A Cautionary Note on the Use of Gaussian Statistics in Satellite-Based UTH Climatologies

V. O. John, S. A. Buehler, and N. Courcoux

Abstract—This letter presents a cautionary note on the assumption of Gaussian behavior for upper tropospheric humidity (UTH) derived from satellite data in climatological studies, which can introduce a wet bias in the climatology. An example study using European Centre for Medium-Range Weather Forecasts reanalysis data shows that this wet bias can reach up to 6 %RH, which is significant for climatological applications. A simple Monte Carlo approach demonstrates that these differences and their link to the variability of brightness temperatures are due to a log-normal distribution of the UTH. This problem can be solved by using robust estimators such as the median instead of the arithmetic mean.

Index Terms—Climatology, Gaussian statistics, median, upper tropospheric humidity (UTH).

I. INTRODUCTION

UPPER tropospheric humidity (UTH) is important for the Earth's climate system because it is one of the prime factors that control the outgoing longwave radiation [1], [2]. Global measurements of UTH are made using satellite-based infrared and microwave remote sensing techniques. In the infrared, UTH is estimated from the 6.7- μm channel measurements, onboard geostationary and polar-orbiting satellites [3]–[5], whereas in the microwave UTH is estimated from the 183.31 ± 1.00 -GHz channel [e.g., Advanced Microwave Sounding Unit B (AMSU-B) channel 18] measurements, onboard polar-orbiting satellites [6], [7]. There is a simple relation, derived by Soden and Bretherton [3], to estimate UTH from radiances (which is applicable to both infrared and microwave measurements)

$$\ln(\text{UTH}) = a + b T_B \quad (1)$$

where $\ln(\text{UTH})$ is the natural logarithm of Jacobian weighted mean of the fractional relative humidity in the upper troposphere, T_B is the radiance expressed in brightness temperature, and a and b are constants. The constants are different for microwave and infrared cases or if one uses different definitions of the Jacobian [7]. It is shown in Buehler and John [7] that this method works well for AMSU-B channel 18 brightness temperatures. The absolute error in UTH obtained by this method is 2 %RH at low UTH values and 6 %RH at high UTH

values. The European Centre for Medium-Range Weather Forecasts (ECMWF) 60-level sampled dataset [8] was used to derive the coefficients a and b in (1). The coefficients were derived for all AMSU-B viewing angles (45 angles from 0.55° to 48.95° from nadir in 1.1° steps) so that the brightness temperatures do not need to be limb corrected. The relation basically provides a tool to transform the brightness temperatures to an intuitive quantity, relative humidity, rather than a retrieval in the normal sense.

Averaging of brightness temperature is generally done in different spatial and temporal scales before making a UTH climatology. The extent of the averaging can vary from a target area of a few tens of kilometers to $2.5^\circ \times 2.5^\circ$ or bigger latitude–longitude grid boxes in the spatial scale [9], [10]. In the temporal, it can vary from a single satellite overpass to several days or up to monthly averages [4], [10]. UTH is then calculated by transforming the averaged brightness temperature using (1). These UTH values can be finally averaged to calculate a UTH climatology. Here, averaging implies taking the simple arithmetic mean. Taking the simple arithmetic mean is correct only when the variable is normally distributed. It is clear from (1) that an exponential function applied to values distributed normally leads to a log-normal distribution.

In reality, neither brightness temperatures nor UTH may be normally distributed. Soden and Bretherton [3] and Zhang *et al.* [11] have already pointed out the non-Gaussian behavior of UTH. Therefore, the use of the simple arithmetic mean here can lead to spurious wet or dry biases in satellite-based UTH climatologies. This study investigates this issue by the use of a simple Monte Carlo approach. Furthermore, the suitability of the median as a simple robust estimator in such problems is also demonstrated.

The letter is organized as follows. Section II describes an evidence of a possible wet bias seen from an analysis using ECMWF data. Section III introduces a Monte Carlo way of quantitatively explaining the results found in Section II. Section IV presents a discussion on the real distribution of AMSU-B brightness temperatures. Section V presents an example of the use of robust estimation, and Section VI gives summary and conclusions.

II. UTH CLIMATOLOGY FROM ECMWF DATA

ECMWF reanalysis fields [12] of temperature and water vapor for one year (December 2001 to November 2002) were used to study this problem. The fields are available four times a day (00, 06, 12, and 18 UTC) on a $1.5^\circ \times 1.5^\circ$ latitude–longitude grid and a 60-level vertical grid which ranges from surface to 10 Pa. A fast radiative transfer model, RTTOV-7 [13], was

Manuscript received June 6, 2005; revised August 22, 2005. This work was supported by the German Federal Ministry of Education and Research (BMBF), within the AFO2000 project UTH-MOS, under Grant 07ATC04. It is a contribution to COST Action 723 "Data Exploitation and Modeling for the Upper Troposphere and Lower Stratosphere."

The authors are with the Institute of Environmental Physics, University of Bremen, D-28359 Bremen, Germany (e-mail: vojoh@uni-bremen.de).

Digital Object Identifier 10.1109/LGRS.2005.859350

Fig. 1. Difference between UOM and MOU (in percent relative humidity).

used to generate AMSU-B channel 18 brightness temperatures from these profiles. Since the atmosphere is very dry in the polar regions and since the ice-covered areas of Greenland and Antarctic are elevated, the application of (1) to brightness temperature is not valid in those regions [5]. Therefore, our analysis is limited to the latitudes from 60°S to 60°N.

It should also be noted that the application of (1) to brightness temperature is valid only if there are no ice clouds or precipitating clouds present because these clouds have a nonnegligible effect on the brightness temperature since the radiation is scattered by ice particles or liquid drops by which the clouds are composed of. But we performed clear sky radiative transfer calculations, therefore the issue of clouds is not relevant here. This is an issue when real AMSU-B data are used, which is discussed in Section IV.

Taking one year of ECMWF data we have 4×365 values at each grid point. The UTH climatology is made in two different ways. In the first case, brightness temperatures at each grid point are first averaged and then transformed to UTH using (1). Therefore, this is the UTH of the mean brightness temperature (UOM). Averaging for one year is an extreme case, not normally taken in practice, but we used a long time period to have enough number of brightness temperatures to make reasonable statistics. In the second case, brightness temperatures are transformed to UTH using (1) and the UTH values are then averaged, giving the mean of UTH (MOU). This represents the other extreme case, not always feasible in practice, since some averaging prior to the UTH transformation may be necessary, for example, for cloud clearance or for noise reduction. The difference between the two climatologies is defined as

$$\begin{aligned} \Delta \text{UTH} &= \text{UTH}(\overline{T_B}) - \overline{\text{UTH}(T_B)} \\ &= \text{UOM} - \text{MOU} \end{aligned} \quad (2)$$

and is shown in Fig. 1. The difference is negative everywhere on the map, and the magnitude of the difference varies from 0 to -6 %RH. There are two main questions to be answered here: 1) Why are the two climatologies different? and 2) Why is the difference always negative? Before answering the questions, it is interesting to note that large differences exist where the variability of the brightness temperature is higher as shown in Fig. 2. This figure shows the standard deviation of the brightness temperature for the entire time period. One can see a good correlation between the variability of the brightness temperature and the difference in the climatologies.

Fig. 2. Standard deviation of brightness temperature (in Kelvin) generated using ECMWF profiles for one year (December 2001 to November 2002).

Fig. 3. Distributions of brightness temperature and UTH. (Left) Gaussian distributions of brightness temperatures with a mean of 245 K and four different standard deviations: 2 K (solid line), 4 K (dotted line), 8 K (dashed line), and 12 K (dashed-dotted line). (Right) UTH distributions for corresponding brightness temperature distributions. Note the log-normal behavior of UTH distributions.

III. MAPPING BRIGHTNESS TEMPERATURE DISTRIBUTIONS TO UTH DISTRIBUTIONS

In order to investigate the differences between UOM and MOU, a simple Monte Carlo type of approach was used to map brightness temperature distributions to UTH distributions. It is visible from (1) that a UTH distribution should behave log-normally if the brightness temperatures are normally distributed, as discussed in the introduction. The reason why we map brightness temperature distributions to UTH distributions as described in this section is to quantitatively understand the difference between UOM and MOU. A similar approach was used to study an apparent supersaturation in the upper troposphere due to temperature measurement errors [7], [14].

As a first step, AMSU-B channel 18 brightness temperatures were generated by creating random ensembles of T_B^{18} values representing a Gaussian distribution with a mean at 245 K and standard deviations varying from 1 to 12 K. Some examples of these brightness temperature distributions are shown in the left panel of Fig. 3. Each of these random ensembles consists of 10^7 values, in order to make smooth distributions. A mean value of 245 K for channel 18 brightness temperatures is reported in John and Buehler [15] for midlatitude regions. These brightness temperature ensembles can be used to simulate UTH ensembles using (1), some examples are shown in the right panel of Fig. 3.

The UTH distributions which are presented in the right panel of Fig. 3 show a log-normal behavior as expected, the distributions get more and more right skewed as the standard deviation of the T_B^{18} distributions increases. Higher values of UTH

Fig. 4. Solid line represents the behavior of the difference between UTH of the mean and the mean of UTH, and the dashed line represents the difference between UTH of the median and the median of UTH as a function of the standard deviation of the T_B^{18} distribution.

occur for colder brightness temperatures. A small decrease in T_B^{18} in the vicinity of the steepest portion of the $UTH-T_B^{18}$ exponential curve will result in a large increase in UTH. As the standard deviation of the T_B^{18} distribution increases, the occurrence of colder brightness temperatures also increases. This results in a large right skewing of the UTH distributions. Note that a log-normal behavior of UTH was reported in Soden and Bretherton [3]. Thus, the means of these distributions are also shifted toward the right, giving a higher value for the mean of UTH.

The reason for the difference between UOM and MOU is the log-normal behavior of the UTH distribution. The differences are always negative due to the right skewed (log-normal) nature of the UTH distribution. It is also clear that the difference will increase if the standard deviation of the T_B^{18} distribution is increased. The solid curve in Fig. 4 shows the difference as a function of the standard deviation of the T_B^{18} distribution.

IV. AMSU DATA

The differences reach up to -20 %RH for a standard deviation of 12 K. But in the case of ECMWF, as shown in Fig. 1, the differences were only about -6 %RH for about 12 K standard deviation. This hints at the fact that real AMSU brightness temperatures are not normally distributed as assumed. Real AMSU data have been used to investigate this. Three years (2001–2003) of AMSU data from NOAA 16 satellite were selected for 42 latitude–longitude grid points over Europe. These grid points correspond to some of the radiosonde stations over Europe. The details of the data can be seen in John and Buehler [15]. The distribution of T_B^{18} is shown (solid line) in Fig. 5.

There are about 1.5 million data points in the distribution. As pointed out in Section II, (1) cannot be applied to brightness temperatures which are contaminated by clouds. In contrast to the clear-sky simulations shown in Section II, the real AMSU-B data can be contaminated by clouds. In order to remove brightness temperatures which are contaminated by clouds we use a simple filter which make use of channel 18 brightness temperature together with that of channel 20. As channel 20 is sensitive to the lower troposphere, T_B^{20} should be warmer than T_B^{18} under

Fig. 5. Distribution of AMSU-B channel 18 brightness temperature (solid curve). The data is taken from NOAA-16 satellite for three years (2001–2003). The dashed curve stands for a Gaussian distribution with the same mean and standard deviation of the real data.

clear-sky conditions. When T_B^{18} is warmer than T_B^{20} , it can be considered as either contaminated by clouds [16], [17] or by surface [7]. Therefore, such cases are excluded from the brightness temperature distribution shown in Fig. 5.

The mean of this distribution is 246.03 K, and the standard deviation is 5.48 K. The minimum brightness temperature is 210.75 K, and the maximum is 271.94 K. Note the non-Gaussian nature of the T_B^{18} distribution. A Gaussian distribution (dashed curve) with the same mean and the standard deviation is also shown in Fig. 5 for a direct comparison. The most important feature is the rare occurrence of very low brightness temperature values in the original T_B^{18} distribution. There are only 117 values which are less than 230 K in the original distribution, whereas there are 2667 values less than 230 K in the Gaussian distribution. The presence of a large number of low brightness temperature values resulted in larger UTH values; thus, ΔUTH was larger when a Gaussian T_B^{18} distribution was assumed. This explains why ΔUTH in Fig. 1 is smaller than that predicted by the Monte Carlo method.

The brightness temperature values shown in Fig. 5 correspond to all possible viewing geometries of the AMSU-B instrument. Therefore, it is not possible to transform all of them to UTH by a single pair of a and b values in (1). Fig. 6 shows the distributions of T_B^{18} and UTH, only for two viewing angles: the nadir (0.55°) and the maximum off-nadir (48.95°). There are 43 048 nadir brightness temperatures and 20 650 maximum off-nadir brightness temperatures. The reason for the difference in number for the nadir and the maximum off-nadir viewing angles is that the pixels were selected from a target area of fixed size around each latitude–longitude grid point, but pixel size at the off-nadir viewing angle is more than double that at nadir. The mean and the standard deviation of the nadir and off-nadir T_B^{18} distributions are 247.63 and 240.58 K and 5.22 and 4.74 K, respectively. The difference between UOM and MOU is -2.54 %RH for the nadir and -2.41 %RH for the maximum-off nadir viewing angles. Therefore, the difference between UOM and MOU can be also observed using the real AMSU-B data. This is also true for the other viewing angles where ΔUTH is about -2.5 %RH.

Fig. 6. Distributions of AMSU-B channel 18 brightness temperature and the transformed UTH values. Distributions are shown only for the nadir (dotted line) and the maximum off-nadir (solid line) angles to avoid clutter.

These numbers are consistent with the results from the ECMWF data, for example, over northern midlatitudes the standard deviation values are about 5 K (see Fig. 2).

It should be noted that Soden and Bretherton [3] reported a log-normal behavior for UTH (see [3, Fig. 16]), and a physical explanation was also given by them making use of a reference model of water vapor transport. Interestingly, we do not find a large log-normal behavior in the UTH distributions, at least over the midlatitude region. This is an important point for further investigation, but it is beyond the scope of this letter.

Another interesting point to note is that although the nadir and maximum off-nadir brightness temperature distributions show a difference of 7 K due to the limb effect (for channel 18 it is limb darkening, brightness temperature decreases as the instrument scans away from the nadir due to a longer path length the radiation has to travel for an off-nadir viewing angle [18]) as shown in the left panel of Fig. 6, the distributions of the retrieved UTH values do not show this difference as demonstrated in the right panel of Fig. 6. This is due to the fact that the coefficients a and b used to retrieve UTH are different for different viewing angles and the limb effect was accounted for while deriving these coefficients. This is an important observation because the method of estimating UTH from AMSU-B brightness temperatures does not demand limb corrected brightness temperatures, which makes the method easy to be used.

V. ADVANTAGE OF THE MEDIAN

In the previous sections, it was demonstrated that the use of the arithmetic mean to describe the central tendency of non-Gaussian distributions can introduce spurious errors due to the nonresistant nature of the mean. It is well-known that the normal arithmetic mean is not a good parameter to describe the central tendency of a non-Gaussian distribution [19]. The performance of the simplest robust estimator, the median, as suggested by Lanzante and Gahrs [20], was tested. The dashed line in Fig. 4 shows the difference between UTH of the median and the median of UTH versus the standard deviation. In contrast to the difference of UOM and MOU, this is zero and shows no deviation with increase in standard deviation. This is due to the monotonic mapping of T_B^{18} to UTH. Thus, the use of median solves the discrepancy of using UOM and MOU. Also, other percentiles can

be used to denote the variability in the data instead of the standard deviation.

VI. SUMMARY AND CONCLUSION

This letter describes a cautionary note on the use of Gaussian statistics in satellite-based UTH climatologies due to non-Gaussian behavior of UTH and also of brightness temperatures. The brightness temperature transformation method acts as a tool for interpreting the brightness temperatures in a more intuitive quantity, the relative humidity. However, if arithmetic means are used the mean UTH after the transformation will depend on to what extent averaging has been done before the transformation. The two different extreme approaches were investigated, either doing all the averaging before the transformation (UOM case) or doing all the averaging after the transformation (MOU case). The second approach is the only theoretically correct one, but difficult in practice. It was shown by an example using ECMWF reanalysis data that the results of the two extreme methods differ significantly, up to 6 %RH. These differences show a dependence on the standard deviation of the brightness temperatures, that is, the differences increases with the standard deviation.

The investigation with the aid of a simple Monte Carlo approach illustrated the relationship of the difference to the logarithmic UTH transformation. Interestingly, the Monte Carlo approach overestimated the differences observed with the ECMWF data. A further investigation with real AMSU-B data showed that this overestimation was due to the assumption of Gaussian distribution for the brightness temperatures in the Monte Carlo method. The differences obtained from the real AMSU data and simulated AMSU data were found to be consistent.

It was demonstrated that the use of a robust estimator, the median, resolves the difference between the two climatologies. Thus, the median has two important advantages. First, it more correctly represents the central tendency for a nonnormally distributed variable such as UTH. Second, the median of the brightness temperatures can be directly transformed to the median of UTH, and so can other percentiles, a property that greatly simplifies the practical calculation of the climatology. Unfortunately, it is the custom to report only mean values in the UTH climatology publications [3]–[5], [21], [22]. The only work which emphasized the use of median in the analysis of UTH data is by Lanzante and Gahrs [20], although it was not for a climatological application. Our study confirms the usefulness of robust statistics in climate applications, therefore we recommend the use of median or other robust estimators over the use of the simple arithmetic mean. If means are desired, we recommend to at least quote in the publication also the median or other percentiles. Although the discussion was only for UTH, the results may also be applicable to other climatological studies.

Although not directly related to this note, it was observed that the transformed UTH values are almost free of limb effect which was clearly present in the brightness temperatures. This is due to the fact that the coefficients used to transform brightness temperatures to UTH were derived for each instrument viewing angle, thus the limb effect was accounted for. This improves

the practical use of the method which does not demand limb correction of brightness temperatures.

ACKNOWLEDGMENT

The authors thank L. Neclos (CLASS, NOAA) for providing AMSU data, ECMWF for the reanalysis data, and T. R. Sreerekha for constructive comments on the manuscript.

REFERENCES

- [1] I. M. Held and B. J. Soden, "Water vapor feedback and global warming," *Annu. Rev. Energy Environ.*, vol. 25, pp. 441–475, 2000.
- [2] J. E. Harries, "Atmospheric radiation and atmospheric humidity," *Q. J. R. Meteorol. Soc.*, vol. 123, pp. 2173–2186, Oct. 1997.
- [3] B. J. Soden and F. P. Bretherton, "Upper tropospheric relative humidity from the GOES 6.7 μ m channel: Method and climatology for July 1987," *J. Geophys. Res.*, vol. 98, no. D9, pp. 16 669–16 688, Sep. 1993.
- [4] —, "Interpretation of TOVS water vapor radiances in terms of layer-average relative humidities: Method and climatology for the upper, middle, and lower troposphere," *J. Geophys. Res.*, vol. 101, no. D5, pp. 9333–9343, Apr. 1996.
- [5] J. J. Bates and D. L. Jackson, "Trends in upper-tropospheric humidity," *Geophys. Res. Lett.*, vol. 28, no. 9, pp. 1695–1698, May 2001.
- [6] R. J. Engelen and G. L. Stephens, "Comparison between TOVS/HIRS and SSM/T-2 derived upper tropospheric humidity," *Bull. Amer. Meteorol. Soc.*, vol. 79, no. 12, pp. 2748–2751, 1998.
- [7] S. A. Buehler and V. O. John, "A simple method to relate microwave radiances to upper tropospheric humidity," *J. Geophys. Res.*, vol. 110, no. D02110, 2005. DOI: 10.1029/2004JD005111.
- [8] F. Chevallier, "Sampled databases of 60-level atmospheric profiles from the ECMWF analysis," ECMWF, Reading, U.K., EUMETSAT SAF Program Res. Rep. 4, 2001.
- [9] J. J. Bates and D. L. Jackson, "A comparison of water vapor observations with AMIP I simulations," *J. Geophys. Res.*, vol. 102, pp. 21 837–21 852, 1997.
- [10] J. J. Bates, X. Wu, and D. L. Jackson, "Interannual variability of upper tropospheric water vapor band brightness temperature," *J. Clim.*, vol. 9, pp. 427–438, 1996.
- [11] C. Zhang, B. E. Mapes, and B. J. Soden, "Bimodality in tropical water vapor," *Q. J. R. Meteorol. Soc.*, vol. 129, no. 594, pp. 2847–2866, 2003.
- [12] P. Kallberg, A. Simmons, S. Uppala, and M. Fuentes, "The ERA-40 Archive," ECMWF, Reading, U.K., ERA-40 Project Rep. Series 17, 2004.
- [13] R. W. Saunders, M. Matricardi, and P. Brunel, "An improved fast radiative transfer model for assimilation of satellite radiances observations," *Q. J. R. Meteorol. Soc.*, vol. 125, pp. 1407–1425, 1999.
- [14] S. A. Buehler and N. Courcoux, "The impact of temperature errors on perceived humidity supersaturation," *Geophys. Res. Lett.*, vol. 30, no. 1759, 2003. DOI: 10.1029/2003GL017691.
- [15] V. O. John and S. A. Buehler, "Comparison of microwave satellite humidity data and radiosonde profiles: A survey of European stations," *Atmos. Chem. Phys.*, vol. 5, pp. 1843–1853, 2005.
- [16] B. A. Burns, X. Wu, and G. R. Diak, "Effect of precipitation and cloud ice on brightness temperatures in AMSU moisture channels," *IEEE T. Geosci. Remote.*, vol. 35, no. 6, pp. 1429–1437, Nov. 1997.
- [17] G. Hong, G. Heygster, J. Miao, and K. Kunzi, "Detection of tropical deep convective clouds from AMSU-B water vapor channels measurements," *J. Geophys. Res.*, vol. 110, no. D05205, 2005. DOI: 10.1029/2004JD004949.
- [18] S. A. Buehler, M. Kuvatrov, V. O. John, U. Leiterer, and H. Dier, "Comparison of microwave satellite humidity data and radiosonde profiles: A case study," *J. Geophys. Res.*, vol. 109, no. D13103, 2004. DOI: 10.1029/2004JD004605.
- [19] J. R. Lanzante, "Resistant, robust and nonparametric techniques for the analysis of climate data: Theory and examples, including applications to historical radiosonde station data," *Int. J. Climatol.*, vol. 16, pp. 1197–1226, 1996.
- [20] J. R. Lanzante and G. E. Gahrs, "The clear-sky bias of TOVS upper-tropospheric humidity," *J. Clim.*, vol. 13, pp. 4034–4041, Nov. 2000.
- [21] R. W. Spencer and W. D. Braswell, "How dry is the tropical free troposphere? implications for global warming theory," *Bull. Amer. Meteorol. Soc.*, vol. 78, no. 6, pp. 1097–1106, Jun. 1997.
- [22] J. J. Bates, D. L. Jackson, F.-M. Breon, and Z. D. Bergen, "Variability of tropical upper tropospheric humidity 1979–1998," *J. Geophys. Res.*, vol. 106, no. D23, pp. 32 271–32 281, Dec. 2001.