

this threshold, suggesting that ICH  $\text{RH}_{\text{liq}}$  might be biased for values under 20 %. However, the observed discrepancy in the  $\text{RH}_{\text{liq}} > 10 \%$  range could also reflect a scenario where no consistent bias exists across specific  $\text{RH}_{\text{liq}}$  ranges. Instead, a broad distribution of measurements around a biased mean state may be present. In such a case, a mean moist bias for a given  $\text{RH}_{\text{liq}}$  range would increase the PDF's representation of  $\text{RH}_{\text{liq}}$  values at higher levels.

Non-linear bias behavior could be attributed to uncertainties in the calibration the sensor's offset drift, which occurs between the routine ICH calibrations conducted every three months (Petzold et al., 2020). Although an in-flight calibration is performed to account for this sensor drift (Smit et al., 2008), uncertainties in the process (as noted in Smit et al. (2008)) can introduce non-linear variations in the bias of individual measurements between calibration intervals. Consequently, while this intercomparison study cannot determine the behavior of the bias for individual measurements, it does provide insights into the bias of mean values used in climatological studies.

The sampling bins with  $\overline{\text{RH}}_{\text{liq}}$  values below 20 % exhibit significant systematic moist biases, with relative differences of 100 % or more for  $\overline{\text{RH}}_{\text{liq}}$  of 10 % and less. For layers closer to the TTP, there is a lower but still noteworthy bias. During summer, certain bins have  $\overline{\text{RH}}_{\text{liq}}$  values of 15 % or less already in the first kilometer above the TTP, whereas the same levels relative to the TTP during other seasons indicate much higher values. This seasonality is attributed to the higher temperatures observed in the respective sampling bins during the summer season, and biases of the mean values are in the range of only about 1 to 3 %  $\text{RH}_{\text{liq}}$  during this season close to the TTP.

For the comparison of CORE and CARIBIC, we used data between 2018 and 2022 only. This specific time frame was chosen because before 2018, a grounding issue between the sensor and the data acquisition unit caused a large noise on the signal and thus a reduced quality flag of the data. Therefore, we decided to utilize only data with the highest quality unaffected by this issue.

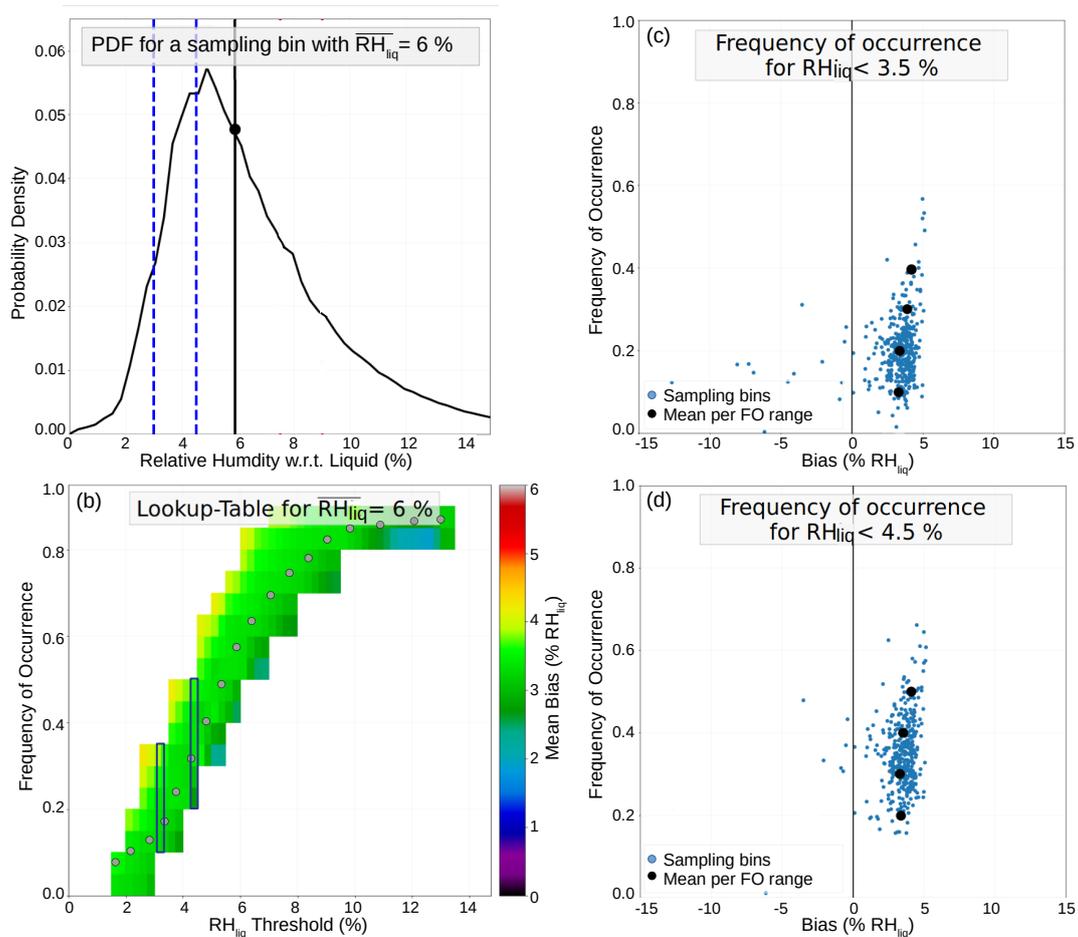
On average, the comparison of CORE and CARIBIC shows similar mean biases in the LMS, similar to what the comparison with MOZAIC revealed. However, for CORE, there is a stronger variation of the bin mean values compared to MOZAIC, which can be attributed to differences in the temporal coverage (2018-2022 and 1995-2022) of CORE and CARIBIC, respectively, due to the large year-to-year variability of UT/LMS  $\text{H}_2\text{O}$  (Kunz et al., 2008).

## 4 Adjustment of IAGOS-MOZAIC&CORE LMS $\text{H}_2\text{O}$ to IAGOS-CARIBIC

### 4.1 Adjustment methodology

Following the mapping approach outlined in Section 3.1, the next step is to apply an algorithm to adjust the biased MOZAIC&CORE  $\text{H}_2\text{O}$  data, using CARIBIC  $\text{H}_2\text{O}$  as a reference. A fixed bias for individual measurements, however, cannot be defined as a function of  $\text{RH}_{\text{liq}}$  due to the sensor offset drift at 0 %  $\text{RH}_{\text{liq}}$  (see Section 3.3). As a result, it is also not feasible to directly adjust ICH data from IAGOS flights using data from campaign flights equipped with ICH and high-precision  $\text{H}_2\text{O}$  instruments, like FISH.

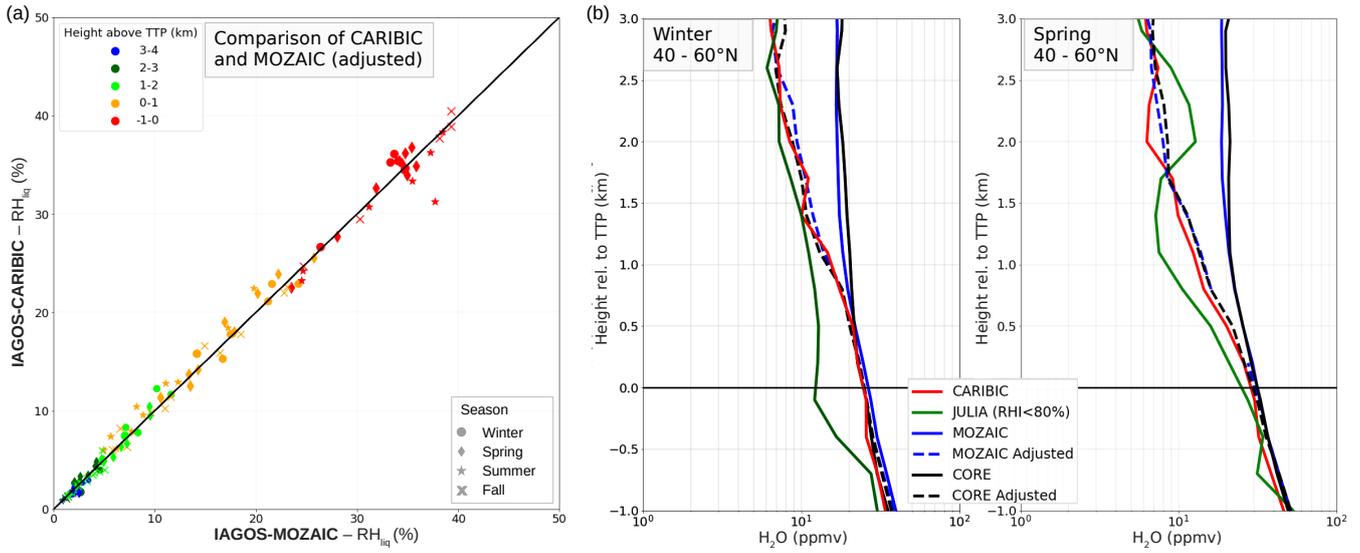
Given the limitations mentioned above, the primary objective of our methodology is to adjust the sampling bin mean values  $\overline{\text{RH}}_{\text{liq}}$ . These mean values, based on thousands of individual measurements, represent a climatological state. One potential



**Figure 7. Application example of the IAGOS adjustment algorithm.** (a) Example  $RH_{liq}$  frequency distribution of a sampling bin with a mean of  $\overline{RH_{liq}} = 6\%$ . For this mean value, (b) shows the lookup-table with the values used as adjustment based on different  $RH_{liq}$  thresholds and the corresponding cumulative distribution (y-axis). For the two thresholds indicated by the blue dotted lines in (a), the plots (c) and (d) show the derivation of the mean bias (black dots) based on the weighted mean of the sampled mean values (blue dots).

approach is to align the probability density functions (PDFs) of the MOZAIC&CORE data with those of the reference dataset (CARIBIC); see Figure 6b for illustration. However, because CARIBIC has fewer measurements, the PDFs for its sampling bin mean values are noisier and more variable, making it difficult to directly match the distributions. As a result, we pursue an alternative approach: analyzing how the probability distribution function of the data influences the mean bias, and adjusting accordingly. The steps of the approach are described below:

1. **Segmenting the distribution and analyzing the bias by frequency of occurrence (FO):** We divide the  $RH_{liq}$  frequency distribution of each sampling bin into smaller segments. For each segment, we compute the frequency of occurrence (FO) of values falling below specific  $RH_{liq}$  thresholds (FO thresholds = 0.5 - 15 %  $RH_{liq}$ , with step size of 0.5 %). This allows



**Figure 8. Application of the adjustment algorithm on the IAGOS data.** Panel (a) shows a comparison of the same sampling bins between IAGOS-MOZAIC and IAGOS-CARIBIC as shown in Figure 6 but with the adjustment algorithm applied to the mean values. Panel (b) shows two mean vertical UT/LMS H<sub>2</sub>O profiles of IAGOS-CARIBIC, JULIA, and IAGOS-MOZAIC&CORE (adjusted and unadjusted).

us to determine how often certain  $\overline{RH_{liq}}$  values occur in the distribution. As an example, in Figure 6a, the FO for the  
 345 10 % threshold is shown. Figure 7a illustrates this concept for a distribution having a mean  $\overline{RH_{liq}}$  of 6 %, with the blue  
 lines indicating two specific RH<sub>liq</sub> thresholds,  $RH_{liq} < 3.5$  % and  $RH_{liq} < 4.5$  %.

We perform this segmentation for the distributions of all MOZAIC&CORE sampling bins that fall in between certain  
 $\overline{RH_{liq}}$  ranges, with a step size of 1 %  $\overline{RH_{liq}}$  ( $\overline{RH_{liq}} = 0.5 - 1.5$  %,  $1.5 - 2.5$  %, ...). For each of the bins in the respective  
 $\overline{RH_{liq}}$  ranges, the bias to the CARIBIC sampling bins is calculated. Next, we also consider the different FO thresholds and sort the biases as a function of the amount of measurements falling below these thresholds. For the two RH<sub>liq</sub>  
 350 thresholds indicated in Figure 7a (blue dashed lines), and based on all sampling bins with  $\overline{RH_{liq}}$  in the range of 5.5  
 – 6.5 %, Figure 7c&d displays the respective biases as a function of the FO (y-axis) is shown (blue dots). Similar plots  
 for  $\overline{RH_{liq}} = 15$  % are presented in Figure A2.

355 2. **Derivation of a mean bias as a function of  $\overline{RH_{liq}}$  and FO:** A regression (black dots in Figure 7c&d) is applied to the  
 correlation between all biases and FO fulfilling the  $\overline{RH_{liq}}$  threshold. The deviations from the regression are on the order  
 of  $\pm 0.5$  %  $\overline{RH_{liq}}$  and thus a robust approximation. By performing this method for various ranges of  $\overline{RH_{liq}}$ , we obtain a  
 series of mean biases as a function of both %  $\overline{RH_{liq}}$  thresholds and the corresponding FO. This enables us to study how  
 the bias varies as a function of the distribution of data points.

360

3. **Interpolation and Construction of Lookup Tables for Bias Correction:** Once the mean biases (black dots) for different FO thresholds are calculated, we apply an interpolation to smooth the fluctuations between the bias values for different  $\overline{\text{RH}}_{\text{liq}}$  ranges. This step is necessary to minimize the effects of distribution uncertainties, leading to more consistent and reliable corrections. The result is a set of lookup tables containing the corrected mean values for specific  $\overline{\text{RH}}_{\text{liq}}$  values and the corresponding FO thresholds. An example of such a lookup table for  $\overline{\text{RH}}_{\text{liq}} = 6\%$  is shown in Figure 7b, where the black boxes highlight the biases corresponding to the two thresholds shown in Figure 7c&d.

4. **Final Bias Calculation and Adjustment:** The final adjusted mean bias is derived by calculating the arithmetic mean of the biases at different FO thresholds. The equation for the final mean bias is given by:

$$\overline{\Delta\text{RH}}_{\text{liq}} = \frac{1}{n} \sum_{i=0}^n \Delta\text{RH}_{\text{liq}}[i], \quad (3)$$

where  $n$  represents the number of FO thresholds. As an example, the biases for different  $\text{RH}_{\text{liq}}$  FO thresholds are illustrated by the grey dots in Figure 7b. The final mean bias,  $\overline{\Delta\text{RH}}_{\text{liq}}$ , is obtained by averaging all the individual biases. This approach ensures that we account for the biases associated with various FO thresholds, leading to a more comprehensive and accurate adjustment of the mean  $\text{RH}_{\text{liq}}$  values.

When the adjustment algorithm is applied to the entire MOZAIC dataset, using the sampled mean values shown in Figure 5, the corrected values exhibit a good agreement with the CARIBIC data, as shown in Figure 8a. The adjusted MOZAIC values now cluster closely around the 1-to-1 line, without any obvious bias. Furthermore, Figure 8b presents vertical profiles of  $\text{H}_2\text{O}$  from MOZAIC (blue), CORE (black), CARIBIC (red), and JULIA (green) during the winter and spring seasons in the 40–60°N latitudinal region. At altitudes of 1 km and more above the TTP, where MOZAIC&CORE exhibited significant biases, the adjusted mean values now align well with the reference data (CARIBIC and JULIA). This confirms that the adjustment methodology is effective and provides reliable mean values across different seasons.

## 4.2 Application and uncertainties of adjusted IAGOS-MOZAIC&CORE $\text{H}_2\text{O}$

### 4.2.1 Uncertainty estimate

While the adjustment improves the agreement between the MOZAIC&CORE and CARIBIC datasets, several sources of uncertainty remain. These include:

- **Measurement uncertainties:** In CARIBIC, uncertainties in the  $\text{RH}_{\text{liq}}$  derivation (from  $\text{H}_2\text{O}$ , temperature, and pressure measurements) introduce a relative bias of about 7 % (see Section 3.2.2).
- **Method uncertainty:** The small number of CARIBIC measurements and differences in the temporal and geographical coverage of the datasets can lead to uncertainties. However, the sampling strategy designed in this study helps reduce such uncertainties.