Atmospheric
Measurement
Techniques

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[1] and **Kostas Eleftheratos**[2]

[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. 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 to 6.5 μm. 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 (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 calculating 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 in 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 timescale 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 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 were installed. The transition from HIRS 2 to HIRS 3 involved a shift of the central wavelength in channel 12, from 6.7 to 6.5 μm. Unfortunately, this is not as small a change as it may appear. The atmosphere is nearly 1.5 times as opaque at 6.5 than at 6.7 μm (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 1 km higher in the atmosphere for HIRS 3 and 4 than for HIRS 2 (see 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; however, the layers strongly overlap due to large half widths of the corresponding weighting kernels of, say, 4 to 5 km. As temperature decreases on average by $6.5\,\mathrm{K\,km^{-1}}$ in the troposphere, the change of the wavelength and the corresponding increase in 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 perform an intercalibration of the channel 12 brightness temperature, T_{12}, of all NOAA satellites, using N12 as a reference. For each satellite they compute monthly and zonal averages, with 10° latitude belts centred on 85° S to 85° N. Thus, they obtain a set of mean brightness temperature values $T_{L,\mathrm{YM}}^{N}$, 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 $T_{L,\mathrm{YM}}^{N} - T_{L,\mathrm{YM}}^{N+1}$, i.e. 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 intercalibrated 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 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 exceed 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 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 on 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 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 intercalibration process is also valid for values found at the low tail of the distribution of T_{12} when it comes to calculating 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 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 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 in Sect. 2. A couple of simple results from the new method are presented, and the new method is discussed in comparison to more traditional methods in Sect. 3. Finally, our results are summarised, conclusions are drawn and an outlook on future research necessities and possibilities is given in Sect. 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. Note that the intercalibration was based on monthly and zonal averages of T_{12} – or, 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 $2.5° \times 2.5°$ grid boxes of the 30 to 70° N zone, and 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 730 473 data pairs.

Figure 1 shows a scatter plot of a random selection of 2 % of the data pairs for the upper-tropospheric humidity with
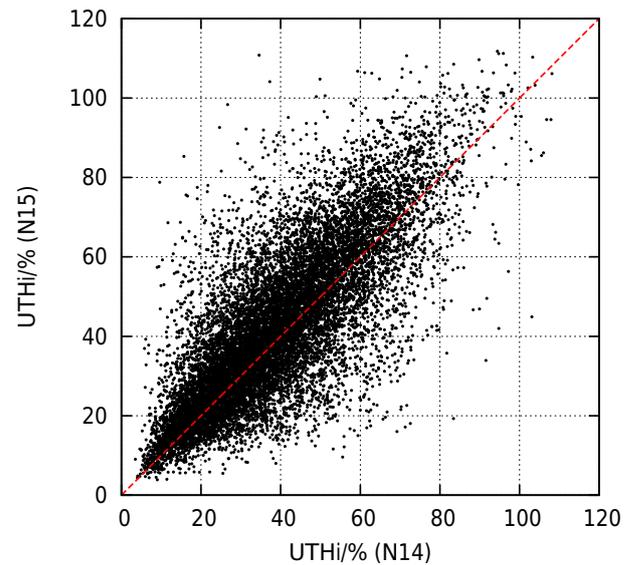

Figure 1. Scatter plot of data of upper-tropospheric humidity with respect to ice (UTHi, in percent), retrieved from channel 12 brightness temperatures from the HIRS 2 instrument on NOAA 14 and from the corresponding HIRS 3 instrument on NOAA 15. The data pairs represent daily average values taken in $2.5° \times 2.5°$ 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.

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 UTHi > 100 %, 2739 records of N15 have 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.
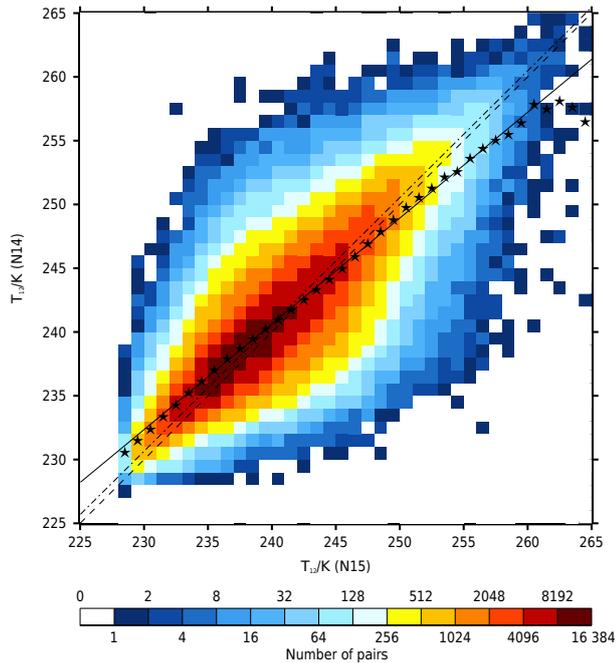
Figure 2. Two-dimensional histogram of $\{T_{12}(N15), T_{12}(N14)\}$ pairs at 1 K resolution. Ideally the gravity centre of the joint distribution (dark-red pixels) would follow the diagonal axis (dashed black line), but it is slightly shifted above the axis. The ordinary least squares linear fit is 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.

$\{T_{12}(N15), T_{12}(N14)\}$ at 1 K resolution. Note 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 ± 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 centred on the diagonal axis. In particular at low $T_{12}(N15)$ there appears to be a tendency of N15 to display lower values than N14, leading to the observed surplus of supersaturation cases relative to N14.

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:

$$(y/\mathrm{K}) = 41.63 + 0.8292\,(x/\mathrm{K}), \tag{1}$$

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 provides barely any improvement.) The relatively small value of the slope is the result of an effect termed "regression dilution" or "regression attenuation" (Cantrell, 2008; Pitkänen 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

$$(y/\mathrm{K}) = 2.05 + 0.994\,(x/\mathrm{K}), \tag{2}$$

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 such a way as if the fit were an OLS fit. Thus, we do not use the bivariate regression for correction of the N15 brightness temperatures.

The stars in Fig. 2 represent the marginal means of $T_{12}(N14)$ per 1 K interval of $T_{12}(N15)$. These represent binwise 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 also useful for the retrieval of upper-tropospheric humidity values has yet to be shown.

We have performed 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

$$\left(\frac{\widehat{T_{12}}}{\mathrm{K}}\right) = 41.63 + 0.8292 \left(\frac{T_{12}}{\mathrm{K}}\right). \tag{3}$$

The resulting scatter plot of the corresponding values of UTHi is shown in Fig. 3 in the same format as in Fig. 1.

It 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.

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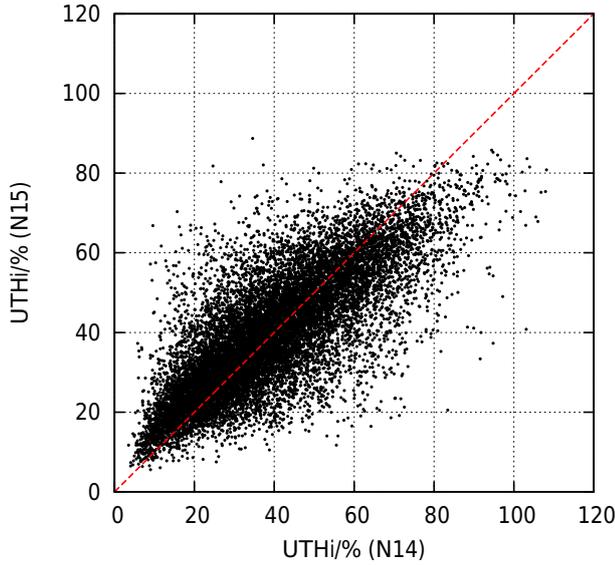
Figure 3. As Fig. 1, but with UTHi retrieved from modified channel 12 brightness temperatures for HIRS 3 on NOAA 15, according to Eq. (3). While the data points no longer have an excess above the (red dashed) diagonal, another problem appears: the range of UTHi retrieved from N15 data is drastically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

events in N14 data, the new regression-based intercalibration eliminates all supersaturation. The comparison of this feature between N14 and N15 has in 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) only lead to minor modifications. The basic problem remains – that is, the strong elimination of high UTHi values and the complete loss of supersaturation. 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.

2.3 Intercalibration via the distribution function

The goal of the new intercalibration exercise is to have a 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. The cumulative distribution functions (cdf's) of the corresponding channel 12 brightness temperatures (Fig. 4) disclose the origin of the difference in supersaturation cases: there are many more (exceeding a factor of 3) 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
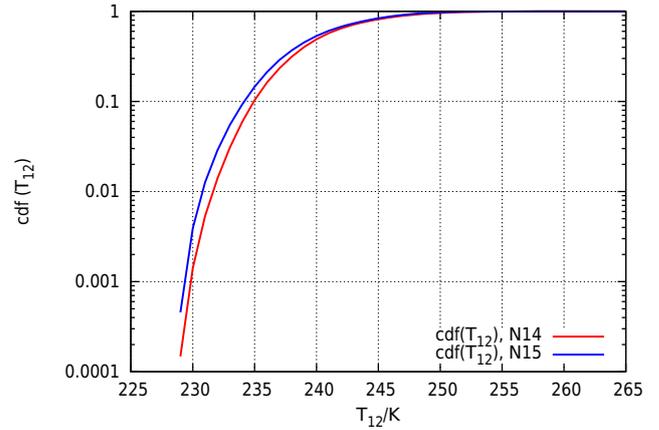

Figure 4. Cumulative distribution functions (cdf) of channel 12 brightness temperatures, measured with HIRS 2 on N14 (red) and with HIRS 3 on N15 (blue). Note the quite large discrepancy (in relative terms) between both cdf's at low values of T_{12}.

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 and the data in each bin are counted, resulting in numbers n_t^s (where the 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 ΔT_1 such that, if all T_{12} in the first bin of the N15 data set are incremented by this value, the surplus δn_1 of these 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 δn_1 additional values that have been shifted from the foregoing bin. The process is stopped either when a bin is reached where the ratio of the two cdf's 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 cdf's, not their difference. This has the consequence that the corrections approach zero as the cdf's 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.

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

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

which gives a Δ of approximately 1 K. Indeed, the maximum correction ΔT_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 become smaller as well, but at a low rate such that the maxi-
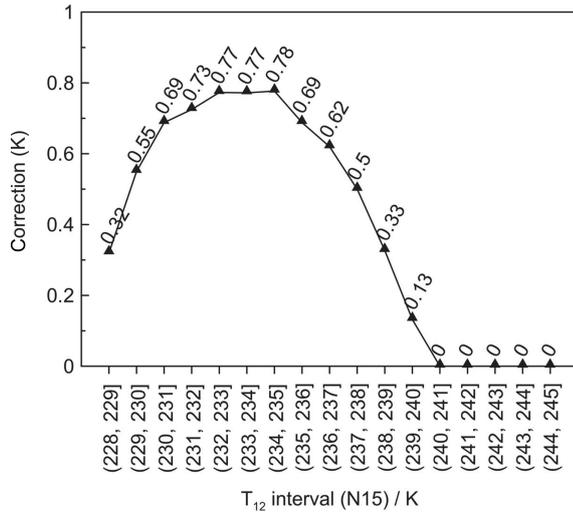
Figure 5. Corrections determined for 1 K bins using the cdf-based procedure described in the text. Note that this procedure leaves all data exceeding 240 K unchanged and that the necessary corrections at lower brightness temperatures are smaller than the regression-based corrections. The respective values are given in each interval for convenience for 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.

mum correction can exceed Δ, 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. (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. The bivariate fit (dash-dotted line) also crosses the middle of the distribution's gravity centre. The best OLS fit (solid black line) is still tilted against the diagonal; its equation is

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

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 bivari-
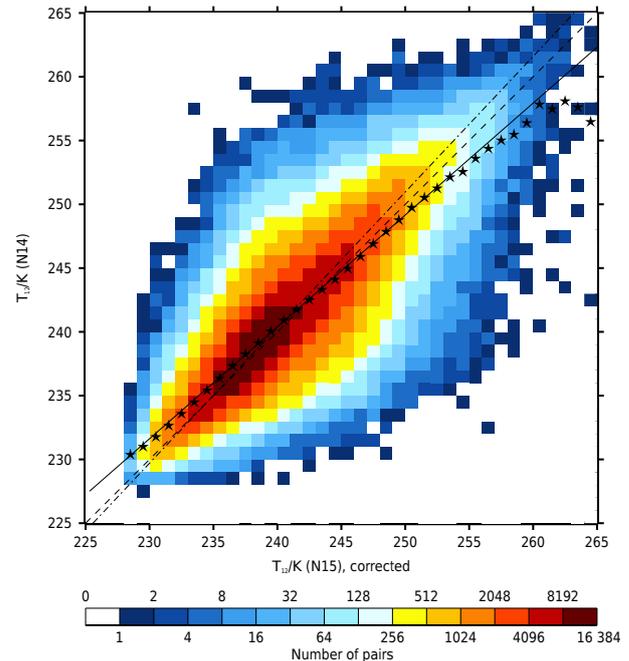
ate fit line has worse parameters than before the correction (see the Appendix), although it fits the data better in the central region. How is this possible? In Fig. 2 (original data) the bivariate fit is nearly parallel to the diagonal but lies above it, clearly reflecting the problem of data pairs concentrating 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} (N14), which, so to speak, rotates the data in the dark-red patch and the lower values clockwise. Accordingly, the bivariate fit became a bit steeper than before (slope 1.06) and its intercept moved 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 (see Figs. 1 and 3).
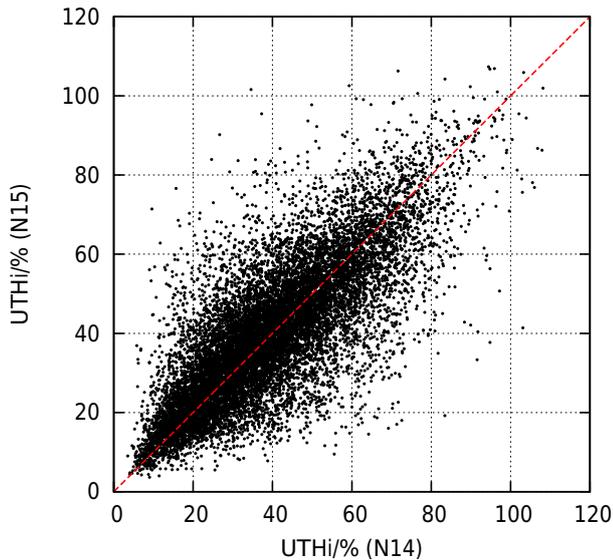
Figure 7. As Fig. 3, but with intercalibration via the cumulative distribution function of brightness temperatures. This procedure leaves supersaturated cases in the N15 data set, and the scatter in the upper UTHi range appears more symmetric around the diagonal than in both Figs. 1 and 3.

It is not necessary to show the T_{12} cumulative distribution functions after the correction; these are almost equal by 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 an improvement indeed 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 ± 2.76) K in the original data (mean and 1 standard deviation). With the correction applied to N15 it reduces to (-0.35 ± 2.70) K.

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

Thus, the mean temperature difference is almost halved, and the mean UTHi difference is even reduced by a factor of 6.

3.2 Simple applications

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
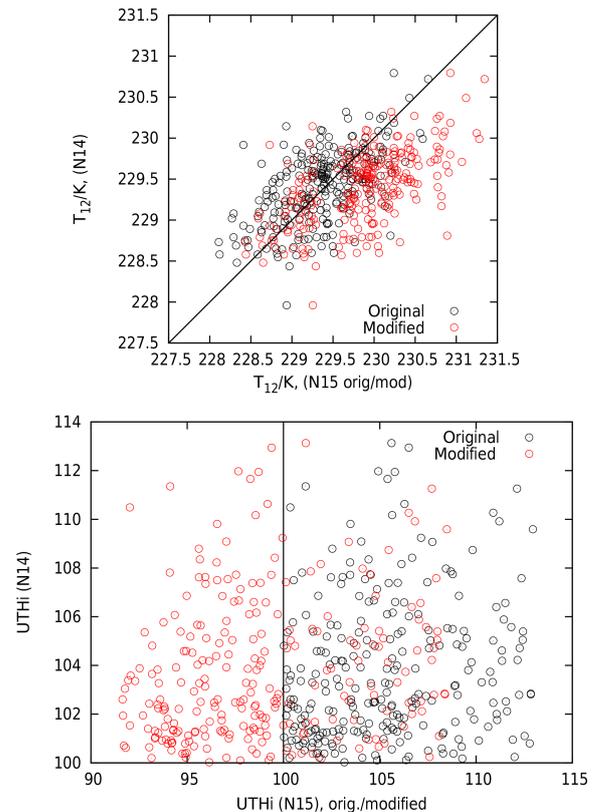

Figure 8. Collection of 256 data pairs where both satellites report ice supersaturation (black points) and their modification after application of the cdf-based intercalibration (red points). Top: effect of the modification on the N15-measured brightness temperature. Bottom: effect of the modification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

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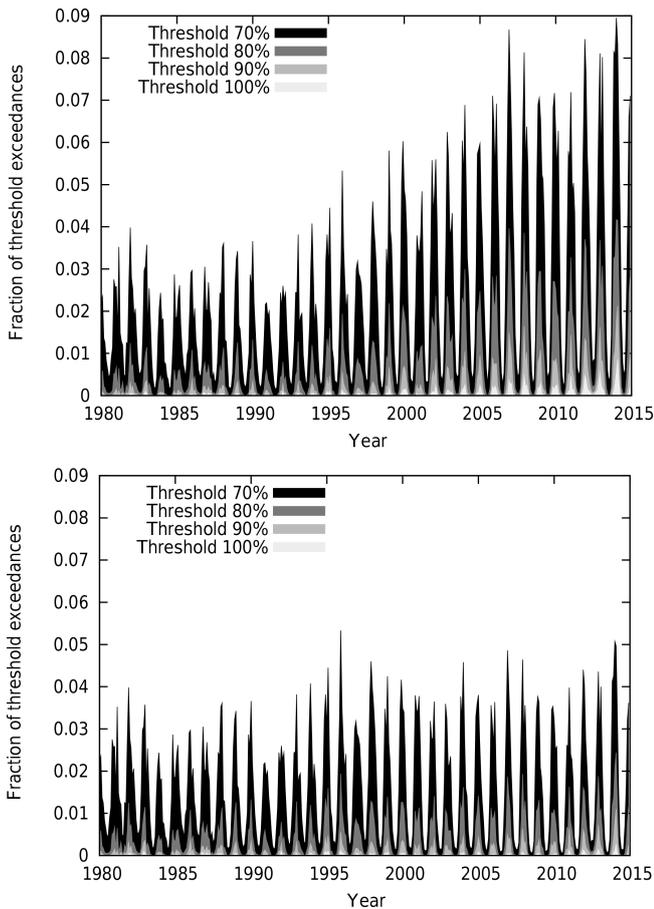
Figure 9. Raw time series of fraction of exceedances for UTHi thresholds from 70 to 100 % before (top) and after (bottom) application of the cdf-based T_{12} intercalibration for all satellites beginning from N15. The data until 1998 are identical in both panels. The raw time series after correction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

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 whether there are any trends in the data at all. A deeper analysis of the four time series will be reported in a forthcoming paper.

3.3 Discussion

There are, in fact, 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 tropo-

sphere, layers that overlap heavily but whose centres are more than 1 km 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 such as "Under which circumstances is it justified or not?" and "Which assumptions have to be made about the structure of temperature and moisture profiles?" This technical note is not the place to answer these questions, but it certainly deserves 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.

In order to discuss the second question, an analysis and comparison of what is effectively done in the cdf-based and the regression-like methods are needed. 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 cdf's, one could just 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 considers 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. Both views have 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. However, 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, no decision can be made regarding 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. Thus, 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 analysing high UTHi values, is provided in Table 1, which shows the average fraction of UTHi 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.

For the case of the 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 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. How-

ever, 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 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 so dramatically just at the transition to HIRS 3. The increase in the frequency of threshold exceedances is not small; it is more than a 3σ increase when we compute the σ from the first 10 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 in UTHi in large regions of the northern midlatitudes 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 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 (e.g. Buehler et al., 2008) or of free-tropospheric humidity data from geostationary satellites (e.g. Schröder et al., 2014). However, this research is beyond the scope 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 the low end of T_{12}. The values are indeed small, but 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 assessing 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.

Finally, we want to stress that it is of great importance 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 intercal-

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. Part (**c**) shows the differences of the means between (**b**) and (**a**).

	1980–1998 (pre-transition period)	1999–2005 (transition period)	2006–2014 (post-transition period)
	(**a**) before cdf-based corrections		
70 %	1.59 ± 1.11	2.85 ± 2.03	3.80 ± 2.70
80 %	0.33 ± 0.36	0.83 ± 0.84	1.31 ± 1.25
90 %	0.07 ± 0.10	0.21 ± 0.28	0.40 ± 0.49
100 %	0.01 ± 0.02	0.04 ± 0.07	0.10 ± 0.15
	(**b**) after cdf-based corrections		
70 %	1.59 ± 1.11	1.68 ± 1.37	1.73 ± 1.50
80 %	0.33 ± 0.36	0.41 ± 0.47	0.51 ± 0.58
90 %	0.07 ± 0.10	0.09 ± 0.12	0.14 ± 0.20
100 %	0.01 ± 0.02	0.01 ± 0.02	0.02 ± 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

ibration 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 information on the upper tail of UTHi is needed for cloud research and for determining how cloudiness will change with climate change. For questions of the radiative balance of the Earth, it is important to know how the very dry regions of the subsidence zones (termed "radiator fins" by Pierrehumbert, 1995) behave with ongoing climate change; thus, the dry end 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 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° × 2.5° grid boxes. The originally intercalibrated data show a very strong increase in very 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 implausible.

The new intercalibration method is constructed in such 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.

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 6.

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 the instruments 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 within the scope of this paper to begin an investigation of this difficult problem, but it is certainly a topic for the near future.

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.

Appendix A: Bivariate regression

Following the suggestion of one reviewer, we added bivariate regression lines to Figs. 2 and 6. These are computed as follows. First, we determine the covariance matrix, $\mathbf{C}$, 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 be for real symmetric matrices. We multiply the larger eigenvalue by a scalar such that its x component becomes unity; then the resulting y component is identical to the slope of the bivariate regression line. The latter crosses the bivariate mean, $(\overline{x}, \overline{y})$; thus, we have a direction and one point which determines the whole line, including its intercept at $x = 0$.

Specifically, for the original data we have

$$\mathbf{C} = \begin{pmatrix} 23.0041 & 19.0753 \\ 19.0753 & 22.7789 \end{pmatrix}. \tag{A1}$$

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 bivariate regression line is 2.04681.

For the cdf-corrected data we have

$$\mathbf{C} = \begin{pmatrix} 20.5694 & 18.0412 \\ 18.0412 & 22.7789 \end{pmatrix}. \tag{A2}$$

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 simply inappropriate to derive corrections to the quantity on the x axis from it. A fixed value of x, as the OLS regression and regression of the first kind assume, is needed to correct x. 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
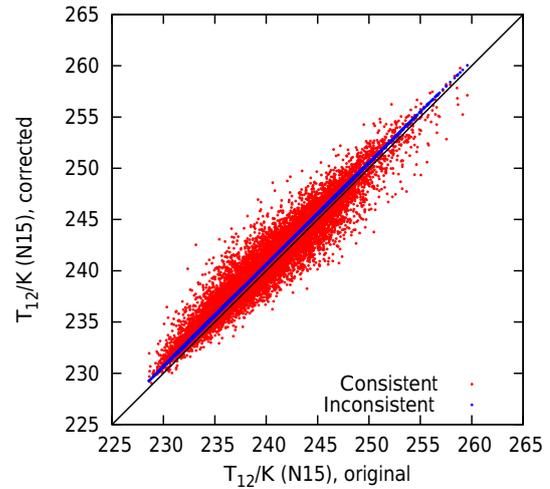

Figure A1. Consistent (red points) and inconsistent (blue points) use of the bivariate regression coefficients for correction of channel 12 brightness temperatures from NOAA 15. The black line is the diagonal $y = x$, included to guide the eye.

formula is

$$\hat{T}_{12/15} = A + \frac{B}{1 + B^2}\left(T_{12/15} + B\,T_{12/14} - AB\right), \tag{A3}$$

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

$$\hat{T}_{12/15} = A + B\,T_{12/15}. \tag{A4}$$

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 be, 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. However, 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; Pitkänen 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 squares regression methods.

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, D14110, doi:10.1029/2007JD009314, 2008.

Cantrell, C. A.: Technical Note: Review of methods for linear least-squares fitting of data and application to atmospheric chemistry problems, Atmos. Chem. Phys., 8, 5477–5487, doi:10.5194/acp-8-5477-2008, 2008.

Chung, E.-S., Soden, B., Sohn, B., and Shi, L.: Upper-tropospheric moistening in response to anthropogenic warming, P. Natl. Acad. Sci. USA, 111, 11636–11641, 2014.

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:10.1002/2015JD024496, 2016.

Dessler, A. and Davis, S.: Trends in tropospheric humidity from reanalysis systems, J. Geophys. Res., 115, D19127, doi:10.1029/2010JD014192, 2010.

Dickson, N. C., Gierens, K. M., Rogers, H. L., and Jones, R. L.: Probabilistic description of ice-supersaturated layers in low resolution profiles of relative humidity, Atmos. Chem. Phys., 10, 6749–6763, doi:10.5194/acp-10-6749-2010, 2010.

Gierens, K. and Eleftheratos, K.: Upper tropospheric humidity changes under constant relative humidity, Atmos. Chem. Phys., 16, 4159–4169, doi:10.5194/acp-16-4159-2016, 2016.

Gierens, K., Kohlhepp, R., Spichtinger, P., and Schroedter-Homscheidt, M.: Ice supersaturation as seen from TOVS, Atmos. Chem. Phys., 4, 539–547, doi:10.5194/acp-4-539-2004, 2004.

Gierens, K., Spichtinger, P., and Schumann, U.: Ice supersaturation, in: Atmospheric Physics, Background – Methods – Trends, edited by: Schumann, U., 9, 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, doi:10.5194/acp-14-7533-2014, 2014.

Irvine, E. A. and Shine, K. P.: Ice supersaturation and the potential for contrail formation in a changing climate, Earth Syst. Dynam., 6, 555–568, doi:10.5194/esd-6-555-2015, 2015.

Jackson, D. and Bates, J.: Upper tropospheric humidity algorithm assessment, J. Geophys. Res., 106, 32259–32270, 2001.

Lamquin, N., Gierens, K., Stubenrauch, C. J., and Chatterjee, R.: Evaluation of upper tropospheric humidity forecasts from ECMWF using AIRS and CALIPSO data, Atmos. Chem. Phys., 9, 1779–1793, doi:10.5194/acp-9-1779-2009, 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.

Pierrehumbert, T.: Thermostats, radiator fins, and the local runaway greenhouse, J. Atmos. Sci., 52, 1784–1806, 1995.

Pitkänen, M., Mikkonen, S., Lehtinen, K., Lipponen, A., and Arola, A.: Artificial bias typically neglected in comparison of uncertain atmospheric data, Geophys. Res. Lett., 43, 10003–10011, doi:10.1002/2016GL070852, 2016.

Roca, R., Guzman, R., Lemond, J., Meijers, J., Picon, L., and Brogniez, H.: Tropical and extra-tropical influences on the distribution of free 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, 11129–11148, doi:10.5194/acp-14-11129-2014, 2014.

Shi, L. and Bates, J.: Three decades of intersatellite-calibrated High-Resolution Infrared Radiation Sounder upper tropospheric water vapor, J. Geophys. Res., 116, D04108, doi:10.1029/2010JD014847, 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, 16669–16688, 1993.

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 best straight line, Am. J. Phys., 72, 367–375, doi:10.1119/1.1632486, 2004.