https://doi.org/10.5194/egusphere-2025-2985
Preprint. Discussion started: 19 August 2025

A novel machine learning retrieval for the detection of ice crystal icing conditions based on geostationary satellite imagery

Matteo Aricò[1], Dennis Piontek[1], Luca Bugliaro[1], Johanna Mayer[1, a], Richard Müller[2], Frank Kalinka[2], and Max Butter[3]

[1]Institut für Physik der Atmosphäre, Deutsches Zentrum für Luft- und Raumfahrt, Oberpfaffenhofen, Germany
[2]Deutscher Wetterdienst, Offenbach am Main, Germany
[3]Deutsche Lufthansa AG, Frankfurt am Main, Germany
[a]now at: European Space Agency, ESRIN, Frascati, Italy

Correspondence: Matteo Aricò (matteo.arico@dlr.de)

Abstract. High ice water content (HIWC) conditions are a concern for aviation as the ingestion of ice particles in the jet engines can induce ice crystal icing (ICI), which results in performance loss and damage. To constantly monitor these conditions, retrievals for the detection of ICI were recently developed based on geostationary satellite imagery, but their calibration is limited to targeted flight campaigns or scattered samplings from ICI events databases. In this work, we close this gap, using exclusively remote sensing data to develop and assess a new retrieval for potential ICI conditions.

Cloud IWC measurements are provided from the synergy of radar and lidar (DARDAR) on board the polar-orbiting satellites CloudSat and CALIPSO. HIWC conditions (IWC $\geq$ 0.5 g m^{-3}) at typical cruise altitudes are used as the proxy for areas with potential ICI formation. The HIWC conditions predictors are taken from a combination of observations and retrievals of the geostationary satellite Meteosat Second Generation (MSG). A random forest is trained and tested based on the collocated dataset of active and passive measurements during the summer months of 2013 and 2015, covering the European domain. The input predictors are the brightness temperature difference between the MSG channels at 6.2 and 10.8 μm wavelengths, the visible channel at 0.6 μm wavelength, the cloud optical thickness at 0.6 μm wavelength, and four convection metrics related to the distance to the closest convective cell, area extent of the convective cells, and convection density in the pixel surroundings. Over Europe, 83 % of HIWC conditions measured in the DARDAR dataset are correctly detected. The associated false alarm rate is 51 %. The retrieval is further tested with the ICI events database reported by Lufthansa. Four out of seven events are correctly detected. In conclusion, the retrieval achieves performances comparable to previously developed retrievals. An operational application would enable aircraft rerouting around areas with high ICI probability.

1 Introduction

Ice Crystal Icing (ICI) is a phenomenon that aircraft may encounter when flying through cloudy regions with high ice crystal concentrations. These regions are mostly found close to deep convection, in particular within tropical mesoscale convective systems (MCSs). In such systems, pilots can easily avoid strong updrafts, as onboard radars can detect embedded precipitation based on its high reflectivity signal, or available satellite-based nowcasting of severe convection (NCS-A, Müller et al., 2022)

https://doi.org/10.5194/egusphere-2025-2985
Preprint. Discussion started: 19 August 2025

can issue early warnings. However, regions outside the main updraft may not be affected by nowcasting warnings and they can still contain high ice concentrations despite having little to no radar reflectivity due to the presence of non-precipitating ice particles (Gayet et al., 2012); this is where ICI events can occur because ice particles can build up inside the engine and lead to performance loss and damage (Grzych, 2010, 2015; Bravin et al., 2015; Haggerty et al., 2019), or they can clog the pitot tube which in turns result into a wrongful transmission of information to the autoflight system; this latter occurrence has caused two fatal accidents in recent years (S. Ayra et al., 2020). In contrast with convection patterns, no clear diurnal trends are found globally; however, a seasonal correlation is observed between local convective active seasons and ICI events (Bravin et al., 2015).

High ice water content (HIWC) conditions are often used as a proxy for potential ICI occurrence. For these conditions, a threshold ranging between 0.5 and 1.0 g m^{-3} is chosen in earlier studies, although a standard value is still under debate because exposure times and engine types might also affect ICI occurrence (de Laat et al., 2017; Yost et al., 2018; Haggerty et al., 2019, 2020; Bedka et al., 2020).

Aircraft manufacturers and airlines have collected ICI events in databases to analyze the importance of the phenomenon. Bravin et al. (2015) present a Boeing database that included 162 events over 12 years. de Laat et al. (2017) construct a database from Airbus containing 59 events, without specifying their time frame. Here, a collection of 100 events from Lufthansa flights during 2016 is considered (Sect. 2.4) to analyze a subset of ICI events as case studies (Sect. 4.2). The worldwide number of ICI events and their impact on engine performance highlights the relevance of the issue to air traffic safety.

The importance of this problem led to the execution of flight campaigns to measure in situ cloud microphysical properties during such events. A combination of specifically designed probes, sensors, and radar instruments was deployed to measure high ice concentrations, particle size distributions, and cloud vertical profiles, respectively. These campaigns are:

- the HAIC-HIWC flight campaign, Darwin, Australia 2014, where HAIC stands for "high altitude ice crystal";
- the HAIC-HIWC II flight campaign, Cayenne, French-Guiana 2015;
- the HAIC-RADAR flight campaign, Fort Lauderdale, Florida 2015;
- the HAIC-RADAR II flight campaign, Fort Lauderdale, Florida 2018.

The problem's relevance and the availability of new in situ measurements triggered activities in the research area of HIWC conditions detection products from satellites. Indeed, the following retrievals were developed:

- Grzych et al. (2015) develop a 3D HIWC mask exploiting infrared (IR) channels from geostationary satellite imagery combined with numerical weather prediction (NWP) wind fields at different heights and the tropopause level (ECMWF-ERA5, Hersbach et al., 2020). The algorithm is tested with the HAIC-HIWC flight campaign case studies, which are used as ground truth. While a clear correlation between the mask and the in situ measured HIWC conditions is found, the algorithm tends to overestimate the areas affected by this phenomenon, but no performance metrics are reported;
- de Laat et al. (2017) approach the problem by manually setting thresholds on retrieved cloud microphysical variables from geostationary satellite imagery. These thresholds are calibrated using case studies in the Airbus dataset and verified

https://doi.org/10.5194/egusphere-2025-2985
Preprint. Discussion started: 19 August 2025

with the synergistic space-borne lidar-radar dataset (DARDAR), derived from active remote sensing measurements on polar-orbiting satellites that include, among others, IWC. The algorithm achieves a probability of detection (POD) of 0.59 but with an associated false alarm rate (FAR) of 0.52;

- Yost et al. (2018) use a combination of geostationary satellite imagery and retrieved cloud optical properties. The considered input variables are associated with a corresponding value of IWC according to a statistical fit performed by collocating the satellite data with flight campaign measurements. This information is translated into a HIWC probability using fuzzy logic. The algorithm is verified with the HAIC-HIWC, HAIC-HIWC II, and HIWC-RADAR flight campaigns, achieving a POD of 0.75 and a FAR of 0.35 during daytime. Reported nighttime performances are inferior (POD: 0.62, FAR: 0.35) because of the lack of cloud optical properties;

- Haggerty et al. (2020) integrate a multitude of data sources, like satellites, on-ground radar, and NWP data. Particle swarm optimization is used to select a subset of variables of interest, which are then combined via fuzzy logic to produce the HIWC probability. The retrieval is verified with the HAIC-HIWC, HAIC-HIWC II, and HIWC-RADAR II flight campaigns, achieving a POD of 0.86 and a FAR of 0.51.

When training potential ICI detection retrievals, a significant amount of in situ HIWC measurements should be considered for statistical significance. Dedicated research flight campaigns are often geographically limited, and they specifically target HIWC conditions. This may introduce a bias when extrapolating from a local to a global context (Haggerty et al., 2020).
While in situ HIWC measurements are the best data to assess potential ICI conditions in convective clouds, alternative approaches exploiting remote sensing measurements can be implemented if one wants to increase the training samples. For operational monitoring, geostationary satellites are used due to their wide field of view and high temporal resolution. Polar-orbiting satellites' active observations cannot be directly applied in operational scenarios because of their small field of view and low repetition time. This work demonstrates the feasibility of a detection method for potential ICI from geostationary satellite observations based on machine learning techniques and trained with the DARDAR dataset as ground truth.
The paper contains a description of the combination of data used to train the ICI detection retrieval in Sect. 2. Next, we describe how the machine learning techniques are applied for the ICI detection task in Sect. 3. In Sect. 4, we present the results validated with active remote sensing data and Lufthansa's ICI database. Finally, in Sect. 5 we summarize the results on the retrieval's performance and discuss its main limitations.

2 Datasets

The ICI retrieval developed in this study relies on physical quantities measured and retrieved by passive instruments on board geostationary satellites, called "predictors" hereafter. The geostationary satellite and the corresponding retrievals employed are presented in Sect. 2.1. The DARDAR dataset is presented in Sect. 2.2 because this contained our ground truth data for IWC measurements of cloud profiles. Lastly, it is important to establish the spatial and temporal distribution of selected in-service ICI events, analyzed in Sect. 2.4.

https://doi.org/10.5194/egusphere-2025-2985
Preprint. Discussion started: 19 August 2025

2.1 MSG and MSG-based retrievals

The predictors' source for this work is the geostationary satellite Meteosat Second Generation (MSG) because it guarantees a continuous spatial coverage of Europe. MSG is equipped with the Spinning Enhanced Visible and Infrared Imager (SEVIRI) that measures reflectance and radiance in the visible and infrared range, thanks to its 11 narrow-band channels and one high-resolution visible (HRV) broadband channel. SEVIRI provides a 3712 $\times$ 3712 pixels image of the Earth disk with a 3 km $\times$ 3 km resolution at the nadir. The temporal resolution is 15 minutes, with a rapid scan service (RSS) available for a subset of the northern hemisphere, where images are produced every 5 minutes (Schmetz et al., 2002). Besides SEVIRI channels, we also use ice cloud properties retrievals based on SEVIRI. These ice cloud properties are used as predictors for our ICI retrieval, so the corresponding geostationary-based retrievals are briefly discussed below.

2.1.1 CiPS

CiPS (Cirrus Properties from SEVIRI), developed and characterized by Strandgren et al. (2017a, 2017b), detects thin cirrus clouds from MSG and determines ice optical thickness, ice water path, and cloud top height. The detection is based on Artificial Neural Networks trained with CALIPSO lidar data as ground truth. The training and validation datasets cover the entire SEVIRI disc and the period between 2007 to 2013, containing close to 50 million data points. The lidar signal experiences strong attenuation when interacting with clouds; therefore, it is considered saturated and thus unreliable whenever there is no backscattering from the surface. This limited CiPS to thin cirrus cloud detection with an optical thickness of approximately below 3. When validated against CALIPSO, CiPS detects correctly 95 % of all cirrus clouds with optical thickness of 1.0, while for thinner cirrus clouds with optical thickness of 0.1, the proportion of detected cirrus over all cirrus is 71 %. The best optical thickness estimation is obtained in the range between 0.35 and 1.7 with a deviation of less than 50 % from CALIPSO's measurements. The detection exploits SEVIRI thermal channels, regional maximum and averaged brightness temperatures in the infrared and water vapor channels, and surface skin temperatures from NWP global reanalysis (Hersbach et al., 2020).

2.1.2 APICS

APICS (Algorithm for the Physical Interpretation of Clouds with SEVIRI Bugliaro et al., 2011) discriminates cloud phase and microphysical properties from MSG. In particular, cloud optical thickness and effective radius (ranging from 5 to 25 μm for water clouds and from 6 to 84 μm for ice clouds) are retrieved using a look-up table approach based on radiative transfer calculations, which exploits the visible channel at 0.6 μm wavelength, and the near-infrared channel at 1.6 μm.

2.1.3 Cb-TRAM

Cb-TRAM (Zinner et al., 2008; 2013) enables the detection and tracking of convective cells from geostationary satellite imagery. It relies on the HRV, infrared 10.8 μm, and water vapor window 6.2 μm channels. Cloud motion and development can be detected through the disparities between two consecutive satellite images. The algorithm can also discriminate different

https://doi.org/10.5194/egusphere-2025-2985
Preprint. Discussion started: 19 August 2025

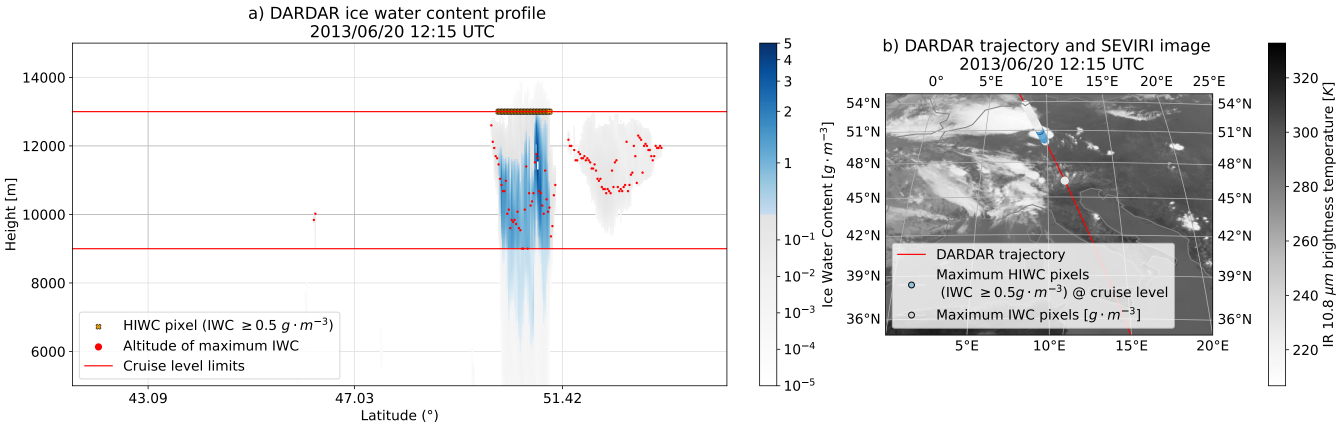

Figure 1. Panel a) mapping from the DARDAR vertically-resolved IWC information to the HIWC flag associated with the geostationary grid pixel. Panel b) SEVIRI brightness temperature at 10.8 μm wavelength with the DARDAR trajectory associated with panel a).

convection development stages: "Stage 1" denotes convection initiation, "Stage 2" rapid vertical development through cloud tops cooling, and "Stage 3" indicates mature convective cells.

2.2 DARDAR

2.2.1 DARDAR description

The DARDAR-CLOUD products developed by Delanoë and Hogan (2008, 2010) exploit the synergy of space-borne data from radar and lidar of the A-train satellite constellation to retrieve ice cloud properties. The A-Train constellation is a group of satellites that use the sun-synchronous orbit at 705 km altitude. CloudSat was equipped with a radar operating in the 94 GHz band, whose aim was to characterize cloud vertical profiles of cloud water and ice contents (Stephens et al., 2002). The lidar on board CALIPSO operated at 532 and 1064 nm wavelengths. CALIPSO provided cloud characterization as a function of height and water and ice content (Winker et al., 2003). These satellites were launched on April 28th 2006 (Delanoë and Hogan, 2010).

The DARDAR-CLOUD products exploit the different sensitivities of the instruments in a synergistic approach. The radar is less sensitive to small particles, but it has a higher penetration capability within thick clouds; the lidar is more sensitive to optically thin clouds, but it is affected by rapid attenuation, while the infrared radiometer can only estimate bulk cloud properties (Delanoë and Hogan, 2010). For this reason, IWC, effective radius, and particle size distributions are retrieved with a variational method that efficiently combines radar and lidar measurements (Delanoë and Hogan, 2008). The DARDAR products are collocated to the CloudSat horizontal resolution of 1.4 km (Stephens et al., 2002) and CALIPSO vertical resolution of 60 m (Delanoë and Hogan, 2010).

https://doi.org/10.5194/egusphere-2025-2985
Preprint. Discussion started: 19 August 2025

2.3 DARDAR-MSG collocation and ICI proxy selection

For our ICI retrieval, we consider DARDAR measurements as ground truth. Therefore, in the first step, we need to collocate the SEVIRI and DARDAR measurements. In the following, we refer to "DARDAR profile" or simply "profile" as the vertical cross-section of clouds as retrieved from the DARDAR dataset. This corresponds to the atmospheric column encompassed in the field of view of one radar-lidar pixel. Instead, we refer to the "DARDAR trajectory" as the DARDAR footprint on the surface in terms of longitude/latitude coordinates.

The DARDAR trajectories have a finer along-track resolution than the geostationary grid. MSG and DARDAR data are combined following the approach described by Mayer et al. (2023). Satellite observations are collocated by exploiting longitude, latitude, cloud top height, and observation times. Cloud top height allows us to correct the parallax effect arising from the different observation geometry of geostationary and polar-orbiting satellites. DARDAR profiles are coarsened to the MSG grid by averaging all profiles within an MSG pixel at each DARDAR height level.

Then, in each averaged profile, we check for HIWC, i.e., IWC ≥ 0.5 g m^{-3}, in an altitude range that is relevant for air traffic. We consider only cruise levels between 9000 m and 13000 m (defined in Sect. 2.4). Figure 1 illustrates the mapping process. Panel a) showcases the DARDAR IWC profiles coarsened to the MSG grid along the satellite track. HIWC areas are represented with the blue shading. If the maximum IWC value within the cruise levels in the DARDAR IWC profile exceeds the HIWC threshold, the HIWC flag is assigned to the corresponding pixel. Panel b) depicts the brightness temperature from SEVIRI at 10.8 μm with the corresponding DARDAR trajectory with its longitude/latitude coordinates, the maximum IWC values for each pixel, and the HIWC flag, if applicable. The HIWC flag was used as the target variable to train the machine learning algorithm (Sect. 3).

We consider June, July, August, and September 2013 and June, July, and August 2015. Summer months were selected because of the seasonal convective activity peak in Europe. Years 2013 and 2015 are selected because they lie within the time window where DARDAR and a single MSG platform (MSG-3) overlap (from 2013 to 2017) to avoid differences that may arise due to different instrument calibrations (Strandgren et al., 2017a; Mayer et al., 2023; Piontek et al., 2023). The collocated dataset results in 165139 collocations, 889 of them flagged as HIWC pixels (see Table 1).

2.3.1 Convection-related metrics from Cb-TRAM

DARDAR trajectories seldom overlap with convective cells as detected by Cb-TRAM. Therefore, additional convection-related metrics are used. The time spent by an aircraft within a HIWC region seems to play a role in the onset of ICI events (Bravin et al., 2015), thus, information about the areal extent of convective cells may be useful during the learning process. To this end, convection-related variables (shown in Fig. 2) are derived from the Cb-TRAM scene:

- distance from the trajectory point to the closest convective cell;
- area size of the closest convective cell, in terms of pixels and km^2. Since organized convective systems, such as MCSs, are defined as cumulonimbus clouds able to generate contiguous precipitation areas in the order of 100 km (Markowski

https://doi.org/10.5194/egusphere-2025-2985
Preprint. Discussion started: 19 August 2025

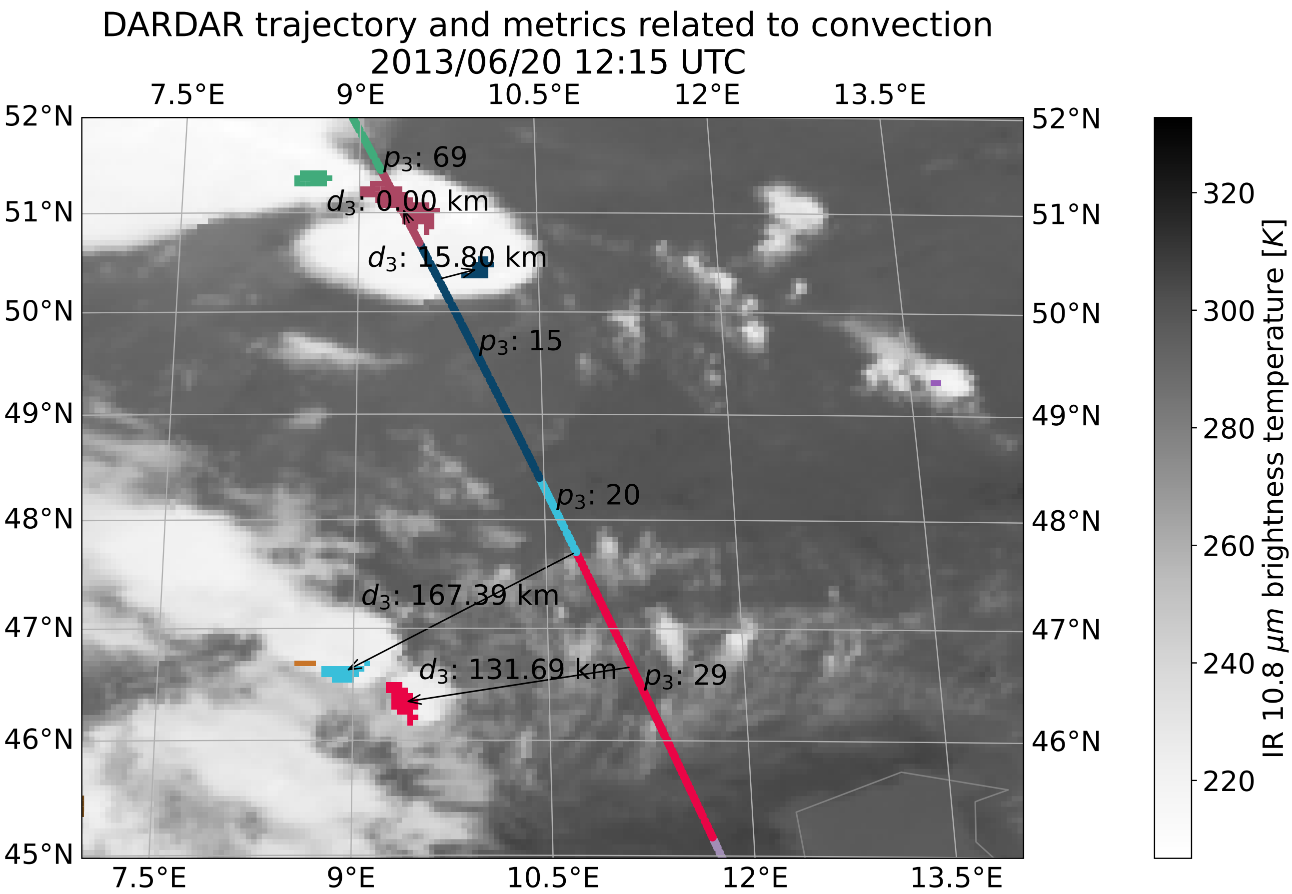

Figure 2. Demonstration of the convection-related variables integrated in the DARDAR dataset. The DARDAR trajectory is color-coded according to the closest convective cell detected. For each color portion in the trajectory, the closest pixel to the respective closest convective cell is indicated by the starting points of the arrows. The distance d_3 is displayed along the arrow. p_3 indicates the areal extent in terms of pixels of the convective cell with the corresponding color. The complete list of convection-related metrics with their definitions is presented in Table 2.

https://doi.org/10.5194/egusphere-2025-2985
Preprint. Discussion started: 19 August 2025

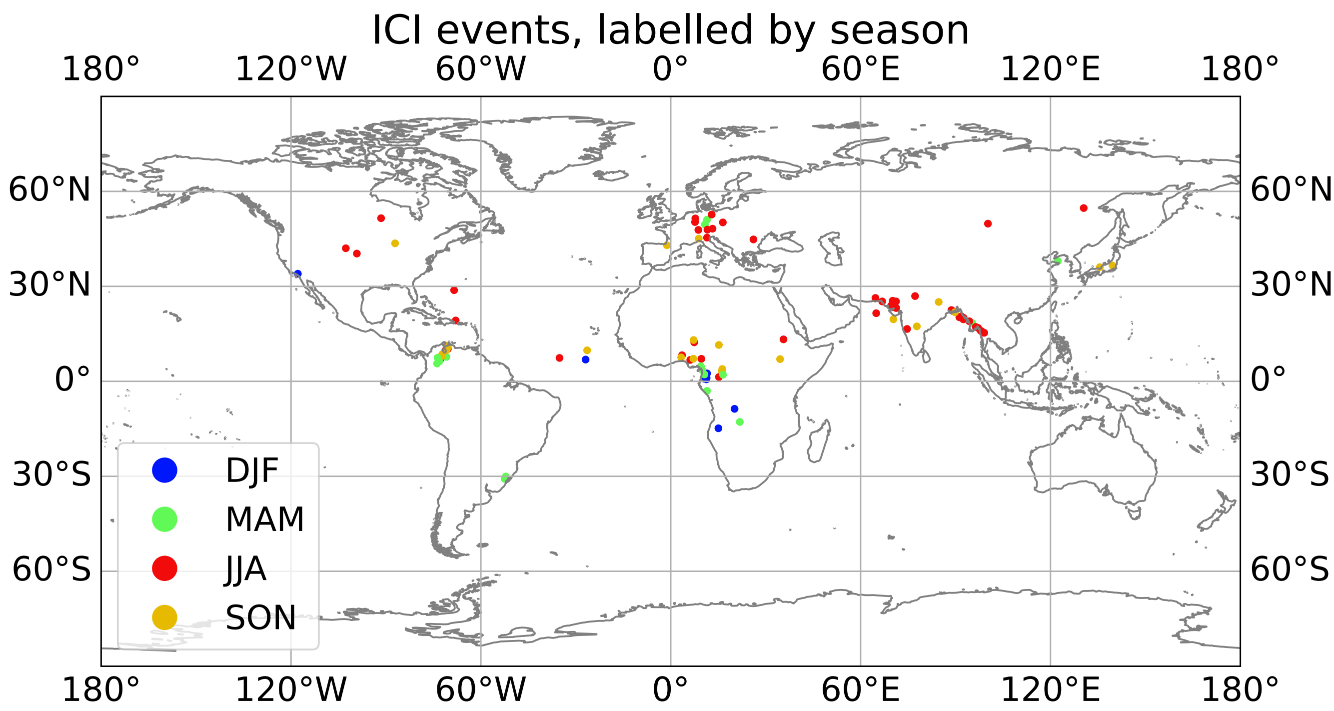

Figure 3. Locations of the Lufthansa ICI events collected in 2016, color-coded according to the season they were recorded.

and Richardson, 2010), this information is useful to assess whether detected cells belong to such organized systems, or if they are associated with single and multi-cell convection, that are generally characterized by a smaller area extent;

– number of convective cells within a 100 km radius. This metric contains the density of convective clouds in the surrounding area, which can be associated with a higher chance of intercepting anvil cirrus;

– pixels within a radius of 10 km, 50 km, and 100 km belong to detected convective cells. This gathers information on cell extent and convection density in the area close to the trajectory point.

The full list of convection-related metrics with their definitions is presented in Table 2.

2.4 Lufthansa ICI database

The Lufthansa ICI database comprises 100 pilot-reported ICI events selected manually based on in situ measured total air temperature anomalies (Kalinka et al., 2023). Figure 3 displays the database's geographical distribution. The database is also important to get an indication of the seasonal occurrence of these events, with a special focus devoted to the European continent. The ICI events distribution agrees with expectations. The majority of them occurred in the northern hemisphere summer

https://doi.org/10.5194/egusphere-2025-2985
Preprint. Discussion started: 19 August 2025

(JJA). In Europe, two events occurred in autumn (SON) and two in spring (MAM). Therefore, we focus on processing DARDAR trajectories during the summer months to maximize the chance of sampling HIWC conditions in Europe.

In Fig. 4, we present the ICI events related to the standard envelope FAA 14 CFR Part 33 Appendix D (Federal Aviation Administration, Department of Transportation, 2023), which depicts where ICI events occur in terms of altitude versus ambient temperature. Most of the events collected by Lufthansa fall within the specified boundaries, except for three cases. 88% events occur between 9000 m (29527 ft) and 13000 m (42650 ft). This altitude band, called "cruise levels" hereafter, indicates the portion of the troposphere that is considered when sampling IWC in the cloud profile from active satellite instruments. Furthermore, while testing for multiple cruise level limits (not shown), we observed that the correct detection of HIWC conditions was more likely when these conditions occur at higher altitudes, as observed also by de Laat et al. (2017). We speculate that this is due to passive sensors mainly measuring emitted and reflected radiation in proximity to cloud tops.

3 ICI retrieval

The core idea behind the ICI retrieval is to combine a broad range of predictors measured by passive instruments on board geostationary satellites (Sect. 2.1) to detect potential ICI events, determined with DARDAR measurements of HIWC conditions. The retrieval, based on a random forest approach, estimates the probability of HIWC conditions. By setting a threshold on the probability of HIWC conditions, one can convert this information to the HIWC binary flag to use it as a deterministic target output variable for the ICI retrieval training and validation. Finally, the ICI retrieval can be applied to geostationary imagery to obtain a probability mask of HIWC conditions.

Random forest classifiers are an ensemble method based on single decision trees. Decision trees divide recursively the predictor space into distinct non-overlapping regions, which are the tree's nodes. The split is aimed at minimizing the output variance within each region. The output probability to predict a certain class is the proportion of that class found in the training dataset for the end nodes, or leaves. The main disadvantage of single trees is that they are very sensitive to the training dataset. Hence, the need for random forests, which reduce the variance by averaging a set of single trees. Because random forests have many hyperparameters that can be tuned by the user, such as the number of trees, the tree depth, and the number of samples allowed in the leaves, it is often necessary to determine them through cross-validation. Cross-validation is a method to estimate the statistical learning method's test error by holding out a subset of the original dataset. K-fold cross-validation consists of dividing the dataset into k groups, or folds, of equal size. One fold is treated as the validation set, while the others are used for training. The procedure is repeated for all the k folds, each time considering a different fold for validation. The dataset where the k-fold cross-validation is performed is split into training and validation sets. This step is used to tune the hyperparameters, to avoid overfitting and unnecessarily complex models. The cross-validation dataset differs from the test set of unseen observations. This is used to evaluate the actual performance of the model with the chosen hyperparameters.

https://doi.org/10.5194/egusphere-2025-2985
Preprint. Discussion started: 19 August 2025

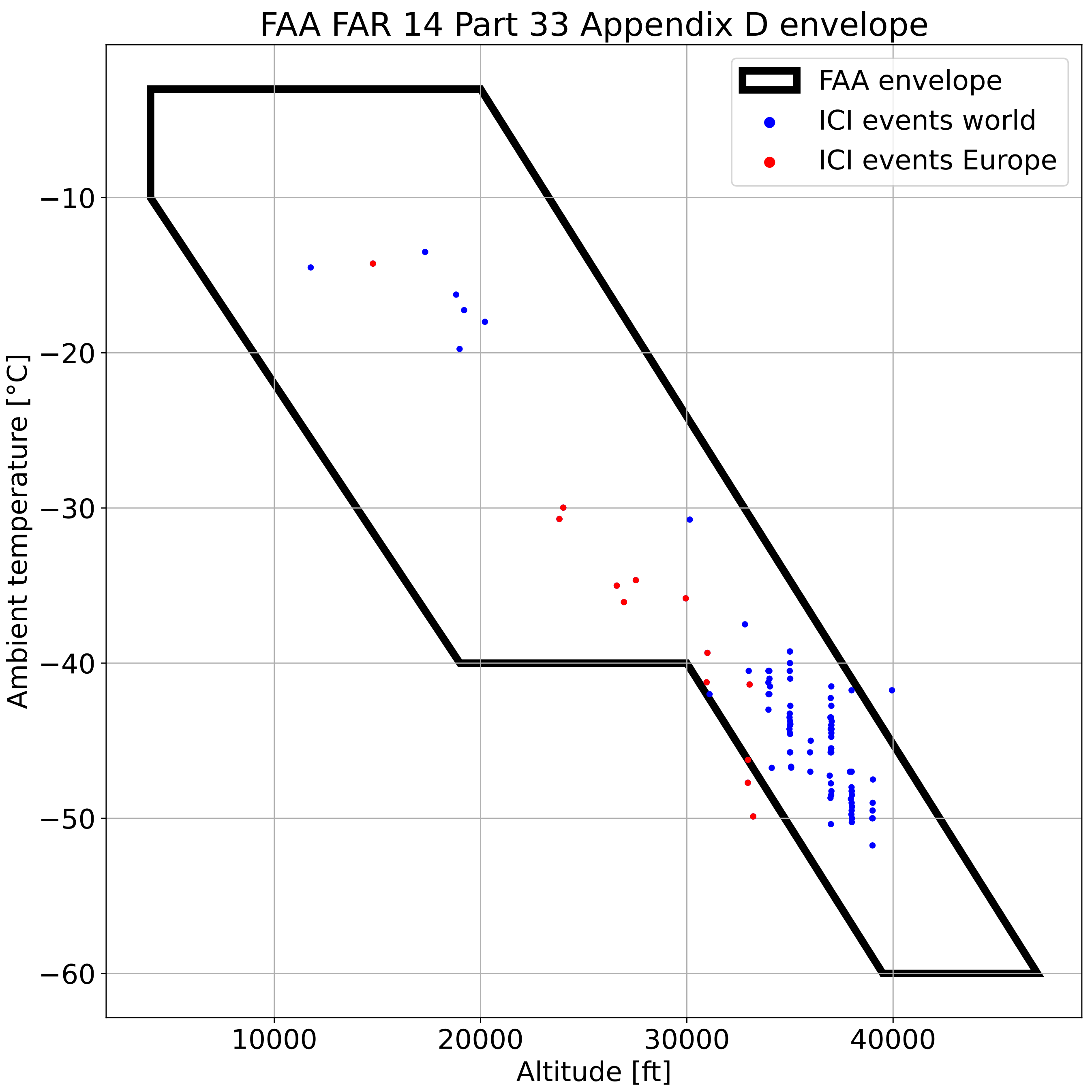

Figure 4. FAA 14 CFR Part 33 Appendix D (Federal Aviation Administration, Department of Transportation, 2023) standard envelope encloses typical air temperature and altitudes found for ICI conditions. The points denote the temperature and altitude associated with the events in the Lufthansa ICI database.

https://doi.org/10.5194/egusphere-2025-2985
Preprint. Discussion started: 19 August 2025

3.1 Dataset imbalance

Many industry and science-related problems are inherently characterized by data imbalance. Imbalanced datasets in classification problems are those datasets that have output quantities skewed toward a specific class. In the case of binary classification problems, the majority class is over-represented compared to the minority class. (Chawla et al., 2004).

The classification of imbalanced datasets significantly challenges the algorithmic approach for several reasons. First, one often wants to predict the minority class. However, the imbalance exposes the classifier to the majority class more frequently during training. For this reason, the minority class can be confused with noise and can be challenging to predict in areas of the data space where both classes overlap (Haixiang et al., 2017). Second, the use of conventional performance metrics, such as accuracy, may reflect the structure of the dataset rather than the classifier's predictive skill. Performance metrics more suitable for imbalanced problems are the Receiver Operating Characteristics (ROC) curve (Chawla et al., 2004) or the Critical Success Index (CSI) (e.g. de Laat et al., 2017). When correcting for the majority and minority class proportion, the ratio between the minority and majority class can be set freely depending on the application, and it is not necessary to exactly balance the two classes (Haixiang et al., 2017).

Table 1 showcases the imbalance between HIWC and no-HIWC sampled pixels by DARDAR. Given the large number of no-HIWC pixels, we undersample the original DARDAR dataset. This has a two-fold effect: first, undersampling a large training dataset is more computationally efficient (Chawla et al., 2004). Second, carefully choosing the undersampling technique reduces the correlation between samples. Indeed, samples of the same DARDAR trajectory correlate in time and space. Correlated samples may induce a bias in the training and validation procedure. This problem is also mentioned by Haggerty et al. (2020) in the case of aircraft measurements.

The undersampling is performed as follows:

- for the DARDAR trajectories with at least one HIWC sample, all HIWC samples belonging to that trajectory are taken, maintaining a buffer distance of 10 pixels if multiple consecutive pixels are flagged with HIWC;
- for DARDAR trajectories with no HIWC samples, pixels are sampled randomly among binned ranges of brightness temperature in the 10.8 μm channel, cloud optical thickness, and distance to the closest convective cells to cover a variety of HIWC-free conditions sufficiently.

Undersampling produces a new proportion between the classes depicted in Table 1. Although still imbalanced, a more aggressive undersampling was tested but led to undermining the subsequent model learning due to a too strong reduction of the variability of the majority class and, consequently, its representativeness. Finally, the test dataset contains five not undersampled DARDAR trajectories with at least one HIWC pixel.

3.2 Feature selection and random forest algorithm

The full list of input features considered for the potential ICI detection from SEVIRI is shown in Table 2.

https://doi.org/10.5194/egusphere-2025-2985
Preprint. Discussion started: 19 August 2025

Table 1. Proportion of HIWC events versus no-HIWC events for the original and undersampled dataset. The MSG slot is a single MSG scene containing one DARDAR trajectory. The MSG slot is flagged with HIWC if the corresponding DARDAR trajectory contains at least one HIWC sample; otherwise, it is flagged as no-HIWC. The undersampled dataset excludes five DARDAR trajectories with at least one HIWC pixel that are left out to test the model.

		# pixels	# MSG slots	% pixels
Original dataset	HIWC	889	83	0.54
	No-HIWC	165139	418	99.46
Undersampled dataset	HIWC	160	78	4.5
	No-HIWC	3424	418	95.5
Test dataset	HIWC	71	5	4.6
	No-HIWC	1477	0	95.4

The high-dimensional dataset produced when considering all the input predictors induces the so-called "curse of dimensionality" (James et al., 2021). In such cases, irrelevant or redundant predictors may act as noise and may lead to inefficient and inaccurate learning (Chawla et al., 2004). Furthermore, a high number of features can affect the variance-bias trade-off characterizing statistical learning methods: having a large set of features, even if relevant, may lead to an increase of variance that eventually outweighs the bias reduction produced by a more sophisticated model (James et al., 2021).

In our case, many features considered for the learning process are correlated by design, e.g., the variables originating from the same geostationary retrieval or the area of the convective cells expressed in km^2 and in the number of pixels. These redundant features may act as noisy features and exacerbate the curse of dimensionality. Therefore, a feature selection approach is chosen to reduce the dimensionality of the input data. This approach selects a subset of input features that optimizes the classifier's performance.

First, to perform the feature selection, the input features' correlation coefficient is determined for all the possible permutations of input predictor couples. This allows building a correlation matrix converted into a correlation-based distance (Ward's distance linkage score on the vertical axis of Fig. 5), which is considered a dissimilarity measure between the predictors. This distance creates a fictitious space within which input predictors are represented as data points. Then, the input feature subsets are obtained using hierarchical clustering, which is a bottom-up approach that assigns, as the first step, one cluster to each sample in the dataset's space. In our case, the samples are the input predictors. Eventually, it progressively identifies affine clusters and merges them until all sampled points end up in a single cluster, corresponding to the full dataset (James et al.,

https://doi.org/10.5194/egusphere-2025-2985
Preprint. Discussion started: 19 August 2025

Table 2. Candidate input predictors for the potential ICI detection. Chosen input predictors for the random forest are highlighted in italics.

SEVIRI		**CiPS/APICS**	
VIS006	*Reflectivity for channel 0.6 μm wavelength*	cth_cips	Cloud top height from CiPS
VIS008	Reflectivity for channel 0.8 μm wavelength	iot_cips	Ice clouds optical thickness from CiPS
IR_016	Reflectivity for channel 1.6 μm wavelength	iwp_cips	Ice water path from CiPS
IR_039	Brightness temperature for channel 3.9 μm wavelength	*ictau*	*Ice clouds optical thickness from APICS*
WV_062	Brightness temperature for channel 6.2 μm wavelength	icref	Ice clouds effective radius from APICS
WV_073	Brightness temperature for channel 7.3 μm wavelength	sza	Solar zenith angle from APICS
IR_087	Brightness temperature for channel 8.7 μm wavelength	wctau_mie	Water cloud optical thickness from APICS
IR_097	Brightness temperature for channel 9.7 μm wavelength	wcreff_mie	Water cloud optical thickness from APICS
IR_108	Brightness temperature for channel 10.8 μm wavelength	phase_apics	Cloud phase from APICS
IR_120	Brightness temperature for channel 12.0 μm wavelength		
IR_134	Brightness temperature for channel 13.4 μm wavelength		
Cb-TRAM stage 2		**Cb-TRAM stage 3**	
Cb2	Cb-TRAM stage 2	Cb3	Cb-TRAM stage 3
D_2	Distance to the closest convective cell of Cb-TRAM stage 2	D_3	Distance to the closest convective cell of Cb-TRAM stage 3
A_2	Area size (km^2) of the closest convective cell of Cb-TRAM stage 2	A_3	Area size (km^2) of the closest convective cell of Cb-TRAM stage 3
p_2	Area size (pixels) of the closest convective cell of Cb-TRAM stage 2	p_3	Area size (pixels) of the closest convective cell of Cb-TRAM stage 3
Cp10_2	Number of pixels within convective cells as detected by Cb-TRAM stage 2 in a radius of 10 km	Cp10_3	Number of pixels within convective cells as detected by Cb-TRAM stage 3 in a radius of 10 km
NC10_2	Number of convective cells as detected by Cb-TRAM stage 2 in a radius of 10 km	NC10_3	Number of convective cells as detected by Cb-TRAM stage 3 in a radius of 10 km
Cp50_2	*Number of pixels within convective cells as detected by Cb-TRAM stage 2 in a radius of 50 km*	Cp50_3	Number of pixels within convective cells as detected by Cb-TRAM stage 3 in a radius of 50 km
NC50_2	Number of convective cells as detected by Cb-TRAM stage 2 in a radius of 50 km	NC50_3	Number of convective cells as detected by Cb-TRAM stage 3 in a radius of 50 km
Cp100_2	Number of pixels within convective cells as detected by Cb-TRAM stage 2 in a radius of 100 km	*Cp100_3*	*Number of pixels within convective cells as detected by Cb-TRAM stage 3 in a radius of 100 km*
NC100_2	Number of convective cells as detected by Cb-TRAM stage 2 in a radius of 100 km	NC100_3	Number of convective cells as detected by Cb-TRAM stage 3 in a radius of 100 km
Features combinations			
BTD_062_108	*Brightness temperature difference between WV_062 and IR_108*		
BTD_062_073	Brightness temperature difference between WV_062 and WV_072		
BTD_039_108	Brightness temperature difference between IR_039 and IR_108		
RD_016_006	Reflectance difference between NIR_016 and VIS006		
D_A-1_2	*Ratio between distance and area of the closest convective cell from Cb-TRAM stage 2*		
D_A-1_3	*Ratio between distance and area of the closest convective cell from Cb-TRAM stage 3*		

https://doi.org/10.5194/egusphere-2025-2985
Preprint. Discussion started: 19 August 2025

2021).

The dendrogram depicts the bottom-up clustering, starting with a cluster containing all the input features at the top and then progressively splitting into multiple branches, each representing one cluster. Feature selection can be implemented by cutting the dendrogram at a certain level of the distance score on the vertical axis. In our case, the cutting level is initially determined through cross-validation to produce 16 clusters. The cutting level line crosses the dendrogram's branches multiple times. Features that can be reached from the same cut point following the branches belong to the same cluster. Features belonging to the same cluster are redundant in the sense that they give access to similar information to the statistical model through the learning process. Therefore, one feature per cluster is selected to obtain a subset of features suitable for learning with imbalanced data.

Furthermore, the permutation importance score allows the estimation of the importance of the selected features by hierarchical clustering. This evaluation requires setting a statistical model and a statistical performance metric that one wishes to optimize, which, in our case, are a random forest and the CSI, respectively. The method consists of shuffling the values of each predictor in the dataset to produce a corrupted dataset, which is fitted to the model chosen. The performance score is then compared with the score of the original dataset. The predictors may be correlated if the model maintains an overall constant predictive skill, but no predictor appears to be important according to the permutation importance score estimation. In this case, applying the permutation to one of them does not lead to a significant performance decrease because the model can access the same information via the correlated feature. This behavior can be seen in Fig. 6. In panel a), the initial choice of 16 input variables reveals that a few features are important according to the permutation score achieved. Still, the model performs well during the training, which can be an indication of correlated features. Shrinking to seven variables (panel b)) does not hinder the model's performance, but all input features become important. This further suggests that panel a) contains correlated features. The input features are selected manually based on the achieved permutation importance score obtained in the cross-validation, commonly used predictors in previous ICI detection retrievals, and the physical knowledge of the ICI phenomenon. From the predictors' list in Table 2, *BTD_062_108* is selected because it is a proxy for updraft speeds (Bravin et al., 2015; Grzych et al., 2015; Yost et al., 2018), *VIS006* and *ictau* are chosen for their ability to highlight optically thick and highly reflective deep convective clouds, while the convection metrics *Cp100_3*, *D_A-1_3*, *Cp50_2*, *D_A-1_2* are selected to account for convective cells density around each pixel and how big and distant the convective cells are from each pixel at different life-cycle stages. It must be noted that the *VIS006* and *ictau* variable choices prevent the retrieval from working in nighttime mode. The selected variables are denoted in italics in Table 2.

A random forest approach is selected to tackle this problem because it is among the most popular approaches to deal with imbalanced classification problems, guarantees interpretability, and can handle large datasets (Haixiang et al., 2017). Finally, the 5-fold cross-validation procedure also led to the random forest hyperparameters choice of 1000 trees inside the forest, 5 minimum allowed samples that can be included in each node at the end of the tree, and a probability threshold of 0.5 to convert from the probabilistic into the deterministic forecast.

https://doi.org/10.5194/egusphere-2025-2985
Preprint. Discussion started: 19 August 2025

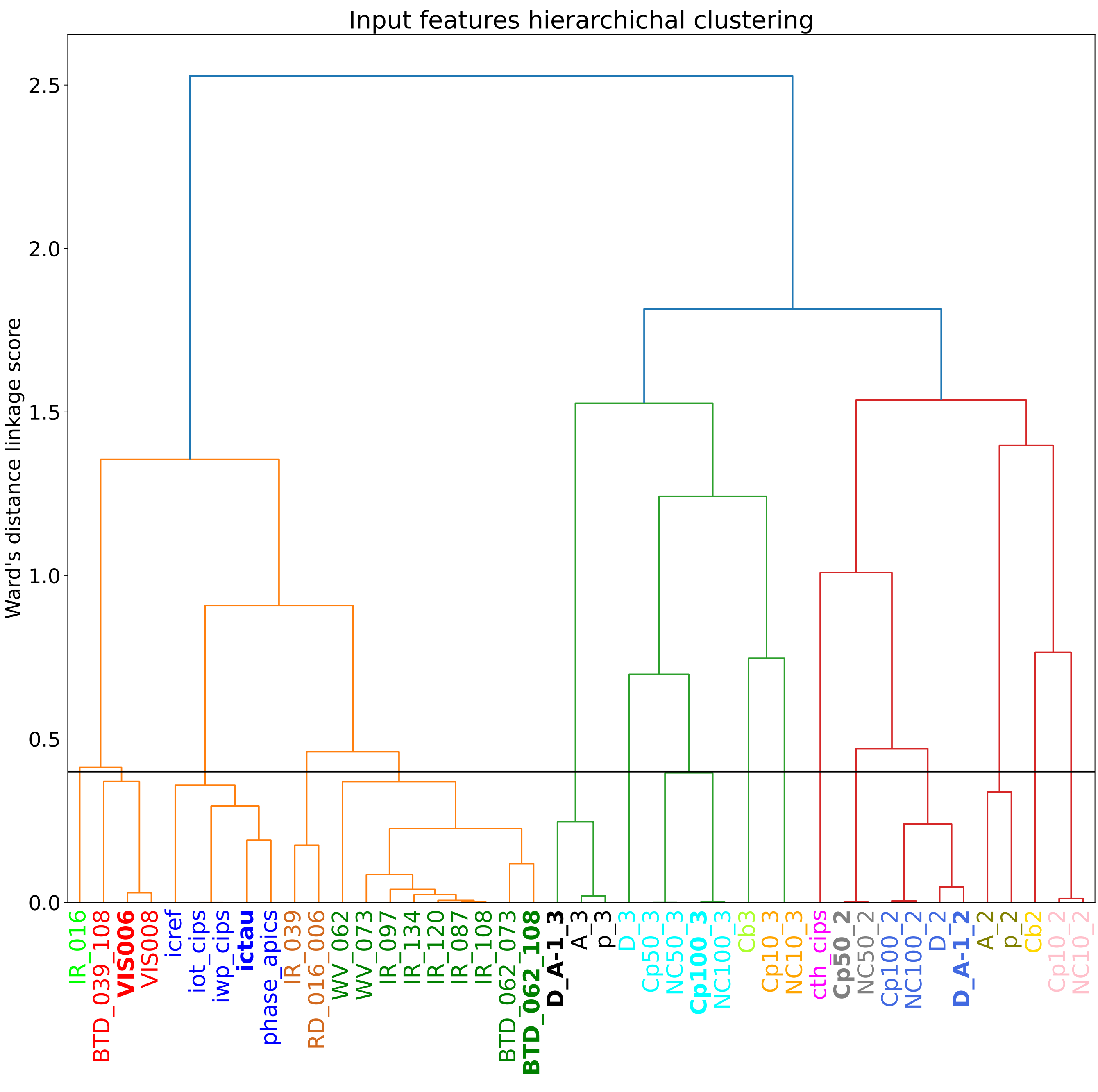

Figure 5. Illustration of the dendrogram depicting the feature extraction procedure through hierarchical clustering, based on the correlation score among the considered input features. The more correlated the clusters, the shorter the vertical branch extent is. The Ward's distance linkage score represents the correlation-based distance between clusters in our dataset's space. The black horizontal line depicts the suggested level by cross-validation at which the dendrogram should be cut to obtain an optimal number of clusters from which input features can be selected. Different colors depict the clusters obtained according to this cut level. The variables in bold are selected for the final version of the model. The choice is based on permutation importance estimation in Fig. 6.

https://doi.org/10.5194/egusphere-2025-2985
Preprint. Discussion started: 19 August 2025

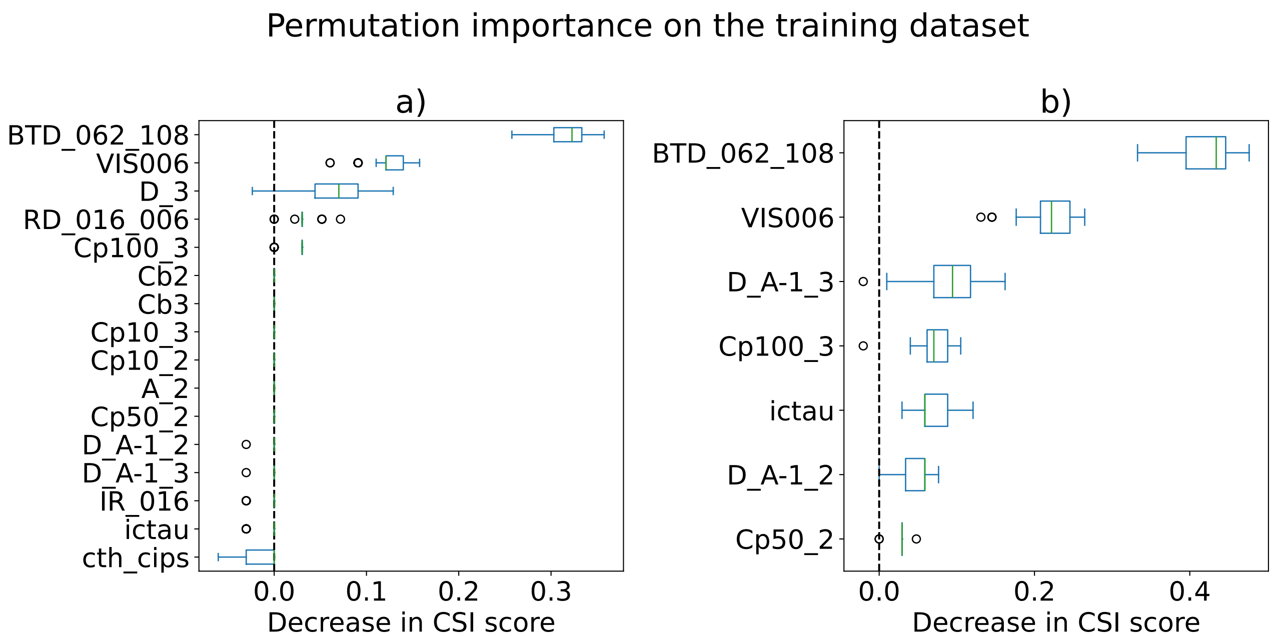

Figure 6. The permutation importance of a variable represents the decrease of the model's performance score (CSI) achieved when that variable is shuffled randomly with respect to the output target. Panel a) shows the permutation importance of 16 input variables achieved during training applied to four out of five folds in one instance of cross-validation. Similarly, panel b) shows the permutation importance of the manually selected input features on an example of a training instance during cross-validation. The box plots depict the distribution obtained when shuffling the variable 50 times randomly. In each box, the green line depicts the median of the distribution, the blue box depicts the interquartile range (IQR), delimited by the 25th and 75th percentiles (first quartile, Q1, and third quartile, Q3, respectively), and the whiskers spread to the last data point within Q1 - 1.5 IQR and Q3 + 1.5 IQR. Any data point beyond the whiskers is shown as an outlier with a white dot.

4 Retrieval application and validation

4.1 Retrieval performance test through DARDAR profiles

The metrics chosen to assess the retrieval's performance followed the ICI detection retrievals literature (de Laat et al., 2017; Yost et al., 2018; Haggerty et al., 2020) to enable comparison:

- the probability of detection (POD) is defined as the number of correctly predicted positive events (true positives, TP) over actual positive events (TP + FN, where FN stands for false negatives)

$$POD = \frac{TP}{TP+FN} \quad (1)$$

https://doi.org/10.5194/egusphere-2025-2985
Preprint. Discussion started: 19 August 2025

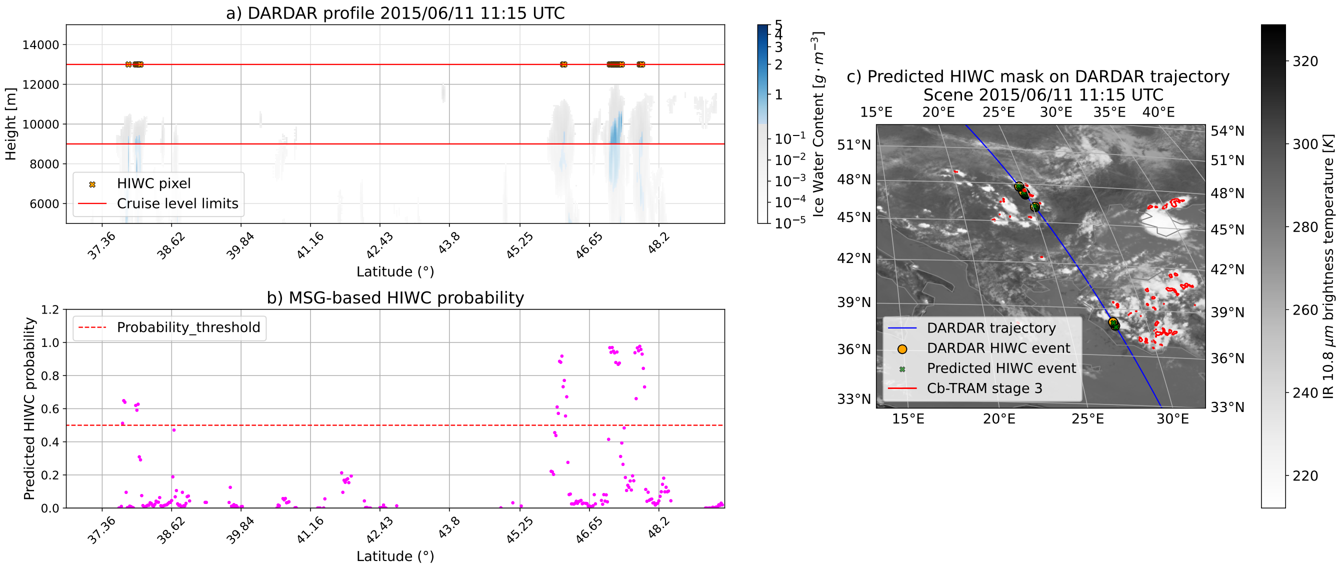

Figure 7. The HIWC conditions retrieval validation based on DARDAR trajectories. Panel a) shows the IWC cloud cross-section profile with the associated HIWC flags assigned with the criteria described in Sect. 2. Panel b) depicts the corresponding HIWC probability predicted by the random forest. Panel c) displays the DARDAR trajectory in the respective MSG image and the prediction of HIWC conditions converted from the probabilistic prediction.

– the false alarm rate (FAR) is defined as the ratio between the falsely predicted positive events (false positive, FP) to all predicted positive events (FP + TP):

$$FAR = \frac{FP}{FP + TP} \tag{2}$$

– the critical success index (CSI) is an index that balances POD and FAR. It is defined as:

$$CSI = \frac{TP}{TP + FN + FP} \tag{3}$$

– the ROC curve is used to display the variation of two performance metrics simultaneously. They are the POD and the FAR. The ROC curve is useful for classifiers because it explores every possible probability threshold to convert probabilistic into deterministic forecasts. In this way, different classifiers can be compared, no matter the chosen threshold. The chance line is depicted as a diagonal. The model lacks predictive skill if the ROC tends to the chance line. Ideally, the ROC should be a step function. The area under the curve, AUC, measures the overall performance of a classifier across multiple probability thresholds. The chance line has an AUC of 0.5, while an ideal ROC curve approaches 1.0 AUC (James et al., 2021).

The retrieval test dataset contains five randomly selected DARDAR trajectories with at least one HIWC pixel. These trajectories are left out of the training and cross-validation procedure. In Fig. 7, we present an example of the DARDAR trajectory used to

https://doi.org/10.5194/egusphere-2025-2985
Preprint. Discussion started: 19 August 2025

test the potential ICI detection. Panel a) shows the cloud's IWC profile where one can see two areas of HIWC conditions: the first, around 37°N, originating from two adjacent deep clouds, and the second, in the proximity of 47°N latitude, composed by a set of three vertically developed clouds that produce three nearby but distinct HIWC areas. Clouds at 37° N are characterized by a notable vertical extent of HIWC conditions with IWC reaching the maximum value of 0.8 g m^{-3} within the cruise levels and by cloud tops extending up to 10410 m. The clouds composing the system at 47° N have more extended HIWC conditions. The central cell is by far the most active in terms of HIWC, with a peak IWC value of 1.1 g m^{-3} and the cloud top at 11940 m altitude. The corresponding MSG-based HIWC probability is displayed in panel b). This is plotted only for icy cloud pixels according to the CiPS mask and it is characterized by a sharp transition from low to high HIWC probabilities. One can also see that the clouds around 37° N have a lower HIWC probability when compared to clouds at 47°N, even though still above the threshold of 0.5. This can be attributed to a relatively higher density of stage 3 convective cells detected by Cb-TRAM near the trajectory.

To assess the overall robustness of the presented approach, the training and subsequent testing are repeated 100 times, each time with another random selection of 5 DARDAR trajectories as test data. The repeated tests generate the distribution in the performance metrics shown in the box plots of Fig. 8. The large variability may suggest the need for further data, as the method seems very sensitive to the training and test datasets used. The median values of POD = 0.83, FAR = 0.51, CSI = 0.45, and AUC = 0.61 are similar to previous retrieval performances, depicted in Table 3. For our retrieval, POD is the least spread metric with 75 % of the tests lying above 0.79, denoting a high probability of detecting positive events correctly. On the other hand, FAR spreads over a much larger range, which is also reflected in the CSI and AUC variability. Focusing on the AUC, this metric lags behind when compared to Yost et al. (2018) and Haggerty et al. (2020) retrievals.

Table 3. Performance metrics comparison of the random forest ICI retrieval presented in this paper versus the previously developed ICI retrievals. Although the training and verification techniques differ, as well as the retrieval's applicability, these results are reported to place this work in the current research context. The metrics shown correspond to the median value of POD, FAR, CSI, and AUC found in Fig. 8. Yost et al. (2018) and Haggerty et al. (2020) developed both daytime and nighttime retrievals, but the metrics reported here refer to daytime only.

	POD	FAR	CSI	AUC
(de Laat et al., 2017)	0.59	0.52	0.36	-
(Yost et al., 2018)	0.75	0.35	-	0.75
(Haggerty et al., 2020)	0.86	0.51	-	0.85
This paper	0.83	0.51	0.45	0.61

This can be linked to the sharp transition from low to high HIWC probability. The sharp probability transition means that acceptable POD can only be achieved if one allows a substantial FAR, suddenly leading the ROC curve to shift from below to above the chance line and eventually producing a low AUC. This sharp transition of predicted HIWC probability can also be observed in Fig. 10, where the two-dimensional HIWC probability mask is shown.

https://doi.org/10.5194/egusphere-2025-2985
Preprint. Discussion started: 19 August 2025

Figure 8. Performance metrics variability in the repeated training and test procedure. The box plots represent the obtained performance scores over 100 iterations. The main central box depicts the interquartile range (IQR), which is the range between the 25th (first quartile, Q1) and 75th percentiles (third quartile, Q3) of the distribution. Whiskers are defined by the last data points lying within Q1 - 1.5 IQR and Q3 + 1.5 IQR. Anything lying outside the whiskers is considered an outlier.

https://doi.org/10.5194/egusphere-2025-2985
Preprint. Discussion started: 19 August 2025

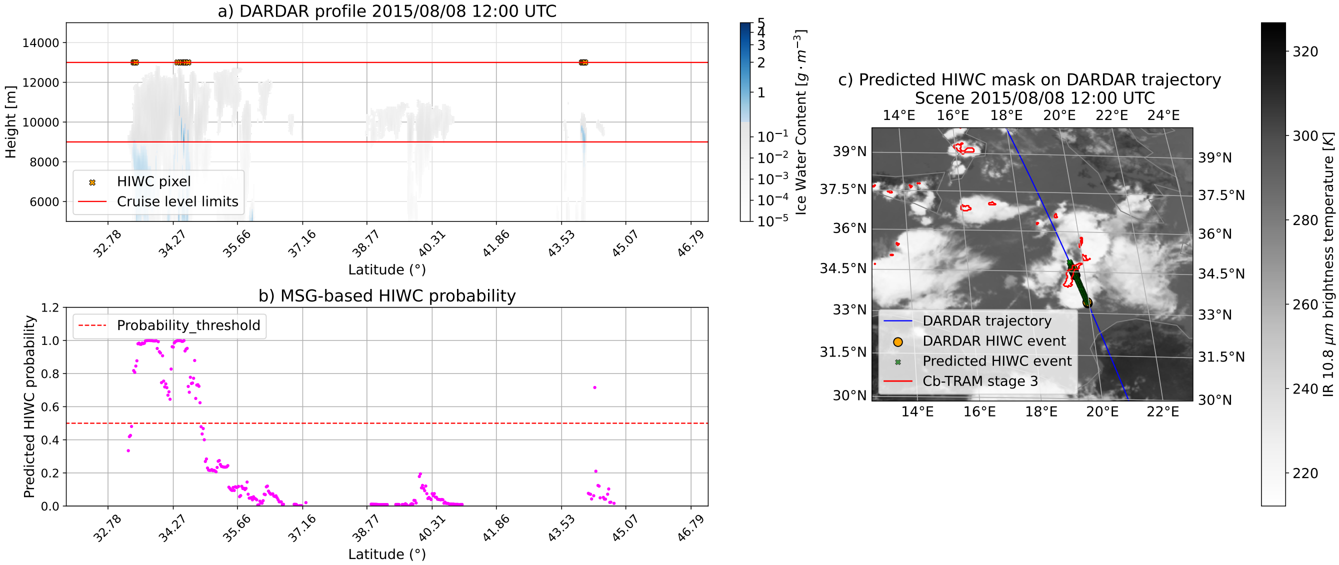

Figure 9. As in Fig. 7 for a different DARDAR trajectory example.

To explain the high FAR incidence, one can observe Fig. 9. Focusing on the convective system between 33°N and 36°N latitude, HIWC conditions within the cruise levels are present in the DARDAR dataset in the southern and northernmost parts of the system. In contrast, the inner parts are characterized by HIWC conditions only below the cruise levels. However, the MSG-based HIWC probability stays above the threshold throughout the horizontal extent of the cloud, though with a small dip in the middle section. This is reflected in panel c), where HIWC conditions are predicted for the entire cloud rather than just the two extremes, giving rise to a high FAR. de Laat et al. (2017) and Haggerty et al. (2020) also observed a relatively high FAR. Haggerty et al. (2020) concludes that most FAR pixels are associated with HIWC conditions occurring at altitudes different than the ones sampled by the aircraft. This is also the case for Fig. 9. However, in this instance, cruise levels are chosen according to the altitude at which ICI events occur in the Lufthansa ICI database. Nonetheless, the best trade-off to retrieve cloud properties remains challenging to find. Cloud properties vary within the cloud structure, while passive sensors can only detect cloud top characteristics or column-integrated quantities. The HIWC conditions detection presented is compared with previously developed retrievals to put this work in the current research context. However, these retrievals have different characteristics. Namely, this method differs by data sources, input features, clouds' microphysical characteristics retrievals, and detection approaches. Furthermore, our retrieval is tested and validated in the Europe domain, and not globally as in e.g. de Laat et al. (2017).

4.2 Lufthansa case studies

The Lufthansa ICI database is presented in Sect. 2.4 and contains 10 case studies in Europe out of the 100 cases available globally. Three events are encountered at night, but nighttime scenes are discarded because of the absence of visible channels

https://doi.org/10.5194/egusphere-2025-2985
Preprint. Discussion started: 19 August 2025

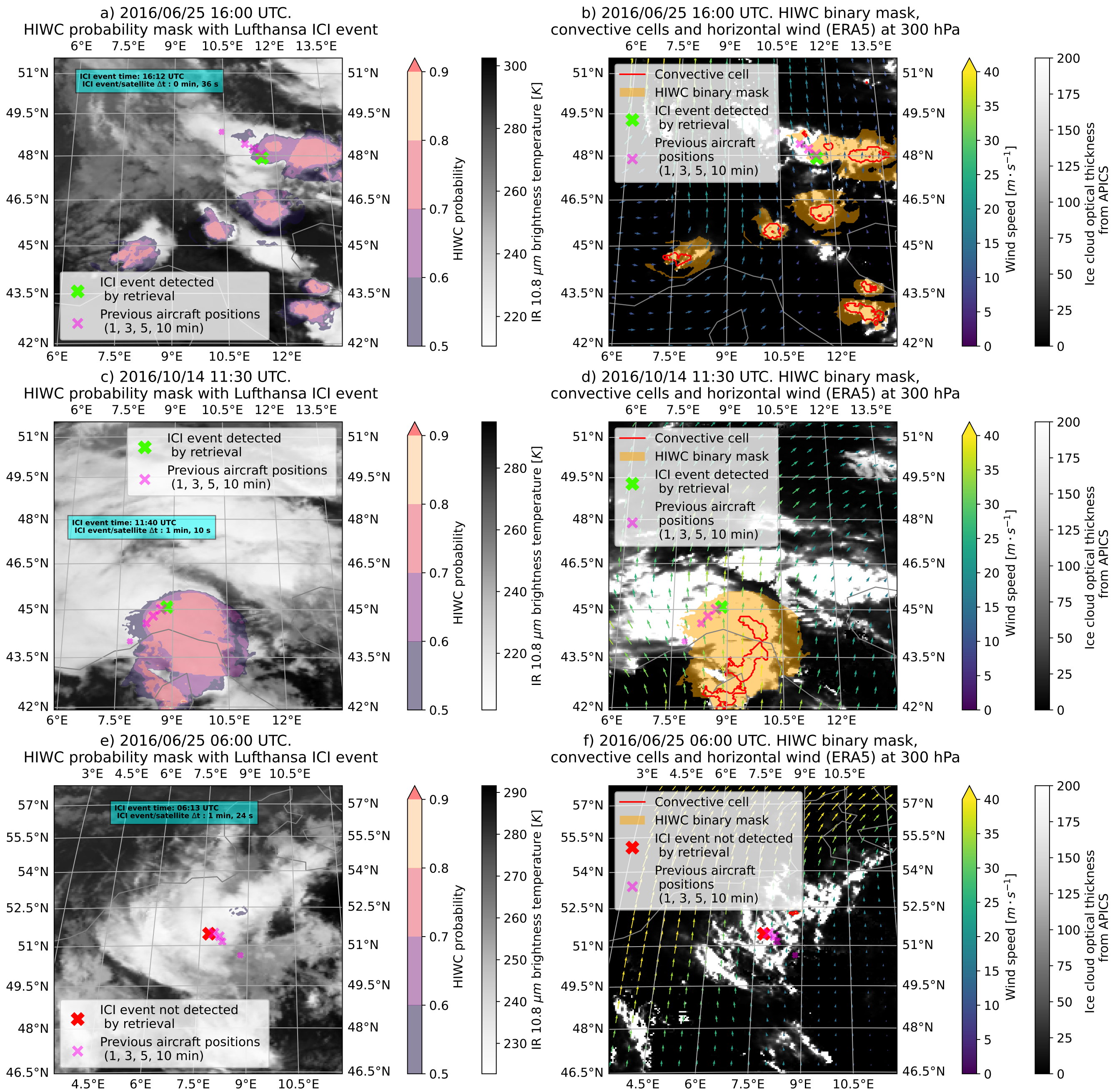

Figure 10. Panels a), c), and e) show the HIWC probability mask with the respective aircraft's positions before and during the in-service ICI events. The ICI event time is shown in the text box with its time delta compared to the actual satellite acquisition time. Panels b), d), and f) show the respective HIWC binary masks, converted using the 0.5 HIWC probability threshold. The mask is combined with Cb-TRAM stage 3 convective cells and horizontal wind field at 300 hPa from ECMWF ERA5 reanalysis data and ice cloud optical thickness from APICS.

https://doi.org/10.5194/egusphere-2025-2985
Preprint. Discussion started: 19 August 2025

and optical thickness data. ICI events are correctly detected by the HIWC mask in four out of the seven remaining daytime scenes.

The criterion for correct detection considers the last available aircraft position, labeled as ICI position, and whether the predicted HIWC probability is larger than our threshold of 0.5. This criterion is applied irrespective of the time difference between the aircraft measurements and the satellite acquisition time, which could be up to 7 minutes and 30 seconds. The three scenes in Fig. 10 are a subset of the processed scenes, selected according to the smallest time delta between the aircraft measurement and the satellite acquisition time. Appendix A contains the remaining Lufthansa case studies.

In Fig. 10, panels a) and c) have large areas of HIWC high probability, often exceeding 0.7–0.8. The mask generally has a sharp transition from 0.5 to 0.7 HIWC probability and seldom approaches 1.0 (a few small areas in panel c)). The HIWC mask is almost completely absent in panel e), with small patches of 0.5 HIWC probability around the Cb-TRAM stage 3 convective cell seen in panel f). The HIWC binary mask shown in panels b), d), and f) is compared with the detected Cb-TRAM convective cells and the ECMWF ERA5 reanalysis wind field at 300 hPa. The model data in panels b), d), and f) have an hourly resolution; therefore, scenes b) and f) have simultaneous satellite images and wind fields, while for scene d), the wind field refers to 11 UTC.

The HIWC masks generally differ from the Cb-TRAM convective cells and the ice optical thickness from APICS. The HIWC masks often stay around detected convective cells with high optical thickness and cold cloud tops, highlighting the need for a dedicated HIWC detection product. In particular, it is possible to observe that the masks propagate downstream of the detected convective cells, even though the wind is not used as an input feature. In panel b), the wind field is relatively weak in correspondence with the biggest HIWC mask patches. The HIWC mask tends to follow the wind field for the convective cells between 45°N and 46.5°N latitude, but this behavior is less evident for the big HIWC patch at 48°N latitude. Instead, panel d) depicts strong wind fields and a large HIWC mask propagating downstream of the main convective cores detected by Cb-TRAM.

From the correctly detected events, airplanes flew inside the HIWC mask before the aircraft's final location, where ICI was reported. This might indicate that aircraft have to fly within ICI conditions for some time to guarantee enough exposure to such conditions for ice to accrete inside the engines. This would also be consistent with one of the failed detections presented in App. A, where the flight flew through the HIWC mask before the final position that falls right outside the mask. For the failed detection in Fig. 10 panel e), we speculate that the HIWC mask is missing due to the absence of large convective cells in panel f). Although ICI events are almost exclusively attributed to convection in the literature (Grzych, 2010; Bravin et al., 2015), ICI events have also been reported in different conditions, such as within extra-tropical cyclones (Gayet et al., 2012). Panel f) could suggest that this retrieval would fail when Cb-TRAM cannot detect deep convective cells.

5 Conclusions & outlook

In this study, our goal is to assess the feasibility of a detection retrieval for potential ICI conditions, based exclusively on remote sensing data and a random forest approach as a machine learning technique. A combination of passive remote sensing

https://doi.org/10.5194/egusphere-2025-2985
Preprint. Discussion started: 19 August 2025

measurements from geostationary satellites is used to detect areas with HIWC conditions at passenger aircraft cruise levels. These conditions are chosen as an indicator for potential ICI formation. For the training phase, HIWC conditions are located from active measurements of IWC from polar-orbiting satellites, i.e. the DARDAR dataset, which is taken as ground truth. The results obtained by testing this approach with DARDAR trajectories lead to median values of the performance metrics POD = 0.83, FAR = 0.51, CSI = 0.45, and AUC = 0.61. This method outperforms the approach presented by de Laat et al. (2017), who validated their algorithm globally with DARDAR data, and it achieved comparable results to Haggerty et al. (2020). However, Haggerty et al. (2020) used multiple input sources, such as ground-based weather radar and numerical weather prediction, in addition to geostationary satellite images. In this study, we used only the latter and we achieved comparable performance.

The validation of the retrieval is also supported by a database of ICI case studies reported by Lufthansa during operational conditions. The retrieval correctly detects four out of seven events, assuming a correct detection whenever the aircraft's final position is within the HIWC mask. From the observation of the case study examples, the HIWC mask surrounds convective cells in areas with optically thick clouds and glaciated cloud tops. Moreover, the mask often follows the wind field downstream of convective cells, which is physically reasonable but not necessarily expected, as the wind field is not explicitly included as an input feature. The failed HIWC conditions detection in the scene without convective cells highlights the importance of convection signatures to obtain a high probability HIWC conditions signal. This is coherent with past literature, where deep convection and ICI events are often found to be correlated. However, in different conditions, such as in extra-tropical cyclones, the failed detection example indicates that the retrieval could have less chance to detect HIWC conditions when there is no deep convection.

Integrating the ICI retrieval with cloud microphysical properties retrievals would enable scientific studies about the possible ice formation pathways during ICI events, especially exploiting the temporal resolution of geostationary satellites, as these are not yet understood (Leroy et al., 2017). For example, one could correlate potential ICI areas with retrieved ice particles' effective radius and updraft speeds that can be either taken from reanalysis data or satellite images. A fully operational ICI detection product would require additional development, as it uses visible channels and optical thickness that prevent this retrieval from detecting HIWC conditions at night. Future development would include a more flexible way to select input predictors according to availability. Furthermore, the training dataset could be enlarged, considering the overlap between MSG-3 and DARDAR between 2013 and 2017. To conclude, the retrieval shows promising performance in detecting potential ICI conditions, using exclusively geostationary satellite imagery as input. This would allow a flexible extension to other geostationary satellite platforms, and its operational implementation would enable airlines to avoid HIWC conditions to mitigate ICI effects on the fleet.

https://doi.org/10.5194/egusphere-2025-2985
Preprint. Discussion started: 19 August 2025

415 **Appendix A: Lufthansa case studies**

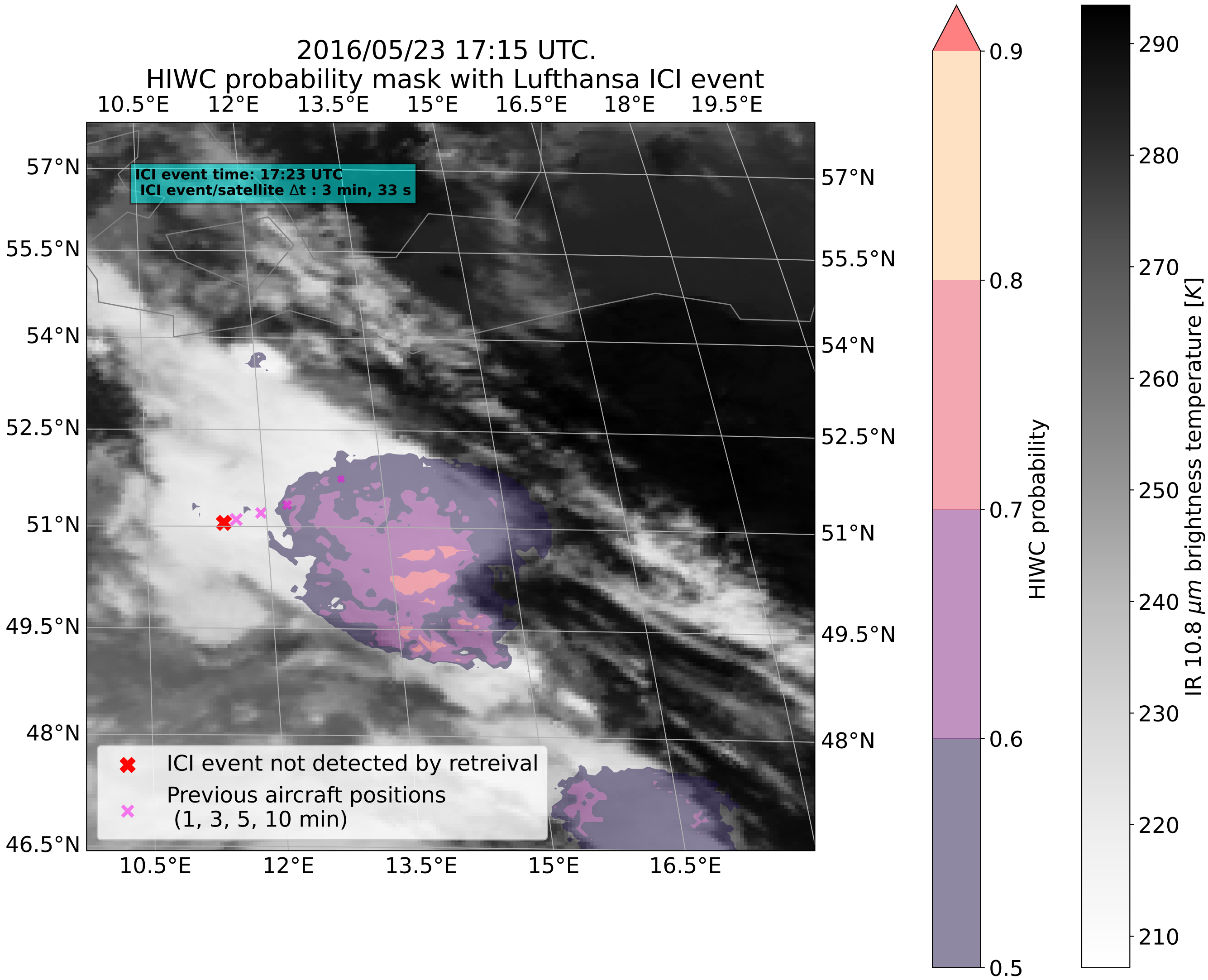

Figure A1. As in Fig. 10 panels a), c), and e) for additional Lufthansa ICI case studies.

https://doi.org/10.5194/egusphere-2025-2985
Preprint. Discussion started: 19 August 2025

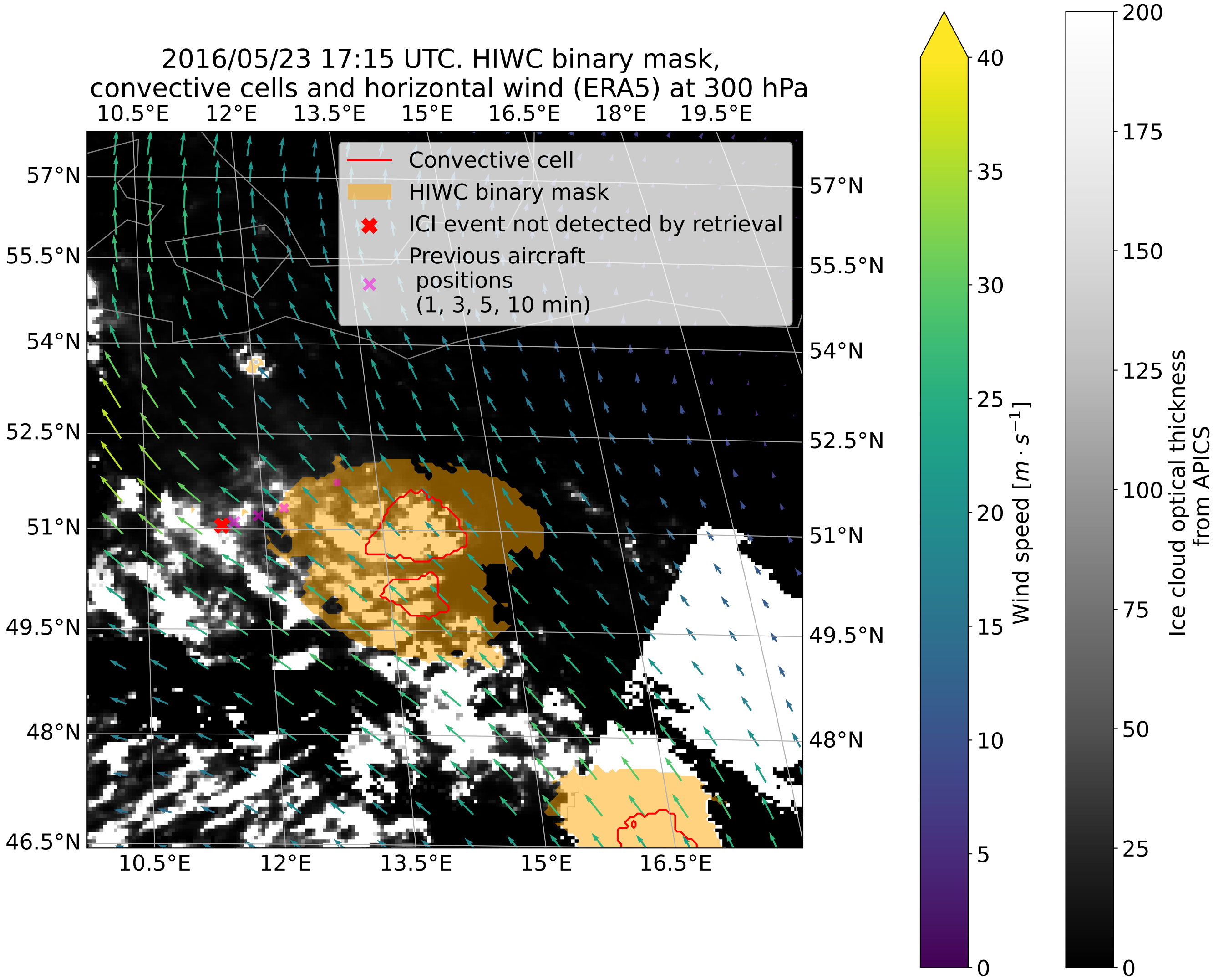

Figure A2. As in Fig. 10 panels b), d), and f) for additional Lufthansa ICI case studies.

https://doi.org/10.5194/egusphere-2025-2985
Preprint. Discussion started: 19 August 2025

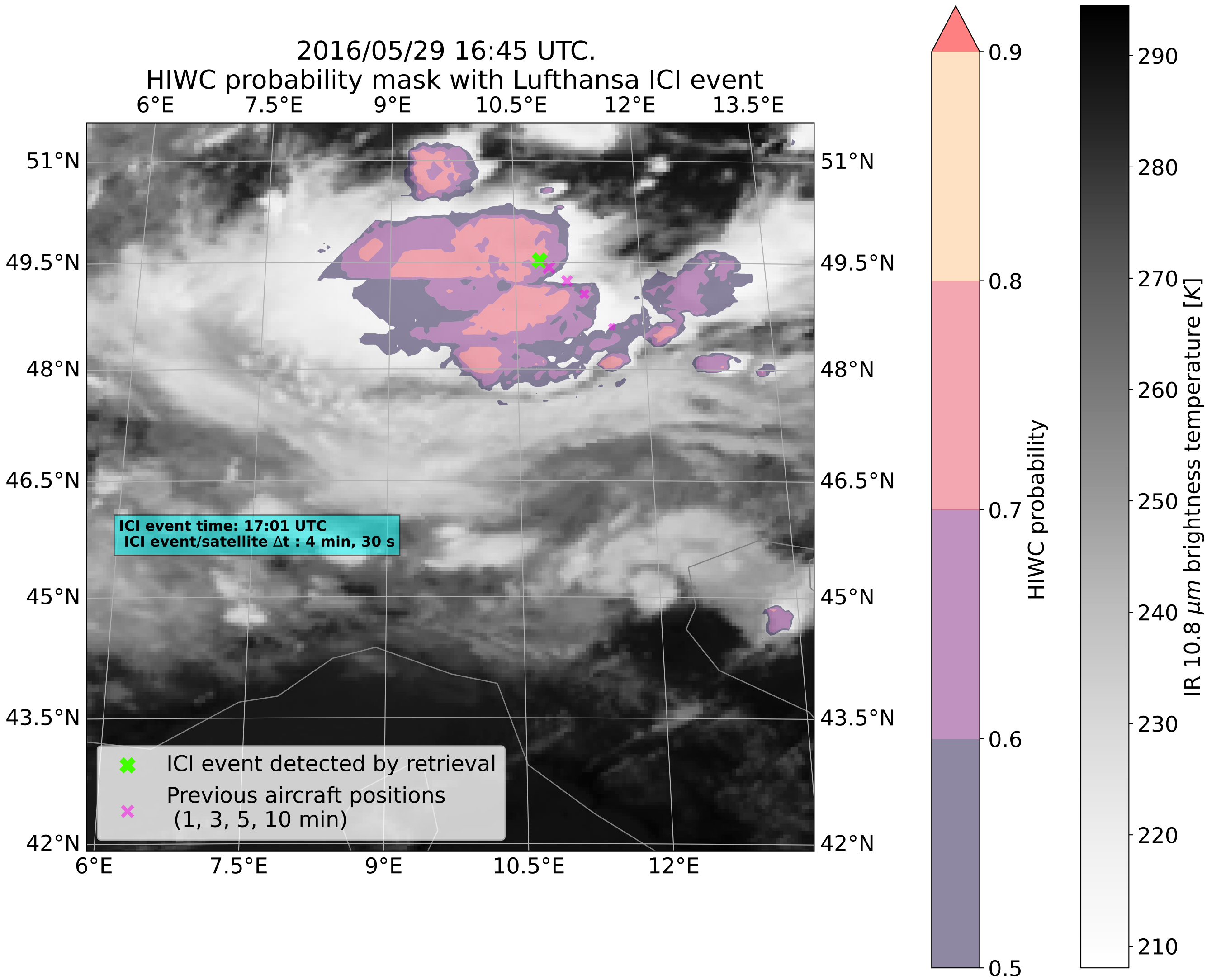

Figure A3. As in Fig. 10 panels a), c), and e) for additional Lufthansa ICI case studies.

https://doi.org/10.5194/egusphere-2025-2985
Preprint. Discussion started: 19 August 2025

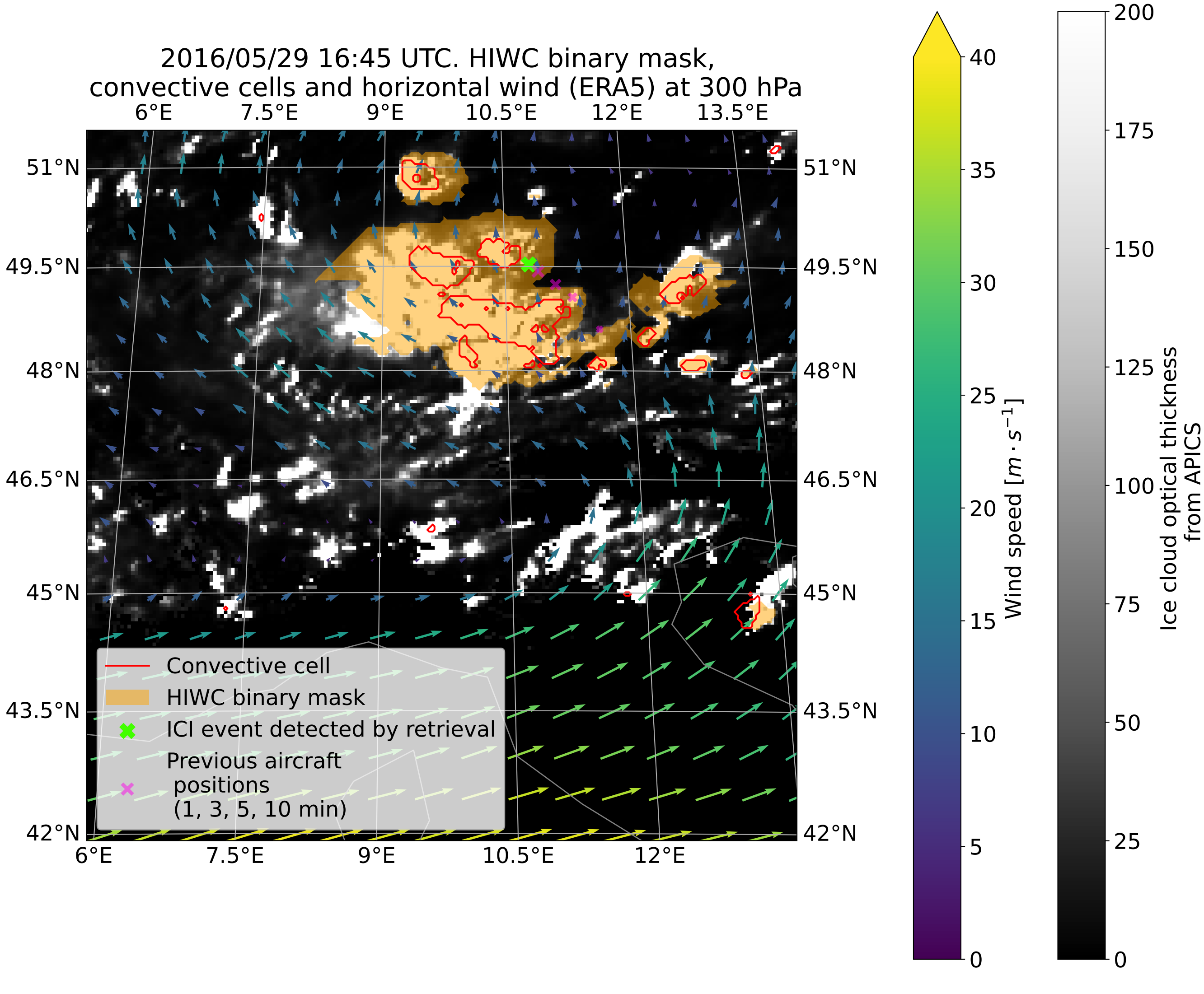

Figure A4. As in Fig. 10 panels b), d), and f) for additional Lufthansa ICI case studies.

https://doi.org/10.5194/egusphere-2025-2985
Preprint. Discussion started: 19 August 2025

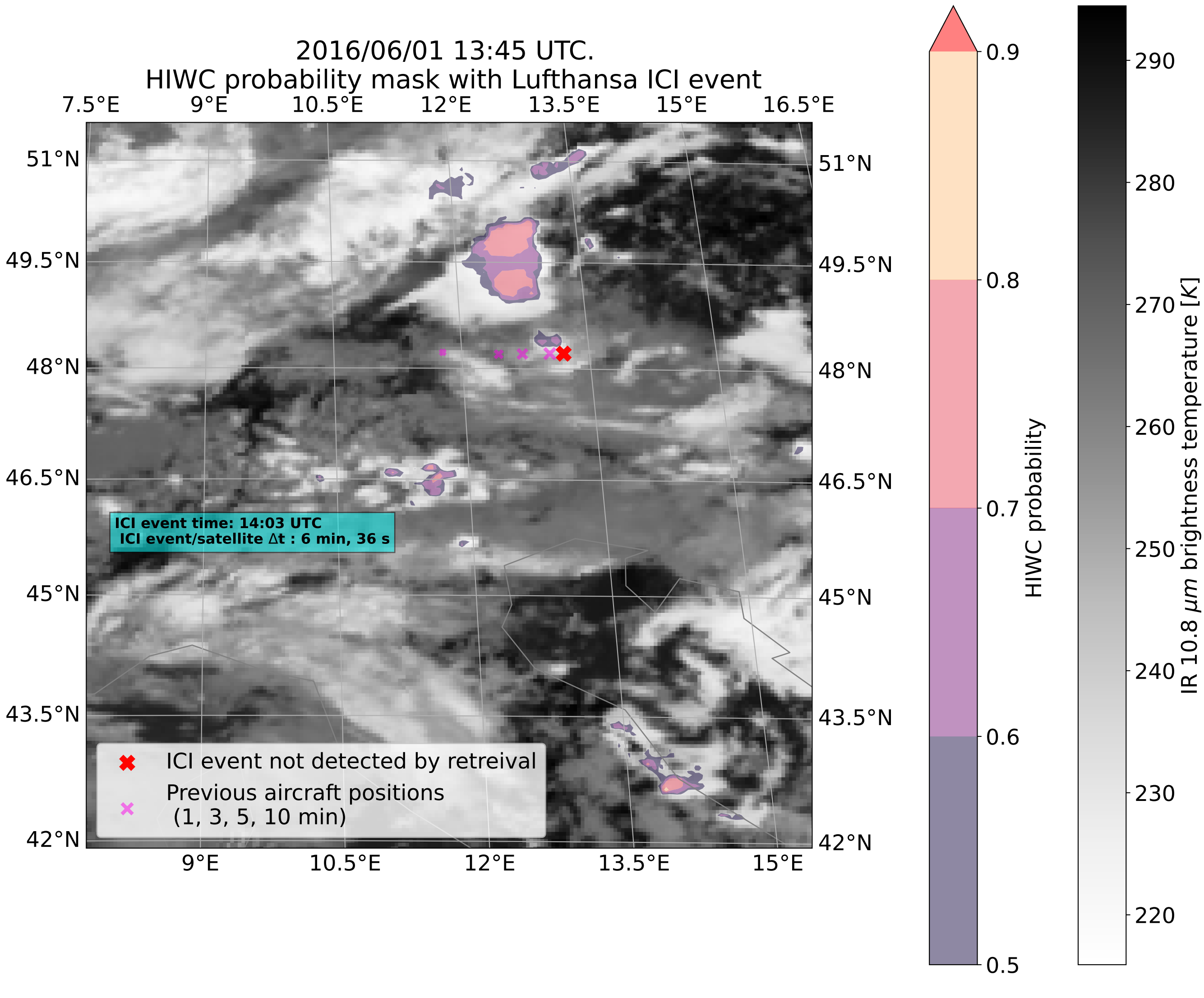

Figure A5. As in Fig. 10 panels a), c), and e) for additional Lufthansa ICI case studies.

https://doi.org/10.5194/egusphere-2025-2985
Preprint. Discussion started: 19 August 2025

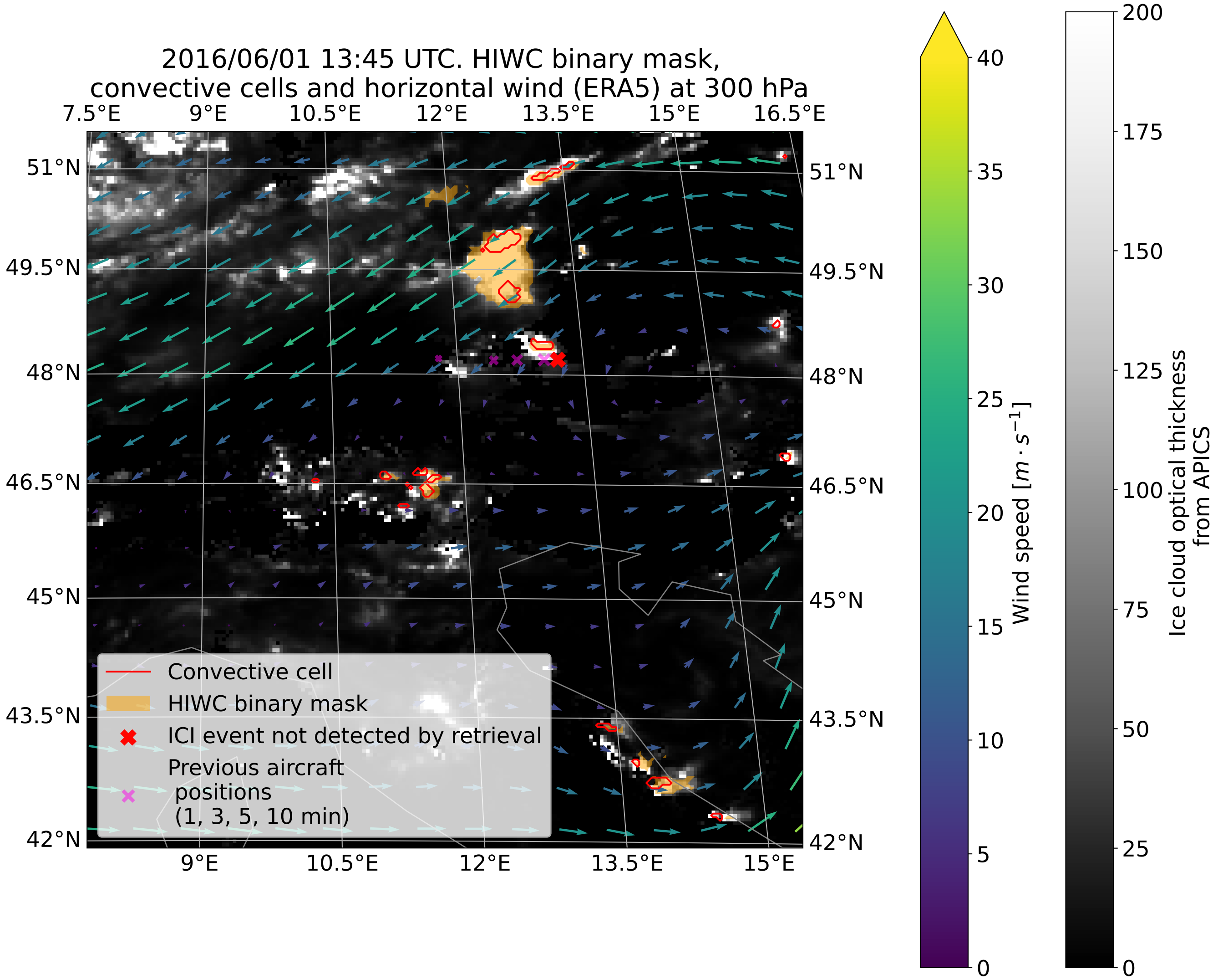

Figure A6. As in Fig. 10 panels b), d), and f) for additional Lufthansa ICI case studies.

https://doi.org/10.5194/egusphere-2025-2985
Preprint. Discussion started: 19 August 2025

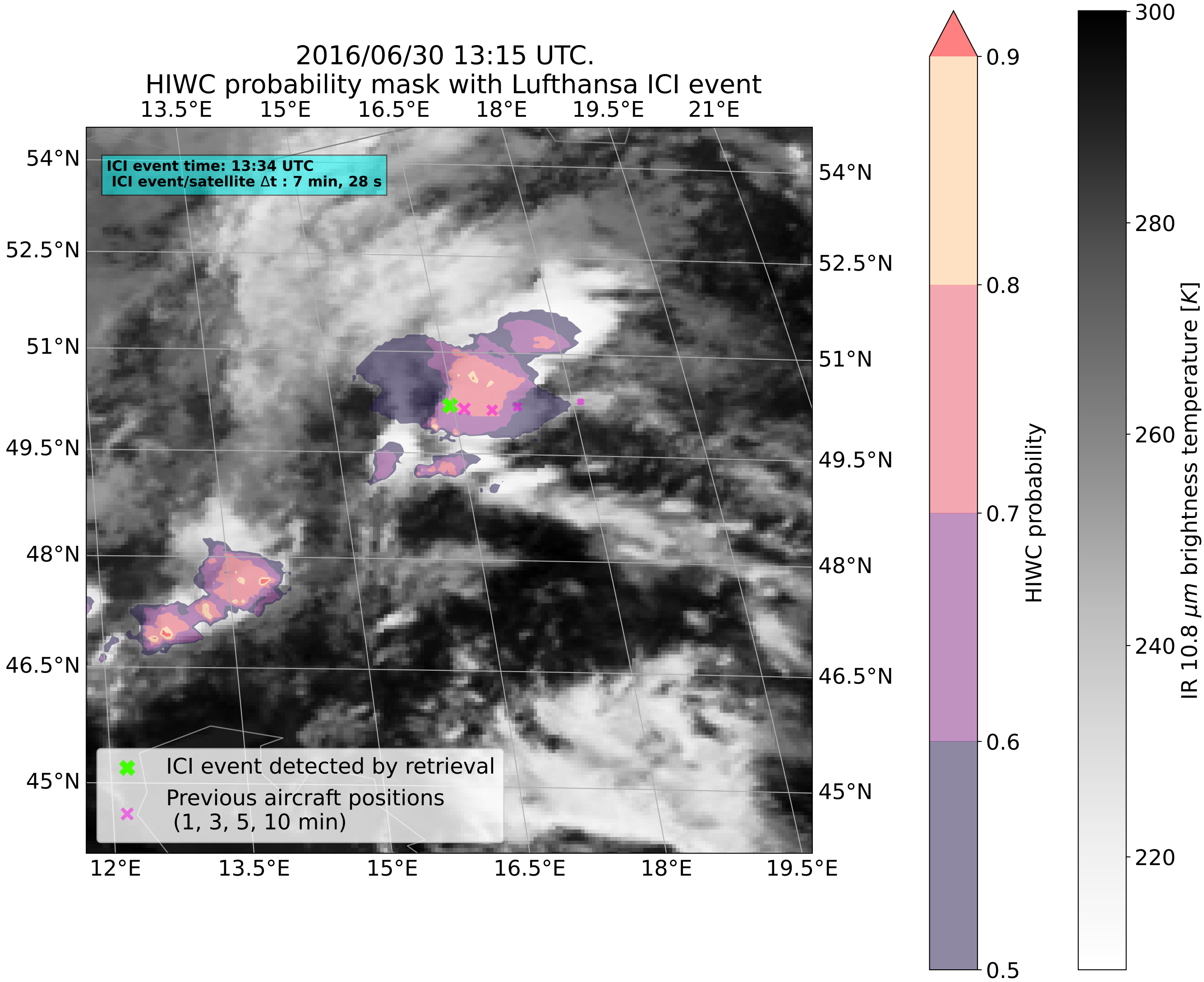

Figure A7. As in Fig. 10 panels a), c), and e) for additional Lufthansa ICI case studies.

https://doi.org/10.5194/egusphere-2025-2985
Preprint. Discussion started: 19 August 2025

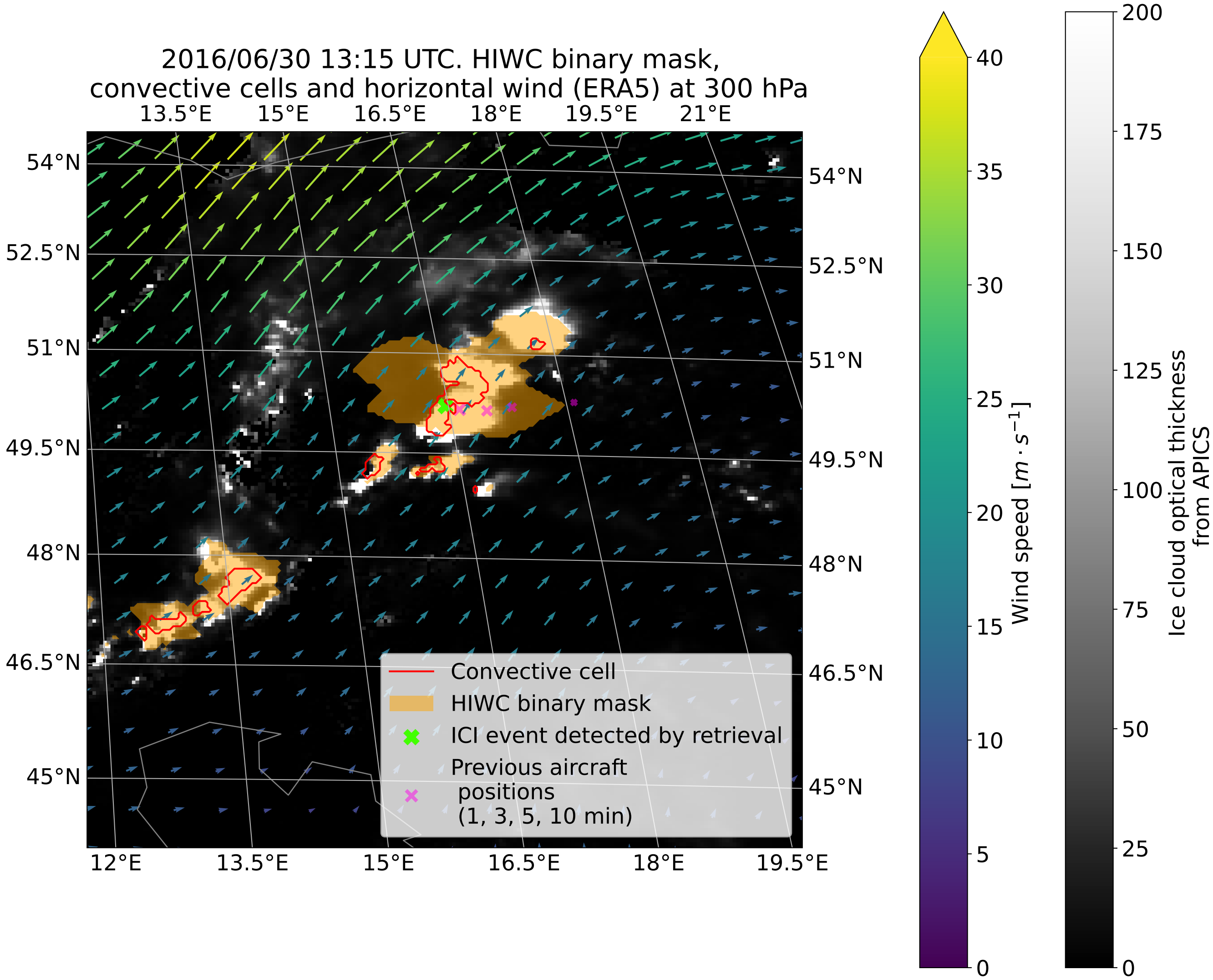

Figure A8. As in Fig. 10 panels b), d), and f) for additional Lufthansa ICI case studies.

https://doi.org/10.5194/egusphere-2025-2985
Preprint. Discussion started: 19 August 2025

Data availability. Archived DARDAR data are available at sftp://sftp.icare.univ-lille.fr, documentation can be found at https://www.icare.univ-lille.fr/dardar/.
MSG data are available via EUMETSAT Data Store at https://user.eumetsat.int/data/satellites/meteosat-second-generation.
Geostationary-based retrievals CiPS, APICS, Cb-TRAM, and HIWC detection data are available on request at DLR.

Author contributions. Conceptualization, DP and LB; methodology, MA; software, MA and JM; validation, MA; formal analysis, MA; investigation, MA; data sources, JM and MB; data curation, MA; writing—original draft preparation, MA; writing—review and editing, DP, LB, JM, RM, FK, MB; visualization, MA; supervision, DP, LB, RM and FK; project administration, DP; funding acquisition, DP, LB, RM and FK. All authors have read and agreed to the published version of the manuscript.

Competing interests. Author MB is employed by Deutsche Lufthansa AG. All other authors declare that they have no conflict of interest.

Acknowledgements. Author MA would like to thank Dino Zardi for the support provided in the context of his master's thesis. We would also like to thank Georgios Dekoutsidis for the manuscript's internal review.
The research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—TRR 301—Project-ID 428312742. The work on which this publication is based was carried out on behalf of the German Federal Ministry of Transport and Digital Infrastructure under FE-No. 50.0391/2021. The responsibility for the content lies exclusively with the authors.

References

Bedka, K., Yost, C., Nguyen, L., Strapp, J. W., Ratvasky, T., Khlopenkov, K., Scarino, B., Bhatt, R., Spangenberg, D., and Palikonda, R.: Analysis and Automated Detection of Ice Crystal Icing Conditions Using Geostationary Satellite Datasets and In Situ Ice Water Content Measurements, SAE International Journal of Advances and Current Practices in Mobility, 2, 35–57, https://doi.org/10.4271/2019-01-1953, 2020.

Bravin, M., Strapp, J. W., and Mason, J.: An Investigation into Location and Convective Lifecycle Trends in an Ice Crystal Icing Engine Event Database, in: SAE Technical Paper Series, SAE Technical Paper Series, SAE International400 Commonwealth Drive, Warrendale, PA, United States, https://doi.org/10.4271/2015-01-2130, 2015.

Bugliaro, L., Zinner, T., Keil, C., Mayer, B., Hollmann, R., Reuter, M., and Thomas, W.: Validation of cloud property retrievals with simulated satellite radiances: a case study for SEVIRI, Atmospheric Chemistry and Physics, 11, 5603–5624, https://doi.org/10.5194/acp-11-5603-2011, 2011.

Chawla, N. V., Japkowicz, N., and Kotcz, A.: Special issue on learning from imbalanced data sets, ACM SIGKDD explorations newsletter, 6, 1–6, 2004.

de Laat, A., Defer, E., Delanoë, J., Dezitter, F., Gounou, A., Grandin, A., Guignard, A., Meirink, J. F., Moisselin, J.-M., and Parol, F.: Analysis of geostationary satellite-derived cloud parameters associated with environments with high ice water content, Atmospheric measurement techniques, 10, 1359–1371, https://doi.org/10.5194/amt-10-1359-2017, 2017.

Delanoë, J. and Hogan, R. J.: A variational scheme for retrieving ice cloud properties from combined radar, lidar, and infrared radiometer, Journal of Geophysical Research: Atmospheres, 113, https://doi.org/https://doi.org/10.1029/2007JD009000, 2008.

Delanoë, J. and Hogan, R. J.: Combined CloudSat-CALIPSO-MODIS retrievals of the properties of ice clouds, Journal of Geophysical Research: Atmospheres, 115, https://doi.org/https://doi.org/10.1029/2009JD012346, 2010.

Federal Aviation Administration, Department of Transportation: 14 CFR Part 33, https://www.ecfr.gov/current/title-14/part-33, 2023.

Gayet, J.-F., Mioche, G., Bugliaro, L., Protat, A., Minikin, A., Wirth, M., Dörnbrack, A., Shcherbakov, V., Mayer, B., Garnier, A., and Gourbeyre, C.: On the observation of unusual high concentration of small chain-like aggregate ice crystals and large ice water contents near the top of a deep convective cloud during the CIRCLE-2 experiment, Atmospheric Chemistry and Physics, 12, 727–744, https://doi.org/10.5194/acp-12-727-2012, 2012.

Grzych, M.: Avoiding convective weather linked to ice-crystal icing engine events, Boeing Aeromagazine, 2010.

Grzych, M., Tritz, T., Mason, J., Bravin, M., and Sharpsten, A.: Studies of Cloud Characteristics Related to Jet Engine Ice Crystal Icing Utilizing Infrared Satellite Imagery, in: SAE Technical Paper Series, SAE Technical Paper Series, SAE International400 Commonwealth Drive, Warrendale, PA, United States, https://doi.org/10.4271/2015-01-2086, 2015.

Haggerty, J., Defer, E., de Laat, A., Bedka, K., Moisselin, J.-M., Potts, R., Delanoë, J., Parol, F., Grandin, A., and DiVito, S.: Detecting Clouds Associated with Jet Engine Ice Crystal Icing, Bulletin of the American Meteorological Society, 100, 31–40, https://doi.org/10.1175/BAMS-D-17-0252.1, 2019.

Haggerty, J. A., Rugg, A., Potts, R., Protat, A., Strapp, J. W., Ratvasky, T., Bedka, K., and Grandin, A.: Development of a Method to Detect High Ice Water Content Environments Using Machine Learning, Journal of Atmospheric and Oceanic Technology, 37, 641–663, https://doi.org/10.1175/JTECH-D-19-0179.1, 2020.

Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., and Bing, G.: Learning from class-imbalanced data: Review of methods and applications, Expert systems with applications, 73, 220–239, 2017.

https://doi.org/10.5194/egusphere-2025-2985
Preprint. Discussion started: 19 August 2025

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, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

James, G., Witten, D., Hastie, T., and Tibshirani, R.: An Introduction to Statistical Learning: with Applications in R, Springer Texts in Statistics, Springer New York, NY, second edn., ISBN 978-1-0716-1418-1, https://doi.org/10.1007/978-1-0716-1418-1, 2021.

Kalinka, F., Butter, M., Jurkat, T., de La Torre Castro, E., and Voigt, C.: A Simple Prototype to Forecast High Ice Water Content Using TAT Anomalies as Training Data, SAE Technical Papers, 2023 1, https://doi.org/10.4271/2023-01-1495, 2023.

Leroy, D., Fontaine, E., Schwarzenboeck, A., Strapp, J. W., Korolev, A., McFarquhar, G., Dupuy, R., Gourbeyre, C., Lilie, L., Protat, A., Delanoe, J., Dezitter, F., and Grandin, A.: Ice Crystal Sizes in High Ice Water Content Clouds. Part II: Statistics of Mass Diameter Percentiles in Tropical Convection Observed during the HAIC/HIWC Project, Journal of Atmospheric and Oceanic Technology, 34, 117–136, https://doi.org/10.1175/JTECH-D-15-0246.1, 2017.

Markowski, P. and Richardson, Y.: Mesoscale Meteorology in Midlatitudes, John Wiley & Sons, Incorporated, Newark, UNITED KINGDOM, ISBN 9780470682098, http://ebookcentral.proquest.com/lib/dlr-ebooks/detail.action?docID=514430, 2010.

Mayer, J., Ewald, F., Bugliaro, L., and Voigt, C.: Cloud Top Thermodynamic Phase from Synergistic Lidar-Radar Cloud Products from Polar Orbiting Satellites: Implications for Observations from Geostationary Satellites, Remote Sensing, 15, 1742, https://doi.org/10.3390/rs15071742, 2023.

Müller, R., Barleben, A., Haussler, S., and Jerg, M.: A Novel Approach for the Global Detection and Nowcasting of Deep Convection and Thunderstorms, Remote Sensing, 14, https://doi.org/10.3390/rs14143372, 2022.

Piontek, D., Bugliaro, L., Müller, R., Muser, L., and Jerg, M.: Multi-Channel Spectral Band Adjustment Factors for Thermal Infrared Measurements of Geostationary Passive Imagers, Remote Sensing, 15, 1247, https://doi.org/10.3390/rs15051247, 2023.

S. Ayra, E., Rodríguez Sanz, Á., Arnaldo Valdés, R., Gómez Comendador, F., and Cano, J.: Detection and warning of ice crystals clogging pitot probes from total air temperature anomalies, Aerospace Science and Technology, 102, 105 874, https://doi.org/10.1016/j.ast.2020.105874, 2020.

Schmetz, J., Pili, P., Tjemkes, S., Just, D., Kerkmann, J., Rota, S., and Ratier, A.: AN INTRODUCTION TO METEOSAT SECOND GENERATION (MSG), Bulletin of the American Meteorological Society, 83, 977 – 992, https://doi.org/10.1175/1520-0477(2002)083<0977:AITMSG>2.3.CO;2, 2002.

Stephens, G. L., Vane, D. G., Boain, R. J., Mace, G. G., Sassen, K., Wang, Z., Illingworth, A. J., O'connor, E. J., Rossow, W. B., Durden, S. L., Miller, S. D., Austin, R. T., Benedetti, A., Mitrescu, C., and the CloudSat science team: THE CLOUDSAT MISSION AND THE A-TRAIN: A New Dimension of Space-Based Observations of Clouds and Precipitation, Bulletin of the American Meteorological Society, 83, 1771 – 1790, https://doi.org/10.1175/BAMS-83-12-1771, 2002.

Strandgren, J., Bugliaro, L., Sehnke, F., and Schröder, L.: Cirrus cloud retrieval with MSG/SEVIRI using artificial neural networks, Atmospheric measurement techniques, 10, 3547–3573, https://doi.org/10.5194/amt-10-3547-2017, 2017a.

Strandgren, J., Fricker, J., and Bugliaro, L.: Characterisation of the artificial neural network CiPS for cirrus cloud remote sensing with MSG/SEVIRI, Atmospheric measurement techniques, 10, 4317–4339, https://doi.org/10.5194/amt-10-4317-2017, 2017b.

https://doi.org/10.5194/egusphere-2025-2985
Preprint. Discussion started: 19 August 2025

Winker, D. M., Pelon, J. R., and McCormick, M. P.: CALIPSO mission: spaceborne lidar for observation of aerosols and clouds, in: Lidar Remote Sensing for Industry and Environment Monitoring III, edited by Singh, U. N., Itabe, T., and Liu, Z., vol. 4893, pp. 1 – 11, International Society for Optics and Photonics, SPIE, https://doi.org/10.1117/12.466539, 2003.

Yost, C. R., Bedka, K. M., Minnis, P., Nguyen, L., Strapp, J. W., Palikonda, R., Khlopenkov, K., Spangenberg, D., Smith, W. L., Protat, A., and Delanoe, J.: A Prototype Method for Diagnosing High Ice Water Content Probability Using Satellite Imager Data, Atmospheric measurement techniques, 11, 1615–1637, https://doi.org/10.5194/amt-11-1615-2018, 2018.

Zinner, T., Mannstein, H., and Tafferner, A.: Cb-TRAM: Tracking and monitoring severe convection from onset over rapid development to mature phase using multi-channel Meteosat-8 SEVIRI data, Meteorology and Atmospheric Physics, 101, 191–210, https://doi.org/10.1007/s00703-008-0290-y, 2008.

Zinner, T., Forster, C., de Coning, E., and Betz, H.-D.: Validation of the Meteosat storm detection and nowcasting system Cb-TRAM with lightning network data – Europe and South Africa, Atmospheric measurement techniques, 6, 1567–1583, https://doi.org/10.5194/amt-6-1567-2013, 2013.