

# Technical Note: Hybrid Machine Learning Model for Bias Correction of UTLS Relative Humidity against IAGOS Observations in ERA5 Reanalysis

Mathieu Antonopoulos<sup>1,2</sup>, Jérémie Juvin-Quarroz<sup>1</sup>, and Olivier Boucher<sup>1</sup>

<sup>1</sup>Institut Pierre-Simon Laplace, Sorbonne Université/CNRS, Paris, France

<sup>2</sup>Institut Polytechnique de Paris, 91120 Palaiseau, France

**Correspondence:** Jérémie Juvin-Quarroz (jeremie.juvin-quarroz@ipsl.fr)

**Abstract.** Persistent contrail cirrus form in Ice-Supersaturated Regions (ISSRs) and are responsible for a large portion of aviation’s non-CO<sub>2</sub> climate impact. Avoiding ISSRs through flight rerouting has been proposed as a short-term mitigation strategy. However, accurate prediction of the Relative Humidity with respect to ice, RH<sub>i</sub>, distribution within ISSRs at cruising altitude remains difficult. Observations are problematic: Satellite-based global measurements carry large uncertainties while *in-situ* measurements offer a limited spatial coverage. On the contrary, reanalyses offer a global estimate of RH<sub>i</sub>, but it is suffer from a dry bias near the tropopause where ISSRs are located as well as significant random errors. In this study, we develop a hybrid machine learning model to improve RH<sub>i</sub> estimates in the upper troposphere and lower stratosphere using ERA5 and aircraft measurements from the In-service Aircraft for a Global Observing System. The model combines a XGBoost regressor for drier conditions (RH<sub>i</sub> < 85%) and an Artificial Neural Network (ANN) for more humid cases (RH<sub>i</sub> > 85%). This hybrid approach significantly outperforms raw ERA5 data, leveraging the ANN’s ability to capture non-linear relationships and the XGBoost’s robustness in handling drier conditions. The mean absolute error (MAE) is reduced from 13.7% to 11.4% and the Equitable Threat Score (ETS) for ISSR detection is improved from 0.36 to 0.44. The greatest improvement is observed in the lower stratosphere, where the ETS increases by 0.18 and the MAE drops to 10.71%. These improvements mark a key step toward more reliable identification of ISSRs, helping reduce the uncertainties that currently limit effective flight-rerouting strategies.

## 1 Introduction

Aviation contributes to climate change through both carbon dioxide (CO<sub>2</sub>) emission and non-CO<sub>2</sub> effects. The latter may have accounted for 66% of aviation’s total net Effective Radiative Forcing (ERF) in 2018 (Lee et al., 2021). The largest contributor to non-CO<sub>2</sub> effective radiative forcing (ERF) is the radiative forcing from condensation trails—contrails—which are linear ice clouds formed behind aircrafts as hot exhaust gases mix with cold ambient air. The physical processes governing contrail formation are well understood (Schmidt, 1941; Appleman, 1953; Schumann, 1996). While most contrails dissipate quickly and have a negligible climatic impact, those that form within Ice Supersaturated Regions (ISSRs) can persist for hours (Kärcher,

2018; Kärcher and Corcos, 2025), eventually developing into contrail cirrus clouds with microphysical and optical properties similar to those of natural cirrus (Li et al., 2023; Wang et al., 2024).

25 ISSRs are regions of the atmosphere where the Relative Humidity with respect to ice is supersaturated,  $RH_i > 100\%$ , which is a meta-stable thermodynamic state. The size of ISSRs has been estimated based on the distance flown within such regions by Measurement of Ozone and Water Vapor by Airbus In-Service Aircraft (MOZAIC) aircraft, showing that they can extend over  $\mathcal{O}(10 - 100)$  kilometers (Gierens and Spichtinger, 2000). ISSRs are typically located at the tropopause level and are rather frequent across the globe (Gierens et al., 1999; Spichtinger et al., 2003a). Their formation is facilitated by the presence of  
30 upward winds and the meeting of cold and warm air streams (Gierens and Brinkop, 2012). The dynamics of ISSRs also present complex pattern as it has been determined that ISSRs' displacement may be slower than the surrounding wind field's (Hofer and Gierens, 2025).

On average, it is estimated that aircraft spends 13.5% of their flight time in ISSRs (Gierens et al., 1999) and that only a small proportion of flights might be responsible for the majority of persistent contrail formation. For instance, it has been estimated  
35 that only 2.7% of total flights are responsible for 80% of the total contrail RF in 2019 (Teoh et al., 2024). These findings have motivated flight rerouting as a short-term mitigation strategy, *e.g.* Rosenow et al. (2018). Being able to meteorologically forecast ISSRs with enough accuracy would allow a slight deviation of the flight route to avoid the formation of persistent contrails, at the cost of a small amount of additional  $\text{CO}_2$  emissions.

Global ISSR forecasts are generally derived for numerical weather forecasting systems, or nowcast from in situ measure-  
40 ments or inferred from remote sensing. Both approaches rely on accurate detections of ISSRs, in the first case to validate models, or in the second through measurements. However, detecting ISSRs remains challenging, particularly through satellite observations (Gierens et al., 2004; Gettelman et al., 2006; Lamquin et al., 2012; Hegglin et al., 2013) due to their relatively narrow vertical extent (Spichtinger et al., 2003b; Wolf et al., 2023). The tropopause region, where ISSRs typically occur, also features steep vertical gradients in humidity (Stenke et al., 2008; Zahn et al., 2014), necessitating high vertical resolution for  
45 accurate characterization. Both remote measurements with lidar (Groß et al., 2014; Krüger et al., 2022) and in-situ observations, such as those made with balloon-borne instruments (Heymsfield et al., 1998) and with on-flight sensors (Diao et al., 2015b; Petzold et al., 2015), provide accurate measurements of  $RH_i$ , but are spatially and temporally limited.

Hence we must rely on numerical weather prediction models to obtain  $RH_i$  values across the global atmosphere. We use the ERA5 reanalysis data (Hersbach et al., 2020) as it is highly trusted source of numerical weather prediction. However,  
50 it is known to suffer from a dry bias in the Upper Troposphere (UT) and Lower Stratosphere (LS) (Teoh et al., 2022; Rolf et al., 2023; Platt et al., 2024), which means that the model often predict lower relative humidity values compared to the observed values. This bias, as well as significant random errors, can lead to considerable errors in  $RH_i$  estimates, particularly in regions where contrails are likely to form. Previous work on the correction of humidity bias in ERA5 have shown potential for improvement of the ISSR detection in the UTLS. Wolf et al. (2025) applied a bivariate Quantile Mapping (QM) correction  
55 to ERA5 temperature and relative humidity fields, significantly reducing the dry bias. They report a reduction of the  $RH_i$  bias from approximately -4.5% to -1.2% on average at pressure levels of 250, 225, and 200 hPa. This bias reduction leads to an

improved classification of different contrail categories. Wang et al. (2025) proposed a machine learning (ML) approach to improve ERA5's RH<sub>i</sub> estimates in the UT using an Artificial Neural Network (ANN).

In this study, we present a hybrid ensemble machine learning approach to correct the dry bias in ERA5's RH<sub>i</sub> in the UTLS. We combine a XGBoost regressor (Chen and Guestrin, 2016) and an ANN, leveraging the complementary strengths of both models. ANNs are well-suited for modeling complex, non-linear relationships in multivariate data (Mihajlović and Nikolić, 2009), while tree-based models like XGBoost excel in handling structured data and often deliver state-of-the-art performance (Chen and Guestrin, 2016). We dynamically chose the model based on the initial RH<sub>i</sub> value provided by ERA5: the XGBoost is used for prediction in drier conditions (RH<sub>i</sub> < 85%), while the ANN is used for more humid cases (RH<sub>i</sub> > 85%). ISSR formation involves more complex and non-linear atmospheric dynamics (Diao et al., 2014, 2015a), which are better captured by ANNs due to their flexibility and non-linear representation capabilities. Effectively, the ANN captures subtle variations relevant to ISSR classification, while the XGBoost model is used to correct bias under simpler thermodynamic conditions.

This paper is organized as follows. In Section 2 we describe the IAGOS and ERA5 datasets, the collocation procedure, filtering, and feature engineering. Section 3 details the ANN and XGBoost architectures, training procedures, and the hybrid model. Section 4 presents our results, showing improvements in regression and classification scores under different meteorological conditions.

## 2 Data

In this study, we will use two complementary datasets. The ERA5 reanalysis data from the European Centre for Medium-Range Weather Forecasts (ECMWF, Copernicus Climate Change Service, 2025), and the *in-situ* observations from the In-service Aircraft for a Global Observing System (IAGOS, Petzold et al., 2015) program. ERA5 provides a global and consistent estimate of atmospheric conditions through a wide range of variables, while IAGOS offers direct measurements of RH<sub>i</sub> at cruise altitudes. Combining these datasets enables a supervised learning approach where the ERA5 features serve as model inputs, while the estimation of RH<sub>i</sub> from IAGOS from temperature and relative humidity measurements serve as the target variable. The following subsections describe each dataset, along with our preprocessing and feature engineering methodology.

Unlike Wang et al., 2025, we focus on the North Atlantic region, spanning latitudes from 40°N to 70°N and longitudes from 65°W to 5°E (Fig. 1). This region shows a relatively uniform distribution of the IAGOS data and is known for frequent contrail formation due to high air traffic and favorable atmospheric conditions. For these reasons, it is thought to be a potential region for contrail avoidance trials. We selected the year 2022 as our study period as it offers recent atmospheric conditions and comes after the maximum COVID disruption of international flights in 2020 and 2021. Our data processing is implemented based on methodology from Wang et al., 2025, with some modifications to ensure proper separation of the training and test datasets to avoid temporal leakage.

## 2.1 ECMWF reanalysis data (ERA5)

ERA5 data are provided on a regular latitude–longitude grid with a horizontal resolution of  $0.25^\circ \times 0.25^\circ$ , corresponding to an approximate grid spacing of 19.5 km at  $45^\circ\text{N}$  latitude. Vertically, the dataset includes 37 pressure levels extending from the surface to the LS. We focus on pressure levels between 500 hPa and 125 hPa, specifically use the following levels: 500, 450, 400, 350, 300, 250, 225, 200, 175, 150, and 125 hPa. The extracted ERA5 variables provide a comprehensive description of the atmospheric conditions (see Table 1 for a summary of all extracted variables and their units). The ERA5 data serve as input for both statistical validation and the training of machine learning models. We use all variables as features for the model, and the  $\text{RH}_i\text{\_ERA5}$  measurement as a baseline to evaluate model performance.

## 2.2 In-service Aircraft for a Global Observing System (IAGOS)

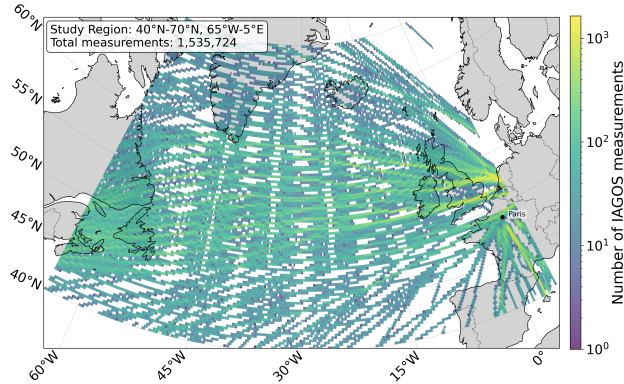
IAGOS is a European research initiative providing long-term, high-resolution *in-situ* atmospheric measurements to support the understanding of atmospheric composition, air quality, and climate processes. Data are collected by equipping commercial aircraft, mostly Airbus A330s, with the 'P1 package' set of scientific instruments. Currently, the IAGOS fleet comprises ten aircraft operating worldwide. The  $\text{RH}_i$  is derived from measured temperature, pressure, and water vapor mixing ratio, using the saturation vapor pressure formula by Sonntag (1994). The measurements are recorded at a temporal resolution of 4 seconds, corresponding to a spatial resolution of approximately 1 km at cruise speeds. However the data points are averaged over a longer period of few minutes because of the time lags involved with the instrument (Borella et al., 2024).

The spatial distribution of IAGOS measurements is inherently non-uniform as commercial flights tend to follow similar routes (cf. Fig. 1). Additionally, the limited number of aircraft in the IAGOS fleet contributes to a concentration of the data in some regions. The highest data density is found along transatlantic flight corridors and Western Europe. The North Atlantic and European regions are the most represented, as most flights from IAGOS are issued by European companies, and transatlantic flights represent an important share of global aviation traffic. This area is of particular interest for contrail studies, given the frequent occurrence of ISSRs (Reutter et al., 2020). Most measurements are collected between 400 and 200 hPa, matching the standard cruise altitude.

To ensure data quality, we retain only samples flagged as 'good' by the IAGOS quality control system. We also restrict the dataset to measurements between 400 and 200 hPa. According to Rolf et al. (2023) and their analysis of the airborne measurements, the P1 sensors used in IAGOS show limited accuracy in dry conditions ( $\text{RH}_i < 10\%$ ). Hence, we exclude those samples from our dataset. After applying all these filtering steps for the year 2022 on the North-Atlantic region, we obtain a dataset comprising 1,535,724 data points from 678 flights.

## 2.3 Data collocation

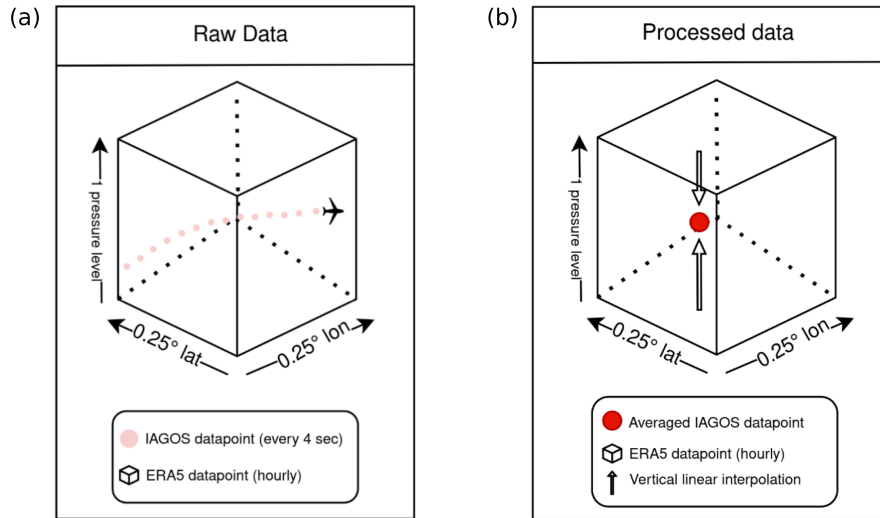
The IAGOS and ERA5 data have different resolutions in both time and space, hence we must proceed to a collocation of the data. Once we collected and filtered the IAGOS data for each flight, we determine the closest point in ERA5 by matching the closest pressure level, longitude and latitude points (at a  $0.25^\circ$  resolution) and timestep (closest hour). Once we get these



**Figure 1.** Density map of IAGOS data points in the North Atlantic region for the year 2022. The color scale indicates the number of data points per  $0.25^\circ \times 0.25^\circ$  gridbox.

indices, we find the closest grid box in ERA5 to each IAGOS data point. The IAGOS points matching the same grid box are  
 120 grouped. Given the four-second resolution of the data and the average grid box length of 19 km, each group contains up to 23  
 IAGOS points and 15 IAGOS points on average. Once grouped, the IAGOS data are averaged within one grid box to obtain  
 a single IAGOS point per ERA5 grid box and timestep. Given that points from same flights are usually located at a similar  
 altitude within a grid box, we effectively horizontally average the IAGOS data for longitudes and latitudes. We further proceed  
 to a vertical linear interpolation of the ERA5 data to match the averaged altitude of the IAGOS point. Future work should  
 125 test log-pressure linear interpolation to reflect the atmosphere’s vertical structure. A visual representation of this process is  
 given in Fig. 2. In order to maintain physical relationships between variables, we recalculate  $RH_i$ \_ERA5 with the interpolated  
 variables  $T$  and  $q$ , which are respectively the air temperature and the specific humidity at the relevant pressure value  $p$ . This  
 is done using the Clausius-Clapeyron equation, with a Magnus form approximation for saturation vapor pressure over ice as  
 described by Alduchov and Eskridge (1996). A comparison between observed  $RH_i$  from IAGOS and  $RH_i$  predictions from  
 130 ERA5 reanalysis is shown in Fig. 4a.

The dataset is then separated into training, validation and test datasets. Careful separation is essential, as atmospheric con-  
 ditions can exhibit strong spatial and temporal correlations. Without proper splitting, the model might train on samples nearly  
 identical to those in the test dataset, potentially leading to overfitting and artificially high performance on validation and test  
 datasets. To avoid this issue while keeping a representative distribution across seasons, days, and hours, we implemented a  
 135 by-day sampling strategy. Specifically, we sample five days for training, one day gap, one day for validation, one day gap, five  
 days for training, one day gap, one day for testing and one day gap, uniformly throughout the year. This differs from Wang  
 et al. (2025) in two ways: i) we use five days for training (four days in Wang et al. (2025)), ii) an additional day gap is used  
 after the validation or testing day to prevent correlations with the following training period. After collocation and split, we  
 obtain 64,712 training samples, 7,025 validation samples and 6,085 test samples. To address the underrepresentation of high  
 140  $RH_i$  values and reduce potential bias, data augmentation is applied to the training dataset: samples with  $RH_i$  above 120% are



**Figure 2.** Illustration of the collocation process between IAGOS and ERA5 datasets. Diagram (a) presents the datasets before collocation, while diagram (b) illustrates the data after preprocessing. IAGOS measurements are averaged within each ERA5 grid box, and the ERA5 variables are interpolated to match the mean altitude of the IAGOS point. The  $RH_i$  is then recomputed based on the interpolated variables.

oversampled by a factor of three, while those below 20%  $RH_i$  are undersampled by a factor of 0.4. This approach results in a  $RH_i$  distribution that is more suitable for our needs.

Finally, we ensure that our processed data are representative of ISSRs characteristics by comparing it to the results in Gierens and Brinkop (2012). We report the distribution of vorticity and divergence in Fig. A1.

## 145 2.4 Feature engineering

ERA5 provides an extensive set of variables, but to further improve the input data and capture new relation (Fig. 3), additional derived variables are added as input features. Following the methodology proposed by Wang et al. (2025), we add for each data point, the selected variables two pressure levels above and below the current pressure level (totaling five levels) as well as data from two hours and six hours prior to the current timestep. In addition we extracted several features from the original ERA5  
 150 and IAGOS data to enrich the model inputs and improve the correction of the  $RH_i$  estimates. These features aim to capture the vertical structure and dynamics of the atmosphere.

To capture the vertical temperature profile, we compute temperature gradients at different scales. These include the gradient between the current level and the two levels immediately above ( $T_{grad\_up}$ ), between the current level and the two levels immediately below ( $T_{grad\_down}$ ), a centered gradient between one level above and one below ( $T_{grad\_centered}$ ), and a  
 155 lower-resolution gradient between two levels above and two below ( $T_{grad\_overall}$ ), in  $K\ hPa^{-1}$ . Similarly, gradients of the vorticity ( $vo\_grad\_up$ ,  $vo\_grad\_down$ ,  $vo\_grad\_centered$ ,  $vo\_grad\_overall$ ), in  $s^{-1}\ hPa^{-1}$ , and the  $RH_i$

(RH<sub>i</sub>\_grad\_up, RH<sub>i</sub>\_grad\_down, RH<sub>i</sub>\_grad\_centered, RH<sub>i</sub>\_grad\_overall), in %, are computed following the same pattern.

To account for the diurnal and seasonal variability in atmospheric conditions, we embedded time information using cosine functions. The local time at the observation point was computed from the UTC hour and longitude, then encoded using sine and cosine transforms to capture daily cycles (`cos_hour`, `sin_hour`). Similarly, the day of the year (`cos_day`, `sin_day`) is encoded to reflect seasonal effects. These four quantity are normalized by  $2\pi$ .

These engineered features were incorporated into the training data to allow the model to better represent the complex atmospheric processes influencing the RH<sub>i</sub> (cf. Table 1). Some of the engineered features proved to be statistically meaningful. In particular, the temperature gradient showed a stronger correlation with RH<sub>i</sub> than the original temperature variable. Time embeddings exhibited almost null direct correlation with RH<sub>i</sub>. However, they were identified as important contributors in the feature importance analysis of the ANN, indicating that their effect may be captured in more complex, non-linear relationships within the model (cf. Fig. 3 and Table 4).

Data is then categorized according to distinct meteorological conditions following methodology from Wang et al. (2025). To determine the cloud presence, we used the cloud ice water content (`ciwc_ERA5`) variable. If `ciwc` equals zero at the current pressure level and two levels above and below it ( $\pm 2$  pressure levels), the data point is classified as clear sky. If any of these five levels has a non-zero `ciwc` value, the point is classified as cloudy. Additionally, we separated the points from the UT and LS. Commercial flights cruise at an atmospheric level called the tropopause, which is the interface between the troposphere and the stratosphere. This separation helps to account for the differing thermodynamic properties and humidity distributions between the troposphere and stratosphere. The classification is based on potential vorticity (`pv_ERA5`). Data points are labeled as UT if potential vorticity is less than 2 PVU (Potential Vorticity Units). Conversely, points with  $PV > 2$  PVU are classified as LS. We use the typical conversion rate of  $1 \text{ PVU} = 10^6 \text{ K kg}^{-1} \text{ m}^2 \text{ s}^{-1}$ . In the test dataset, 2641 points are classified as Cloudy-Sky, 3444 as Clear-Sky, 4566 as LS and 1519 as UT.

## 2.5 Evaluation Metrics

The performance of the model is assessed using different metrics. Since we primarily solve a regression task, we use the Mean Absolute Error (MAE), expressed as RH<sub>i</sub> percentage points. However, MAE alone does not fully capture the model’s ability to correct the biases in ERA5. The aim of the model is to accurately identify ISSRs, which requires a classification perspective. Therefore, following common practices in ISSR prediction (such as in Wang et al., 2025 and Wolf et al., 2025), the Equitable Threat Score (ETS) is used as a complementary evaluation metric. The ETS is a skill score that quantifies the accuracy of a binary classification (e.g., detecting ISSRs where RH<sub>i</sub>  $\geq$  100%) while accounting for hits that could occur by random chance. The ETS is defined as

$$\text{ETS} = \frac{TP - r}{TP + FP + FN - r}, \quad (1)$$

where  $TP$  denotes true positives (correctly predicted ISSRs),  $FP$  denotes false positives (non-ISSRs incorrectly predicted as ISSRs),  $FN$  denotes false negatives (ISSRs missed by the model),  $TN$  denotes true negatives (correctly predicted non-ISSRs),

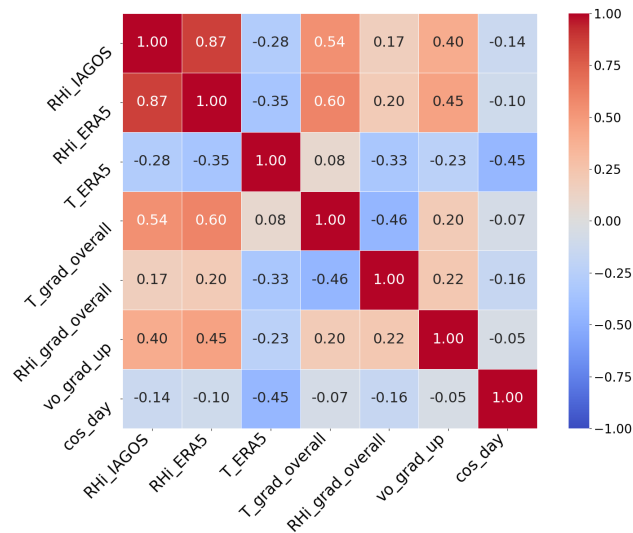
Category	Feature Name	Description	Unit
ERA5 variables	T_ERA5	Air temperature	K
	q_ERA5	Specific humidity	$\text{kg} \cdot \text{kg}^{-1}$
	w_ERA5	Vertical wind speed	$\text{Pa} \cdot \text{s}^{-1}$
	u_ERA5	Zonal (west-east) wind component	$\text{m} \cdot \text{s}^{-1}$
	v_ERA5	Meridional (south-north) wind component	$\text{m} \cdot \text{s}^{-1}$
	pv_ERA5	Potential vorticity	$\text{K} \cdot \text{m}^2 \cdot \text{kg}^{-1} \cdot \text{s}^{-1}$
	ciwc_ERA5	Cloud ice water content	$\text{kg} \cdot \text{kg}^{-1}$
	geopt_ERA5	Geopotential	$\text{m}^2 \cdot \text{s}^{-2}$
	vo_ERA5	Relative Vorticity	$\text{s}^{-1}$
	d_ERA5	Horizontal divergence	$\text{s}^{-1}$
	RHi_ERA5*	RH <sub>i</sub>	%
Engineered features	T_grad_up	Temp. gradient: current/upper level	$\text{K} \cdot \text{hPa}^{-1}$
	T_grad_down	Temp. gradient: current/lower level	$\text{K} \cdot \text{hPa}^{-1}$
	T_grad_centered	Temp. gradient: $\pm 25$ hPa	$\text{K} \cdot \text{hPa}^{-1}$
	T_grad_overall	Temp. gradient: $\pm 50$ hPa	$\text{K} \cdot \text{hPa}^{-1}$
	RHi_grad_up	RH <sub>i</sub> gradient: current/upper level	$\% \text{hPa}^{-1}$
	RHi_grad_down	RH <sub>i</sub> gradient: current/lower level	$\% \text{hPa}^{-1}$
	RHi_grad_centered	RH <sub>i</sub> gradient: $\pm 25$ hPa	$\% \text{hPa}^{-1}$
	RHi_grad_overall	RH <sub>i</sub> gradient: $\pm 50$ hPa	$\% \text{hPa}^{-1}$
	vo_grad_up	Vorticity gradient: current/upper level	$\text{s}^{-1} \cdot \text{hPa}^{-1}$
	vo_grad_down	Vorticity gradient: current/lower level	$\text{s}^{-1} \cdot \text{hPa}^{-1}$
	vo_grad_centered	Vorticity gradient: $\pm 25$ hPa	$\text{s}^{-1} \cdot \text{hPa}^{-1}$
	vo_grad_overall	Vorticity gradient: $\pm 50$ hPa	$\text{s}^{-1} \cdot \text{hPa}^{-1}$
	PVU	Potential vorticity in PVU	PVU
	cloudy	Ice-cloud flag	{0, 1}
	cos_hour	Cosine of local time	—
sin_hour	Sine of local time	—	
cos_day	Cosine of day in year	—	
sin_day	Sine of day in year	—	
IAGOS variables	p_IAGOS	IAGOS pressure	hPa
	<b>RHi_IAGOS</b>	IAGOS RH <sub>i</sub>	%

**Table 1.** Summary of all features used for training, grouped by category. ERA5 variables include values at current pressure level and context,  $\pm 2$  pressure levels, and 2 h and 6 h prior current timestep (i.e. 7 versions for each ERA5 variable). The target IAGOS value is indicated in **bold**. RHi\_ERA5 is recomputed after linear interpolation to maintain consistency and physical dependencies

190 and  $r$  is the number of hits expected by random chance, computed as

$$r = \frac{(TP + FP)(TP + FN)}{TP + FP + FN + TN}, \quad (2)$$

The ETS adjusts for random hits and is particularly suited for evaluating rare event prediction. It ranges from  $-1/3$  (worse than random chance) to 1 (perfect prediction), with 0 indicating no better results than random guessing. Our baseline for evaluating model improvement is the ERA5 RH<sub>i</sub> prediction, which shows a 0.36 ETS and 13.7% MAE (see Table 3).



**Figure 3.** Correlation matrix showing the relationships between selected engineered features, the target variable RHi\_IAGOS, and the initial ERA5 estimate. The engineered gradient features show a relatively strong correlations with the target RHi\_IAGOS. The time embeddings (cos\_day) exhibit low direct correlation with RHi\_IAGOS.

### 195 3 Hybrid Machine Learning Model development

Correcting the dry bias in ERA5’s RH<sub>i</sub> estimates in the UTLS presents a significant challenge due to the inherently non-linear nature of atmospheric processes. To address this, we propose a hybrid machine learning approach that combines two complementary models: an ANN and a XGBoost. The ANN is well-suited for capturing complex, non-linear dependencies, particularly under high humidity conditions, while XGBoost performs robustly in drier regimes.

200 The hybrid strategy dynamically selects the appropriate model for prediction based on the ERA5-estimated RH<sub>i</sub> value. Specifically, for samples where ERA5 RH<sub>i</sub> is below 85%, predictions are taken from the XGBoost model, which has demonstrated superior performance in lower-humidity environments. Conversely, for RH<sub>i</sub> values exceeding 85%, the ANN model is employed, leveraging its capacity to extract highly non-linear patterns, such as those associated with ISSRs. The 85% threshold was chosen based on two key considerations. First, ERA5 is known to exhibit a MAE of approximately 15% in humid  
 205 conditions. By subtracting this from the 100% threshold typically used to identify ISSRs, we ensure that the ANN model is prioritized in critical high-humidity cases. Second, validation dataset performance confirmed that this hybrid architecture consistently outperforms either model used in isolation.

Firstly, the ANN component is trained to predict RH<sub>i</sub> using the input features summarized in Table 1. Data are normalized using min–max scaling, with the RH<sub>i</sub> range extended to 200% to reduce sensitivity to outliers and mitigate dry bias. The  
 210 final network architecture includes three hidden layers with 100 neurons each, employing He initialization and ReLU activation functions. Batch normalization and dropout regularization are applied between layers to stabilize training and reduce overfitting. The output layer uses a linear activation function to predict RH<sub>i</sub>, denoted as RHi\_ANN.

Training is conducted using the Adam optimizer (Kingma and Ba, 2014) with a learning rate of 0.001, decay rate of  $5 \cdot 10^{-3}$ , and momentum of 0.98. The model is trained over 150 epochs with a batch size of 1024, and early stopping is employed to halt training if validation loss does not improve over 20 consecutive epochs. A custom loss function, based on the Mean Squared Error (MSE), is used to emphasize accuracy in high-RH<sub>i</sub> conditions. The loss function is defined as:

$$\mathcal{L}(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n w_i \cdot (y_i - \hat{y}_i)^2, \quad \text{with} \quad w_i = 1 + y_i^\alpha. \quad (3)$$

In Eq. 3,  $y_i$  is the target value at index  $i$ ,  $\hat{y}_i$  is the predicted value at index  $i$ ,  $n$  is the number of samples, and  $\alpha$  is a weight factor. The weight parameter ( $\alpha$ ) that defines the weight of the RH<sub>i</sub> value is denoted WEIGHT\_FACTOR. This weighted approach prioritizes performance in ISSRs, addressing the ERA5 dry bias. Hyperparameters were optimized via grid search on the train dataset as detailed in Table 2.

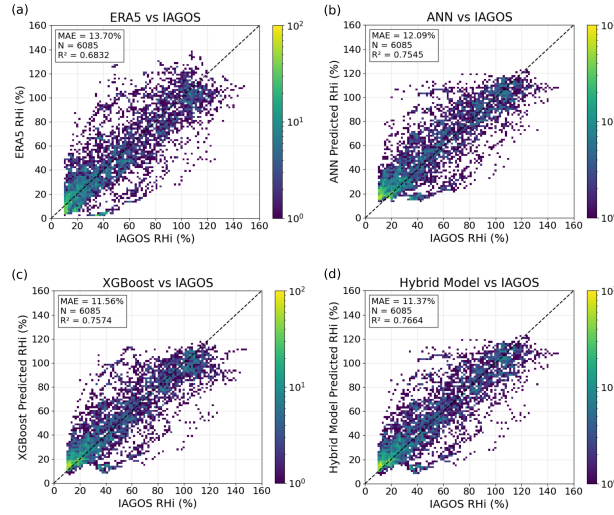
Hyperparameter	Values
BATCH_SIZE	[512, <b>1024</b> , 2048]
DROPOUT_RATE	[0.0, <b>0.1</b> , 0.3]
EPOCHS	[100, <b>150</b> ]
HIDDEN_DIM	[64, <b>100</b> , 128, 256, 512]
LEARNING_RATE	[0.1, 0.01, <b>0.001</b> , 0.0001]
NUM_LAYERS	[2, <b>3</b> , 4, 5]
WEIGHT_DECAY	[ <b>0.005</b> , 0.0001, 0.00001]
MOMENTUM	[0.85, 0.95, <b>0.98</b> , 0.99]
PATIENCE	[10, <b>20</b> , 30]
MIN_DELTA	[ <b>0.0001</b> ]
WEIGHT_FACTOR	[3, 10, <b>30</b> ]
OPTIMIZER	[ <b>adam</b> , sgd]
SCHEDULER_FACTOR	[ <b>0.5</b> ]
SCHEDULER_PATIENCE	[10, <b>20</b> ]

**Table 2.** Hyperparameter grid used for tuning the ANN model. Each row lists the set of values tested for a specific hyperparameter during the grid search process. The values in **bold** indicate the configuration that achieved the best performance on the validation dataset and was selected for final training.

Secondly, the XGBoost model uses the same input feature set as the ANN, including engineered variables. XGBoost is an efficient implementation of gradient boosting known for its scalability, accuracy, and interpretability. The final model was trained with 100 estimators, a learning rate of 0.1, a maximum tree depth of 4, a subsampling ratio of 0.9, and a column sampling ratio per tree of 0.8. In Table A1, we report the performance comparison for a set of ML models.

## 4 Results

The hybrid model achieved an ETS of 0.44 and a MAE of 11.37%. Compared to ERA5, it showed an improvement of +0.08 in ETS and -2.33% in MAE. Fig. 4 compares scatter plots of our different models predictions on the test dataset to the baseline



**Figure 4.** Comparison of RH<sub>i</sub> predictions between different models with respect to the target value (RH<sub>i\_IAGOS</sub>). (a) presents the baseline ERA5 reanalysis, (b) ANN model, (c) XGBoost model, and (d) the hybrid ensemble model. All predictions are made on the test dataset between 400 and 200 hPa over the North Atlantic region (cf. Fig. 1) for the year 2022.

Model	ETS	MAE (%)	Description
ERA5 (Baseline)	0.36	13.70	Original reanalysis data (baseline)
ANN	0.44 (+0.08)	12.09 (-1.61)	Strong ISSR classification, moderate error reduction
XGBoost	0.35 (-0.01)	11.56 (-2.14)	Weak ISSR classification, strong error reduction
Hybrid Model	<b>0.44</b> (+0.08)	<b>11.37</b> (-2.33)	Strong ISSR classification, strong error reduction

**Table 3.** Performance comparison of models for RH<sub>i</sub> correction. Improvement relative to ERA5 baseline shown in parentheses.

ERA5 estimates, see Fig. 4a. Fig. 4c shows that the XGBoost model was able to correct low RH<sub>i</sub> values with great accuracy, consequently showing strong MAE improvement. However, it struggles with higher RH<sub>i</sub> values, resulting in poor ETS. The ANN, Fig. 4b, shows a strong ETS, indicating its ability to classify ISSRs, but it overestimates the lower RH<sub>i</sub> values leading to a higher MAE. The hybrid model shown in Fig. 4d combines the strengths of both models, achieving a good balance between ETS and MAE.

The RH<sub>i</sub> estimates are effectively improved by combining the strengths of both models, leveraging the ANN's ability to capture non-linear relationships and the XGBoost's robustness in handling drier conditions. The hybrid model achieves a higher ETS comparable to the ANN's while maintaining a lower MAE than both models. This hybrid approach demonstrates the potential of combining different machine learning techniques to improve predictions in complex atmospheric conditions. Table 3 summarizes the performance of all models in terms of ETS and MAE.

ANN Model		XGBoost Model	
Feature	Importance	Feature	Importance
RHi_ERA5	0.14	RHi_ERA5_prior_2h	0.31
cos_day	0.09	RHi_ERA5	0.13
RHi_ERA5_up1	0.06	Cloudy	0.12
cos_hour	0.05	T_grad_down	0.09
Cloudy	0.04	RHi_ERA5_up1	0.06

**Table 4.** Top five most important features for the ANN and XGBoost models. As expected, RHi\_ERA5 ranks highest in both models due to its strong correlation with the target variable. Several engineered features demonstrate strong importance, specifically the temperature gradient, the binary cloud presence indicator, and the time embeddings. Time embedding features show strong importance for the ANN model but not for XGBoost, indicating that each model captures different aspects of the underlying atmospheric patterns.

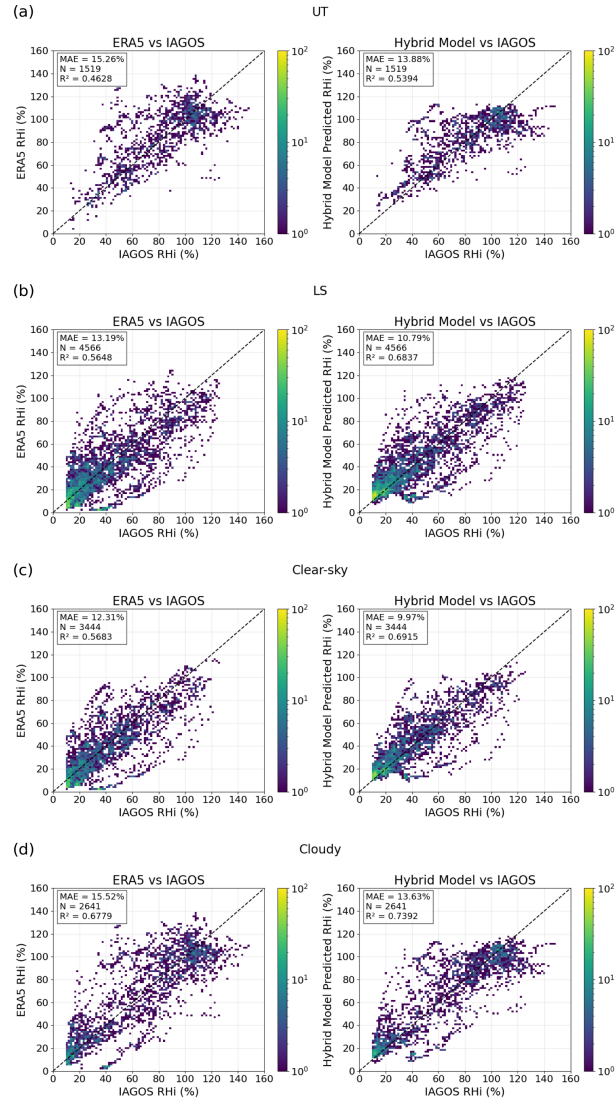
Table 4 presents the most important features used by the ANN and XGBoost models. As expected we find RHi\_ERA5 as the most important feature. This table shows that the models relied on different features for predictions: the XGBoost model relies more heavily on features with stronger linear relationships to the target, such as the vertical temperature gradients, while the ANN is able to exploit more complex, non-linear patterns, like the time embeddings. This observation highlights the complementary nature of the two models, with the ANN capturing complex relationships (often associated with more humid regions) and the XGBoost focusing on more linear patterns.

The performance of the hybrid model relative to the baseline ERA5 are assessed for both UT and LS under clear- and cloudy-sky conditions. Results are shown in Fig. 5. The best improvement in MAE is observable in the LS, from 13.19% to 10.71%. This is explained because the model performs a better correction on lower RH<sub>i</sub> values and the LS counts significantly less ISSRs (see Fig. 6). In Table 5, we report the predictions performance for each of the cases. In comparison, while QM successfully removes bias from the distribution, it does not improve point-wise predictive accuracy. Wolf et al. (2025) explicitly report that after applying QM, the “RMSE, MAE, and MSE increase unnoticeably”. In contrast, our model significantly reduces the MAE from 13.70% to 11.37%.

Finally, Fig. 6 presents the number of correctly predicted ISSRs for the various cases. We observe that the hybrid model correctly predicts significantly more ISSRs than ERA5 in the LS, 92 out of 204, compared to 54 out of 204 for ERA5. Overall, the hybrid model improves ISSR detection by approximately 10% across all conditions. In the UT, the improvement is less significant, with only 4 additional ISSRs correctly classified.

## 5 Conclusions

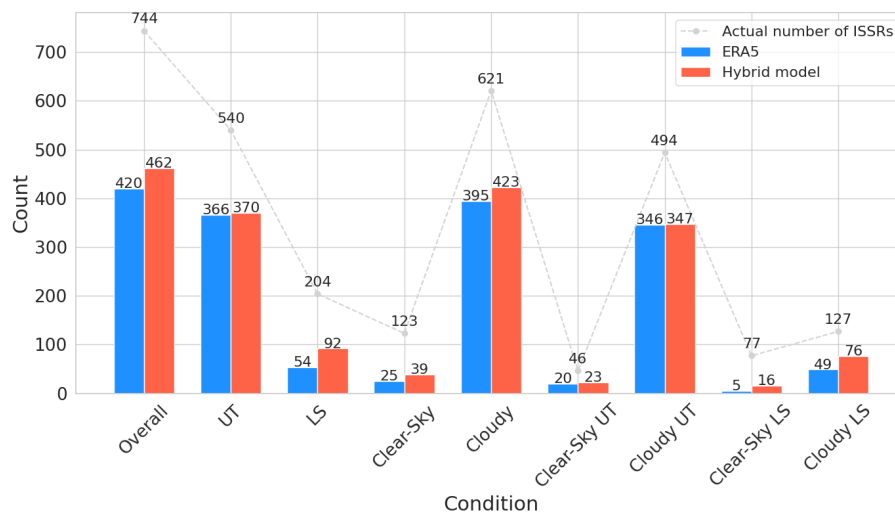
In this study, we used a hybrid ensemble model to improve RH<sub>i</sub> from ERA5 reanalysis in the UTLS using *in-situ* measurement from IAGOS. The dry bias of ERA5 is considerably reduced and the number of correctly predicted ISSRs is increased, especially in the LS where it nearly doubles.



**Figure 5.** Comparison of ISSR prediction performance between ERA5 and the hybrid model across four meteorological conditions. These are the predictions made on the test dataset between 400 and 200 hPa over the North Atlantic region (cf. Fig. 1) for the year 2022.

Condition	Model	TP	FN	FP	TN	ETS
UT	ERA5	24.09	11.45	13.50	50.95	0.30
	HYB	24.36	11.19	10.66	53.79	0.35 (+0.05)
LS	ERA5	1.18	3.29	1.34	94.20	0.19
	HYB	2.01	2.45	0.72	94.81	0.37 (+0.18)
Cloudy	ERA5	14.96	8.56	9.66	66.83	0.33
	HYB	16.02	7.50	7.08	69.41	0.42 (+0.09)
Clear-sky	ERA5	0.73	2.85	0.32	96.11	0.18
	HYB	1.13	2.44	0.23	96.20	0.29 (+0.11)

**Table 5.** Comparison of ERA5 and hybrid model (HYB) prediction performance against IAGOS observations across different meteorological conditions on the test dataset. We show the percentage of True Positive (TP), False Negative (FN), False Positive (FP) and True Negative(TN) per condition.



**Figure 6.** Number of correctly predicted ISSRs by ERA5 and the hybrid model compared to IAGOS ground truth across different meteorological conditions on the 2022 test dataset. Blue bars indicate true positives from ERA5, red bars from the hybrid model, and the dashed curve represents the total number of actual ISSRs in each condition based on IAGOS observations. The largest improvement is observed in the LS. The far left bars show the overall comparison across all conditions.

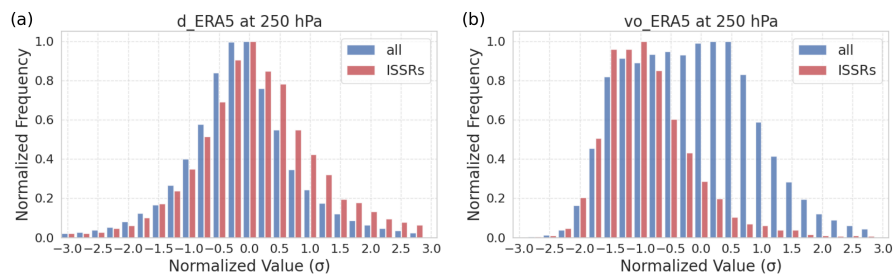
260 The hybrid model is built on two complementary models: a XGBoost regressor that excels in dry conditions ( $RH_i < 85\%$ ), and an ANN that performs better in more humid regions ( $RH_i > 85\%$ ). The hybrid model demonstrated significant improvements compared to the initial ERA5 estimates, both in terms of MAE and ETS. It achieved an MAE of 11.37%, compared

to 13.7% for ERA5, and improved the ETS from 0.36 to 0.44. The best improvement was observed in the LS, where the hybrid model increased ETS by +0.18 and reduced MAE from 13.19% to 10.71%. The hybrid model consistently outperformed  
265 ERA5 across all defined meteorological conditions, detecting approximately 10% more ISSRs overall.

However, the model presents limitations, especially in predicting very high RH<sub>i</sub> values (above 120%), likely due to the very limited availability of such samples in the dataset. Although ISSR detection has improved, reanalysis products still exhibit limitations to support potential contrail avoidance strategies, especially in the absence of clearly established performance requirements. Future work could explore advanced downsampling techniques to improve collocation between IAGOS and  
270 ERA5 datasets despite their different spatial and temporal resolutions.

*Data availability.* The IAGOS data are available on the IAGOS Data Portal (Boulanger et al., 2019). The ERA5 reanalysis data are available from the Copernicus Climate Change Service (C3S) Climate Data Store (CDS) (Hersbach et al., 2023).

## Appendix A



**Figure A1.** Histograms of vorticity (a) and divergence (b) for the North-Atlantic region in 2022 on the 250 hPa level. This pressure level corresponds to the UT, where ISSRs are frequently observed. The blue bars refer to all points region while the red ones refer to ISSRs only. The data are normalized using the standard deviation of all grid boxes. We compared these distributions with studies of dynamical characteristics of ISSRs from Gierens and Brinkop (2012). By ensuring that our processed data exhibit comparable statistical properties, we confirmed the reliability of our data processing pipeline.

Model	Train $R^2$	Test $R^2$	Train MAE	Test MAE	Train MSE	Test MSE	Train ETS	Test ETS
Decision Tree	0.87	0.66	8.76	13.64	12.79	19.45	0.47	<b>0.40</b>
Random Forest	<b>0.99</b>	0.74	<b>1.37</b>	12.12	<b>2.54</b>	17.00	<b>0.92</b>	0.29
Gradient Boosting	0.87	0.75	9.28	11.84	12.90	16.56	0.46	0.31
Hist GB	0.89	0.74	7.36	11.76	11.73	17.00	0.58	0.34
AdaBoost	0.69	0.61	17.0	17.79	19.88	20.84	0.06	0.02
XGBoost	0.87	<b>0.76</b>	9.36	11.56	13.05	<b>16.25</b>	0.48	0.35
LightGBM	0.87	0.76	9.19	<b>11.55</b>	12.82	16.30	0.48	0.32
Linear Regression	0.77	0.73	12.30	12.74	16.98	17.41	0.29	0.35
Ridge	0.77	0.73	12.25	12.62	16.94	17.34	0.29	0.37
Lasso	0.75	0.72	13.53	13.71	17.95	17.69	0.08	0.11
ElasticNet	0.75	0.72	13.13	13.38	17.77	17.56	0.12	0.14
Bayesian Ridge	0.77	0.73	12.25	12.62	16.95	17.37	0.29	0.37
SGD Regressor	0.77	0.72	12.23	12.62	17.07	17.41	0.32	0.39
KNN	0.99	0.60	1.76	15.18	3.54	21.09	0.88	0.32

**Table A1.** Model performance comparison across regression and classification metrics for a set of ML model available in Scikit-learn (Pedregosa et al., 2011). The train and test dataset are described in Section 2.3. Best values per metric are in **bold**.

*Author contributions.* OB and JJQ conceived and designed the experiments; MA conducted the experiments and analyzed the data; MA and JJQ wrote the manuscript; OB provided critical review.

*Competing interests.* None of the author has any competing interests.

*Acknowledgements.* The authors thank Ziming Wang for useful discussions on her work which has initiated our study.

## References

- 280 Alduchov, O. A. and Eskridge, R. E.: Improved Magnus Form Approximation of Saturation Vapor Pressure, *Journal of Applied Meteorology*,  
35, 601–609, [https://doi.org/10.1175/1520-0450\(1996\)035<0601:imfaos>2.0.co;2](https://doi.org/10.1175/1520-0450(1996)035<0601:imfaos>2.0.co;2), 1996.
- Appleman, H.: The formation of exhaust condensation trails by jet aircraft, *Bulletin of the American Meteorological Society*, 34, 14–20,  
<https://doi.org/10.1175/1520-0477-34.1.14>, 1953.
- Borella, A., Vignon, É., Boucher, O., and Rohs, S.: An Empirical Parameterization of the Subgrid-Scale Distribution of Water Vapor in the UTLS for Atmospheric General Circulation Models, *Journal of Geophysical Research: Atmospheres*, 129,  
285 <https://doi.org/10.1029/2024JD040981>, 2024.
- Boulangier, D., Thouret, V., and Petzold, A.: IAGOS Data Portal, <https://doi.org/10.25326/20>, 2019.
- Chen, T. and Guestrin, C.: XGBoost: A Scalable Tree Boosting System, in: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16*, p. 785–794, ACM, <https://doi.org/10.1145/2939672.2939785>, 2016.
- Copernicus Climate Change Service: ERA5 hourly data on pressure levels from 1940 to present, <https://doi.org/10.24381/cds.bd0915c6>,  
290 2025.
- Diao, M., Zondlo, M. A., Heymsfield, A. J., Avallone, L. M., Paige, M. E., Beaton, S. P., Campos, T., and Rogers, D. C.: Cloud-scale ice-supersaturated regions spatially correlate with high water vapor heterogeneities, *Atmospheric Chemistry and Physics*, 14, 2639–2656,  
<https://doi.org/10.5194/acp-14-2639-2014>, 2014.
- Diao, M., Jensen, J. B., Pan, L. L., Homeyer, C. R., Honomichl, S., Bresch, J. F., and Bansemmer, A.: Distributions of ice supersaturation  
295 and ice crystals from airborne observations in relation to upper tropospheric dynamical boundaries, *Journal of Geophysical Research: Atmospheres*, 120, 5101–5121, <https://doi.org/https://doi.org/10.1002/2015JD023139>, 2015a.
- Diao, M., Jensen, J. B., Pan, L. L., Homeyer, C. R., Honomichl, S., Bresch, J. F., and Bansemmer, A.: Distributions of ice supersaturation  
and ice crystals from airborne observations in relation to upper tropospheric dynamical boundaries, *Journal of Geophysical Research: Atmospheres*, 120, 5101–5121, <https://doi.org/10.1002/2015JD023139>, 2015b.
- 300 Gattelman, A., Fetzer, E. J., Eldering, A., and Irion, F. W.: The global distribution of supersaturation in the upper troposphere from the  
Atmospheric Infrared Sounder, *Journal of Climate*, 19, 6089 – 6103, <https://doi.org/10.1175/JCLI3955.1>, 2006.
- Gierens, K. and Brinkop, S.: Dynamical characteristics of ice supersaturated regions, *Atmospheric Chemistry and Physics*, 12, 11 933–11 942,  
<https://doi.org/10.5194/acp-12-11933-2012>, 2012.
- Gierens, K. and Spichtinger, P.: On the size distribution of ice-supersaturated regions in the upper troposphere and lowermost stratosphere,  
305 *Annales Geophysicae*, 18, 499–504, <https://doi.org/10.1007/s00585-000-0499-7>, 2000.
- Gierens, K., Schumann, U., Helten, M., Smit, H., and Marengo, A.: A distribution law for relative humidity in the upper troposphere and lower  
stratosphere derived from three years of MOZAIC measurements, *Annales Geophysicae*, 17, 1218–1226, <https://doi.org/10.1007/s00585-999-1218-7>, 1999.
- Gierens, K., Kohlhepp, R., Spichtinger, P., and Schroedter-Homscheidt, M.: Ice supersaturation as seen from TOVS, *Atmospheric Chemistry  
310 and Physics*, 4, 539–547, <https://doi.org/10.5194/acp-4-539-2004>, 2004.
- Groß, S., Wirth, M., Schäfler, A., Fix, A., Kaufmann, S., and Voigt, C.: Potential of airborne lidar measurements for cirrus cloud studies,  
*Atmospheric Measurement Techniques*, 7, 2745–2755, <https://doi.org/10.5194/amt-7-2745-2014>, 2014.
- Hegglin, M. I., Tegtmeier, S., Anderson, J., Froidevaux, L., Fuller, R., Funke, B., Jones, A., Lingenfelser, G., Lumpe, J., Pendlebury,  
D., Remsberg, E., Rozanov, A., Toohey, M., Urban, J., von Clarmann, T., Walker, K. A., Wang, R., and Weigel, K.: SPARC Data

- 315 Initiative: Comparison of water vapor climatologies from international satellite limb sounders, *J. Geophys. Res.*, 118, 11,824–11,846, <https://doi.org/10.1002/jgrd.50752>, 2013.
- 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.,  
320 Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.: The ERA5 global reanalysis, *Quarterly Journal of the Royal Meteorological Society*, 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 hourly data on pressure levels from 1940 to present, Copernicus Climate  
325 Change Service (C3S) Climate Data Store (CDS), <https://doi.org/10.24381/cds.bd0915c6>, accessed on 01-June-2025, 2023.
- Heymsfield, A. J., Miloshevich, L. M., Twohy, C., Sachse, G., and Oltmans, S.: Upper-tropospheric relative humidity observations and implications for cirrus ice nucleation, *Geophysical Research Letters*, 25, 1343–1346, <https://doi.org/https://doi.org/10.1029/98GL01089>, 1998.
- Hofer, S. M. and Gierens, K. M.: Kinematic properties of regions that can involve persistent contrails over the North Atlantic and Europe  
330 during April and May 2024, *Atmospheric Chemistry and Physics*, 25, 13, <https://doi.org/10.5194/acp-25-6843-2025>, 2025.
- Kingma, D. and Ba, J.: Adam: A Method for Stochastic Optimization, *International Conference on Learning Representations*, <http://arxiv.org/abs/1412.6980>, 2014.
- 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, *Atmospheric Chemistry and Physics*, 22, 15 559–15 577, <https://doi.org/10.5194/acp-22-15559-2022>, 2022.  
335
- Kärcher, B.: Formation and radiative forcing of contrail cirrus, *Nature Communications*, 9, 1824, <https://doi.org/10.1038/s41467-018-04068-0>, 2018.
- Kärcher, B. and Corcos, M.: On the lifetimes of persistent contrails and contrail cirrus, manuscript submitted to *JGR: Atmospheres*, 2025.
- Lamquin, N., Stubenrauch, C. J., Gierens, K., Burkhardt, U., and Smit, H.: A global climatology of upper-tropospheric ice supersaturation  
340 occurrence inferred from the Atmospheric Infrared Sounder calibrated by MOZAIC, *Atmospheric Chemistry and Physics*, 12, 381–405, <https://doi.org/10.5194/acp-12-381-2012>, 2012.
- Lee, D., Fahey, D., Skowron, A., Allen, M., Burkhardt, U., Chen, Q., Doherty, S., Freeman, S., Forster, P., Fuglestedt, J., Gettelman, A., De León, R., Lim, L., Lund, M., Millar, R., Owen, B., Penner, J., Pitari, G., Prather, M., Sausen, R., and Wilcox, L.: The contribution of global aviation to anthropogenic climate forcing for 2000 to 2018, *Atmospheric Environment*, 244, 117 834,  
345 <https://doi.org/10.1016/j.atmosenv.2020.117834>, 2021.
- 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, *Atmospheric Chemistry and Physics*, 23, 2251–2271, <https://doi.org/10.5194/acp-23-2251-2023>, 2023.
- Mihajlović, I. and Nikolić, D.: Artificial neural network method applied on the nonlinear multivariate problems, *serbian journal of management, sjm*, 143, 2009.  
350

- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E.: Scikit-learn: Machine Learning in Python, *Journal of Machine Learning Research*, 12, 2825–2830, 2011.
- 355 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 Team, I.: Global-scale atmosphere monitoring by in-service aircraft – current achievements and future prospects of the European Research Infrastructure IAGOS, *Tellus B: Chemical and Physical Meteorology*, <https://doi.org/10.3402/tellusb.v67.28452>, 2015.
- Platt, J. C., Shapiro, M. L., Engberg, Z., McCloskey, K., Geraedts, S., Sankar, T., Stettler, M. E. J., Teoh, R., Schumann, U., Rohs, S., Brand, E., and Van Arsdale, C.: The effect of uncertainty in humidity and model parameters on the prediction of contrail energy forcing, *Environmental Research Communications*, 6, 095 015, <https://doi.org/10.1088/2515-7620/ad6ee5>, 2024.
- 360 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, *Atmospheric Chemistry and Physics*, 20, 787–804, <https://doi.org/10.5194/acp-20-787-2020>, 2020.
- Rolf, C., Rohs, S., Smit, H., Krämer, M., Bozóki, Z., Hofmann, S., Franke, H., Maser, R., Hoor, P., and Petzold, A.: Evaluation of compact hygrometers for continuous airborne measurements, *Meteorologische Zeitschrift*, 33, <https://doi.org/10.1127/metz/2023/1187>, 2023.
- 365 Rosenow, J., Fricke, H., Luchkova, O., et al.: Minimizing contrail formation by rerouting around dynamic ice-supersaturated regions, *Aeronautics and Aerospace Open Access Journal*, 2, 105–111, <https://doi.org/10.15406/aaaj.2018.02.00039>, 2018.
- Schmidt, E.: Die Entstehung von Eisnebel aus den Auspuffgasen von Flugmotoren, eintrag von Ulrich Schumann, zur Sicherstellung des Zugangs zu diesem wissenschaftshistorisch wichtigen Dokument; mit Zustimmung des Rechteinhabers (Nachfolger des Oldenbourg Verlags), 1941.
- 370 Schumann, U.: On conditions for contrail formation from aircraft exhausts, *DLR Forschungsbericht DLR-FB 1996-10*, Deutsches Zentrum für Luft- und Raumfahrt (DLR), <https://elib.dlr.de/8302/>, 1996.
- Sonntag, D.: Advancements in the field of hygrometry, *Meteorologische Zeitschrift*, 3, 51–66, <https://doi.org/10.1127/metz/3/1994/51>, 1994.
- Spichtinger, P., Gierens, K., Leiterer, U., and Dier, H.: Ice supersaturation in the tropopause region over Lindenberg, Germany, *Meteorologische Zeitschrift*, 12, 143–156, <https://doi.org/10.1127/0941-2948/2003/0012-0143>, 2003a.
- 375 Spichtinger, P., Gierens, K., and Read, W.: The global distribution of ice-supersaturated regions as seen by the Microwave Limb Sounder, *Quarterly Journal of the Royal Meteorological Society*, 129, 3391–3410, <https://doi.org/10.1256/qj.02.141>, 2003b.
- Stenke, A., Grewe, V., and Ponater, M.: Lagrangian transport of water vapor and cloud water in the ECHAM4 GCM and its impact on the cold bias, *Climate Dynamics*, 31, 491–506, <https://doi.org/10.1007/S00382-007-0347-5>, 2008.
- 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, *Atmospheric Chemistry and Physics*, 22, 10919–10935, <https://doi.org/10.5194/acp-22-10919-2022>, 2022.
- 380 Teoh, R., Engberg, Z., Schumann, U., Voigt, C., Shapiro, M., Rohs, S., and Stettler, M. E. J.: Global aviation contrail climate effects from 2019 to 2021, *Atmospheric Chemistry and Physics*, 24, 6071–6093, <https://doi.org/10.5194/acp-24-6071-2024>, 2024.
- Wang, X., Wolf, K., Boucher, O., and Bellouin, N.: Radiative Effect of Two Contrail Cirrus Outbreaks Over Western Europe Estimated Using Geostationary Satellite Observations and Radiative Transfer Calculations, *Geophysical Research Letters*, 51, <https://doi.org/10.1029/2024g1108452>, 2024.
- 385

- Wang, Z., Bugliaro, L., Gierens, K., Hegglin, M. I., Rohs, S., Petzold, A., Kaufmann, S., and Voigt, C.: Machine learning for improvement of upper-tropospheric relative humidity in ERA5 weather model data, *Atmospheric Chemistry and Physics*, 25, 2845–2861, <https://doi.org/10.5194/acp-25-2845-2025>, 2025.
- 390 Wolf, K., Bellouin, N., and Boucher, O.: Long-term upper-troposphere climatology of potential contrail occurrence over the Paris area derived from radiosonde observations, *Atmospheric Chemistry and Physics*, 23, 287–309, <https://doi.org/10.5194/acp-23-287-2023>, 2023.
- Wolf, K., Bellouin, N., Boucher, O., Rohs, S., and Li, Y.: Correction of ERA5 temperature and relative humidity biases by bivariate quantile mapping for contrail formation analysis, *Atmospheric Chemistry and Physics*, 25, 157–181, <https://doi.org/10.5194/acp-25-157-2025>, 2025.
- 395 Zahn, A., Christner, E., van Velthoven, P. F. J., Rauthe-Schöch, A., and Brenninkmeijer, C. A. M.: Processes controlling water vapor in the upper troposphere/lowermost stratosphere: An analysis of 8 years of monthly measurements by the IAGOS-CARIBIC observatory, *J. Geophys. Res.*, 119, 11,505–11,525, <https://doi.org/10.1002/2014JD021687>, 2014.