou, S. and Liou, K.: Ice microphysics and climatic temperature feedback, *Atmos. Res.*, 35, 127–138, 1995.
- Paltridge, G., Arking, A., and Pook, M.: Trends in middle- and upper-level tropospheric humidity from NCEP reanalysis data, *Theor. Appl. Climatol.*, 98, 351–359, 2009.
- 30 Pierrehumbert, T.: Thermostats, radiator fins, and the local runaway greenhouse, *J. Atmos. Sci.*, 52, 1784–1806, 1995.
- Pitkäinen, M., Mikkonen, S., Lehtinen, K., Lipponen, A., and Arola, A.: Artificial bias typically neglected in comparison of uncertain atmospheric data, *Geophys. Res. Lett.*, 43, 10 003–10 011, doi: 10.1002/2016GL070 852, 2016.
- Roca, R., Guzman, R., Lemond, J., Meijers, J., Picon, L., and Brogniez, H.: Tropical and extra-tropical influences on the distribution of free  
35 tropospheric humidity over the intertropical belt, *Surv. Geophys.*, 33, 565 583, doi:10.1007/s10712–011–9169–4, 2011.
- Schröder, M., Roca, R., Picon, L., Kniffka, A., and Brogniez, H.: Climatology of free-tropospheric humidity: extension into the SEVIRI era, evaluation and exemplary analysis, *Atmos. Chem. Phys.*, 14, 11 129–11 148, 2014.

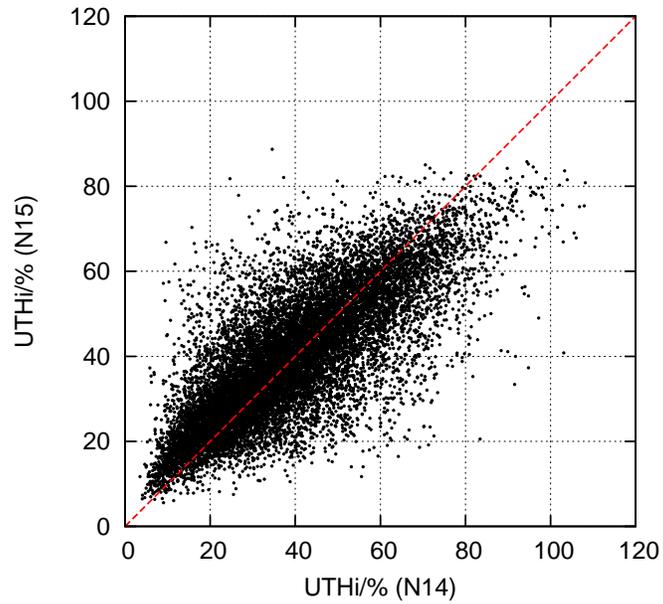
- Shi, L. and Bates, J.: Three decades of intersatellite-calibrated High-Resolution Infrared Radiation Sounder upper tropospheric water vapor, *J. Geophys. Res.*, 116, D04 108, doi:10.1029/2010JD014 847, 2011.
- Soden, B. and Bretherton, F.: Upper tropospheric relative humidity from the GOES 6.7  $\mu\text{m}$  channel: Method and climatology for July 1987, *J. Geophys. Res.*, 98, 16 669–16 688, 1993.
- 5 Soden, B., Jackson, D., Ramaswamy, V., Schwarzkopf, M., and Huang, X.: The radiative signature of upper tropospheric moistening, *Science*, 310, 841–844, 2005.
- Stephens, G.: Cloud feedbacks in the climate system: A critical review, *J. Climate*, 18, 237–273, 2005.
- von Storch, H. and Zwiers, F.: *Statistical analysis in climate research*, Cambridge University Press, Cambridge, UK, 1. edn., 2001.
- York, D., Evensen, N., Lopez Martinez, M., and De Basabe Delgado, J.: Unified equations for the slope, intercept, and standard errors of the  
10 best straight line, *Am. J. Phys.*, 72, 367–375, doi: 10.1119/1.1632 486, 2004.



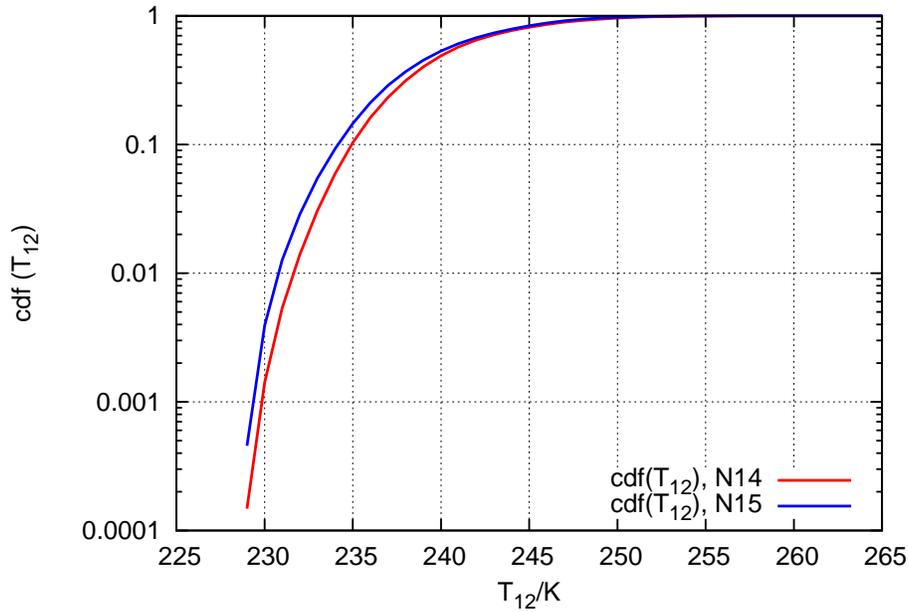
**Figure 1.** Scatter plot of data of upper-tropospheric humidity with respect to ice (UTHi, in percent), retrieved from channel 12 brightness temperatures from the HIRS 2 instrument on NOAA 14 and from the corresponding HIRS 3 instrument on NOAA 15. The data pairs represent daily average values taken in  $2.5^\circ \times 2.5^\circ$  grid boxes in the northern latitude belt of 30 to 70 °N. The red dashed diagonal ( $y = x$ ) and the grid serve only to guide the eye. Ideally the data pairs would be arranged on the diagonal or at least symmetrically to it. However, it is evident that, in particular at high values of UTHi from NOAA 14 there are more data pairs above the diagonal, showing a tendency of the HIRS 3 instrument on NOAA 15 to give higher UTHi values than HIRS 2 on NOAA 14.



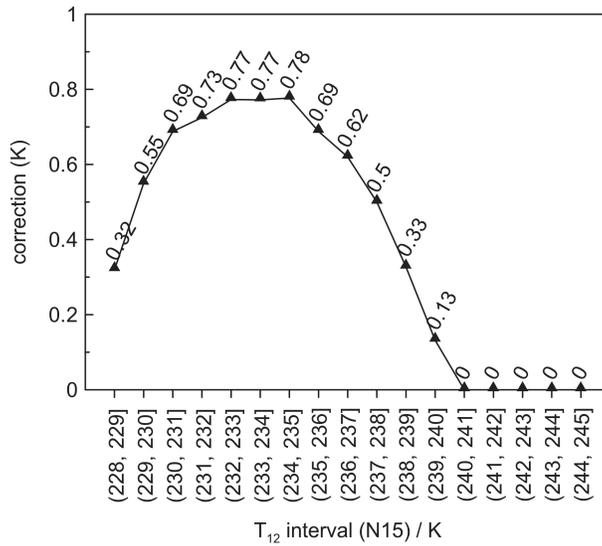
**Figure 2.** Two-dimensional histogram of  $\{T_{12}(N15), T_{12}(N14)\}$ -pairs in 1 K resolution. Ideally the gravity centre of the joint distribution (dark red pixels) would follow the diagonal axis (dashed black line), but it is slightly shifted above the axis. The **ordinary least squares** linear fit is given by the solid black line, **and the bivariate regression is the dash-dotted line**. Marginal means of  $T_{12}(N14)$  for each 1 K interval of  $T_{12}(N15)$  are represented by stars; they closely resemble the **ordinary least squares** regression. Both show that  $T_{12}$  measured by HIRS 3 is lower at the low end of the data range which causes an excess of supersaturation in the UTHi retrieval.



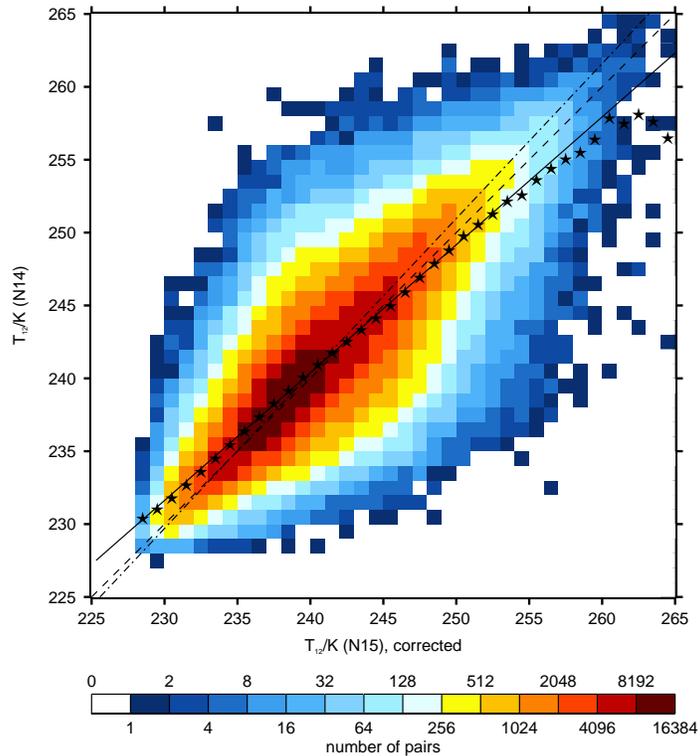
**Figure 3.** As Fig. 1, but with UTHi retrieved from modified channel 12 brightness temperatures for HIRS 3 on NOAA 15, according to eq. 3. While the data points no longer have an excess above the (red dashed) diagonal, another problem appears: The range of UTHi retrieved from N15 data is drastically reduced to values slightly exceeding 90%, that is, instead of reducing the number of supersaturation cases, they have been eliminated, an undesired effect.



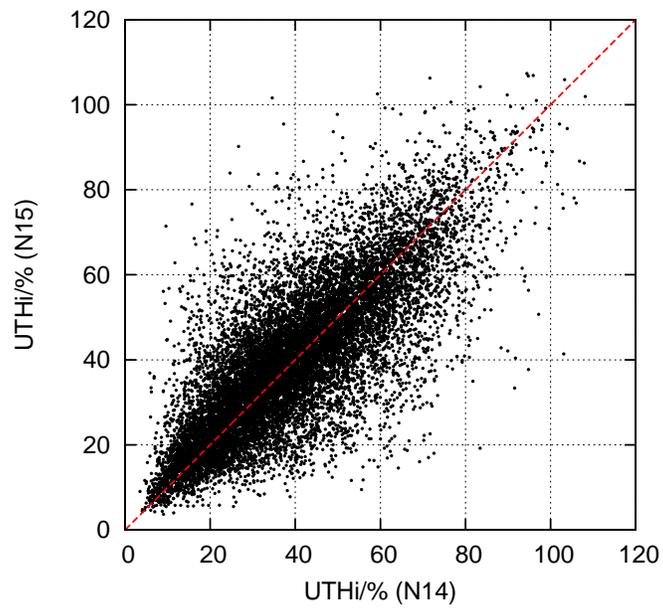
**Figure 4.** Cumulative distribution functions (cdf) of channel 12 brightness temperatures, measured with HIRS 2 on N14 (red) and with HIRS 3 on N15 (blue). Note the quite large discrepancy (in relative terms) between both cdfs at low values of  $T_{12}$ .



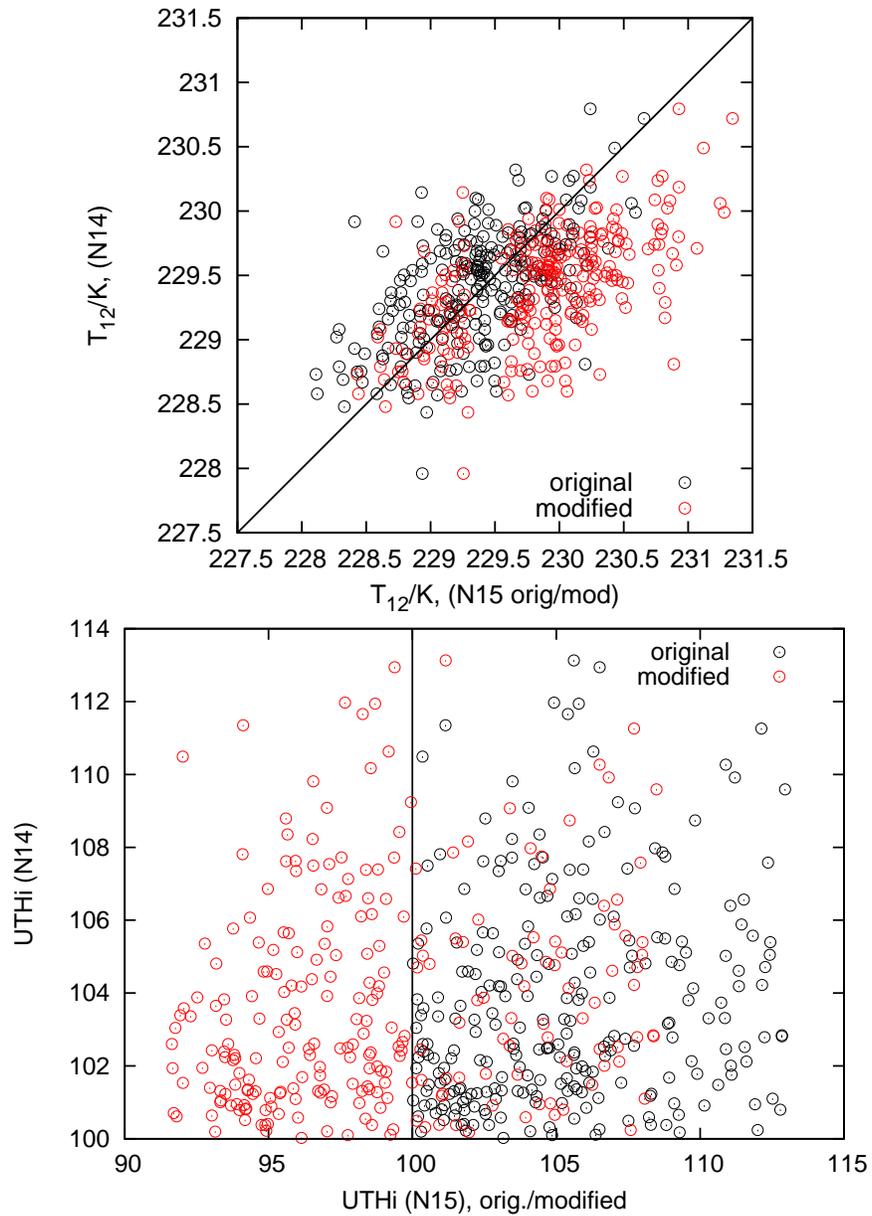
**Figure 5.** Corrections determined for 1 K bins using the cdf-based procedure described in the text. Note that this procedure leaves all data exceeding 240 K unchanged and that the necessary corrections at lower brightness temperatures are smaller than the regression based corrections. The respective values are given in each interval for convenience of the potential user.



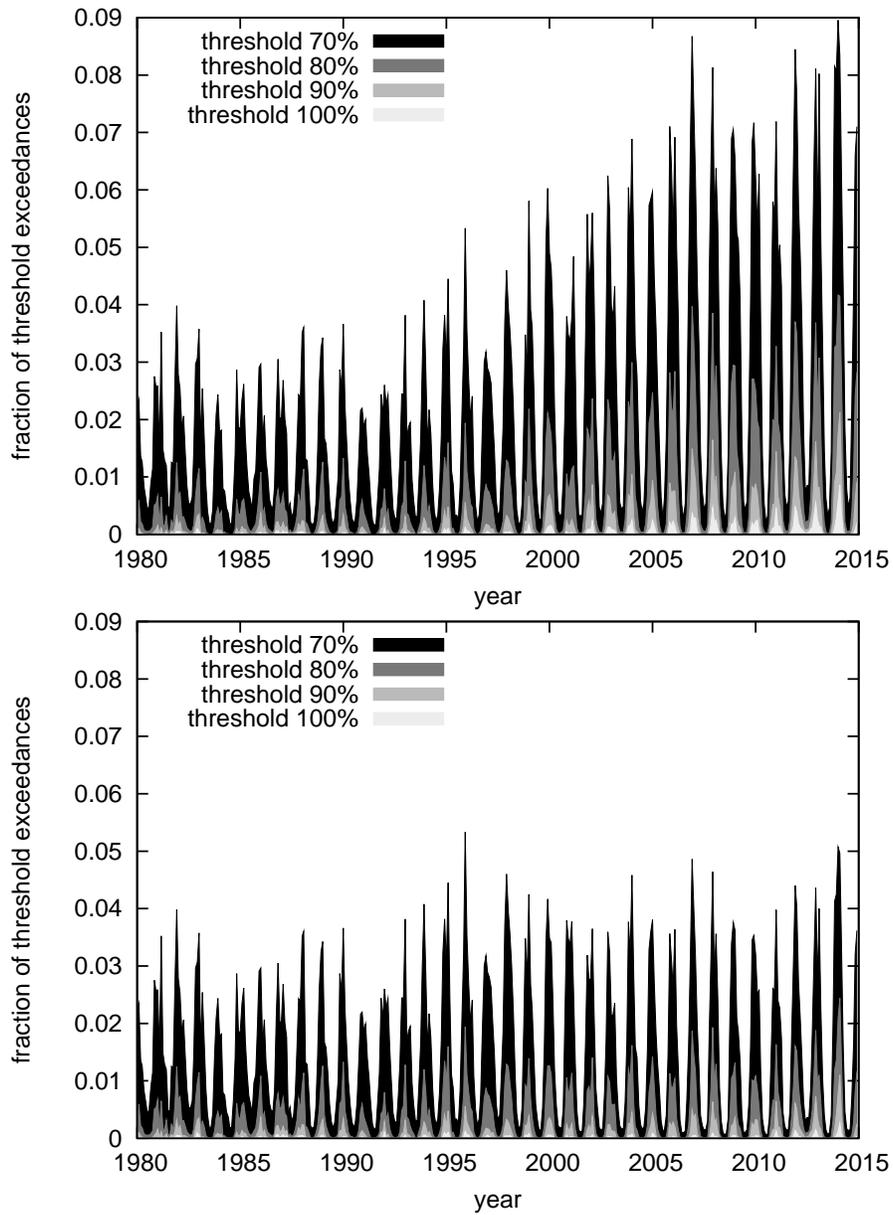
**Figure 6.** As Fig. 2, but after correction of  $T_{12}$  (N15) with the cdf-based procedure described in the text. The gravity centre of the joint distribution (dark red pixels) is now following the diagonal axis (dashed black line), however **both regression lines, the ordinary least squares (solid) and the bivariate (dash-dotted), are tilted against the diagonal.** Marginal means of  $T_{12}$  (N14) for each 1 K interval of the corrected  $T_{12}$  (N15) are represented by stars; they again closely resemble the **ordinary least squares** linear regression. The tilt between the **ordinary least squares** fit and the diagonal is smaller than in Fig. 2, which means that the **cdf-based** correction brings  $T_{12}$  (HIRS 3) levels closer to  $T_{12}$  (HIRS 2) levels.



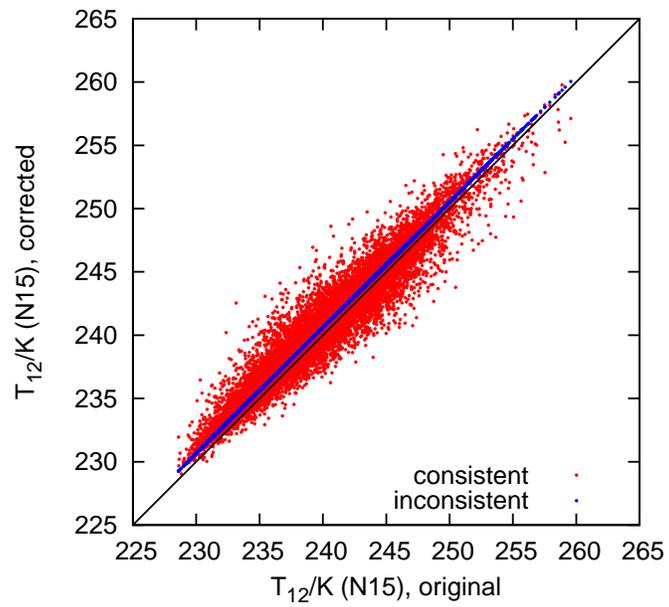
**Figure 7.** As Fig. 3, but with intercalibration via the cumulative distribution function of brightness temperatures. This procedure leaves supersaturated cases in the N15 data set, and the scatter in the upper UTHi range appears more symmetric around the diagonal than in both figures 1 and 3.



**Figure 8.** 256 data pairs where both satellites report ice supersaturation (black points) and their modification after application of the cdf-based intercalibration (red points). Top: Effect of the modification on the N15-measured brightness temperature. Bottom: Effect of the modification on UTHi. More than two thirds of all N15-supersaturation cases are shifted to a UTHi value between 90 and 100%.



**Figure 9.** Raw time series of fraction of exceedances for UTHi thresholds from 70% to 100% before (top) and after (bottom) application of the cdf-based  $T_{12}$  intercalibration for all satellites beginning from N15. **The data until 1998 are identical in both panels.** The raw time series after correction (bottom) does not show peculiar jumps and sudden increases anymore.



**Figure A1.** Consistent (red points) and inconsistent (blue points) use of the bivariate regression coefficients for correction of channel 12 brightness temperatures from NOAA 15. The black line is the diagonal  $y = x$ , included to guide the eye.

**Table 1.** Average fraction of exceedances and corresponding standard deviations (both in percent) for UTHi thresholds from 70% to 100% during 1980–1998 (period before the transition from HIRS 2 to HIRS 3/4), 1999–2005 (transition period) and 2006–2014 (post–transition period), before (a) and after (b) application of the cdf–based  $T_{12}$  intercalibration. (c) shows the differences of the means between (b) and (a).

|                                     | 1980–1998<br>(pre–transition period) | 1999–2005<br>(transition period) | 2006–2014<br>(post–transition period) |
|-------------------------------------|--------------------------------------|----------------------------------|---------------------------------------|
| (a) before cdf–based corrections    |                                      |                                  |                                       |
| 70%                                 | $1.59 \pm 1.11$                      | $2.85 \pm 2.03$                  | $3.80 \pm 2.70$                       |
| 80%                                 | $0.33 \pm 0.36$                      | $0.83 \pm 0.84$                  | $1.31 \pm 1.25$                       |
| 90%                                 | $0.07 \pm 0.10$                      | $0.21 \pm 0.28$                  | $0.40 \pm 0.49$                       |
| 100%                                | $0.01 \pm 0.02$                      | $0.04 \pm 0.07$                  | $0.10 \pm 0.15$                       |
| (b) after cdf–based corrections     |                                      |                                  |                                       |
| 70%                                 | $1.59 \pm 1.11$                      | $1.68 \pm 1.37$                  | $1.73 \pm 1.50$                       |
| 80%                                 | $0.33 \pm 0.36$                      | $0.41 \pm 0.47$                  | $0.51 \pm 0.58$                       |
| 90%                                 | $0.07 \pm 0.10$                      | $0.09 \pm 0.12$                  | $0.14 \pm 0.20$                       |
| 100%                                | $0.01 \pm 0.02$                      | $0.01 \pm 0.02$                  | $0.02 \pm 0.04$                       |
| (c) differences between (b) and (a) |                                      |                                  |                                       |
| 70%                                 | 0                                    | –1.17                            | –2.07                                 |
| 80%                                 | 0                                    | –0.42                            | –0.80                                 |
| 90%                                 | 0                                    | –0.12                            | –0.26                                 |
| 100%                                | 0                                    | –0.03                            | –0.08                                 |