Correction of temperature and relative humidity biases in ERA5 by bivariate quantile mapping for contrail prediction and classification.

Kevin Wolf[1,a], Nicolas Bellouin[1,2], Olivier Boucher[1], Susanne Rohs[3], and Yun Li[3]

[1]Institut Pierre-Simon Laplace, Sorbonne Université / CNRS, Paris, France
[2]Department of Meteorology, University of Reading, Reading, United Kingdom
[3]Institute of Energy and Climate Research – Troposphere (IEK-8), Forschungszentrum Jülich, Jülich, Germany
[a]now at Leipzig Institute for Meteorology (LIM), Leipzig University, Leipzig, Germany

Correspondence: Kevin Wolf (kevin.wolf@ipsl.fr)

Abstract.

The skill of the atmospheric reanalysis ERA5 from the European Centre for Medium-Range Weather Forecasts (ECMWF) at simulating temperature and relative humidity in the upper troposphere and lower stratosphere is assessed by using five years of In-service Aircraft for a Global Observing System (IAGOS) observations. IAGOS flight trajectories are used to extract co-located meteorological conditions from ERA5, namely temperature and relative humidity, which are compared with the IAGOS measurements. This assessment is particularly relevant to the study of contrail formation, so focuses on the highly frequented airspace that spans the Eastern United States over the North Atlantic and to central Europe. The comparison is performed in terms of mean, median, probability density functions, and a confusion matrix. For temperature a good agreement is identified with a maximum bias of -0.4 K at the 200 hPa level. Larger biases are found for relative humidity up to -5.5 % at the 250 hPa level. A bias correction method based on a multivariate quantile mapping is proposed and applied to remove the systematic biases that are characterized by an atmosphere that is too dry and too cold. After the correction the bias in temperature is reduced to less than 0.1 K and in relative humidity to less than -1.5 %. To estimate potential contrail formation on the basis of ERA5, data points from IAGOS as well as corrected and uncorrected data points from ERA5 - temperature and relative humidity - are flagged for contrail formation using the Schmidt-Appleman criterion. In the IAGOS data set, 44.0 % and 12.1 % of the samples represent conditions for non-persistent contrails and persistent contrails, respectively. The original ERA5 analyses show corresponding numbers of 50.3 % and 7.9 % for non-persistent and persistent contrails, respectively. This suggests an overestimation of non-persistent contrails at the expense of an underestimation of persistent contrails. To improve the representation of both contrail types, the proposed quantile mapping correction method was applied, which largely removed biases in temperature and relative humidity. As a result, the representation of non-persistent and persistent contrail formation potential in corrected ERA5 data has improved, now flagging 44 % and 10.9 % of the data points as non-persistent and persistent, respectively. Despite this improvement, differences in contrail occurrence persist after the correction, which are traced back to the underlying biases in temperature and relative humidity, as well as to the non-linearities in the Schmidt-Appleman criterion. This suggests that the original ERA5 is either too dry and warm or too cold and moist compared to IAGOS. The differences depend on the region, with largest differences at 250 hPa but decreasing biases with with increasing altitude.

Copyright statement. TEXT

1 Introduction

Aviation contributes to global climate warming (Lee et al., 2021). The total contribution by aviation is commonly split into two parts. One fraction is directly attributable to carbon-dioxide (CO_2) and is well quantified. For the year 2018, aviation was estimated to be responsible for 2.5 to 2.6 % of global CO_2 emissions (Friedlingstein et al., 2019; Lee et al., 2021; Boucher et al., 2021). The other contributing fraction to aviation-induced climate change comes from byproducts resulting from fossil fuel combustion, like nitrogen oxides (NO_x), sulfur dioxide (SO_2), and aerosol particles. Furthermore, the combustion of all fuels, regardless whether they are fossil or synthetic, lead to the emission of water vapor (WV) as long as they contain hydrogen.

The effects of WV are receiving increasing attention in recent years as the emitted WV in the engine exhaust allows and triggers the formation of condensation trails, also called contrails (Schumann, 1996; Kärcher, 2018). Optically thin cirrus and contrails are known to have a net warming effect on the climate (Burkhardt and Kärcher, 2011; Schumann et al., 2015; Lee et al., 2021). The influence of a perturbation, e.g., clouds, aerosols, or gases, on the Earth's atmosphere and its radiative transfer is quantified by the radiative forcing (RF). By definition, RF is defined as the difference in the net irradiance at the top of atmosphere under perturbed and unperturbed conditions (Ramanathan et al., 1989). Effective radiative forcing (ERF) includes in addition the radiative impact of adjustments in the troposphere and stratosphere (Bickel et al., 2020). The aviation-induced global CO_2-related ERF is estimated to be around 30 mWm^{-2} (Boucher et al., 2021). Contrail RF is estimated to be stronger, at about 60 mWm^{-2} but is subject to much larger uncertainties (Burkhardt and Kärcher, 2011).

Contrail formation depends on the ambient conditions, which have to be sufficiently cold and moist. The thresholds of temperature, below which a contrail forms, and relative humidity above which a contrail can form, are estimated with the Schmidt–Appleman criterion (SAc, Schmidt, 1941; Appleman, 1953). For a contrail to be persistent (with the common meaning that is has a lifetime longer than 10 minutes), the ambient air has to fulfill the SAc and must also be supersaturated with respect to ice. When these criteria are fulfilled and persistent contrails have formed, they can remain for hours, spread, merge, and increase the total cirrus cloud cover. Employing climate simulations and analyzing satellite observations, Burkhardt and Kärcher (2011) and Quaas et al. (2021) estimated an increase in total cloud cover due to contrail formation of 6 to 10 % in the mid-latitudes of the northern hemisphere, where most of the flights occur.

To lower the climate impact of aviation it is important to reduce CO_2 as well as non-CO_2 effects. An approach to minimize non-CO_2 effects is active flight re-routing to avoid areas where contrails are likely to form and persist, which would require accurate numerical weather predictions. A useful prerequisite is to identify and document flight levels and regions of the Earth's atmosphere that are particularly prone to contrail formation due to meteorological and dynamical conditions that favor contrail formation. Such a statistical data base might be obtained in four different ways.

The first approach builds on ground-based observations. For example, Schumann et al. (2013) used a roof-top camera to infer cirrus properties and contrail occurrence. However, this approach is limited to a single or few locations. In a second approach, satellite observations provide a top-down view with the required global coverage but come with some drawbacks

(Meyer et al., 2002; Minnis et al., 2013). Depending on the sensor and the satellite platform, the temporal or spatial resolutions are often insufficient to detect young contrails with low cloud optical thickness (Kärcher et al., 2009). Furthermore satellite observations, similarly to ground-based observations, can be compromised by underlying cloud layers between the surface and the cirrus.

In a third approach, contrail occurrence can be assessed by model simulations. However, the assessment strongly relies on the accurate representation of the temperature and humidity fields at high altitudes, as well as that of ice cloud amount and microphysical properties, in the model. Contrail modeling can be done interactively or offline. Interactive contrail models are typically implemented in climate models, e.g., Bock and Burkhardt (2016), by simulating ice supersaturated regions and calculating contrail cirrus cover based on aircraft emission inventories. Offline contrail models, such as CoCiP (Schumann, 2012), use meteorological fields to predict contrail formation and evolution to contrail cirrus. A frequent source of meteorological data is ERA5 (Hersbach et al., 2020), a state-of-the-art global modeling system from the European Centre for Medium-Range Weather Forecasts (ECMWF). ERA5 builds on the decade-long improvements of the Integrated Forecasting System (IFS) of ECMWF and replaces its predecessor ERA-interim (Dee et al., 2011). Previous studies showed that the IFS scheme and the associated data assimilation predict well the temperature field, as verified against radiosonde and satellite observations (Dyroff et al., 2015; Carminati et al., 2019). Slightly less accurate is the prediction and re-analysis of relative humidity, which is generally challenging due to the high temporal and spatial variability of WV. Specific issues have been identified in the upper troposphere and lower stratosphere, as well as with the general representation of ice supersaturation. For example, Bland et al. (2021) compared radiosonde observations with operational ECMWF IFS weather forecast and identified a lower stratosphere moist-bias. Similarly, Krüger et al. (2022) compared measurements from a differential absorption Lidar with ECMWF ERA5-reanalysis data (on a relative-tropopause coordinate) and identified a small moist bias in the upper troposphere that increases to a moderate to significant moist bias in the lower stratosphere. Contrarily, studies that compared water vapor concentrations and ice supersaturation in ERA-interim and ERA5 with aircraft in-situ observations found that conditions of ice supersaturation are not frequent enough in those reanalysis products, suggesting a dry bias (Kunz et al., 2014; Dyroff et al., 2015; Gierens et al., 2020; Reutter et al., 2020; Schumann et al., 2021). Consequently, there is no consensus whether ECMWF re-analysis products are subject to a moist or dry bias in the upper troposphere. It is noted that in-situ aircraft observations are potentially biased by avoiding deep-convective clouds and the outflow of such clouds. However, cirrus clouds are typically not avoided (Petzold et al., 2020) and, therefore, a potential sampling issue with respect to cirrus clouds is not expected to play more than a minor role.

In situ measurement campaigns are a potential fourth approach, during which contrails are directly probed and contrail properties are investigated. Dedicated measurement campaigns, for instance by Krämer et al. (2009, 2020) and Voigt et al. (2017), are rare. Furthermore, they may lack spatial representation by targeting specific atmospheric features as well as cloud conditions, which may bias the results (Petzold et al., 2020). Fortunately, the IAGOS data set is different in the way that it covers large areas of North America, the North Atlantic, and Europe, which have now been sampled for around two decades including its predecessor Measurement of OZone and water vapour on Airbus In-service airCraft (MOZAIC; Marenco et al., 1998; Petzold et al., 2017). Comparing and bias correcting ERA5 with IAGOS is insofar important because: i) IAGOS data

has been shown to be reliable; ii) IAGOS samples temperature and relative humidity at exactly at the locations and pressure levels that are relevant to aviation studies; and iii) ERA5 is often used for predicting potential contrail formation. It is important to stress that we do not seek to make a universally applicable correction of humidity in ERA5 but rather provide a corrected humidity to enable better estimates of contrail occurrence. Relying on IAGOS data allows us to do so exactly at the locations and pressure levels that are relevant for aviation studies.

To mitigate the dry bias under conditions close to ice-supersaturation in ERA-interim and ERA5, studies have applied either multiplication factors (Schumann and Graf, 2013; Schumann et al., 2015) or parameterized corrections (Teoh et al., 2022a). However, these proposed corrections do not consider the temperature dependence of humidity nor the spatial variations in the bias, particularly at different pressure levels. In this study we propose a correction for ERA5 data that is based on a bivariate quantile mapping (QM), which is a standard method of model bias correction (Cannon et al., 2015; Cannon, 2016, 2018). The QM method allows the removal of biases based on the statistical distributions of an observed and modeled quantity, for example temperature and relative humidity, with the aim to better estimate the contrail formation potential in air traffic regions. Here, the QM is trained on 3.5 years of IAGOS observations and collocated ERA5 data of temperature and relative humidity. The QM method is then applied on 5.5 years of ERA5 data and compared with IAGOS. Subsequently, we determine the impact of the correction on the representation of non-persistent and persistent contrails with respect to IAGOS. In case of false classifications the underlying differences in simulated and observed temperature and relative humidity are determined to identify systematic shortcoming in ERA5.

Subsequent to this introduction, Sec. 2 describes the data and methods used in this study. After that the results are presented in Sec. 3 and summarized in Sec. 4. The appendices A–C provide detailed information about the IAGOS data analysis.

2 Data and Methods

2.1 In-service Aircraft for a Global Observing System

The In-service Aircraft for a Global Observing System (IAGOS; Petzold et al., 2015) is a framework of commercial aircraft that are equipped with a set of sensors for in situ measurements of meteorological conditions, trace gas concentrations, and cloud properties. Since 2015, all aircraft within the IAGOS framework have been equipped with the 'Package 1' (P1) instrument package system that includes a backscatter cloud probe (BCP) to measure the particle number concentration N_{ice}, and a dedicated sensor 'ICH' that measures temperature T_{P1} and relative humidity r_{P1}. The BCP is a single particle backscattering optical spectrometer to detect cloud particles with sizes between 5 and 75 μm. Light with 658 nm wavelength is emitted by a light emitting diode and directed through a quartz window to the outside of the aircraft fuselage. The light is focused on a narrow range of 4 cm that represents the target area. Cloud particles within the focus backscatter the radiation to a sensor. The intensity of the radiation is proportional to the size, the refractive index, and the shape of the particles as well as the angle under which the particles were hit by the beam. Directly from these measurement the particles size and the particle number concentration N can be derived. More details on the BCP can be found in Beswick et al. (2014). Measurements of N are used to separate for in-cloud ($N \geq 0.015\ \mathrm{cm}^{-3}$) and cloud-free measurements ($N < 0.001\ \mathrm{cm}^{-3}$) following the thresholds given

by Petzold et al. (1997). For intermediate conditions, where $0.001 \leq N < 0.015\ \mathrm{cm}^{-3}$, the measurements cannot be clearly attributed to in-cloud or cloud-free conditions, so they are assigned to the intermediate category (Petzold et al., 2017; Sanogo et al., 2023).

The ICH package is comprised of a capacitive sensor (Humicap-H, Vaisala, Finland) for measurements of relative humidity (defined over liquid water) and a collocated PT-100 platinum sensor for temperature measurements. Both sensors are mounted within a Model 102 BX housing of Rosemount Inc. (Aerospace Division, USA) to minimize heating from solar radiation and thermodynamic effects. The recorded data is post-processed by the IAGOS consortium to correct the raw data following Helten et al. (1998) and Boulanger et al. (2018, 2020). Hereby an "in-flight calibration method" (IFC) correcting an offset drift during the course of the deployment period is applied (Smit et al., 2008; Petzold et al., 2017).

Post-processed data of T_{P1} and r_{P1} are stored every four seconds. However, the response time $t_{1-1/\mathrm{e}}$ of a sensor is an important characteristic as it directly affects the measurements. $t_{1-1/\mathrm{e}}$ is commonly defined as the time that is required by a sensor to adapt to $1 - \frac{1}{e} = 0.6\bar{3}$ of an abrupt change in the measured quantity. The temperature sensor is characterized by a response time $t_{1-1/\mathrm{e}}$ of 4 s and an accuracy of ± 0.5 K . The IAGOS humidity sensor is characterized by an average uncertainty of $\pm 6\ \%$. Including uncertainties from sensor calibration and data post-processing, the uncertainty ranges between 5 % and up to 10 % and increases with decreasing temperature (Helten et al., 1998). The humidity sensor's response time $t_{1-1/\mathrm{e}}$ was determined to be 1 s at 293 K and increases to several minutes at 233 K (Neis et al., 2015). $t_{1-1/\mathrm{e}}$ of the relative humidity sensor increases due to reduced molecular diffusion into and out of the sensors polymer substrate. In a first order approximation, the distance between two IAGOS measurements of T_{P1} and r_{P1} is 0.96 km at a cruise speed of 240 m s^{-1}. However, $t_{1-1/\mathrm{e}}$ of the relative humidity sensor averages these measurements over a distance that ranges between 15 km (253 K) and 50 km (233 K) at cruise altitude.

IAGOS measurements in the lower stratosphere that are typically characterized by low values of relative humidity $\approx r_{\mathrm{P1}} < 10\ \%$, are subject to a moist bias. This moist bias is a non-linear function of the relative humidity and requires a multi-dimensional regression correction that is currently under development (Konjari et al., 2022). Therefore, this known moist bias in IAGOS is not corrected in our analysis and it should be kept in mind that subsequent differences between ERA5 and IAGOS for low values of relative humidity may also be attributable to artifacts in the IAGOS measurements. However, since the focus of this analysis is to investigate contrail formation and persistence, only high values of relative humidity are relevant. Consequently, the moist bias for low relative humidity values in the IAGOS observations has little impact, on our analysis.

In this study, we use only the IAGOS measurements that fulfill the following criteria :

- IAGOS quality flag of T_{P1} and r_{P1} is 'good' and 'limited'
- measurements are located between 30°N and 70°N; 110°W and 30°E
- measurements are between 325 and 150 hPa
- r_{P1} (w.r.t liquid water) is between 0 and 100 %

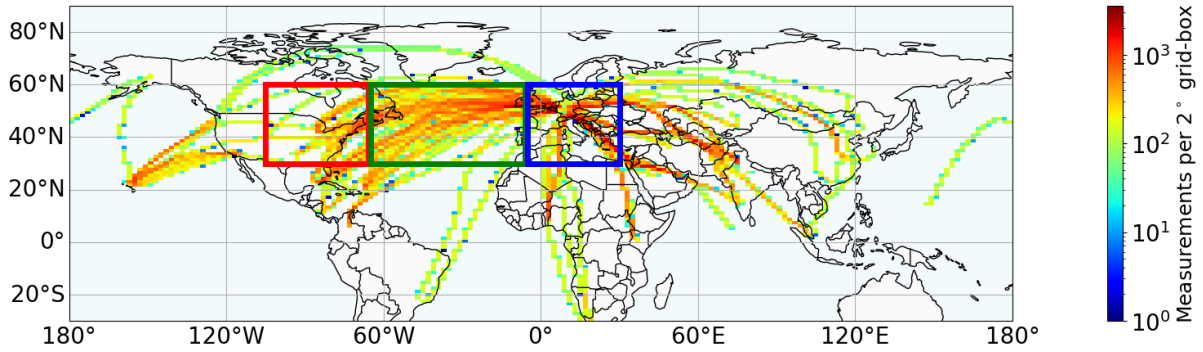

Figure 1. Number of measurements per 2°x2° grid-box of analyzed IAGOS measurements for the years 2015 to 2021 (inclusive). The measurements are filtered for data quality and pressure levels. This study uses measurements in the three boxes: United States (US, red), North Atlantic (NA, green), and continental Europe (EU, blue). Longitude coordinates of the bounding boxes are selected to follow Petzold et al. (2020).

Table 1. ERA5 pressure levels (in hPa) and pressure ranges used to collocate the IAGOS observations.

Pressure level (hPa)	Pressure range (hPa)
300	$275.0 \leq p < 325.0$
250	$237.5 \leq p < 275.0$
225	$212.5 \leq p < 237.5$
200	$187.5 \leq p < 212.5$
175	$150.0 \leq p < 187.5$

While IAGOS has been operated for many years, the global horizontal and vertical coverage remains heterogeneous. Figure 1 shows a density plot of all IAGOS measurements from the years 2015 to 2021 fulfilling the above criteria. Due to the history of IAGOS and the contributing airlines, the highest measurement density is found across the North Atlantic domain (Fig. 1, green, 65°W–5°W). A slightly reduced density is found over North America (Fig. 1, red, 105°W–65°W) and Europe (Fig. 1, blue, 5°W–30°E), particularly towards the western and eastern boundaries of the respective boxes. Outside of the boxes, the coverage is lower and, therefore, we focus our analysis on these three domains. These domains also follow the selection from Petzold et al. (2020).

Figure 2a–e shows the total numbers of measurements per pressure level (p-level) and the fractions attributable to the three sub-domains, which can also be understood as a proxy for the altitude distribution of commercial air traffic in the North Atlantic corridor. The largest number of samples (35.3 %) are found at the 200 hPa level (Fig. 2d). Slightly fewer samples are obtained at the 225 hPa level (Fig. 2c) with 32.0 % and at the 250 hPa level (Fig. 2b) with 26.5 %. Contributions from p-level 300 hPa and 175 hPa are small with 2.5 % and 3.3 %, respectively. Due to the typical flight profiles (for an example see Fig. 10 in Petzold et al. (2015)) the majority of measurements at low p-levels (Fig. 2e) are sampled over Europe, where aircraft reach

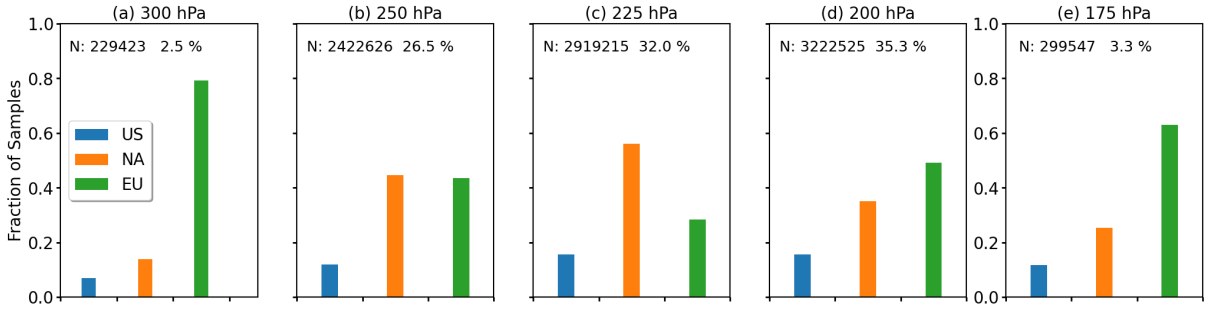

Figure 2. (a–e) Fraction of analyzed IAGOS observations by pressure level separated by sub-domain (see Table 1). The total number of samples per pressure level as well as the fraction with respect to the total sample size are indicated at the top.

their maximum cruising altitude when returning to their main hubs. For intermediate p-levels, the fraction of samples over the North Atlantic is largest compared to the EU and US domain (Fig. 2c), while for the highest p-levels th EU domain dominates again (Fig. 2a), corresponding to where the majority of IAGOS-contributing airlines have their main hubs.

The measurement density is a function of longitude, latitude, and p-level. In addition, the sampling is biased, i.e., by avoiding severe weather and by avoiding or favoring specific atmospheric circulation patterns, such as the jet stream (Petzold et al., 2020). The North Atlantic Flight Tracks (routes typically used by aircraft to cross the Atlantic) are selected on a daily basis to avoid (westbound) or take advantage (eastbound) of the jet stream. This might cause a bias in the sampling of certain atmospheric conditions that might be associated with the jet stream and mid-latitude storm activity (Pasquier et al., 2019).

2.2 ERA5

Meteorological data are obtained from the ECMWF Copernicus Climate Data Store (ECMWF CDS, 2021) using output of the High Resolution component (HRES, T639) of ERA5 (Hersbach et al., 2020). The maximal spatial resolution on the Cartesian grid with $0.25° \times 0.25°$ and maximal temporal resolution of one hour are used. We also use the native vertical resolution with a 50-hPa spacing between 350 and 300 hPa, and a 25-hPa spacing between 300 and 150 hPa. Along-track temperature T_{ERA}, relative humidity r_{ERA}, and cloud fraction $\mathrm{CF}_{\mathrm{ERA}}$ are extracted using the nearest neighbor method, i.e., selecting the ERA5 grid points that are temporally and spatially closest to the IAGOS observations. Spatial and temporal interpolation of relative humidity is not done because the relative humidity depends on the underlying temperature and absolute humidity field, which are both related through the Clausius–Clapeyron relationship. Due to the exponential nature of the Clausius–Clapeyron equation, linear interpolation, for example, could lead to incorrect values of relative humidity.

The current version of the ERA5 data set was generated with the ECMWF Integrated Forecasting System (IFS) cycle Cy41r2, which was operational in 2016. Within ERA5 the relative humidity is provided with respect to liquid water or ice depending on T_{ERA} of the grid box. In general, relative humidity (unitless) is defined as the ratio of the water vapor pressure $e(T)$ to the

saturation water vapor pressure $e_{\mathrm{sat}}(T)$ as:

$$r = \frac{e}{e_{\mathrm{sat}}(T)}. \tag{1}$$

In ERA5, $e_{\mathrm{sat}}(T)$ is given by:

$$e_{\mathrm{sat}}(T) = \alpha \cdot e_{\mathrm{sat,l}}(T) + (1-\alpha) \cdot e_{\mathrm{sat,i}}(T), \tag{2}$$

with $e_{\mathrm{sat,l}}(T)$ and $e_{\mathrm{sat,i}}(T)$ the saturation water vapor pressure over liquid water and ice, respectively. $e_{\mathrm{sat,l}}(T)$ and $e_{\mathrm{sat,i}}(T)$ are given by:

$$e_{\mathrm{sat}}(T) = a_1 \cdot \exp\left\{a_3 \cdot \left(\frac{T-T_0}{T-a_4}\right)\right\}, \tag{3}$$

with a_1 = 611.21 Pa, a_3 = 17.502, and a_4 = 32.19 K for liquid water and a_1 = 611.21 Pa, a_3 = 22.587, and a_4 = -0.7 K over ice; and in both cases T_0 is 273.16 K (Buck, 1981; Alduchov and Eskridge, 1996; ECMWF, 2020). The scaling factor α in Eq. 2 is a piecewise linear function of temperature T determined by:

$$\alpha = \begin{cases} 0 & \text{for} \quad T < T_{\mathrm{ice}} \\ \frac{T-T_{\mathrm{ice}}}{T_0-T_{\mathrm{ice}}} & \text{for} \quad T_{\mathrm{ice}} \leq T < T_0 \\ 1 & \text{for} \quad T_0 \leq T \end{cases} \tag{4}$$

with T_{ice} = 250.16 K and T_0 = 273.16 K. For consistency and comparability with IAGOS all extracted values of relative humidity are converted to be defined solely over liquid water ($\alpha = 1$) or ice ($\alpha = 0$) and are subsequently referred to as r_{ERA} and $r_{\mathrm{ERA,ice}}$, respectively. For consistency, IAGOS relative humidity defined over liquid water is labeled with r_{P1} and defined over ice with $r_{\mathrm{P1,ice}}$.

The fixed (Cartesian) grid resolution of 0.25° of ERA5 does not correspond to a constant longitudinal grid box size in km which instead depends on the latitude. Considering the three sub-domains between 30°N and 70°N, the spatial resolution of one ERA5 grid-box ranges between 24 km (30°N) and 14 km (70°N). Therefore, we assume an average grid box size of 19 km. However, it should be noted that ERA5 is a spectral model with an internal Gaussian resolution of around 31 km and, thus, the effective resolution is coarser than the Cartesian grid resolution (Hersbach et al., 2020). While the IAGOS relative humidity measurements are already smoothed due to the response time of the relative humidity sensor, we additionally smooth the IAGOS measurements by applying a Gaussian filter to account for the mismatch in spatial resolution between IAGOS and ERA5. The standard deviation σ of a Gaussian filter is approximated with:

$$\sigma = \frac{k-1}{6}, \tag{5}$$

which can be regarded as an approximation for a Gaussian distribution, as $3 \cdot \sigma$ includes 99.7 % of the Gaussian distribution. k is the window length of the smoothing filter and achieved by setting σ = 3, based on the assumption of an average cruise speed of around 240 m s^{-1} and a resulting segment length (distance between two measurements) of around 1 km.

2.2.1 In-cloud representation of supersaturation in ERA5

Previous studies have shown that the upper troposphere is frequently supersaturated with respect to ice under cloud-free (Gierens et al., 1999; Petzold et al., 2020) as well as cloudy conditions (Spichtinger et al., 2004; Dekoutsidis et al., 2023). While ice supersaturation (ISS) in cloud-free conditions is represented in state-of-the-art numerical weather models, they currently lack in the appropriate representation of ISS under cloudy conditions. Often, the ISS is clipped to r_{ice} = 100 %, applying the so called "saturation adjustment" (McDonald, 1963). This adjustment is also applied in the IFS ice cloud microphysical scheme. The adjustment is a necessity of a missing diagnostic variable that would track the time-dependent in-cloud saturation (Tompkins et al., 2007). As a consequence of the adjustment, all available 'excess' water vapor, which is beyond the threshold, is deposited on existing ice particles within one time step, forcing r_{ice} back to 100 %. While the adjustment approach proved to be suitable for most atmospheric conditions (Gierens et al., 1999; Tompkins et al., 2007; Lamquin et al., 2009), the adjustment results in an underestimation of ISS in the upper troposphere (Gierens et al., 2020), which is problematic for contrail and cirrus representation.

To compensate for the dry bias in ERA5 for contrail detection applications, $r_{\mathrm{ERA,ice}}$ values are sometimes scaled by multiplication factors between 0.8 or 0.9, particularly in Schumann and Graf (2013) and Schumann et al. (2015). An updated scaling-method was proposed by Teoh et al. (2022a, T22 thereafter) that enhances $r_{\mathrm{ERA,ice}}$ > 100 % and reduces $r_{\mathrm{ERA,ice}}$ < 100 % by a a factor which depends on the original $r_{\mathrm{ERA,ice}}$. Within our study, we use T22-corrected values of $r_{\mathrm{ERA,ice}}$ as a benchmark.

2.3 Quantile mapping

In this study we propose to use a quantile mapping (QM) method to remove the lack of ISS in ERA5. QM is a correction method that it is frequently used to correct model biases in comparison to observations in a way that imposes the observed statistical distribution (Maraun et al., 2010; Maraun, 2012; Cannon et al., 2015; Cannon, 2018). Within our study, the QM technique is applied on ERA5 data and IAGOS measurements, which are regarded as the reference. Subsequently, we provide a brief overview of the mathematical concept of QM for which we follow the notations from Cannon et al. (2015) and Cannon (2018).

The basis of QM algorithms is to consider cumulative distribution functions (CDFs), $F_{\mathrm{o,h}}$ and $F_{\mathrm{m,h}}$, of the observed ($x_{\mathrm{o,h}}$) and simulated ($x_{\mathrm{m,h}}$) quantity, respectively. The CDFs describe the probability that the value of a quantity (or random variable) x, for example temperature or relative humidity, has a value that is lesser or equal to x. In our case $x_{\mathrm{o,h}}$ are the IAGOS T_{P1} or r_{P1} measurements, and $x_{\mathrm{m,h}}$ the corresponding along-track data from ERA5. The subscript 'h' commonly refers to historical data, in our case the reference period from 2018 to 2021. This can also be understood as the training data. Based on the relationship of $F_{\mathrm{o,h}}$ and $F_{\mathrm{m,h}}$, the biased model output $x_{\mathrm{m,p}}(t)$ at any given time t is corrected. The corrected value is represented by $\hat{x}_{\mathrm{m,p}}(t)$ (Cannon et al., 2015; Cannon, 2018). This is written in mathematical notation as:

$$\hat{x}_{\mathrm{m,p}}(t) = F_{\mathrm{o,h}}^{-1}\{F_{\mathrm{m,h}}[x_{\mathrm{m,p}}(t)]\}. \tag{6}$$

Equation 6 therefore couples a (potentially biased) model output to the most likely value that is observed in reality by the convolution of $F_{\mathrm{m,h}}$ and $F_{\mathrm{o,h}}^{-1}$. Here we apply the QM technique to the full period, i.e., the inference period, from January 2015 to June 2021, which includes but exceeds the training period.

Equation 6 describes the basic QM bias correction that depends only on one variable. Here, we propose a bivariate QM version for T_{ERA} and $r_{\mathrm{ERA,ice}}$ as the bias between ERA5 and IAGOS might depend on latitude. Such a multi-variate QM is similar to the suggested versions by Cannon (2016), Cannon (2018), or François et al. (2020).

For the temperature bias correction, $F_{\mathrm{o,h}}$ and $F_{\mathrm{m,h}}$ are determined at each p-level and for two latitude bands. The latitude bands are defined by the outer boundaries of the investigated area with 30°N and 70°N, with the split center point given by the 50 % percentile of the measurements per pressure level. Thus $F_{\mathrm{o,h}}(p,\Phi)$ and $F_{\mathrm{m,h}}(p,\Phi)$ are determined for different classes of pressure p and latitude Φ. $F_{\mathrm{o,h}}(p,\Phi)$ and $F_{\mathrm{m,h}}(p,\Phi)$ spans a temperature range from 190 and 273 K. Similarly, r_{ice} is corrected with $F_{\mathrm{o,h}}(p,\Phi,T)$ and $F_{\mathrm{m,h}}(p,\Phi,T)$, which are calculated for each p-level, two latitude bands Φ, and five temperature bins. As above, T ranges from 190 to 273 K with 5 temperature bins defined by 20 % percentile steps so that each temperature bin contains an equal number of observations at each p-level and latitude bin. Consequently, $F_{\mathrm{o,h}}(p,\Phi,T)$ and $F_{\mathrm{m,h}}(p,\Phi,T)$ are calculated for a total of 80 bins. A visualization of the resulting CDFs of temperature and relative humidity are given in Appendix A.

It is important to note that this basic version of QM assumes a time invariant bias between model and observations. We assume the ERA5 data set to be invariant in time as ERA5 is generated with one version of the IFS Cycle 41r2 and we only consider ERA5 data from the years 2015 to 2022. However, the IAGOS reference observations might vary over time as the spatial distribution of the sampling, instrument calibration, and data post-processing can change.

The temporal consistency of IAGOS relative humidity measurements was investigated by means of monthly climatologies. A constant bias in temperature and relative humidity between ERA5 and IAGOS was found. An exception are IAGOS relative humidity measurements from the year 2017, when IAGOS observations tend towards elevated relative humidity observations with respect to the other years, while the bias in temperature remained constant (see Appendix B). Biases between IAGOS, ERA5, and corrected ERA5 were further separated for their dependencies on latitude and longitude. While the bias in the temperature was found to be independent of longitude and latitude, the bias in relative humidity was smallest in North America and increased towards continental Europe (see Appendix C).

2.4 Schmidt–Appleman criterion, potential contrail formation, and contrail persistence

To allow for contrail formation the ambient air must be sufficiently cold and moist. The formation is typically estimated using relative humidity r_{crit} and a critical temperature T_{crit} threshold that is derived from the Schmidt–Appleman criterion (SAc, Appleman, 1953). The SAc is based solely on thermodynamic principles and has been tested to be a valid approximation although it does not inform on the fate of the contrail, which is a more complicated function of the ambient conditions but also the interactions of the vortex phase with the environment. The SAc is a necessary but insufficient criterion for persistent contrails. For contrails to be persistent (lifetime > 10 min), the ambient air must be additionally supersaturated with respect to ice (r_{ice} > 100 %) in so called ice supersaturated regions (ISSR). However, even under slightly sub-saturated conditions

contrails can form, but the persistence is uncertain. In weakly sub-saturated conditions the dissipation of ice crystals is slow and, hence, contrails can remain for hours (Li et al., 2023). Within this study, we use the revised version of the SAc, following Schumann (1996) and Rap et al. (2010). Calculations are performed for kerosene with a fuel specific energy $\mathcal{Q} = 43.2\ \mathrm{MJ\,kg^{-1}}$ and an emission index of water EI_{H20} = 1.25. The overall propulsion efficiency η is set to a typical value of 0.3 (Rap et al., 2010). For details on the SAc and equations used to calculate T_{crit} and r_{crit} the reader is referred to Wolf et al. (2023).

The SAc and the requirement for ice supersaturation separate the water-vapor-pressure–temperature diagram (see Fig.2 in Wolf et al. (2023)) in four different areas. The first area represents conditions, where the ambient air fulfills the SAc but is sub-saturated with respect to ice. In our study, contrails that form under these conditions are regarded as non-persistent and are labeled as NPC. Within the second area the SAc is fulfilled and ambient air is additionally supersaturated with respect to ice and persistent contrails (PCs) can form. The third area is treated as a special case, in which the ambient air does not fulfill the SAc but is ice supersaturated. Contrails that might have formed under conditions 'R1-NPC' or 'R2-PC' and that are mixed in area 3 may persist and spread. Therefore, area 3 can be understood as a potential 'Reservoir' (R) for contrails (Wolf et al., 2023). The SAc and the ISS threshold are used to flag the IAGOS measurements and the along-track ERA5 for NPC, PC, and R conditions. Samples that belong to none of these three categories are flagged as "no contrails" (NoC).

3 Results

3.1 Distributions of temperature and relative humidity from ERA5 and IAGOS

In a first step, along-track temperature and relative humidity from IAGOS and ERA5 are compared in terms of probability density functions (PDFs), mean values, and mean difference (MD). The performances of the QM–correction and the T22–correction are further quantified by the root-mean square error (RMSE), the mean absolute error (MAE), the mean square error, and the mean difference (MD). The analysis is limited to p-levels 250, 225, and 200 hPa, representing the most frequented p-levels (Fig. 2b–d).

Beginning with the temperature distributions, Fig. 3 (first column) shows that at p-levels 200 and 225 hPa, measured T_{P1} and simulated T_{ERA} agree well in terms of the MD (dashed lines) and the overall shape of the distributions. Only minor deviations in the MD of -0.4 K (200 hPa) and -0.1 K (225 hPa) are found, with a negative MD suggesting ERA5 is colder than observed on average ($T_{\mathrm{ERA}} < T_{\mathrm{P1}}$, see Fig. 3 second column). After the bias correction, the MD is reduced at all p-levels to below 0.1 K and also the shape of the PDFs of $T_{\mathrm{ERA}}^{\mathrm{cor}}$ are adjusted to better match the distributions of T_{P1} (Fig. 3 second column).

Turning to the relative humidity, which is plotted in the third column of Fig. 3. The distributions of r_{ice} are bimodal although the two modes have different magnitudes. The bimodal shape in the PDFs of upper-air r_{ice} matches with previous studies, e.g., Ruzmaikin et al. (2014), who used satellite observations from the Atmospheric Infrared Sounder (AIRS). The first mode at low r_{ice} is caused by dry atmospheric conditions related to dry air intrusions from the stratosphere into the upper troposphere, e.g., behind frontal zones (Browning, 1997), and flight sections within the lower stratosphere. The second mode at r_{ice} = 100 % is related to regions of high humidity or measurements inside clouds. With the general decrease in absolute humidity and

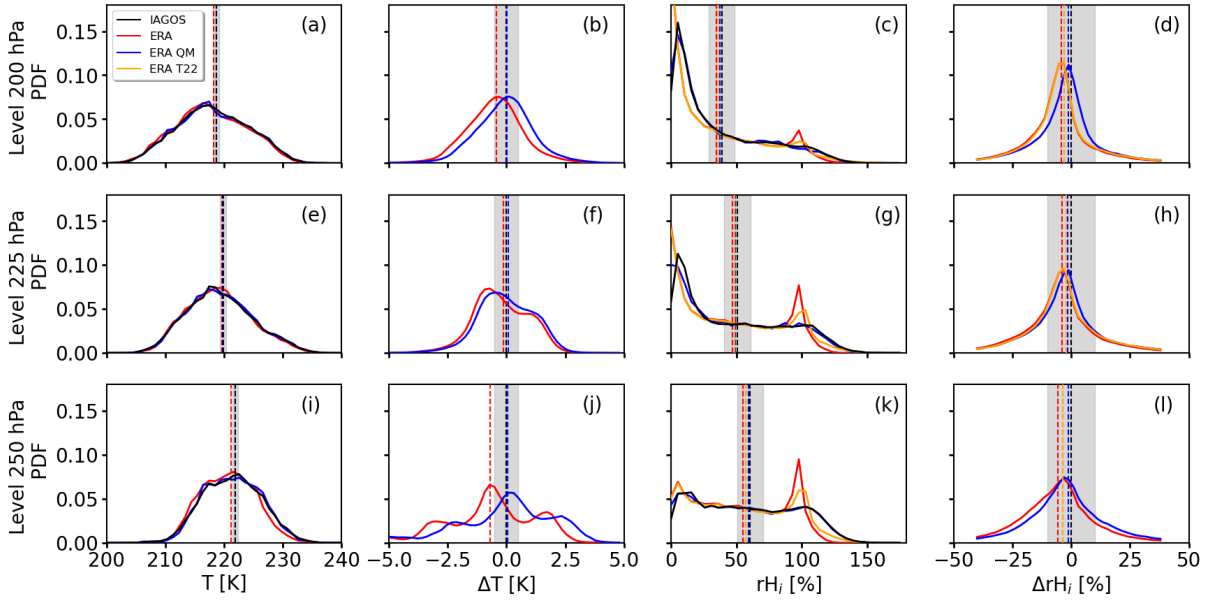

Figure 3. Probability density functions (PDFs) of temperature T (in K) and relative humidity r_{ice} w.r.t ice (in %) from IAGOS (black), ERA5 (red), and bias-corrected ERA5 data (blue). From top to bottom, rows represent pressure levels 200, 225, and 250 hPa. The first column shows PDFs of temperature from IAGOS T_{P1}, ERA5 T_{ERA}, and the bias-corrected ERA5 $T_{\mathrm{ERA}}^{\mathrm{cor}}$. The second column presents absolute difference of T_{ERA} and $T_{\mathrm{ERA}}^{\mathrm{cor}}$ with respect to T_{P1}. Columns three and four are the same as columns one and two but for relative humidity r_{ice}. In addition, bias-corrected $r_{\mathrm{ERA,ice}}^{\mathrm{T22}}$ using the correction method after T22 is given in orange. Differences are calculated by subtracting the IAGOS reference from the model output. In each plot, the median values of the distributions are indicated by the vertical dashed lines with the black line indicative of the IAGOS data. For reference, the average measurement uncertainties for T_{P1} and $r_{\mathrm{ERA,ice}}$ with ± 0.5 K and ± 10 % are indicated around mean T_{P1} and $r_{\mathrm{ERA,ice}}$, respectively.

possible intrusion of dry air from the stratosphere, the first mode becomes more and more pronounced with decreasing p, while the second mode flattens and almost vanishes.

Comparing the PDFs of $r_{\mathrm{ERA,ice}}$ and $r_{\mathrm{P1,ice}}$ minor differences are found for the first mode. However, larger differences appear for the second mode at r_{ice} = 100 %, where the occurrence frequency of large $r_{\mathrm{ERA,ice}}$ well exceeds $r_{\mathrm{P1,ice}}$, while $r_{\mathrm{ERA,ice}}$ > 100 % are underrepresented. The PDF of $r_{\mathrm{ERA,ice}}$ close to 100 % is characterized by a triangular shape, while the distribution of $r_{\mathrm{P1,ice}}$ is smaller in magnitude, broader in width, and skewed towards $r_{\mathrm{P1,ice}}$ > 100 %. Furthermore, at all p-levels, mean $r_{\mathrm{ERA,ice}}$ (red line, column three in Fig. 3), is generally shifted to lower values compared to mean $r_{\mathrm{P1,ice}}$ (black line). This indicates a lack of ISSR in ERA5 that is expected from its use of saturation adjustment (Sec. 2.2.1). The resulting MDs are determined to be −4.3 % (200 hPa), −3.8 % (225 hPa), and −5.5 % (250 hPa).

Smoothing the IAGOS data, as explained in Section 2.2, leads to mean values of T_{P1} and $r_{\mathrm{P1,ice}}$ for the native and the smoothed data that are similar by 0.1°C and 1 %, respectively. As smoothing did not change the mean values significantly, the differences in the PDFs of ERA5 and IAGOS, as well as the bias in mean $r_{\mathrm{ERA,ice}}$ compared to $r_{\mathrm{P1,ice}}$ cannot be attributed to

Table 2. Mean values of temperature T and relative humidity r_{ice} from IAGOS and ERA5 calculated from the original and the corrected values using the QM–correction and the scaling from T22. The data includes filtered measurements from January 2015 to January 2021. Values in parentheses are the differences relative to IAGOS.

Pressure level (hPa)	$\overline{T}_{\mathrm{P1}}$ (K)	$\overline{T}_{\mathrm{ERA}}$ (K)	$\overline{T}^{\mathrm{cor}}_{\mathrm{ERA}}$ (K)	
250	221.9	221.2 (-0.7)	221.9 (0.0)	
225	219.8	219.7 (-0.1)	219.9 (0.1)	
200	218.7	218.3 (-0.4)	218.7 (0.0)	
Pressure level (hPa)	$\overline{r}_{\mathrm{P1,ice}}$ (%)	$\overline{r}_{\mathrm{ERA,ice}}$ (%)	$\overline{r}^{\mathrm{cor}}_{\mathrm{ERA,ice}}$ (%)	$\overline{r}^{\mathrm{T22}}_{\mathrm{ERA,ice}}$ (%)
250	60.4	54.9 (-5.5)	59.4 (-0.9)	56.8 (-3.7)
225	50.6	46.8 (-3.8)	49.1 (-1.5)	48.6 (-2.0)
200	38.8	34.5 (-4.3)	37.5 (-1.3)	35.8 (-3.0)

differences in the spatial resolutions. However, the smoothing of the IAGOS data leads to a reduction in the variability as well as in the extreme values in measured T_{P1} and $r_{\mathrm{P1,ice}}$ (not shown here).

To correct the lack of ISS, i.e., the mismatch in the PDFs (Fig. 3 third column), the QM-technique is applied. After the QM–correction the MDs are reduced almost by half to -1.3 % (200 hPa), -1.5 % (225 hPa), and -0.9 % (250 hPa), which indicates a remaining slight dry-bias in $r_{\mathrm{ERA,ice}}$ compared to IAGOS as the MD remains negative (see Fig. 5j). However, the QM–correction leads to an adjustment of the entire PDFs such that the shape of the PDFs of corrected $r^{\mathrm{cor}}_{\mathrm{ERA,ice}}$ match the IAGOS observations. For comparison, we apply the T22–correction that only partly removes the dry-bias, resulting in MDs between -3.7 % (250 hPa) and -2.0 % (225 hPa) (see Fig. 3 fourth column and Fig. 5j). Furthermore, differences in the second mode in relation to the IAGOS observations remain as the T22–correction only scales values above a certain threshold, which primarily shifts the bulk of data points from the 100 % to higher r_{ice}. An overview of the original and corrected mean T and r_{ice} is given in Table 2.

The individual PDFs of r_{ice} are used to compile joined 2–dimensional (2D) histograms that are shown in Fig. 4a–c. In general, the frequency distribution of $r_{\mathrm{ERA,ice}}$ and $r_{\mathrm{P1,ice}}$ follow the diagonal line of 'ideal' agreement (Fig. 4a). However, the distribution is slightly shifted to below the 1:1-line, indicating a lower $r_{\mathrm{ERA,ice}}$ and therefore dryer conditions in ERA5 compared to the IAGOS observations. Particularly striking is the elongated feature of the $r_{\mathrm{ERA,ice}}$ distribution positioned close to 100 % (second mode) and a flattening for $r_{\mathrm{ERA,ice}}$ > 130 % as a result of the saturation adjustment. Gierens et al. (2020) presented a similar comparison of r_{ice} between ERA5 and IAGOS, providing only a scatter plot and not a density distribution. They found a strong scattering around the 1:1-line and described the distribution as "scattered all over the place" with a poor agreement among $r_{\mathrm{ERA,ice}}$ and $r_{\mathrm{P1,ice}}$. While we agree that the distributions in Fig. 4a–c are subject to scattering, the majority of the points (read to dark-red colors) show a reasonable alignment along the 1:1-line. For the individual pressure fields of 250, 225, and 200 hPa, r^2-score of 0.74, 0.79, and 0.75 are determined, respectively (also see Fig. 5h).

After the application of the QM–correction the alignment with the 1:1-line is improved (see Fig. 4b). As expected from Fig. 3, the artificially pronounced second mode in $r_{\mathrm{ERA,ice}}$ is removed in $r^{\mathrm{cor}}_{\mathrm{ERA,ice}}$ and the distribution extends further towards

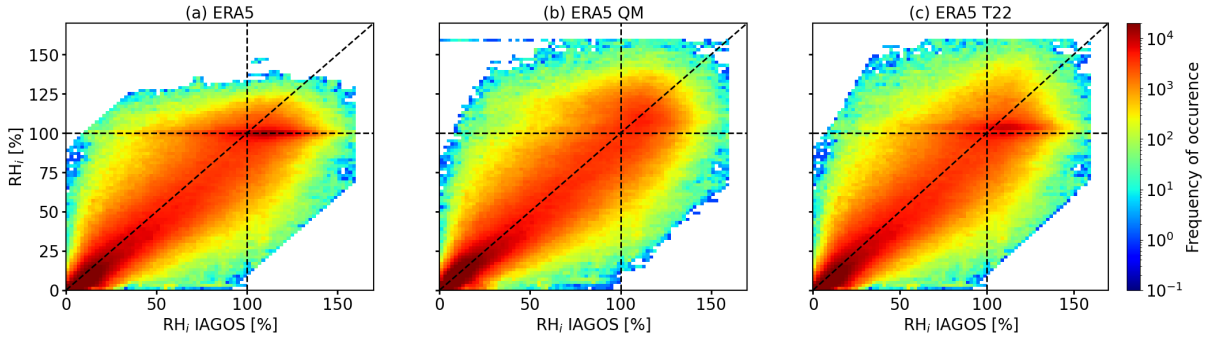

Figure 4. (a–c) Bidimensional histogram of original $r_{\mathrm{ERA,ice}}$ (left), QM-corrected $r^{\mathrm{cor}}_{\mathrm{ERA,ice}}$ (center), and T22-corrected $r^{\mathrm{T22}}_{\mathrm{ERA,ice}}$ (right) as a function of IAGOS-observed $r_{\mathrm{P1,ice}}$. Relative humidity is binned in intervals of 2 %. r_{ice} is given as relative humidity with respect to ice. Pressure levels 250, 225, and 200 hPa are combined. Perfect agreement is indicated by the diagonal dashed line and ice saturation is indicated by the horizontal and vertical dashed lines.

r_{ice} > 130 %, better representing the conditions observed by IAGOS. The QM–correction leads to r^2-values of 0.73, 0.78, and 0.75 at 250, 225, and 200 hPa, respectively, that are similar to the uncorrected ones (also see Fig. 5h).

For reference, the T22-corrected r_{ice} is compared with observed $r_{\mathrm{P1,ice}}$ and shown in Fig. 4c. The scaling of the T22-method enhances r_{ice} that are close or above 100 % and shifts the elongated feature towards higher r_{ice} but does not eliminate it. For this correction r^2 of 0.73 (250 hPa), 0.78 (225 hPa), and 0.74 (200 hPa) are calculated. So this type of correction leads to a small decrease in the r^2-score compared to the original ERA5 data.

To quantify the performance and the impact of the QM-method, five metrics are calculated, namely the root mean square error (RMSE), the mean absolute error (MAE), the r–square score, the mean square error (MSE), and the mean deviation (MD). This set of metrics has been selected to account for the different sensitivity of the metrics to outliers.

The top row in Fig. 5 visualizes the calculated metrics for the temperature. In general, the 250 hPa p-level is characterized by the largest RMSE of 2.1 K, MAE of 1.6 K, and MSE of 4.3 K^2 in relation to the other p-levels, which is explained by the enhanced natural variability in the temperature field with increasing p-level. A larger natural variability leads to larger differences among the IAGOS measurements and the ERA5 grid-box mean values. At the 225 and 200 hPa levels, in a more stratified atmosphere, the RMSE, MAE, and MSE are generally lower and similar for both p-levels with values around 1.2 K, 1 K, and 1.5 K^2, respectively. The QM–correction leads to a minimal increase in the r^2-score at all p-levels, while RMSE, MAE, and MSE increase unnoticeable. However, as expected and as it was demonstrated before, the MD is significantly reduced.

Similarly, the bottom row in Fig. 5 visualizes the calculated metrics for the original, the QM-corrected, and the T22-corrected r_{ice} against the IAGOS observations. As for the temperature, the RMSE, MAE, and MSE are largest for the 250 hPa p-level followed by the 225 and 200 hPa p-levels. At all p-levels, the QM- and T22–corrections lead to a constant or marginally increased RMSE, MAE, MSE, while the r^2-score remains almost constant. The increase in RMSE, MAE, and MSE appears counter-intuitive from the results shown in Fig. 4, with an improvement in the mean values and the distributions. However, both correction methods are purely statistical and do not remove differences in temperature and relative humidity of individual

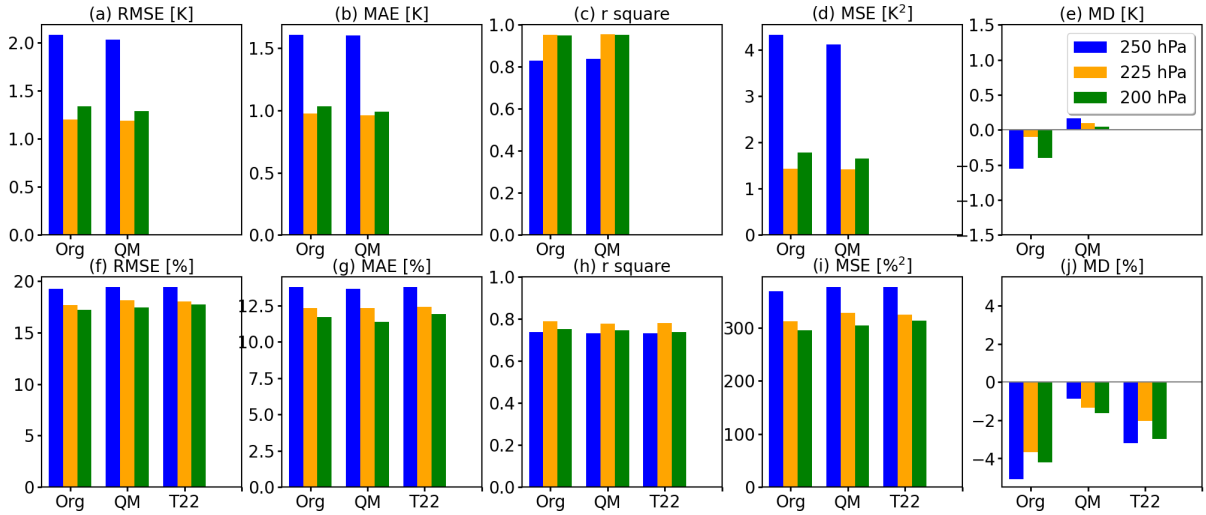

Figure 5. Bar plots of: **(a, f)** root mean square error (RMSE); **(b, g)** mean absolute error (MAE); **(c, h)** r^2–score; **(d, i)** mean square error (MSE); and **(e, j)** mean difference (MD) of ERA5 against IAGOS. First row shows metrics for T and the second row for r_{ice}. The first set of bars represent the original ERA5 output (label Org) while the second set represents the data set after the quantile-mapping correction (label QM). In the second row a third set of bars indicates the T22–correction. The metrics are calculated for pressure levels of 250 (blue), 225 (orange), and 200 hPa (green).

data points. Instead singular data points might be falsely adjusted by the QM–correction, which then creates outliers on which the RMSE and MSE respond very sensitively, thus the large RMSE and MSE for relative humidity. In contrast, MAE is less susceptible to outliers.

3.2 Distribution of relative humidity under cloud-free and in-cloud conditions

The IFS ISS adjustment partly depends on the ERA5 cloud fraction $\mathrm{CF_{ERA}}$ as only cloud-containing grid-boxes are clipped in r_{ice} (Tompkins et al., 2007). The effect of $\mathrm{CF_{ERA}}$ on the distribution of r_{ice} is investigated by separating $r^{\mathrm{cor}}_{\mathrm{ERA,ice}}$ for grid-boxes with $\mathrm{CF_{ERA}} < 0.2$ (cloud-free), $0.2, <= \mathrm{CF_{ERA}} < 0.8$ (intermediate), and $0.8, <= \mathrm{CF_{ERA}} <= 1$ (cloudy). Accordingly, IAGOS measurements of $r_{\mathrm{P1,ice}}$ are separated for cloud-free, intermediate, and in-cloud measurements using the cloud particle number concentration N as described in Sec. 2.1. This is only a subset of the IAGOS data used in this study, because BCP data are not always available for these flights. Data from pressure levels 250, 225, and 200 hPa are considered here.

Figure 6a shows PDFs of r_{ice} from IAGOS (black), ERA5 (red), and QM-corrected ERA5 (blue). The data is filtered separately for conditions, where IAGOS measures outside of clouds or where grid-boxes are almost cloud-free in ERA5 (CF <= 0.2). This category includes 88.3 % of the ERA5 data and 97.7 % of the IAGOS data. For cloud-free conditions, the distributions of $r_{\mathrm{ice}} < 60$ % are similar to the one presented Fig. 3c, g, and k. As expected, the three PDFs of IAGOS, ERA5, and QM-corrected ERA5 are characterized by a peak towards small r_{ice}, which is attributed to measurements and grid-boxes

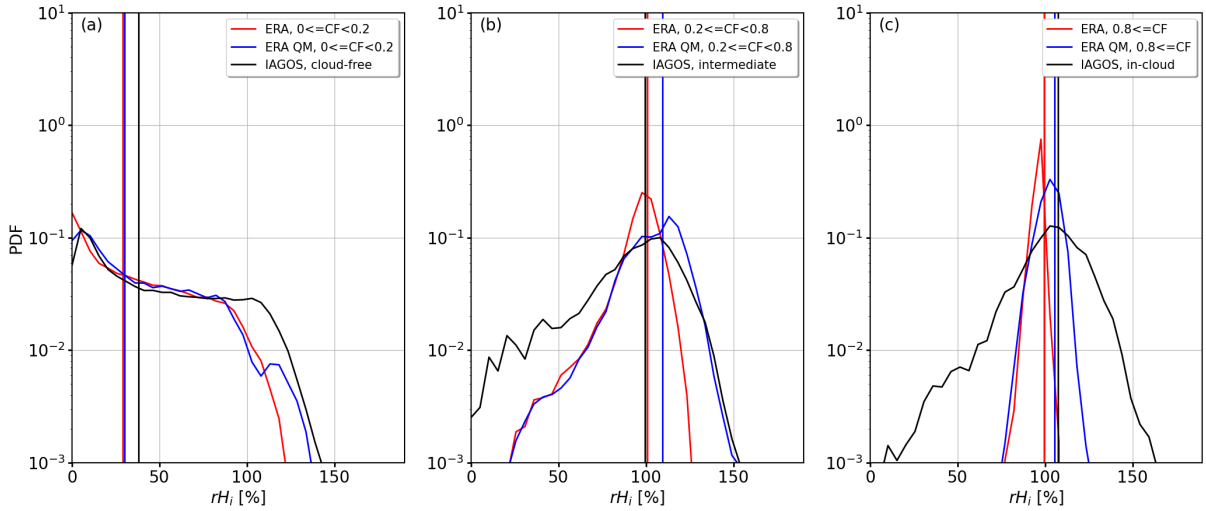

Figure 6. a–c Probability density functions of relative humidity r_{ice} w.r.t ice (in %) from IAGOS (black), original ERA5 (red), and corrected ERA5 (blue) using quantile mapping. Panels **a–c** show the PDFs separated for cloud-free, intermediate, and in-cloud conditions, respectively. Mean values of the distributions are indicated by the vertical lines.

with dry conditions. Differences in the PDFs appear, when r_{ice} approaches 100 %, where the probability of occurrence in $r_{\mathrm{P1,ice}}$ is higher compared to $r_{\mathrm{ERA,ice}}$. This resembles the scatter plot in Fig. 4a, where $r_{\mathrm{ERA,ice}}$ systematically tends to smaller r_{ice}. Recall that IAGOS is subject to a slight moist bias under conditions with low absolute humidity that are often encountered in the lower stratosphere (see Sec. 2.1). It can be seen in Fig. 6a that the QM–correction has only a limited impact on $r^{\mathrm{cor}}_{\mathrm{ERA,ice}}$. This is due to the nature of, and is an advantage of, the QM-technique, which respects the probability of occurrence by giving less weight to rare conditions (logarithmic y-scale). For cloud-free or almost cloud-free conditions mean values for $r_{\mathrm{P1,ice}}$, $r_{\mathrm{ERA,ice}}$, and $r^{\mathrm{cor}}_{\mathrm{ERA,ice}}$ of 28.9, 30.1, and 38 % were determined.

PDFs of r_{ice} of intermediate values are shown in Fig. 6b. This category includes 7.3 % of the ERA5 data and 1.3 % of the IAGOS data. Within this category it is not clear whether a grid-box should be considered as cloudy or cloud-free. The same is the case for IAGOS measurements. The shape of the PDFs changed compared to Fig. 6a, as the shape is now dominated by a peak in r_{ice} between 90 and 100 %. The shape of $r_{\mathrm{P1,ice}}$ from IAGOS revels the largest variability (width of the distribution), partly due to the intermediate detection of in-cloud and cloud-free conditions. QM-corrected values of $r^{\mathrm{cor}}_{\mathrm{ERA,ice}}$ lead to a distribution, where the left tail resembles the distribution from the original data $r_{\mathrm{ERA,ice}}$, while the right tail approaches the distribution of $r_{\mathrm{P1,ice}}$. Under intermediate cloud conditions the QM–correction skews the distribution to the right, which leads to a mean of $r^{\mathrm{cor}}_{\mathrm{ERA,ice}}$ = 109.3%. For $r_{\mathrm{ERA,ice}}$ and $r_{\mathrm{P1,ice}}$ mean values of 100.9 and 99.4 % are determined, respectively. The improved representation of $r_{\mathrm{ERA,ice}}$ in the intermediate category is particularly important, as NPC and PC formation is relevant from a radiative perspective in cloud-free or almost cloud-free air.

Figure 6c shows PDFs of r_{ice}, when $\mathrm{CF_{ERA}}$ is larger than 0.8, i.e., almost the entire grid-box is considered as cloudy, or when IAGOS measurements are from inside of clouds. This category includes 4.4 % of the ERA5 data and 0.9 % of the

Table 3. Fractions of measurement points (in %) labeled as non-persistent, persistent, and no contrail formation, as well as reservoir conditions. The results using the scaling method after Teoh et al. (2022b) is labeled with T22.

	IAGOS	ERA5				
		Original	T correction	r_{ice} correction	T+r_{ice} correction	r_{ice} correction T22
Condition						
NPC	44.0	50.3	47.8	46.4	44.0	46.9
PC	12.1	7.9	7.7	11.2	10.9	10.5
R	1.2	0.8	1.0	1.2	1.5	1.2
None	42.6	41.0	43.6	41.2	43.6	41.3

IAGOS data. The relatively limited number of samples (<1 %) from within clouds causes a less robust PDF compared to the PDF based on measurements conducted outside of clouds. This difference might be due to the fact that, compared to research aircraft measurements, the backscatter cloud probe misses clouds with N smaller than its detection limit of N = 0.001 cm^{-3} (Beswick et al., 2014; Petzold et al., 2017). All three distributions of r_{ice} are smaller compared to the cloud-free or intermediate conditions, with $r_{\mathrm{P1,ice}}$ being broadest. This is partly due to the larger natural variability in the IAGOS measurements compared to the ERA5 simulations of $r_{\mathrm{ERA,ice}}$. The distribution of $r_{\mathrm{ERA,ice}}$ is centered between 75 and 100 % with a steep slope on either sides. Particularly towards $r_{\mathrm{ERA,ice}}$ of 100 % the cutoff of $r_{\mathrm{ERA,ice}}$ for in-cloud conditions is prominent, which is not represented in $r_{\mathrm{P1,ice}}$. The QM–correction slightly broadens the distribution of $r^{\mathrm{cor}}_{\mathrm{ERA,ice}}$ towards values above 100 %. The bias in $r_{\mathrm{ERA,ice}}$ under cloudy conditions is reduced resulting in a mean $r^{\mathrm{cor}}_{\mathrm{ERA,ice}}$ of 105.6 %. This is closer to the measured mean $r_{\mathrm{IAGOS,ice}}$ of 107.6 % compared to the original output of $r_{\mathrm{IAGOS,ice}}$ with a mean of 99.6 %. In addition, the distribution of QM-corrected $r^{\mathrm{cor}}_{\mathrm{ERA,ice}}$ is slightly broadened but does not resemble the IAGOS but better agrees with measurements from, e.g., Krämer et al. (2016, 2020) and Li and Groß (2022), who reported in-cloud r_{ice} between 90 and 110 % due to the slow sublimation or growth of ice particles in cloudy conditions.

3.3 Along-track contrail formation potential and the effect of applied corrections

Along-track time series of uncorrected and corrected ERA5 data, and IAGOS measurements are flagged for non-persistent (NPC) and persistent (PC) contrails, and reservoir (R) conditions using the method described in Sec. 2.4. All data points not belonging to any of the categories are flagged for no contrail formation (NoC). Considering all data points from the years 2015 to 2021 at p-levels 250–200 hPa, it is found that 44 % of the IAGOS observations show a potential for NPC formation. PC appear to be less frequent with about 12.1 % and R conditions are rare with an occurrence of only 1.2 %. Using the original along-track ERA5 output, the contrail formation potential for NPC, PC, and R are estimated to be 50.3 %, 7.9 %, and 0.8 %, respectively. Due to the dry bias and the clipping of $r_{\mathrm{ERA,ice}}$ at 100 % the NPC category is enhanced in ERA5 at the expense of the PC category compared to the IAGOS measurements.

For reference, Teoh et al. (2020) (see Table 1 therein) estimated that 18.4 % of the flights form contrails (i.e., at least one contrail section during a flight) with only 7.4 % of the total, analyzed flight distance led to contrails. While there is a reasonable agreement in the occurrence of contrails, we identified more than twice the chance of forming non-persistent contrails. There are two main potential sources of disagreement. First, the account of aircraft characteristics is different. The estimates of our study solely rely on the SAc including constant values for fuel properties (specific heat capacity Q) and the overall propulsion efficiency η given in Sec. 2.4. Contrarily, the more elaborate method by Teoh et al. (2020) uses a fleet data set that includes flight specific information of aircraft engine type, thrust settings during flight stages, and estimates of black carbon (soot) emissions. These information were ingested into the contrail cirrus prediction model from (CoCiP, Schumann, 2012) to determine contrail formation and the related radiative effect. It is noted that CoCiP only considers flight sections as a contrail, when a certain contrail radiative effect is exceeded, i.e., the ice particle number is larger than 10^3 m^{-3} and the cirrus optical thickness is larger than 10^{-6} (Schumann, 2012). Thus, the results from Teoh et al. (2020) consider the potential for contrail formation, actual aircraft emissions, the synoptic conditions, and the contrail radiative effect. For our approach, with the IAGOS data set, no such aircraft-performance data are available. Secondly, the way flight distance is counted as contrail-forming is different between the two studies. In our study, the SAc accounts only for thermodynamic properties.

Subsequently, the impact of corrected $T_{\mathrm{ERA}}^{\mathrm{cor}}$ and $r_{\mathrm{ERA,ice}}^{\mathrm{cor}}$ on the along-track classification of NPC, PC, and R is investigated. The individual contributions of T and r are separated by applying the QM–correction separately on T and r. The scaling method from T22 is shown as a benchmark.

Applying the QM correction only on T_{ERA} leads to a reduction in the fraction of NPC from 50.3 % to 47.8 % and for PC from 7.9 % to 7.7 %, respectively. The correction increases the mean $T_{\mathrm{ERA}}^{\mathrm{cor}}$ (ambient temperature), allowing fewer ERA5 samples to pass the T and r_{ice} thresholds for NPC and PC formation. Consequently, the fraction of NoC or R conditions increases, where supersaturation is reached but the SAc is not fulfilled. The remaining differences in the distributions of NPC, PC, and R between the T_{ERA}-only corrected and the IAGOS measurements indicate that temperature correction alone is insufficient to better represent NPC and PC (see Table 3).

Applying the QM correction only to $r_{\mathrm{ERA,ice}}$ reduces the frequency of NPC to 46.3 %. At the same time the number of PC increases to 11.4 % and R conditions are slightly increased to 1.2 %. Thus, this correction leads to an increase in PC, mostly on the expense of the NPC category. This is simply because of the higher mean $r_{\mathrm{ERA,ice}}^{\mathrm{cor}}$ and correspondingly more samples that pass the thresholds given by the SAc. Compared to the T_{ERA}–only correction, the $r_{\mathrm{ERA,ice}}$–only correction has the largest impact on the categorization and is, therefore, the main driver of potential miss-classification and needs to be correctly represented.

The scaling-based T22–correction is most similar to the QM correction of $r_{\mathrm{ERA,ice}}$–only. After the T22–correction, 46.9 % of the samples were identified as NPC, which is slightly above the IAGOS reference and similar to the estimated occurrence after the QM $r_{\mathrm{ERA,ice}}$–only correction. With the T22 correction, PC and R conditions are found in 10.5 % and 1.2 % of the cases, which is also comparable to the result from the QM $r_{\mathrm{ERA,ice}}$–only correction.

Applying the QM correction to both T_{ERA} and $r_{\mathrm{ERA,ice}}$ results in a decrease of NPC to 44.0 %, which corresponds to the occurrence of NPC that is found in the IAGOS data set and is below the original ERA5 data. It is also slightly lower than after

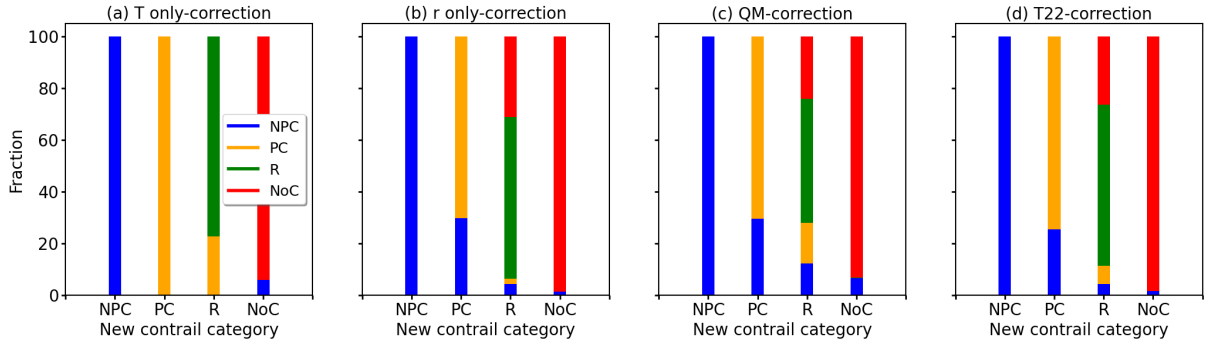

Figure 7. a–d Redistributed fractions (in %) of original ERA5 contrail classification with respect to the classification after applying the T-only correction, the r-only correction, the QM–correction, and the correction after T22, respectively. The original classification of non-persistent contrail (NPC) are given in blue, persistent contrails (PC) are given in orange, reservoir conditions (R) are colored in green, and samples that do not allow for contrails formation (NoC) are given in red.

the QM $r_{\mathrm{ERA,ice}}$-only correction. This is due to the simultaneous correction of T and $r_{\mathrm{ERA,ice}}$ as some samples become to warm to form contrails. PC conditions are found in 10.9 % of the samples, which is slightly less compared to the QM correction of $r_{\mathrm{ERA,ice}}$. The frequency of R conditions as well as NoC increase slightly. It is found that the combined correction of T_{ERA} and $r_{\mathrm{ERA,ice}}$ leads to the best agreement with the IAGOS observations. While the improvement is primarily caused by the correction of $r_{\mathrm{ERA,ice}}$, it is emphasized that also T has to be corrected as the calculation of $r_{\mathrm{ERA,ice}}$ depends on the underlying temperature field. The conversion of specific humidity to relative humidity and the conversion between relative humidity w.r.t ice and w.r.t liquid water via the saturation curves becomes very sensitive to T, when approaching low temperatures that exist at typical flight levels, see, e.g., Ambaum (2020).

For a detailed understanding on how the QM–correction modifies the classification of NPC, PC, R, and NoC, the redistribution among the contrail categories is determined by tracking the classification before and after the corrections. The distributions have to be interpreted qualitatively as the statistics include a potential yearly variation. Figure 7a–d shows the contribution of the pre-correction categories to the classification after applying a specific correction method.

For example, Fig. 7a shows that for the majority of the QM T_{ERA}–only corrected ERA5 samples that are now classified as 'NPC' were already NPC before the correction. Only a minority of the new NPC samples were previously identified as PC or belonged to the NoC category before. The QM T_{ERA}-only correction does not significantly affect the PC category but leads to the largest redistribution in the R category. Due to the increase in mean T_{ERA}, previous PC-flagged samples now contribute by 21 % to the R category. However, the proportion of R conditions relative the total number of samples is small and thus the overall relevance is small. Similarly, samples previously classified as NPC contribute little to the NoC category.

Similar to Fig. 7a, the QM $r_{\mathrm{ERA,ice}}$-only correction, given in Fig. 7b, as well as the T22–correction, given in Fig. 7d, lead to only minor changes in the NPC category. In the case of the QM $r_{\mathrm{ERA,ice}}$-only correction the newly flagged samples in the R category are composed of samples that already belong to the R category (62.5 %) or the NoC category (31.3 %). In the case

Table 4. Schematic contingency table for a binary event. Adapted from Stephenson (2000).

IAGOS	ERA5 detection	
detection	Yes	No
Yes	True Positive (TP)	False Negative (FN)
No	False Positive (FP)	True Negative (TN)

of the T22–correction a similar pattern is found but with an additional share of previously flagged PC samples (7.0 %). Both corrections show similar patterns for the newly formed PC category, which now consist of about of 75 % and 25 % of former PC and NPC data points, respectively.

The QM–correction, shown in Fig. 7c, is a superposition of the QM T_{ERA}-only and QM $r_{\mathrm{ERA,ice}}$-only correction. No redistribution within the NPC category is found. The R category is subject to the strongest re-distribution but keeping in mind that the R category represents the smallest proportion of all data points. The newly formed PC category now consists of 29 % and 71 % of former NPC and PC data points, respectively.

3.4 Analysis of collocated contrail formation potential from ERA5 and IAGOS

Beyond the comparison of bulk statistics, the collocated temporal and spatial along-track representation of NPC and PC in ERA5 are validated against IAGOS observations using a confusion matrix, where we consider NPC, PC, R, and NoC conditions as single binary events. A confusion matrix is a special type of a two dimensional contingency table for which a schematic is given in Table 4. In our case the classification is based on: i) the IAGOS observations and ii) the ERA5 data. Perhaps persistent contrails form only in a minority of situations, we computed the equitable threat score (ETS; Mason, 2012), following the reasoning of Gierens et al. (2020) given in the Appendix A of their paper. The ETS can be regarded as equitable to the four entries of the contingency table, when the total number of samples is sufficiently large (Hogan et al., 2010; Gierens et al., 2020), which is the case considering our data set. The ETS ranges between 0 for random relations and 1 for perfect correlation, and is calculated on the basis of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) by:

$$ETS = \frac{TP - r}{TP + FN + FP - r}, \tag{7}$$

with:

$$r = \frac{(TP + FP) \cdot (TP + FN)}{(TP + FP + FN + TN)}. \tag{8}$$

The ETS is calculated for the original as well as corrected ERA5 data against IAGOS estimates (see Table 5 and Fig. 8).

Statistics based on a confusion matrix like the one given in Table 4 are a tough test for ERA5 because even small spatial or temporal errors in the temperature or humidity fields can lead to mis-classifications. To estimate the effect of a possible

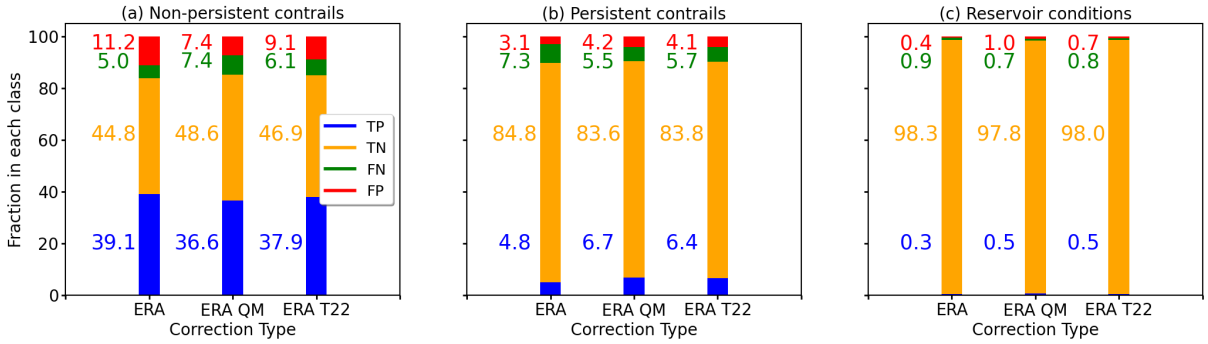

Figure 8. (a–b) Fraction of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) predictions from ERA5 data classifications based on IAGOS observations (reference) for non-persistent and persistent contrails, as well as reservoir conditions, respectively. ERA5 data is compared in its original form 'ERA' (first column), after the QM–correction 'ERA cor2d' (second column), and using the correction after Teoh et al. (2022b) 'ERA T22' (third column).

Table 5. Equitable threat score (ETS) calculated from the confusion matrix between IAGOS (reference) and the original ERA5 as well as the corrected ERA5 output.

	ERA5	ERA5 (3h)	ERA5 QM	ERA5 T22
PC	0.27	0.26	0.36	0.35
NPC	0.51	0.50	0.54	0.53
R	0.19	0.17	0.24	0.23

pattern shift, we use the three-hourly (3h) ERA5 data of T_{ERA} and $r_{\mathrm{ERA,ice}}$. Evaluating the NPC and PC formation as well as the R condition with the confusion matrix between IAGOS and the coarsened ERA5 (3h), the ETS remains almost constant. This indicates that sensitivity of the confusion matrix and the ETS to temporal and spatial decorrelations that occur within three hours is low. In addition, the differences in ETS between the original ERA5 and the three-hourly ERA5 data provides a reference for estimating the impact of the corrections in relation to temporal-spatial mismatches. Thus, differences in the calculated ETS that are larger than that reference are truly due to incorrect values of T_{ERA} and $r_{\mathrm{ERA,ice}}$.

The application of the QM–correction modifies the distributions of temperature and relative humidity in such a way that PC conditions are correctly detected more frequently, resulting in an increase in ETS from 0.27 to 0.36. Similarly, the NPC and R categories an increase in ETS from 0.51 to 0.54 and 0.19 to 0.24, respectively, is observed. Thus, the QM–correction leads to an improvement across all categories. The QM–correction appears to be most effective for the PC category, which is also the most relevant category considering the longevity and the potential radiative effects of contrails. Similar improvements in the contrail representation are observed for the T22–correction. For all categories, an increase in the ETS is observed compared to the original ERA5 data. The T22–correction related ETS are comparable to those of the QM–correction and are listed in Table 5. Based on the original ERA5 data, an ETS of 0.27 and 0.51 is calculated for PC and NPC conditions, respectively. For

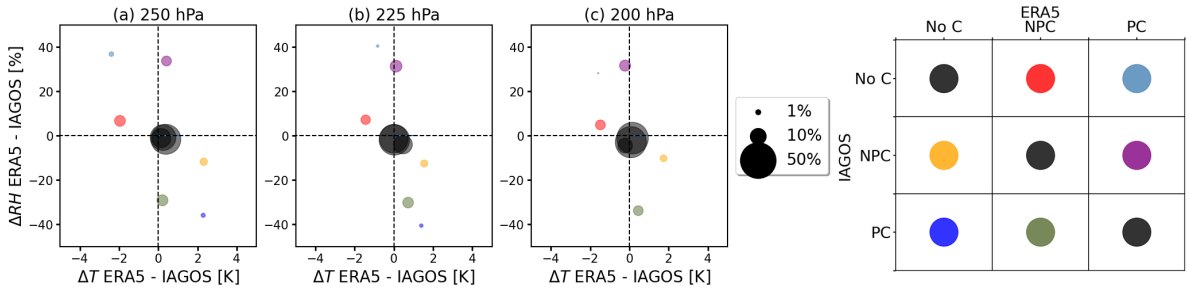

Figure 9. (a–c) Mean difference in temperature T (in K) and relative humidity r_{ice} (in %) between IAGOS and ERA5 corrected by quantile mapping for three pressure levels: from left to right, 250, 225, and 200 hPa. Colored dots represent a combination of mean ΔT and Δr for one of the nine categories of the contingency table (right). The area of the dots is proportional to the fraction of measurement–simulation pairs with respect to the total number per pressure level. Colors indicate the classification, using the legend shown in the right-hand panel with categories no contrail formation (NoC), non-persistent contrails (NPC), and persistent contrails (PC).

reference, we provide values obtained by Gierens et al. (2020), who compared ERA5 with MOZAIC measurements. Gierens et al. (2020) compared MOZAIC data and ERA5 for individual months, while we calculate the ETS on the basis of several years. Therefore, the ETS from Gierens et al. (2020) are subject to significant monthly variations. For comparability, the ETS values from Gierens et al. (2020), given for the months of January, April, July, and October in their Table 1, were used to calculate mean ETS of 0.12 and 0.74 for PC and NPC conditions, respectively. Evaluating the contingency tables by means of ETS it can be concluded that the QM–correction as well as the T22–correction lead to a better representation of all contrail types compared to the original ERA5 data. Compared to the mean values calculated from Gierens et al. (2020) we found a lower performance of uncorrected and corrected ERA5 for the NPC category, while there is a better performance for the PC category, especially after the QM–correction and the T22–correction.

3.5 Disentangling of classification with respect to temperature and relative humidity

Even after QM–correction, about 16 % of the NPC and 11.5 % of the PC observation–measurement pairs are classified as 'false positive' and 'false negative'. The sensitivity study using 3 hourly ERA5 data showed that this is unrelated to spatial mismatches but is rather due to actual deviations in temperature and relative humidity between IAGOS and ERA5. Subsequently, we aim to quantify the mean differences in temperature and relative humidity that remained after the QM–correction and that contribute to the misclassification of potential contrail formation. Within the following section all ERA5 values are QM-corrected.

The along track-samples from IAGOS and ERA5 are categorized by a contingency table with categories: NoC, NPC, and PC taking IAGOS as the reference. The created contingency table is visualized in the legend of Fig. 9. The diagonal elements of the contingency table represent combinations of IAGOS and ERA5 that agree in terms of contrail occurrence, while all off-diagonal values are incorrectly classified. For each of the nine contingency table combinations the corresponding mean

differences in the temperature:

$$\overline{\Delta T} = \frac{1}{n}\sum_{i=0}^{n} T_{\mathrm{ERA5,i}} - T_{\mathrm{IAGOS,i}} \tag{9}$$

and relative humidity:

$$\overline{\Delta r_{\mathrm{ice}}} = \frac{1}{n}\sum_{i=0}^{n} r_{\mathrm{ERA5,ice,i}} - r_{\mathrm{IAGOS,ice,i}} \tag{10}$$

are calculated, with n the number of data points in each category. Figure 9a–c presents the 2D–space spanned by $\overline{\Delta T}$ and $\overline{\Delta r}$ for each of the contingency table combinations at p-levels 250, 225, and 200 hPa, respectively. In the following, a notation of 'A–B' with $\mathrm{A,B} \in \{\mathrm{NoC,NPC,PC}\}$ is used as an abbreviation for the classification of A from IAGOS and B in ERA5. For example, a notation of 'NPC-PC' means a combination of IAGOS NPC conditions and ERA5 PC conditions.

In general, $\overline{\Delta T}$ and $\overline{\Delta r_{\mathrm{ice}}}$ are similar at all three p-levels and the three p-levels are discussed simultaneously and the fraction of each category compared to the total number of samples are given for the middle layer at 225 hPa. 81.9 % of the observation–model combinations are correctly categorized and represented along the contingency table-diagonal. As expected, corresponding $\overline{\Delta T}$ and $\overline{\Delta r_{\mathrm{ice}}}$ (black dots) are close to the origin.

Contrarily, the off-diagonal groups are mostly located in the top-left and lower-right quadrants. Misclassifications for PC–NPC (green, 4.5 %) and NPC–PC (violet, 5.4 %) due to errors in $\overline{\Delta r_{\mathrm{ice}}}$ mostly. Samples in the PC–NPC group (green) were incorrectly categorized due to a too low relative humidity in ERA5, while the NPC–PC samples (violet) were too moist. But of course, since r_{ice} depends on T, mis-classifications are also caused by errors in T, even if they do not dominate in these cases.

Misclassifications for the combinations 'NoC–NPC' (red, 3.7 %) and 'NPC–NoC' (yellow, 1.9 %) are mostly due to errors in T. For 'NoC–NPC' and 'NPC–NoC', $T_{\mathrm{ERA}}^{\mathrm{cor}}$ was colder or warmer than T_{P1}, respectively.

Least frequent are the misclassifications 'NoC–PC' (light blue, 0.3 %) and PC–NoC (dark blue, 0.5 %). These two groups are subject to the largest $\overline{\Delta T}$ and $\overline{\Delta r_{\mathrm{ice}}}$. Samples in these categories were only found at the 250 and 225 hPa p-level, while the PC–NoC (dark blue) is not found at the 200 hPa level. It is likely that data points in the two categories result from small scale variations captured by IAGOS that are not represented by ERA5 due to temporal and spatial resolution.

It is worth identifying whether the misclassification in ERA5 with respect to IAGOS is most often due to biases in temperature or in humidity. Focusing on the PC representation in ERA5, the primary reason for a misclassification after the correction is the deviation in r_{ice}. This is shown by the proximity of the violet and green dots to the y-axis (small ΔT), while the differences in r_{ice} are larger than ± 20 %. Hence, the underestimation (green dot) or overestimation (violet dot) of potential contrail formation is primary related to the underlying humidity field in ERA5.

4 Summary

In this study we proposed a temperature and relative humidity correction method for ERA5 based on a bivariate quantile mapping (QM) technique to better estimate the contrail formation potential. The QM–correction was trained on 3.5 years of

IAGOS observations and collocated ERA5 data of T_{ERA} and $r_{\mathrm{ERA,ice}}$. The QM–correction was then applied on 5.5 years of ERA5 data and compared with IAGOS. The target region covers the eastern United States, the North Atlantic, and continental Europe, spanning 30°N to 70°N and 110°W to 30°E for pressure (p) levels 250 to 200 hPa, where the majority of IAGOS observations are available (93.8 %).

Alongside the IAGOS data post-processing and the calculation of cumulative distribution functions (CDF) for the QM–correction, the along-track biases in temperature and relative humidity among ERA5 and IAGOS were analyzed. In general, biases in temperature and relative humidity are characterized by a dependence at p-level with the largest differences typically for the lowest p-level, i.e., 200 hPa. Biases were further separated for their dependencies on latitude and longitude. While the bias in the temperature was found to be independent of longitude and latitude, the bias in relative humidity was smallest in North America and increased towards continental Europe. The temporal consistency of IAGOS relative humidity measurements was investigated by means of monthly climatologies. A constant bias in temperature and relative humidity between ERA5 and IAGOS was found. An exception are IAGOS relative humidity measurements from the year 2017, when IAGOS observations tend towards elevated relative humidity observations with respect to the other years, while the bias in temperature remained constant. It is again noted that in-situ observations from IAGOS are potentially biased by avoiding deep-convective clouds and their outflow, while cirrus clouds are generally not avoided.

Using the bivariate QM–correction, the bias in T_{ERA} was reduced from -0.7, -0.1, and -0.4 K, at p-levels 250, 225, and 200 hPa, respectively, to below 0.1 K at all p-levels. The relative humidity bias was reduced from -5.5, -3.8, and -4.3 % to -0.9, -1.5, and -1.3 % at 250, 225, and 200 hPa, respectively. While a slight dry-bias compared to IAGOS remains, a significant improvement in terms of the probability density functions (PDFs) of the relative humidity distribution is achieved. PDFs of corrected relative humidity are almost identical in shape with the PDFs determined from the IAGOS observations. A previously existing artificial peak at $r_{\mathrm{ERA,ice}}$ = 100 % in the PDFs of ERA5, which is caused by the saturation adjustment in ERA5, was removed. Consequently, corrected values of $r^{\mathrm{cor}}_{\mathrm{ERA,ice}}$ better represent the actual conditions in terms of mean value and frequency of occurrence.

Subsequently, the impact of the QM correction on the detection and classification of NPC, PC, and R with respect to IAGOS was evaluated. Measurements from IAGOS and along-tack ERA5 data were flagged for NPC, PC, R, and NoC conditions. Based on the original ERA5 data set, 50.3, 7.9, and 0.8 % of all data points were identified as NPC, PC, and R, respectively. Compared to the IAGOS estimates with 44.0, 12.1, and 1.2 % for NPC, PC, and R, an overestimation of NPC and underestimation of PC was identified in ERA5. After the ERA5 QM correction, 44.0, 10.9, and 1.5 % of the samples were identified as NPC, PC, and R conditions, indicating a general improvement of the contrail representation with respect to the original ERA5 data. Using a parameterized relative humidity correction from Teoh et al. (2022a), here used as a reference for comparison, led to 46.9, 10.5, and 1.2 % of NPC, PC, and R conditions, respectively, which is comparable to the performance from the QM–correction.

The temporal and spatial representation of NPC, PC, and R in ERA5 with respect to IAGOS was assessed with a contingency table. Based on the contingency table the equitable threat score (ETS) was calculated. The largest improvement is found for

the PC category with an increase in ETS from 0.27 to 0.36. Smaller improvements were found for the NPC and the R category with an increase in ETS from 0.51 to 0.54 and 0.2 to 0.24, respectively.

The contingency table further revealed that 81.9 % of the data samples were coherently flagged in IAGOS and ERA5 after QM–correction. In these cases almost no biases in temperature and relative humidity between IAGOS and ERA5 remain. The remaining 18.1 % of the data points, which were incorrectly classified for NPC, PC, and R conditions by ERA5, are caused by remaining biases in temperature and relative humidity of varying magnitude. The misclassifications were insensitive to the applied correction method. False classifications of NPC as PC were primarily dominated by a relative humidity bias, while false classifications of NPC as NoC were dominated by a bias in the temperature. However, the majority of misclassifications were caused by combinations of temperature and relative humidity biases with ERA5 either being cold-moist or a warm-dry biased compared to IAGOS. Furthermore, the relative humidity bias between IAGOS and ERA5 was found to depend on the temperature.

Overall, the presented QM–correction allowed to remove the systematic bias in temperature and relative humidity in ERA5 with IAGOS as the reference. As a result, the QM-correction leads to a representation of NPC, PC, and R in ERA5 that is comparable to the distribution identified in the IAGOS observations.

Code availability. The python code that was used to perform the analysis and the quantile correction is provided following: https://doi.org/10.5281/zenodo.8418565

Data availability. ERA5 data can be obtained from the European Centre for Medium-Range Weather Forecasts (ECMWF) data catalog at https://doi.org/10.24381/cds.f17050d7 (Hersbach et al., 2023).

The IAGOS data can be downloaded from the IAGOS data portal at https://doi.org/10.25326/20 (Boulanger et al., 2020).

Appendix A: Cumulative distribution functions for quantile mapping

Here we provide an example for calculated cumulative distribution functions (CDFs) of relative humidity r defined with respect to ice. IAGOS CDFs ($F_{\mathrm{o,h}}$) and ERA CDFs ($F_{\mathrm{m,h}}$) are calculated on basis of the observed IAGOS relative humidity ($x_{\mathrm{o,h}}$) and simulated, along-track ERA5 relative humidity ($x_{\mathrm{m,h}}$), respectively, following the description in Sec. 2.3. Figure A1 shows $F_{\mathrm{o,h}}$ (dashed lines) and $F_{\mathrm{m,h}}$ (solid lines) for individual pressure (p) levels between 350 and 200 hPa. As described in Sec. 2.3 the full domain (specified in Sec. 2.1) is subdivided into two latitude bands. The split point is determined by the 50th-percentile at each p-level such that both latitude bands contain equal numbers of data points. For legibility, only CDFs of the northern most latitude band are shown here. The selection is arbitrary and conclusions are transferable between the two bands.

The black lines in Fig. A1 indicate $F_{\mathrm{o,h}}$ and $F_{\mathrm{m,h}}$ from the quantile mapping (QM) approach that considers only for p-level dependence and the latitude band. For the majority of the p-levels $F_{\mathrm{o,h}}$ and $F_{\mathrm{m,h}}$ are similar in shape. An exception is r between 100 and 110 % at levels $350 \leq \mathrm{p} \leq 250$ hPa, where $F_{\mathrm{m,h}}$ (ERA5) shows a dominant mode, while $F_{\mathrm{o,h}}$ (IAGOS)

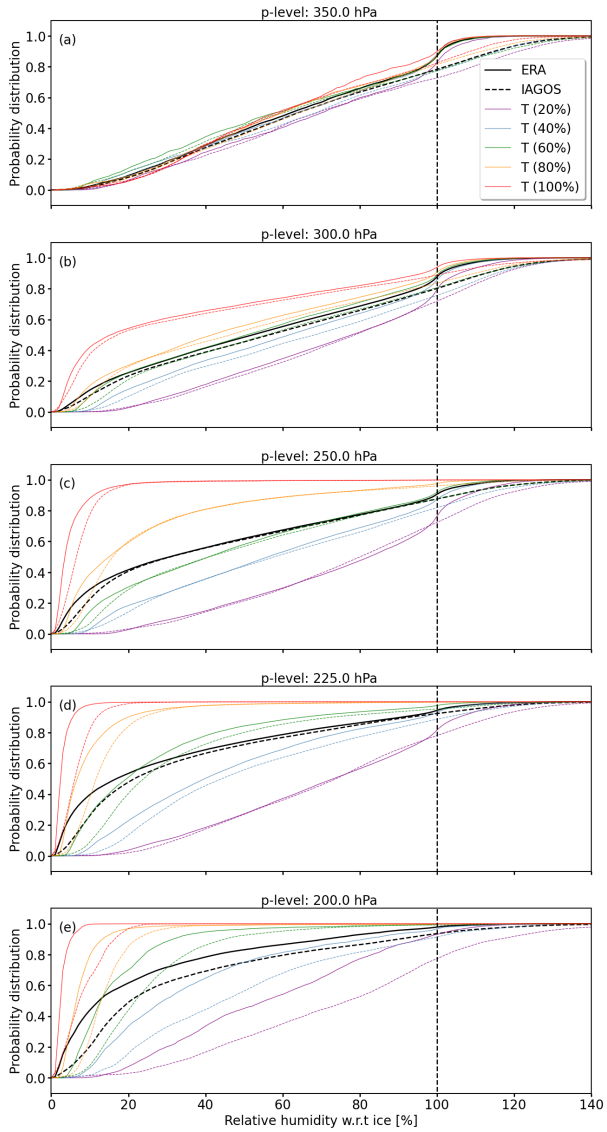

Figure A1. (a–e) Cumulative distribution functions (CDFs) F of relative humidity w.r.t ice (in %). CDFs from ERA5 $F_{\mathrm{m,h}}$ and IAGOS $F_{\mathrm{o,h}}$ are given by solid and dashed lines, respectively. The black lines represent $F_{\mathrm{o,h}}$ and $F_{\mathrm{m,h}}$ that depend at p-level and latitude Φ. Color-coded are $F_{\mathrm{m,h}}$ and $F_{\mathrm{o,h}}$ that additionally consider for five temperature bins with bin sizes defined by 20th-percentiles.

remains flat. The mode in $F_{\mathrm{m,h}}$ is a superposition of two effects. While the peak is of natural origin, as reported by Krämer et al. (2016, 2020), it is also caused by the saturation adjustment in ERA5 (see Section 2.2.1). This mode becomes less prominent with decreasing p as the atmosphere gets drier with altitudes, so supersaturation is less likely. Simultaneously, the differences between $F_{\mathrm{o,h}}$ and $F_{\mathrm{m,h}}$ increase for $r < 20$ %, where both $F_{\mathrm{o,h}}$ and $F_{\mathrm{m,h}}$ are further characterized by a steep slope. The largest effect in this regard is found at the 200 hPa p-level, where $F_{\mathrm{m,h}}$ contains a larger fraction of high relative humidity values

compared to $F_{\mathrm{o,h}}$, indicating an underestimation of r that is not attributable to the saturation adjustment. For example, 50 % of the ERA relative humidity are smaller than around 15 %, while the respective value for IAGOS is around 22 %, indicating a general dry-bias unrelated to the saturation adjustment.

The color-coded lines in Fig. A1 represent the bivariate QM approach, where r is additionally separated for five temperature (T) bins that are defined by 20 %-steps. $F_{\mathrm{m,h}}$ and $F_{\mathrm{o,h}}$ that result from the bivariate QM reveal a strong dependence in T, which becomes visible in the deviating shapes of $F_{\mathrm{m,h}}$ and $F_{\mathrm{o,h}}$ at constant p-level and latitude band. The systematic order of the colored lines further indicates that T-bins with low T (0–20th-percentile, violet lines) are mostly dominated by high relative humidity values, while bins with higher T (80–100th-percentile, red lines) are dominated by low r. The CDFs with lower T are generally flat with a continuous slope, while T bins with higher temperatures are dominated by a steep slope for r < 10 %, particularly for p < 250 hPa. However, for the bivariate QM correction the actual shape of $F_{\mathrm{m,h}}$ and $F_{\mathrm{o,h}}$ is less relevant but the difference. These difference between $F_{\mathrm{m,h}}$ and $F_{\mathrm{o,h}}$ are increasing with decreasing p-level. The importance to consider the T-dependence is further highlighted by the fact that the simpler, univariate QM approach (black) and related $F_{\mathrm{m,h}}$ and $F_{\mathrm{o,h}}$ do not consider the shape and the shape difference that is required to adequately correct r under different ambient conditions, particularly with decreasing p-level.

Appendix B: Temporal consistency in temperature and relative humidity of IAGOS and ERA5

Applying the quantile mapping (QM)–correction in the presented form requires a time-invariant bias in temperature (T) and relative humidity (r_{ice}) between the IAGOS and the ERA5. The bias among both might vary due to variations in the instrument calibration procedure or changes in the sampling distribution due to seasonal flight schedules.

We tested for time invariance by calculating mean values of T and r_{ice} from ERA5 as well as IAGOS over all samples for each month spanning January 2015 to 2021 on p-levels of 250, 225, and 200 hPa.

Figure B1a shows that monthly mean T_{ERA} (red) and T_{P1} (black) agree well, which is expected from the small bias presented in Fig. 3. Furthermore, the monthly mean difference between T_{P1} and T_{ERA}, given in Fig. B1b, remains constant with values around -0.5 K and maximal -1 K, except for some individual spikes. Figure B1b also shows that QM-corrected $T^{\mathrm{cor}}_{\mathrm{ERA}}$ (blue) better match with T_{P1}, which is indicated by maximal differences of ± 0.5 K.

Similarly, Fig. B1c shows monthly mean of $r_{\mathrm{P1,ice}}$ (black), original ERA5 $r_{\mathrm{ERA,ice}}$ (red), and QM-corrected ERA5 $r^{\mathrm{cor}}_{\mathrm{ERA,ice}}$ (blue) ranging between 40 % and 50 % for the majority of the period. An exception is the period after 2020, which is due to low data availability (see Fig. B1e). Figure B1c illustrates that $r_{\mathrm{ERA,ice}}$ (red) follows $r_{\mathrm{P1,ice}}$ (black) with an offset between 3 % and up to -12 % that has been shown before (fourth column in Fig. 3). Like for the temperature correction, Fig. B1d clearly shows that the QM–correction increases mean $r^{\mathrm{cor}}_{\mathrm{ERA,ice}}$ such that the bias between ERA5 and IAGOS is reduced, bringing the monthly means of $r^{\mathrm{cor}}_{\mathrm{ERA,ice}}$ closer to the 0.

Even though the bias Δr_{ice} = $r_{\mathrm{ERA,ice}}$ - $r_{\mathrm{P1,ice}}$ remains fairly constant for the majority of the presented time series, the differences are particularly pronounced for the years 2016 and 2017. However, their temperature bias ΔT = T_{ERA}- T_{P1}

remains constant (Fig. B1b), which suggests that changes in the sampling, e.g., due to modified aircraft operations, are not the cause but the known grounding problem of IAGOS acquisition at this time period (see Sanogo et al. (2023)).

In the absence of alternative observations to compare against IAGOS, we turn to the interannual variation in r_{ice} to confirm that relative humidity measurements for years 2016 and 2017 are anomalous. Multi-year, monthly climatological means of $r_{\mathrm{ERA,ice}}$ and $r_{\mathrm{P1,ice}}$ are calculated spanning the years 2015 to 2021. Using uncorrected $r_{\mathrm{ERA,ice}}$ as the reference, anomalies of $r_{\mathrm{P1,ice}}$, $r_{\mathrm{ERA,ice}}$, and $r^{\mathrm{cor}}_{\mathrm{ERA,ice}}$ are determined by subtracting the monthly mean of an individual year from the multi-year, monthly climatological mean. Figure B2 shows mean anomalies of $r_{\mathrm{P1,ice}}$ that range from -11.9 % (2017) to 0.8 % (2020). Similarly, mean anomalies of $r_{\mathrm{ERA,ice}}$ range between -4.2 % (2017) and 5.5 % (2020). The mean anomalies between ERA5 and IAGOS are largest for the years 2016 and particularly 2017 the difference between the anomalies for year 2017 exceeds all other years with -6.8 %. Slightly smaller mean anomaly difference between ERA5 and IAGOS anomalies are found to the years 2016 with -4.7 % and 2020 with -4.7 %. Therefore, years 2017 and parts of 2016 are special cases compared to the other years in terms of anomalies during which $r_{\mathrm{P1,ice}}$ is likely biased towards too moist values.

Appendix C: Latitudinal and longitudinal dependent deviations in temperature and relative humidity between ERA5 and IAGOS

The bias between IAGOS and ERA5 might depend on the geographic position, e.g., due to characteristic spatial distributions of water vapor in the atmosphere. Such spatial-dependent biases in T and r_{ice} among ERA5 and IAGOS are identified by calculating mean differences for bins of 10° in latitude and longitude, at p-levels 300, 250, 225, and 200 hPa. The calculations include all samples from years 2015 to 2021 and from within the defined subdomain (30°N–70°N, 105°W–30°E).

First, the longitudinal variation in ΔT is analyzed (Fig. C1a). In general, a tendency toward more negative ΔT is found for decreasing p-levels, reaching a maximum on the 200 hPa p-level, where ΔT mostly reaches values of up to -2 K. Large ΔT on 200 hPa westwards of 80° must be cautiously interpreted due to the low number of available samples in this pressure level and longitude bin (see Fig. C1e). The general negative ΔT indicates that the mean temperature from ERA5 is predominantly lower than measured by IAGOS. ΔT at pressure levels 250 hPa (green) and 225 hPa (red) is almost constant over the entire longitude range with ΔT being smaller than -0.5. An exception is the 300 hPa level, where ΔT exceeds -0.5 K and reaches values of up to -1 K east of 50°W. Separating ΔT for latitudes between 30°N and 70°N does not reveal any latitudinal dependencies. An exception is the 200 hPa p-level, where ΔT increases towards the equator and reaches up to -1.7 K at 30°N.

Similar to T, the longitudinal and latitudinal dependence of r_{ice} is analyzed (see Fig. C1c, d). In general, Δr_{ice} increase from west, with Δr_{ice} around 0 %, towards east, reaching Δr of up to -25 % at the 300 and 250 hPa p-levels. No systematic offset among the p-levels is found. While Δr_{ice} is largest at the 200 hPa level at 110°W, Δr_{ice} is among the smallest levels at 30°E. Conversely, Δr_{ice} is small at the 250 hPa level at 110°W and is the second largest Δr_{ice} at 30°E. Similar to T, separating Δr_{ice} for latitudes shows not strong latitudinal sensitivity with the smallest values between -10 % and -4 % at the 225 and 200 hPa p-levels. Largest Δr_{ice}, of up to $\Delta r_{\mathrm{ice}} = -25$ %, are found at the 300 hPa level particularly between 40°N and 60°N.

Separating biases in T and r clearly shows the necessity to consider the p-level in the QM–correction. In contrast, binning for latitudes appears to be of minor importance, which relaxes the requirement for more than two bins in the proposed QM–correction. In contrast, the dependency of r_{ice} on the longitude is much more pronounced and would require individual cumulative distribution functions but could not be considered for in the QM–correction as dividing the data in three sub-domains would lead to insufficient data in rarely sampled combinations of T and p.

Author contributions. **KW** performed the data analysis and prepared the manuscript. **NB** and **OB** contributed equally to the preparation of the manuscript. **SR** and **YL** provided constructive comments and helped with the interpretation of the IAGOS measurements.

Competing interests. The authors declare no competing interest.

Acknowledgements. This research has been supported by the French Ministère de la Transition écologique et Solidaire (N° DGAC 382 N2021-39), with support from France's Plan National de Relance et de Resilience (PNRR) and the European Union's NextGenerationEU. NB, AP and YL acknowledge funding from the Horizon 2020 project ACACIA (grant no. 875036). Furthermore, we acknowledge the support from Garry Lloyd for post-processing the IAGOS P1 data of the single particle backscattering optical spectrometer.

Table A1. Notations

Symbol	Long-name	Unit
α	Scaling factor in ERA5	-
η	Overall propulsion efficiency	0-1
Φ	Latitude	$^\circ$
σ	Standard deviation of Gaussian distribution	-
$\mathrm{CF_{IAGOS}}$	Fraction of in-cloud measurements by IAGOS	0–1
$\mathrm{CF_{ERA}}$	Cloud fraction from ERA5	0–1
c_p	Isobaric heat capacity of air	$\mathrm{J\,kg^{-1}\,K^{-1}}$
$e(T)$	Water vapor pressure, temperature dependent	Pa
$e_{\mathrm{sat,l}}(T)$	Saturation water vapor pressure over water, temperature dependent	Pa
$e_{\mathrm{sat,i}}(T)$	Saturation water vapor pressure over ice, temperature dependent	Pa
EI	Emission index of water vapor for the fuel	$\mathrm{kg\,kg^{-1}}$
F	Cumulative distribution function for quantile mapping	-
N_ice	Particle number concentration	$\mathrm{cm^{-3}}$
p	Pressure	hPa
$\mathcal{P}$	Probability for contrail occurrence	0–1
$r_{\mathrm{P1,ice}}$	Relative humidity with respect to ice from IAGOS package 1 (P1)	%
r_{P1}	Relative humidity with respect to liquid water from IAGOS package 1 (P1)	%
$r_{\mathrm{ERA,ice}}$	Relative humidity with respect to ice from ERA5	%
r_crit	Critical relative humidity from Schmidt–Appleman criterion	[0–1]
$r^{\mathrm{cor}}_{\mathrm{ERA,ice}}$	Relative humidity with respect to ice from ERA5 bias corrected	%
$r^{\mathrm{T22}}_{\mathrm{ERA,ice}}$	Relative humidity with respect to ice from ERA5 corrected with method T22	%
$r_{\mathrm{ERA,liq}}$	Relative humidity with respect to liquid water from ERA5	%
$r^{\mathrm{cor}}_{\mathrm{ERA,liq}}$	Relative humidity with respect to liquid water from ERA5	%
$t_{1-1/\mathrm{e}}(T)$	Temperature dependent sensor response time to adjust to a signal change by 63%	s
T_0	Freezing temperature in ERA5	K
T_ice	Lower temperature limit for scaling of relative humidity conversion in ERA5	K
T_crit	Critical temperature from Schmidt–Appleman–criterion	K
T_{P1}	Temperature measured by IAGOS package 1 (P1)	K
T_{ERA}	Temperature from ERA5	K
$T^{\mathrm{cor}}_{\mathrm{ERA}}$	Temperature from ERA5 bias corrected	K
$q_{\mathrm{sat,liq}}$	Saturation specific humidity with respect to a liquid water surface	$\mathrm{kg\,kg^{-1}}$
$q_{\mathrm{sat,ice}}$	Saturation specific humidity with respect to a ice surface	$\mathrm{kg\,kg^{-1}}$
Q_heat	Specific heat capacity	$\mathrm{J\,kg^{-1}}$
$\hat{x}_{\mathrm{m,p}}(t)$	Transfer function for quantile mapping	

References

Alduchov, O. A. and Eskridge, R. E.: Improved magnus form approximation of saturation vapor pressure, J. Appl. Meteorol., 35, 601–609, https://doi.org/10.1175/1520-0450(1996)035<0601:IMFAOS>2.0.CO;2, 1996.

Ambaum, M. H. P.: Accurate, simple equation for saturated vapour pressure over water and ice, Q. J. Royal Meteorol. Soc., 146, 4252–4258, https://doi.org/10.1002/qj.3899, 2020.

Appleman, H.: The formation of exhaust condensation trails by jet aircraft, B. American Meteorol. Soc., 34, 14–20, https://doi.org/10.1175/1520-0477-34.1.14, 1953.

Beswick, K., Baumgardner, D., Gallagher, M., Volz-Thomas, A., Nedelec, P., Wang, K.-Y., and Lance, S.: The backscatter cloud probe - a compact low-profile autonomous optical spectrometer, Atmos. Meas. Tech., 7, 1443–1457, https://doi.org/10.5194/amt-7-1443-2014, 2014.

Bickel, M., Ponater, M., Bock, L., Burkhardt, U., and Reineke, S.: Estimating the effective radiative forcing of contrail cirrus, J. Climate, 33, 1991–2005, https://doi.org/10.1175/JCLI-D-19-0467.1, 2020.

Bland, J., Gray, S., Methven, J., and Forbes, R.: Characterising extratropical near-tropopause analysis humidity biases and their radiative effects on temperature forecasts, Q. J. Royal Meteorol. Soc., 147, 3878–3898, https://doi.org/10.1002/qj.4150, 2021.

Bock, L. and Burkhardt, U.: Reassessing properties and radiative forcing of contrail cirrus using a climate model, J. Geophys. Res. Atmos., 121, 9717–9736, https://doi.org/10.1002/2016JD025112, 2016.

Boucher, O., Borella, A., Gasser, T., and Hauglustaine, D.: On the contribution of global aviation to the CO_2 radiative forcing of climate, Atmos. Environ., 267, 118 762, https://doi.org/10.1016/j.atmosenv.2021.118762, 2021.

Boulanger, D., Blot, R., Bundke, U., Gerbig, C., Hermann, M., Nédélec, P., Rohs, S., and Ziereis, H.: IAGOS final quality controlled Observational Data L2 – Time series, Aeris [data set], https://doi.org/10.25326/06, 2018.

Boulanger, D., Thouret, V., and Petzold, A.: IAGOS Data Portal, AERIS [data set], https://doi.org/10.25326/20, 2020.

Browning, K. A.: The dry intrusion perspective of extra-tropical cyclone development, Meteorol. Appl., 4, 317–324, https://doi.org/10.1017/S1350482797000613, 1997.

Buck, A. L.: New equations for computing vapor pressure and enhancement factor, J. Appl. Meteorol., 20, 1527–1532, https://doi.org/10.1175/1520-0450(1981)020<1527:NEFCVP>2.0.CO;2, 1981.

Burkhardt, U. and Kärcher, B.: Global radiative forcing from contrail cirrus, Nat. Clim. Change, 1, 54–58, https://doi.org/10.1038/nclimate1068, 2011.

Cannon, A. J.: Multivariate bias correction of climate model output: Matching marginal distributions and intervariable dependence structure, J. Clim., 29, 7045–7064, https://doi.org/10.1175/JCLI-D-15-0679.1, 2016.

Cannon, A. J.: Multivariate quantile mapping bias correction: an N-dimensional probability density function transform for climate model simulations of multiple variables, Clim. Dyn., 50, 31—49, https://doi.org/10.1007/s00382-017-3580-6, 2018.

Cannon, A. J., Sobie, S. R., and Murdock, T. Q.: Bias correction of GCM precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes?, J. Climate, 28, 6938–6959, https://doi.org/10.1175/JCLI-D-14-00754.1, 2015.

Carminati, F., Migliorini, S., Ingleby, B., Bell, W., Lawrence, H., Newman, S., Hocking, J., and Smith, A.: Using reference radiosondes to characterise NWP model uncertainty for improved satellite calibration and validation, Atmos. Meas. Tech., 12, 83–106, https://doi.org/10.5194/amt-12-83-2019, 2019.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. Royal Meteorol. Soc., 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

Dekoutsidis, G., Groß, S., Wirth, M., Krämer, M., and Rolf, C.: Characteristics of supersaturation in midlatitude cirrus clouds and their adjacent cloud-free air, Atmos. Chem. Phys., 23, 3103–3117, https://doi.org/10.5194/acp-23-3103-2023, 2023.

Dyroff, C., Zahn, A., Christner, E., Forbes, R., Tompkins, A. M., and van Velthoven, P. F. J.: Comparison of ECMWF analysis and forecast humidity data with CARIBIC upper troposphere and lower stratosphere observations, Q. J. Royal Meteorol. Soc., 141, 833–844, https://doi.org/10.1002/qj.2400, 2015.

ECMWF: IFS Documentation CY47R1 - Part IV: Physical Processes, 4, ECMWF, https://doi.org/10.21957/cpmkqvhja, 2020.

François, B., Vrac, M., Cannon, A. J., Robin, Y., and Allard, D.: Multivariate bias corrections of climate simulations: which benefits for which losses?, Earth Sys. Dyn., 11, 537–562, https://doi.org/10.5194/esd-11-537-2020, 2020.

Friedlingstein, P., Jones, M. W., O'Sullivan, M., Andrew, R. M., Hauck, J., Peters, G. P., Peters, W., Pongratz, J., Sitch, S., Le Quéré, C., Bakker, D. C. E., Canadell, J. G., Ciais, P., Jackson, R. B., Anthoni, P., Barbero, L., Bastos, A., Bastrikov, V., Becker, M., Bopp, L., Buitenhuis, E., Chandra, N., Chevallier, F., Chini, L. P., Currie, K. I., Feely, R. A., Gehlen, M., Gilfillan, D., Gkritzalis, T., Goll, D. S., Gruber, N., Gutekunst, S., Harris, I., Haverd, V., Houghton, R. A., Hurtt, G., Ilyina, T., Jain, A. K., Joetzjer, E., Kaplan, J. O., Kato, E., Klein Goldewijk, K., Korsbakken, J. I., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lenton, A., Lienert, S., Lombardozzi, D., Marland, G., McGuire, P. C., Melton, J. R., Metzl, N., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S.-I., Neill, C., Omar, A. M., Ono, T., Peregon, A., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rödenbeck, C., Séférian, R., Schwinger, J., Smith, N., Tans, P. P., Tian, H., Tilbrook, B., Tubiello, F. N., van der Werf, G. R., Wiltshire, A. J., and Zaehle, S.: Global carbon budget 2019, Earth Sys. Sci. Data, 11, 1783–1838, https://doi.org/10.5194/essd-11-1783-2019, 2019.

Gierens, K., Schumann, U., Helten, M., Smit, H., and Marenco, A.: A distribution law for relative humidity in the upper troposphere and lower stratosphere derived from three years of MOZAIC measurements, Ann. Geophys., 17, 1218–1226, https://doi.org/10.1007/s00585-999-1218-7, 1999.

Gierens, K., Matthes, S., and Rohs, S.: How well can persistent contrails be predicted?, Aerospace, 7, https://doi.org/10.3390/aerospace7120169, 2020.

Helten, M., Smit, H. G. J., Sträter, W., Kley, D., Nedelec, P., Zöger, M., and Busen, R.: Calibration and performance of automatic compact instrumentation for the measurement of relative humidity from passenger aircraft, J. Geophys. Res. Atmos., 103, 25 643–25 652, https://doi.org/10.1029/98JD00536, 1998.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Q. J. Royal Meteorol. Soc., 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., and Thépaut, J.-N.: ERA5 monthly averaged data on single levels from 1940 to present.,

https://doi.org/10.24381/cds.f17050d7, copernicus Climate Change Service (C3S) Climate Data Store (CDS), (Accessed on 08-Sep-2023), 2023.

Hogan, R. J., Ferro, C. A. T., Jolliffe, I. T., and Stephenson, D. B.: Equitability revisited: Why the "Equitable Threat Score" is not equitable, Weather Forecasting, 25, 710 – 726, https://doi.org/10.1175/2009WAF2222350.1, 2010.

Kärcher, B.: Formation and radiative forcing of contrail cirrus, Nat. Commun., 9, https://doi.org/10.1038/s41467-018-04068-0, 2018.

Kärcher, B., Burkhardt, U., Unterstrasser, S., and Minnis, P.: Factors controlling contrail cirrus optical depth, Atmos. Chem. Phys., 9, 6229–6254, https://doi.org/10.5194/acp-9-6229-2009, 2009.

Konjari, P., Krämer, M., Brast, N., Reutter, P., Petzold, A., Rohs, S., and Rolf, C.: UTLS Water Vapor Climatologies derived from combined In-Situ Passenger and Research Aircraft Measurements, EGU22-9841, EGU General Assembly 2022, Vienna, Austria, https://doi.org/10.5194, 2022.

Krämer, M., Schiller, C., Afchine, A., Bauer, R., Gensch, I., Mangold, A., Schlicht, S., Spelten, N., Sitnikov, N., Borrmann, S., de Reus, M., and Spichtinger, P.: Ice supersaturations and cirrus cloud crystal numbers, Atmos. Chem. Phys., 9, 3505–3522, https://doi.org/10.5194/acp-9-3505-2009, 2009.

Krämer, M., Rolf, C., Luebke, A., Afchine, A., Spelten, N., Costa, A., Meyer, J., Zöger, M., Smith, J., Herman, R. L., Buchholz, B., Ebert, V., Baumgardner, D., Borrmann, S., Klingebiel, M., and Avallone, L.: A microphysics guide to cirrus clouds – Part 1: Cirrus types, Atmos. Chem. Phys., 16, 3463–3483, https://doi.org/10.5194/acp-16-3463-2016, 2016.

Krämer, M., Rolf, C., Spelten, N., Afchine, A., Fahey, D., Jensen, E., Khaykin, S., Kuhn, T., Lawson, P., Lykov, A., Pan, L. L., Riese, M., Rollins, A., Stroh, F., Thornberry, T., Wolf, V., Woods, S., Spichtinger, P., Quaas, J., and Sourdeval, O.: A microphysics guide to cirrus – Part 2: Climatologies of clouds and humidity from observations, Atmos. Chem. Phys., 20, 12 569–12 608, https://doi.org/10.5194/acp-20-12569-2020, 2020.

Krüger, K., Schäfler, A., Wirth, M., Weissmann, M., and Craig, G. C.: Vertical structure of the lower-stratospheric moist bias in the ERA5 reanalysis and its connection to mixing processes, Atmos. Chem. Phys., 22, 15 559–15 577, https://doi.org/10.5194/acp-22-15559-2022, 2022.

Kunz, A., Spelten, N., Konopka, P., Müller, R., Forbes, R. M., and Wernli, H.: Comparison of Fast In situ Stratospheric Hygrometer (FISH) measurements of water vapor in the upper troposphere and lower stratosphere (UTLS) with ECMWF (re)analysis data, Atmos. Chem. Phys., 14, 10 803–10 822, https://doi.org/10.5194/acp-14-10803-2014, 2014.

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, https://doi.org/10.5194/acp-9-1779-2009, 2009.

Lee, D. S., Fahey, D. W., Skowron, A., Allen, M. R., Burkhardt, U., Chen, Q., Doherty, S. J., Freeman, S., Forster, P. M., Fuglestvedt, J., Gettelman, A., De León, R. R., Lim, L. L., Lund, M. T., Millar, R. J., Owen, B., Penner, J. E., Pitari, G., Prather, M. J., Sausen, R., and Wilcox, L. J.: The contribution of global aviation to anthropogenic climate forcing for 2000 to 2018, Atmos. Environ., 244, 117 834, https://doi.org/10.1016/j.atmosenv.2020.117834, 2021.

Li, Q. and Groß, S.: Satellite observations of seasonality and long-term trends in cirrus cloud properties over Europe: investigation of possible aviation impacts, Atmos. Chem. Phys., 22, 15 963–15 980, https://doi.org/10.5194/acp-22-15963-2022, 2022.

Li, Y., Mahnke, C., Rohs, S., Bundke, U., Spelten, N., Dekoutsidis, G., Groß, S., Voigt, C., Schumann, U., Petzold, A., and Krämer, M.: Upper-tropospheric slightly ice-subsaturated regions: frequency of occurrence and statistical evidence for the appearance of contrail cirrus, Atmos. Chem. Phys., 23, 2251–2271, https://doi.org/10.5194/acp-23-2251-2023, 2023.

Maraun, D.: Nonstationarities of regional climate model biases in European seasonal mean temperature and precipitation sums, Geophys. Res. Lett., 39, https://doi.org/10.1029/2012GL051210, 2012.

Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., Brienen, S., Rust, H. W., Sauter, T., Themeßl, M., Venema, V. K. C., Chun, K. P., Goodess, C. M., Jones, R. G., Onof, C., Vrac, M., and Thiele-Eich, I.: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, Revi. Geophys., 48, https://doi.org/10.1029/2009RG000314, 2010.

Marenco, A., Thouret, V., Nédélec, P., Smit, H., Helten, M., Kley, D., Karcher, F., Simon, P., Law, K., Pyle, J., Poschmann, G., Von Wrede, R., Hume, C., and Cook, T.: Measurement of ozone and water vapor by Airbus in-service aircraft: The MOZAIC airborne program, an overview, J. Geophys. Res. Atmos., 103, 25 631–25 642, https://doi.org/10.1029/98JD00977, 1998.

Mason, I. B.: Forecast verification - a practitioner's guide in atmospheric science, John Wiley & Sons, 2012.

McDonald, J. E.: The Saturation Adjustment in Numerical Modelling of Fog, J. Atmos. Sci., 20, 476–478, https://doi.org/10.1175/1520-0469(1963)020<0476:TSAINM>2.0.CO;2, 1963.

Meyer, R., Mannstein, H., Meerkötter, R., Schumann, U., and Wendling, P.: Regional radiative forcing by line-shaped contrails derived from satellite data, J. Geophys. Res. Atmos., 107, ACL 17–1–ACL 17–15, https://doi.org/10.1029/2001JD000426, 2002.

Minnis, P., Bedka, S. T., Duda, D. P., Bedka, K. M., Chee, T., Ayers, J. K., Palikonda, R., Spangenberg, D. A., Khlopenkov, K. V., and Boeke, R.: Linear contrail and contrail cirrus properties determined from satellite data, Geophys. Res. Lett., 40, 3220–3226, https://doi.org/10.1002/grl.50569, 2013.

Neis, P., Smit, H. G. J., Rohs, S., Bundke, U., Krämer, M., Spelten, N., Ebert, V., Buchholz, B., Thomas, K., and Petzold, A.: Quality assessment of MOZAIC and IAGOS capacitive hygrometers: insights from airborne field studies, Tellus B: Chem. Phys. Meteorol., 67, 28 320, https://doi.org/10.3402/tellusb.v67.28320, 2015.

Pasquier, J. T., Pfahl, S., and Grams, C. M.: Modulation of atmospheric river occurrence and associated precipitation extremes in the North Atlantic region by European weather regimes, Geophys. Res. Let., 46, 1014–1023, https://doi.org/10.1029/2018GL081194, 2019.

Petzold, A., Busen, R., Schröder, F. P., Baumann, R., Kuhn, M., Ström, J., Hagen, D. E., Whitefield, P. D., Baumgardner, D., Arnold, F., Borrmann, S., and Schumann, U.: Near-field measurements on contrail properties from fuels with different sulfur content, J. Geophys. Res. Atmos., 102, 29 867–29 880, https://doi.org/10.1029/97JD02209, 1997.

Petzold, A., Thouret, V., Gerbig, C., Zahn, A., Brenninkmeijer, C. A. M., Gallagher, M., Hermann, M., Pontaud, M., Ziereis, H., Boulanger, D., Marshall, J., Nédélec, P., Smit, H. G. J., Friess, U., Flaud, J.-M., Wahner, A., Cammas, J.-P., Volz-Thomas, A., and IAGOS TEAM: Global-scale atmosphere monitoring by in-service aircraft – current achievements and future prospects of the European Research Infrastructure IAGOS, Tellus: Chem. Phys. Meteorol., 67B, 28 452, https://doi.org/10.3402/tellusb.v67.28452, 2015.

Petzold, A., Krämer, M., Neis, P., Rolf, C., Rohs, S., Berkes, F., Smit, H. G. J., Gallagher, M., Beswick, K., Lloyd, G., Baumgardner, D., Spichtinger, P., Nédélec, P., Ebert, V., Buchholz, B., Riese, M., and Wahner, A.: Upper tropospheric water vapour and its interaction with cirrus clouds as seen from IAGOS long-term routine in situ observations, Faraday Discuss., 200, 229–249, https://doi.org/10.1039/C7FD00006E, 2017.

Petzold, A., Neis, P., Rütimann, M., Rohs, S., Berkes, F., Smit, H. G. J., Krämer, M., Spelten, N., Spichtinger, P., Nédélec, P., and Wahner, A.: Ice-supersaturated air masses in the northern mid-latitudes from regular in situ observations by passenger aircraft: vertical distribution, seasonality and tropospheric fingerprint, Atmos. Chem. Phys., 20, 8157–8179, https://doi.org/10.5194/acp-20-8157-2020, 2020.

Quaas, J., Gryspeerdt, E., Vautard, R., and Boucher, O.: Climate impact of aircraft-induced cirrus assessed from satellite observations before and during COVID-19, Environ. Res. Let., 16, 064 051, https://doi.org/10.1088/1748-9326/abf686, 2021.

Ramanathan, V., Cess, R. D., Harrison, E. F., Minnis, P., Barkstrom, B. R., Ahmad, E., and Hartmann, D.: Cloud-radiative forcing and climate: Results from the Earth radiation budget experiment, Science, 243, 57–63, http://www.jstor.org/stable/1703174, 1989.

Rap, A., Forster, P. M., Jones, A., Boucher, O., Haywood, J. M., Bellouin, N., and De Leon, R. R.: Parameterization of contrails in the UK Met Office Climate Model, J. Geophys. Res. Atmos., 115, D10 205, https://doi.org/10.1029/2009JD012443, 2010.

Reutter, P., Neis, P., Rohs, S., and Sauvage, B.: Ice supersaturated regions: properties and validation of ERA-Interim reanalysis with IAGOS in situ water vapour measurements, Atmos. Chem. Phys., 20, 787–804, https://doi.org/10.5194/acp-20-787-2020, 2020.

Ruzmaikin, A., Aumann, H. H., and Manning, E. M.: Relative humidity in the troposphere with AIRS, J. Atmos. Sci., 71, 2516 – 2533, https://doi.org/10.1175/JAS-D-13-0363.1, 2014.

Sanogo, S., Boucher, O., Bellouin, N., Borella, A., Wolf, K., and Rohs, S.: Variability of the properties of the distribution of the relative humidity with respect to ice: Implications for contrail formation, EGUsphere, 2023, 1–24, https://doi.org/10.5194/egusphere-2023-2601, 2023.

Schmidt, E.: Die Entstehung von Eisnebel aus den Auspuffgasen von Flugmotoren, in Deutschen Akademie der Luftfahrtforschung, Verlag R. Oldenbourg, 44, 1–15, 1941.

Schumann, U.: On conditions for contrail formation from aircraft exhausts, Meteorologische Zeitschrift, 5, 4–23, https://doi.org/10.1127/metz/5/1996/4, 1996.

Schumann, U.: A contrail cirrus prediction model, Geosci. Model Dev., 5, 543–580, https://doi.org/10.5194/gmd-5-543-2012, 2012.

Schumann, U. and Graf, K.: Aviation-induced cirrus and radiation changes at diurnal timescales, J. Geophys. Res. Atmos., 118, 2404–2421, https://doi.org/10.1002/jgrd.50184, 2013.

Schumann, U., Hempel, R., Flentje, H., Garhammer, M., Graf, K., Kox, S., Lösslein, H., and Mayer, B.: Contrail study with ground-based cameras, Atmos. Meas. Tech., 6, 3597–3612, https://doi.org/10.5194/amt-6-3597-2013, 2013.

Schumann, U., Penner, J. E., Chen, Y., Zhou, C., and Graf, K.: Dehydration effects from contrails in a coupled contrail–climate model, Atmos. Chem. Phys., 15, 11 179–11 199, https://doi.org/10.5194/acp-15-11179-2015, 2015.

Schumann, U., Poll, I., Teoh, R., Koelle, R., Spinielli, E., Molloy, J., Koudis, G. S., Baumann, R., Bugliaro, L., Stettler, M., and Voigt, C.: Air traffic and contrail changes over Europe during COVID-19: a model study, Atmos. Chem. Phys., 21, 7429–7450, https://doi.org/10.5194/acp-21-7429-2021, 2021.

Smit, H. G. J., Volz-Thomas, A., Helten, M., Paetz, W., and Kley, D.: An in-flight calibration method for near-real-time humidity measurements with the airborne MOZAIC sensor, J. Atmos. Ocean. Technol., 25, 656 – 666, https://doi.org/10.1175/2007JTECHA975.1, 2008.

Spichtinger, P., Gierens, K., Smit, H. G. J., Ovarlez, J., and Gayet, J.-F.: On the distribution of relative humidity in cirrus clouds, Atmos. Chem. Phys., 4, 639–647, https://doi.org/10.5194/acp-4-639-2004, 2004.

Stephenson, D. B.: Use of the “odds ratio” for diagnosing forecast skill, Weather Forecasting, 15, 221–232, https://doi.org/10.1175/1520-0434(2000)015<0221:UOTORF>2.0.CO;2, 2000.

Teoh, R., Schumann, U., Majumdar, A., and Stettler, M. E. J.: Mitigating the climate forcing of aircraft contrails by small-scale diversions and technology adoption, Environ. Sci. Technol., 54, 2941–2950, https://doi.org/10.1021/acs.est.9b05608, pMID: 32048502, 2020.

Teoh, R., Schumann, U., Gryspeerdt, E., Shapiro, M., Molloy, J., Koudis, G., Voigt, C., and Stettler, M. E. J.: Aviation contrail climate effects in the North Atlantic from 2016 to 2021, Atmos. Chem. Phys., 22, 10 919–10 935, https://doi.org/10.5194/acp-22-10919-2022, 2022a.

Teoh, R., Schumann, U., Voigt, C., Schripp, T. Shapiro, M., Engberg, Z., Molloy, J., Koudis, G., and Stettler, M. E. J.: Targeted Use of Sustainable Aviation Fuel to Maximize Climate Benefits, Environ. Sci. Technol., https://doi.org/10.1021/acs.est.2c05781, 2022b.

Tompkins, A. M., Gierens, K., and Rädel, G.: Ice supersaturation in the ECMWF integrated forecast system, Q. J. Royal. Meteorol. Soc., 133, 53–63, https://doi.org/https://doi.org/10.1002/qj.14, 2007.

Voigt, C., Schumann, U., Minikin, A., Abdelmonem, A., Afchine, A., Borrmann, S., Boettcher, M., Buchholz, B., Bugliaro, L., Costa, A., Curtius, J., Dollner, M., Dörnbrack, A., Dreiling, V., Ebert, V., Ehrlich, A., Fix, A., Forster, L., Frank, F., Fütterer, D., Giez, A., Graf, K., Grooß, J.-U., Groß, S., Heimerl, K., Heinold, B., Hüneke, T., Järvinen, E., Jurkat, T., Kaufmann, S., Kenntner, M., Klingebiel, M., Klimach, T., Kohl, R., Krämer, M., Krisna, T. C., Luebke, A., Mayer, M., Mertes, S., Molleker, S., Petzold, A., Pfeilsticker, K., Port, M., Rapp, M., Reutter, P., Rolf, C., Rose, D., Sauer, D., Schäfler, A., Schlage, R., Schnaiter, M., Schneider, J., Spelten, N., Spichtinger, P., Stock, P., Walser, A., Weigel, R., Weinzierl, B., Wendisch, M., Werner, F., Wernli, H., Wirth, M., Zahn, A., Ziereis, H., and Zöger, M.: ML-CIRRUS – The airborne experiment on natural cirrus and contrail cirrus with the high-altitude long-range research aircraft HALO, Bull. Amer. Meteor. Soc., https://doi.org/10.1175/BAMS-D-15-00213.1, 2017.

Wolf, K., Bellouin, N., and Boucher, O.: Long-term upper-troposphere climatology of potential contrail occurrence over the Paris area derived from radiosonde observations, Atmos. Chem. Phys., 23, 287–309, https://doi.org/10.5194/acp-23-287-2023, 2023.

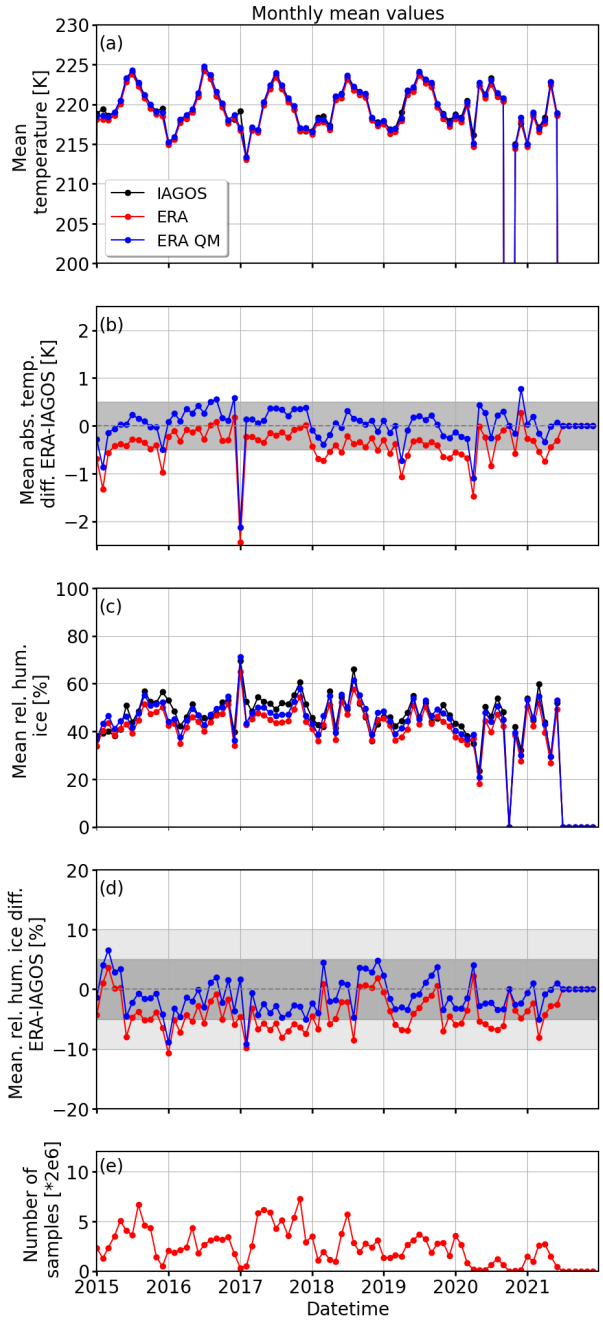

Figure B1. **(a)** Time series of monthly mean temperature (in K) from IAGOS (black), ERA5 (red), and corrected ERA5 (blue). **(b)** Time series of temperature difference (in K) from ERA5 minus IAGOS (red) as well as corrected ERA5 minus IAGOS (blue). Panels **(c)** and **(d)** are similar to **(a)** and **(b)** but for relative humidity with respect to ice r_{ice} in $\%$. **(e)** Total number of samples available to calculate the monthly mean.

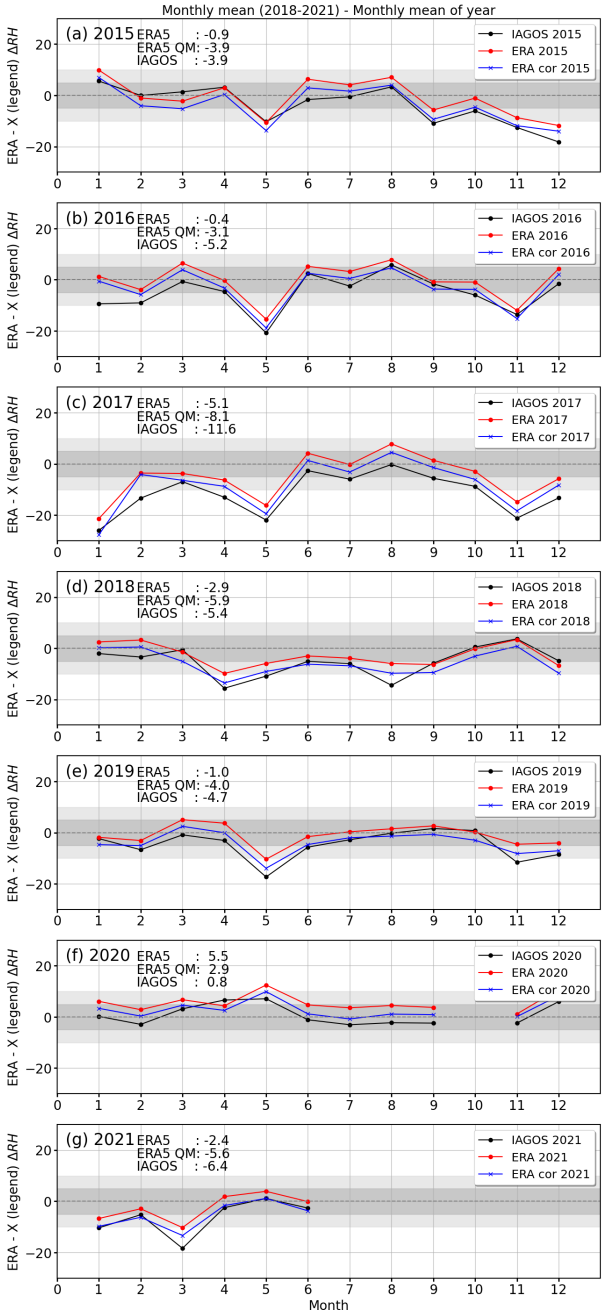

Figure B2. (a–g) Anomalies of relative humidity from ERA5 $r_{\mathrm{ERA,ice}}$ (red), corrected ERA5 $r^{\mathrm{cor}}_{\mathrm{ERA,ice}}$ (blue), and IAGOS $r_{\mathrm{P1,ice}}$ (black) with respect to the multi-year $r_{\mathrm{ERA,ice}}$ for the years 2015 to 2021. Differences are given in unit of relative humidity.

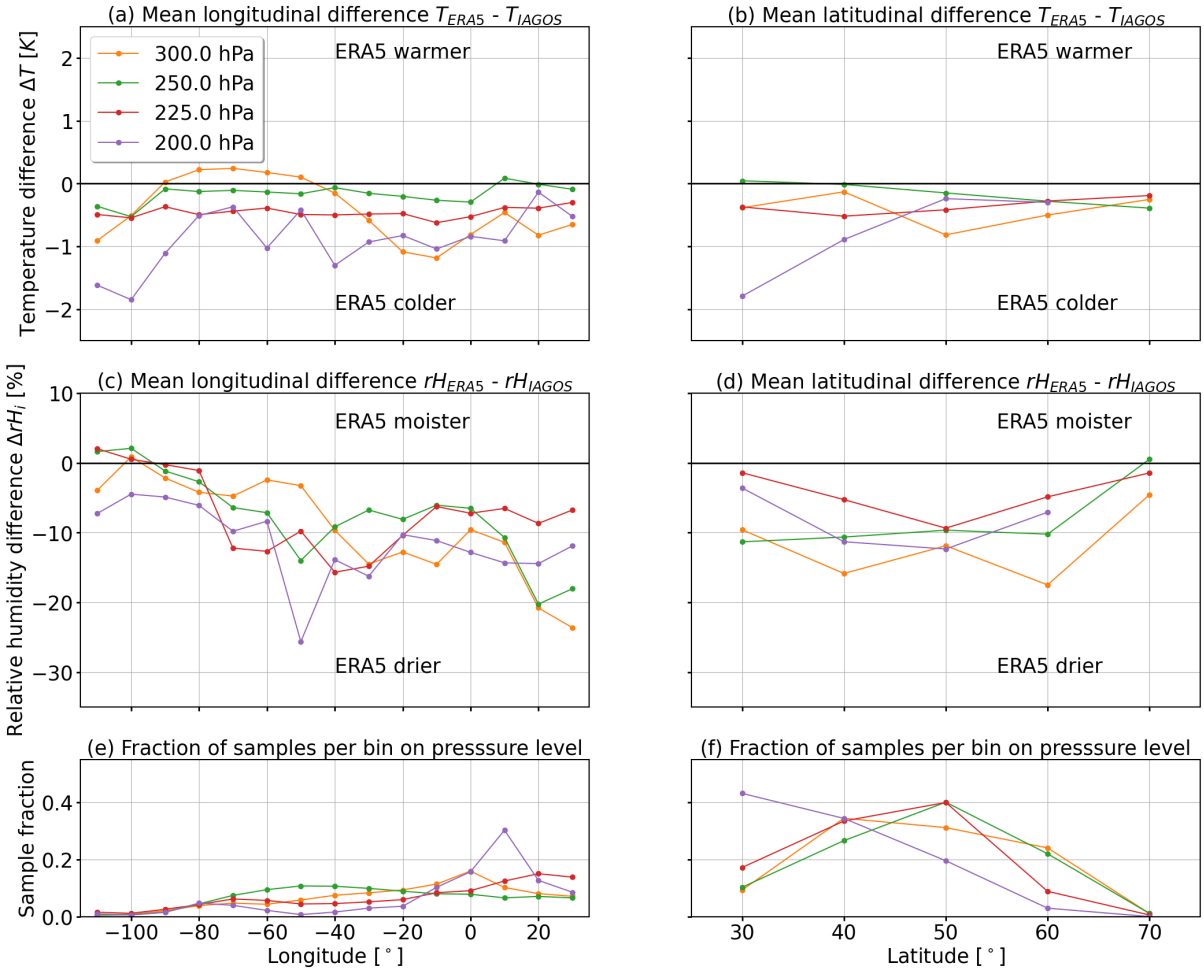

Figure C1. (a–b) Temperature difference ΔT (in K) between ERA5 and IAGOS as a function of Longitude and Latitude, respectively. **(c–d)** Same as top row but for difference in relative humidity r_{ice} (in %). Pressure levels of 300, 250, 225, and 200 hPa are indicated in orange, green, red, and purple, respectively. **(e–f)** Fraction of available samples per longitude or latitude bin with respect to the total number of samples per pressure level.