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

Summary statement – Anonymous reviewer 1

This paper describes a method for assessing the likelihood of ICI derived from geostationary satellite observations and derived products, coupled with cloud water content estimates derived from the CloudSat/CALIPSO-based DARDAR product. ICI and HIWC often occurs within deep convection, though it has also been observed within mid-latitude frontal cloud bands, and represents a significant hazard to aviation. Machine learning identified the most important metrics for diagnosing ICI that are combined to estimate a (daytime only) ICI/HIWC likelihood for the MSG SEVIRI imager, that has been validated with a subset of DARDAR not used in model training. Performance is fairly comparable to existing methods, though with slightly weaker validation stats. The product is then validated with DARDAR and compared with a small sample of European ICI events encountered by in-service Lufthansa aircraft.

The authors have a clear understanding of the ICI/HIWC hazard, existing satellite-based methods from the literature focused on diagnosing ICI/HIWC, machine learning best practices, and the most appropriate geostationary parameters for diagnosing this hazard. The paper is clear and well written. My concerns with the paper begin with the exclusive focus over Europe. Convection over Europe is relatively infrequent compared with Africa or the tropical Atlantic and typically weaker (with warmer tops) than these regions. Figure 3 shows several African/Atlantic Lufthansa ICI events that could have been studied, which if included would increase confidence in the method's global applicability. DARDAR likely viewed many intense storms over Africa as well that would serve as excellent training for the model. Another concern is the fact that there is no attempt to try to develop/validate a model that can operate at night. Aside from tau and visible reflectance, all other parameters are available at night. I would like to see a night-time product demonstration. Third, there was an international HIWC/HAIC field campaign based in Cayenne, French Guyana in 2015 that was within the SEVIRI field of view. IWC data was collected at 5 sec intervals from 2 aircraft which would be an extremely robust dataset for model validation. The authors should explore this data as it has been a number of years since collection and the data should be freely available by now. I have a number of other more minor comments/concerns listed below

Given that the paper and its writing are of high quality, but due to the significant concerns mentioned above, I say the paper is acceptable but with major revisions to address these concerns.

Specific Minor Comments/Concerns

Sections 2.1.1 and 2.1.2, the differences between CIPS and APICS, and the ramifications of these differences on the analysis are not very clear. I see both produce optical depth but you use the optical depth from one model for water cloud and the other for ice cloud. Additionally it is not explained why you are not using the cloud product data operationally generated by EUMETSAT which would make your method more easy to apply by others in the community.

Section 4.1, I don't think you need to use paper space to define very commonly used validation metrics. You could simply cite the Wilks meteorological statistics book and move one Wilks, D. S., 2006*: Statistical Methods in the Atmospheric Sciences*. 2nd ed. International Geophysics Series, Vol. 100, Academic Press, 648 pp.

Validation stats in general, it would be interesting to see the validation applied to > 1.0 g m-3 data in addition to > 0.5 as the higher value is likely to be more consequential for aircraft.

Figure 7 and many other mapped data figures (i.e. Figure 10), the mapped product is very hard to see details of. For Fig 7, I recommend you make the map much larger and place below the curtain plot data. For Fig 10, consider enlarging the graphics as I cannot see details when printed out on paper.

Figure A.3, there is an extremely odd look to the HIWC product with a discontinuity at 49.3 N latitude. What is the reason for this? Figure A.7 has an odd diagonal discontinuity too.

All Figures in Appendix, what is the purpose of plotting the wind information on the maps? It seems like an unnecessary detail that adds clutter to the map.

Comment 1+3: My concerns with the paper begin with the exclusive focus over Europe. Convection over Europe is relatively infrequent compared with Africa or the tropical Atlantic and typically weaker (with warmer tops) than these regions. Figure 3 shows several African/Atlantic Lufthansa ICI events that could have been studied, which if included would increase confidence in the method's global applicability.
Third, there was an international HIWC/HAIC field campaign based in Cayenne, French Guyana in 2015 that was within the SEVIRI field of view. IWC data was collected at 5 sec intervals from 2 aircraft which would be an extremely robust dataset for model validation. The authors should explore this data as it has been a number of years since collection and the data should be freely available by now.

Author reply:

The focus over Europe is due to the products that we have used as high ice water content predictors. Convection products (Cb-TRAM) are limited to the upper part of the SEVIRI HRV channel, which remains still over Europe and North Africa. For this reason, we could not readily extend our study to larger domain including Africa, and the Lufthansa ICI cases thereof. This is the same reason why we did not include HIWC/HAIC case studies as well. However, the algorithm can be extended to any similar product, if it provides a similar piece of information. This has been demonstrated in the added Section 4.4 combining the discussion about one ICI case outside Europe and the validation with in-situ measurements.

Manuscript changes:

Page 7, Line 190: *We focused on this region because the products that we have used as high ice water content predictors are limited to the upper part of the SEVIRI HRV channel, which remains still over Europe and North Africa.*

Addition of Section 4.4.

Page 23, Line 406: *Finally, the algorithm was validated with a case study of the HIWC-HIAC II flight campaign. However, Cb-TRAM is not available for the tropical regions covered by this campaign (see Sect. 2.1.3). Therefore, to cover this domain, alternative data are retrieved. Deep convective systems are provided by the TOOCAN database (Fiolleau and Roca, 2013, 2019). To prove the adaptability of the method to any equivalent product than the ones presented in the Sect. 2.1, cloud optical thickness was retrieved via the Optimal Cloud Analysis data record (EUMETSAT, 2022). The aforementioned data are displayed in Fig. 12.*

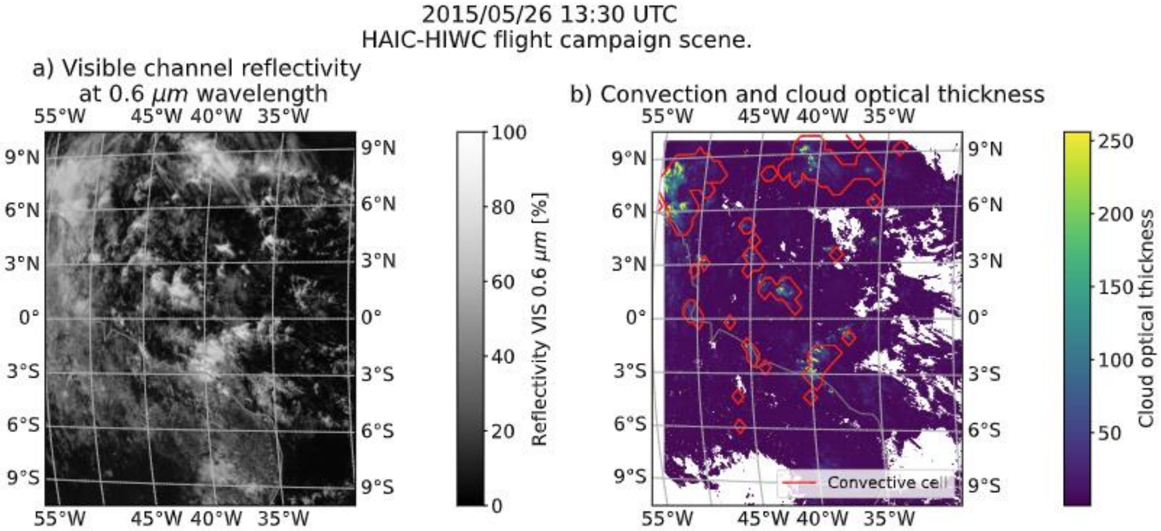

Figure 13. Satellite scene for the HAIC-HIWC II campaign case study. We considered flight 23, flying from French Guyana the 26^{th} May 2015. Panel a) shows the SEVIRI visible channel at 0.6 μm wavelength reflectivity. Panel b) depicts the cloud optical depth from Optimal Cloud Analysis (EUMETSAT, 2022) and deep convective cells from the TOOCAN database (Fiolleau and Roca, 2019).

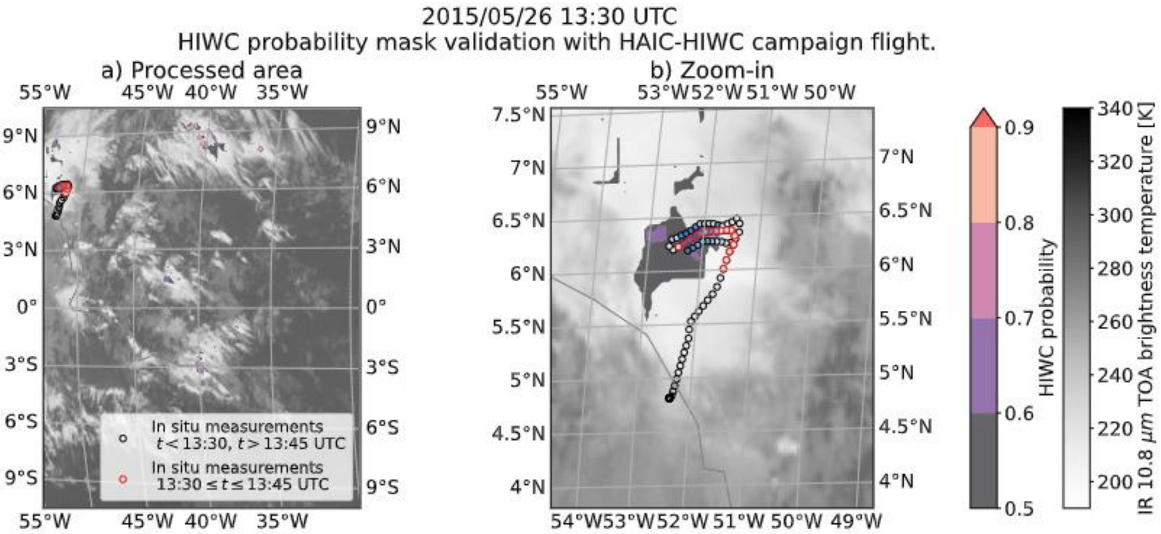

Figure 14. HIWC mask validation with flight 23 of the HAIC-HIWC II flight campaign. The flight crossed a HIWC region (IWC$\geq 0.5 g\cdot m^{-3}$) from 13:37 to 13:45 peaking at 13:41 with IWC=$3.3 g\cdot m^{-3}$. HIWC measurements are depicted in blue shades, while no-HIWC are shown in grey shades. Markers with red borders depict the flight position within the satellite scanning times. Flight waypoints are corrected for parallax effect, shifted to the corresponding satellite grid coordinates.

Figure 13 shows the corresponding computed HIWC mask. Although convection is widespread throughout the domain in panel b) of Fig. 12, the HIWC mask is relatively limited in extent in panel a) of Fig. 13. It features HIWC probabilities higher than 0.9 for convective cells around 40°W and 9°N, and 40°W and 3°S, while HIWC probabilities closer to the flight (52°W and 6°N) are relatively lower, peaking at 0.7. Panel b) shows a good agreement between the measured HIWC and the HIWC probability mask. IWC≥0.5g · m−3 sampled points mostly fall within the mask, whose values increase together with the measured IWC. The retrieval shows promising results even outside the domain where it was trained, and using input data equivalent to the ones discussed in Fig. 2.1. Given the results obtained with this case study, we speculate that only little calibration would be required to adapt the retrieval to input parameters coming from different data sources.

References:

Fiolleau, T. and Roca, R.: An Algorithm for the Detection and Tracking of Tropical Mesoscale Convective Systems Using Infrared Images From Geostationary Satellite, IEEE Transactions on Geoscience and Remote Sensing, 51, 4302–4315, https://doi.org/10.1109/TGRS.2012.2227762, 2013.

Fiolleau, T. and Roca, R.: TOOCAN – Tracking Of Organized Convection Algorithm using a 3-dimensional segmentatioN, https://doi.org/10.14768/20191112001.1, 2019.

EUMETSAT and European Organisation For The Exploitation Of Meteorological Satellites: Optimal Cloud Analysis Climate Data Record Release 1 - MSG - 0 degree, https://doi.org/10.15770/EUM_SEC_CLM_0049, 2022.

Comment 2: Another concern is the fact that there is no attempt to try to develop/validate a model that can operate at night. Aside from tau and visible reflectance, all other parameters are available at night. I would like to see a night-time product demonstration.

Author reply:

Although a dedicated nighttime product is out of the scope of the paper, it is a good point, and we have added a night-time product demonstration for few Lufthansa ICI cases.

Manuscript changes:

Section 4.3, and additional Lufthansa ICI case in the Appendix.
Page 23, Line 389: *The retrieval is here tested during nighttime. In this scenario, the random forest model does not have access to visible channel information and cloud optical thickness. Furthermore, it has been trained exclusively with day-time samples. Nevertheless, it can access infrared channels and convection related variables. In night-time mode, we decided to use instrumental values to fill the missing optical information required by the random forest approach.*

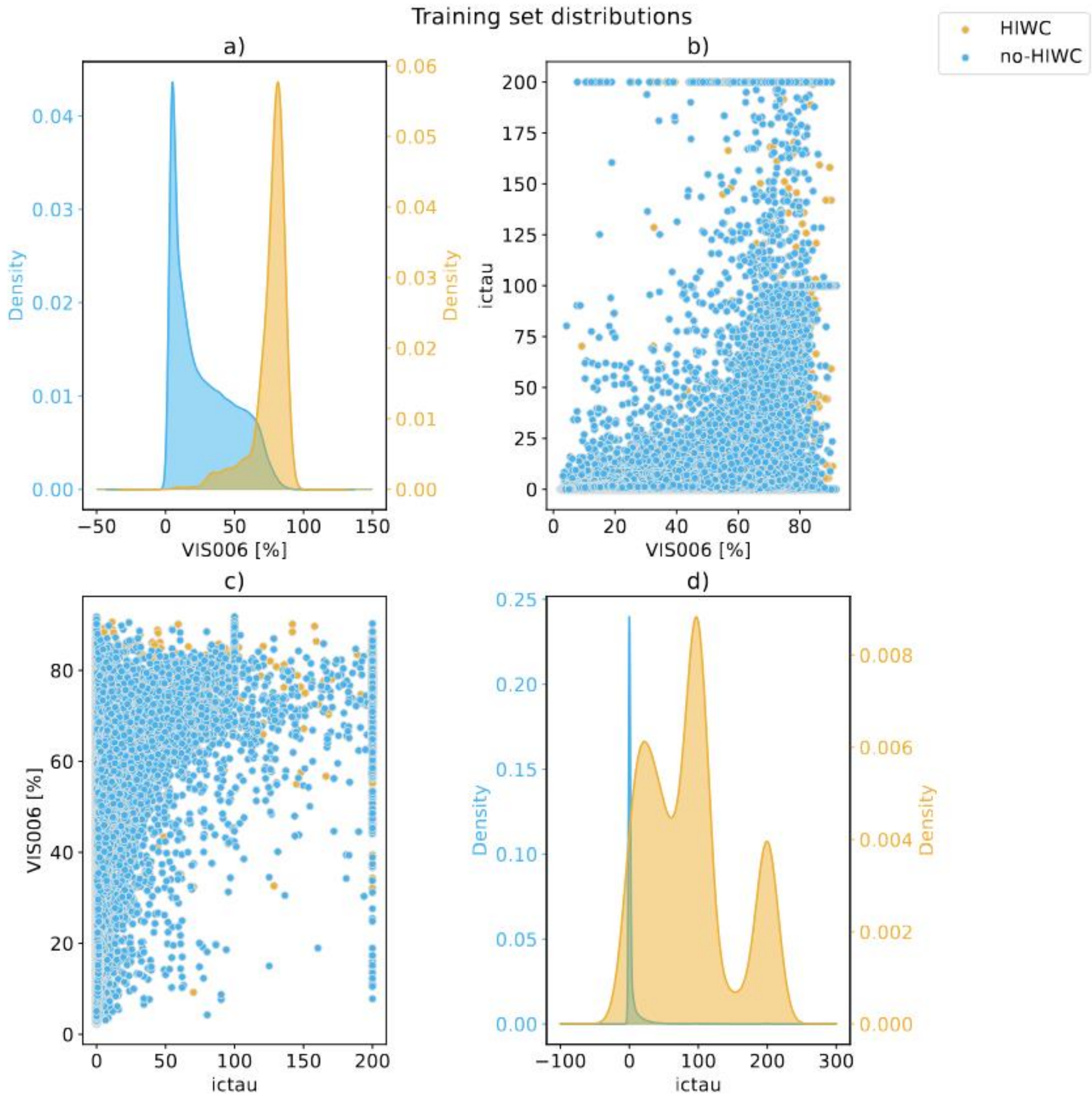

Figure 11. a) *VIS006* and d) *ictau* samples distributions in the training dataset, for HIWC (orange) and no-HIWC (blue). Negative values in both distributions arise because of the curve smoothing for plot purposes. The real sample locations are visible in panels b), and d).

In Fig. 11, the distribution in the training dataset of VIS006 and ictau for HIWC and no-HIWC is shown. These distributions allowed us to select a bias-free value with which we filled the missing information in nighttime mode. In particular, this bias-free value is selected such that it favours neither HIWC prediction, nor non-HIWC, i.e. the instrumental value should be in a range where HIWC and no-HIWC training samples distributions overlap. The values are set to VIS006= 80% and ictau= 50. The significance of this choice is shown in Fig. 12.

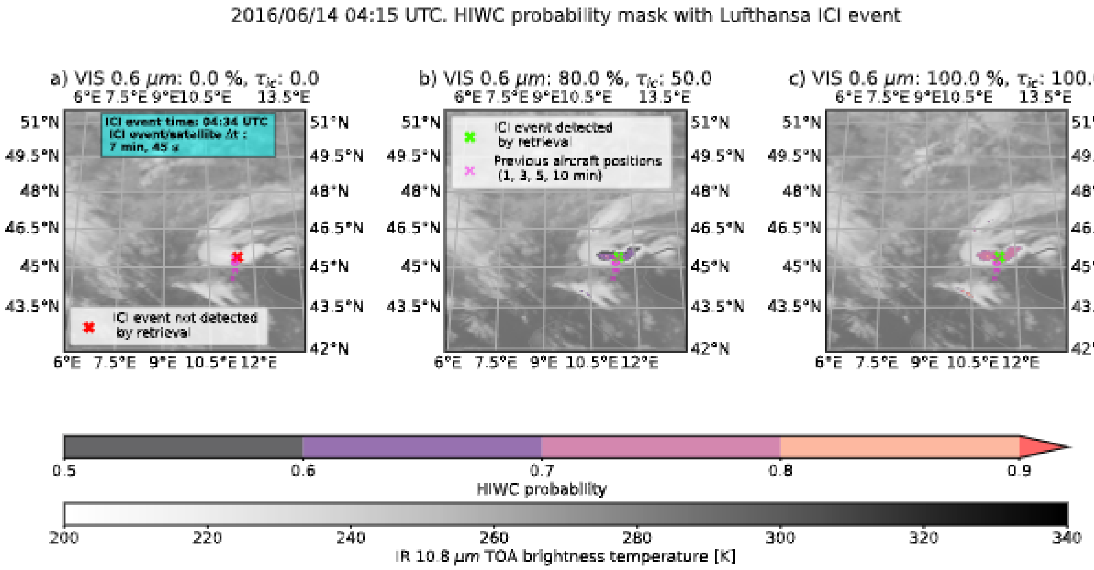

Figure 12. ICI retrieval nighttime mode demonstration example for a Lufthansa ICI event. Panel a) shows the HIWC mask setting instrumental values that favor no-HIWC prediction (see Fig. 11. Panel b) depicts the HIWC mask with bias free instrumental values. Panel c) shows the HIWC mask with instrumental values favoring HIWC predictions.

The mask in panel a), where we set VIS006= 0% and ictau= 0, is absent because the HIWC probability never exceeds 0.5. In panel b), the bias-free choice of VIS006= 80% and ictau= 50 leads to a smooth transition of HIWC probability between areas without detected HIWC and areas where HIWC is detected. Panel c) displays instead a sharp transition to high HIWC probabilities, as soon as this is detected by overcoming the 0.5 probability threshold. We observe that the constant instrumental values with which we fill missing information modulate the HIWC probability mask significantly. The choice made for panel b) is the best to achieve realistic results even with missing solar information. This demonstrates the good performance of the model even during nighttime.

Comment 4: Sections 2.1.1 and 2.1.2, the differences between CIPS and APICS, and the ramifications of these differences on the analysis are not very clear. I see both produce optical depth but you use the optical depth from one model for water cloud and the other for ice cloud. Additionally it is not explained why you are not using the cloud product data operationally generated by EUMETSAT which would make your method more easy to apply by others in the community.

Author reply:

CiPS is developed to detect thin cirrus cloud properties only, as it is a neural-network-based retrieval, and it is trained with ice clouds with optical thickness up to a value of 3.

Page 4, Line 106*: 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.*

APICS detects both ice and water cloud optical thickness with a rule-based approach considering visible and near infrared channels.

Page 4, Line 116: *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.*

CiPS and APICS thus analyze similar cloud optical and microphysical characteristics, but they perform best in different situations. For this work both were considered initially as input variables, but redundant variables have been eventually filtered out by the training process (Sect. 3.2)

Manuscript changes:

To address the reason why we did not use EUMETSAT products:
Page 4, Line 97: *The considered retrievals for this study are developed in-house, because of our expertise in their strengths and limitations and because of their availability to us. Nevertheless, in one example we have applied our algorithm using alternative products as input: optical thickness from EUMETSAT and convective cloud information from TOOCAN. This is demonstrated in Sect. 4.4.*
To address CiPS and APICS differences:
Page 5, Line 119*: CiPS and APICS thus analyze similar cloud optical and microphysical characteristics, but they perform best in different situations. CiPS is better suited for thin cirrus clouds analysis, both during day and nighttime. APICS has a wider scope, covering both ice and water clouds of any thickness, but it is limited to daytime due to its rule-based approach on visible and near-infrared channels. Both retrievals are used in this study because they may provide candidate precursors of high ice water content conditions. The suitability of these retrievals for this task has been discussed in Sect. 3.2.*

Comment 5: Section 4.1, I don't think you need to use paper space to define very commonly used validation metrics. You could simply cite the Wilks meteorological statistics book and move one Wilks, D. S., 2006: Statistical Methods in the Atmospheric Sciences. 2nd ed. International Geophysics Series, Vol. 100, Academic Press, 648 pp.

Author reply:
We understand the point and we moved the validation metrics definition in the Appendix

Manuscript changes:
Page 17, Line 309: *The statistical metrics chosen to assess the retrieval's performance are well established in the atmospheric science literature (Wilks, 2019). Their definitions can be found in Sect. A1.*

References:

Wilks, D. S.: Statistical Methods in the Atmospheric Sciences, Elsevier, fourth edn., ISBN 9780128158234, https://doi.org/10.1016/c2017-0-03921-6, 2019.

Comment 6: Validation stats in general, it would be interesting to see the validation applied to > 1.0 g m-3 data in addition to > 0.5 as the higher value is likely to be more consequential for aircraft.

Author reply:
To discriminate HIWC with low IWC, two values were mentioned in the literature:
- 0.5 g · m−3 used in (Yost et al., 2018; Bedka et al., 2020; J. A. Haggerty et al., 2020). This choice is mainly motivated by a lower sensitivity of passive remote sensing platforms to discriminate moderate to high IWC.
- 1.0 g · m−3 used in (de Laat et al., 2017). The choice is mainly motivated by historical reasons, where the community accepted this threshold because a necessary condition for in-service ICI is to have a high ice concentration.

For this reason, we opted for the 0.5 g · m−3 threshold.
However, a discussion about the 1.0 g · m−3 threshold is now added to improve the significance of the results.

Manuscript changes:
Performance evaluation for IWC > 1.0 g m-3 is included in Tab.3 with associated discussion:
Page 18, Line 331: *The model has been tested with HIW C = IW C ≥ 1.0g · m−3. The original version is used, trained with samples labeled as HIWC if IW C ≥ 0.5g · m−3 and adapted with a higher probability threshold of 0.7, to compensate the lower occurrence of HIWC when those are defined with the higher*

threshold of 1.0g · m−3. Table 3 shows that, in this case, FAR is reduced significantly, at the expenses of a decreased POD. CSI and AUC do not vary compared to the test settings consistent with training settings.

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.

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.

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

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.

Comment 7: Figure 7 and many other mapped data figures (i.e. Figure 10), the mapped product is very hard to see details of. For Fig 7, I recommend you make the map much larger and place below the curtain plot data. For Fig 10, consider enlarging the graphics as I cannot see details when printed out on paper.

Author reply:

Thanks for pointing that out and for providing a suggestion for a new layout. They have been modified in the manuscript.

Manuscript changes:

Figure 1, 7, 9, 10 were adjusted as suggested.

Comment 8: Figure A.3, there is an extremely odd look to the HIWC product with a discontinuity at 49.3 N latitude. What is the reason for this? Figure A.7 has an odd diagonal discontinuity too.

Author reply:

This is true and the reason of those discontinuities is now discussed in the Appendix.

Manuscript changes:

Addition of Section A3.

Page 29, Line 485: *Some HIWC probability masks display a discontinuity, as in Fig. A2, and A3. Those discontinuities may be explained with the convection related metrics. Those metrics, such as the distance to the closest convective cell and the areas of the closest convective cell, present such discontinuities, as in Fig. A6. Convective pixels in the surrounding radius of 100 km introduces rounded discontinuities, as in panel a), while distance and area extent of the closest convective cells introduce linear discontinuities, as in panel b) and c). Those discontinuities may be further* *emphasized by the random forest approach, which does not enforce smooth outputs, but only takes the majority vote from single decision trees. We speculate that the discontinuities might be more pronounced when the other supporting input features, such as visible channels and optical thickness, lie in a region where the split between HIWC and no-HIWC is clear (see Fig. 11). Thus, this artifact might be more pronounced during nighttime, though this evidence was not found in Lufthansa ICI cases in Fig. 12, and A5. However, this statement is supported by Fig. A7, where the nighttime demonstration approach (Fig. 12) was applied to a daytime scene (panel a) of Fig. 10). There, the rounded artifacts due to distance-related convection metrics are emphasized by the artificial unavailability of solar channels information that we introduced as demonstration.*

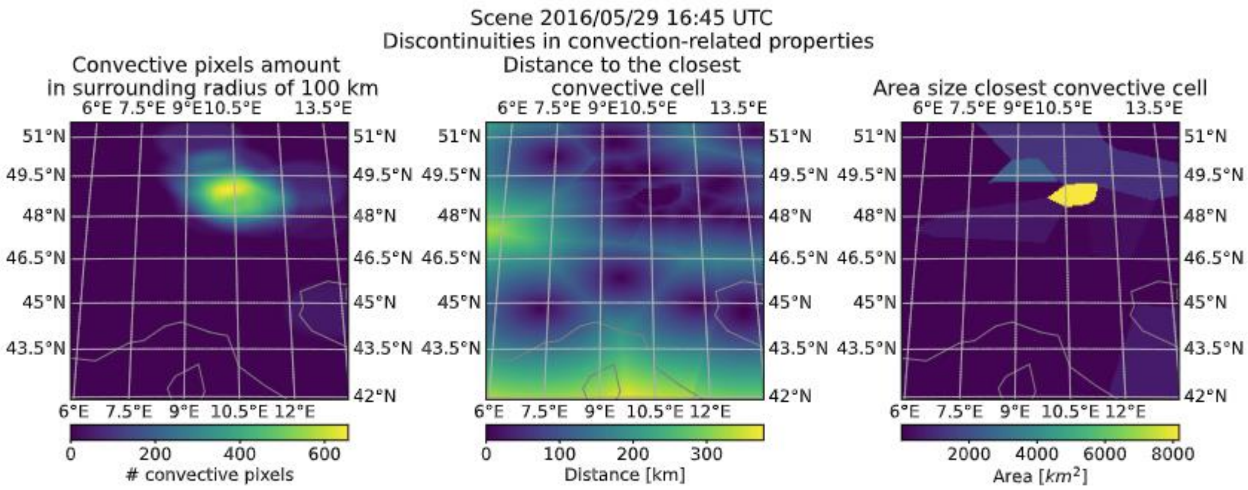

Figure A6. Convection related metrics for the Lufthansa ICI case of Fig. A2 associated with discontinuities.

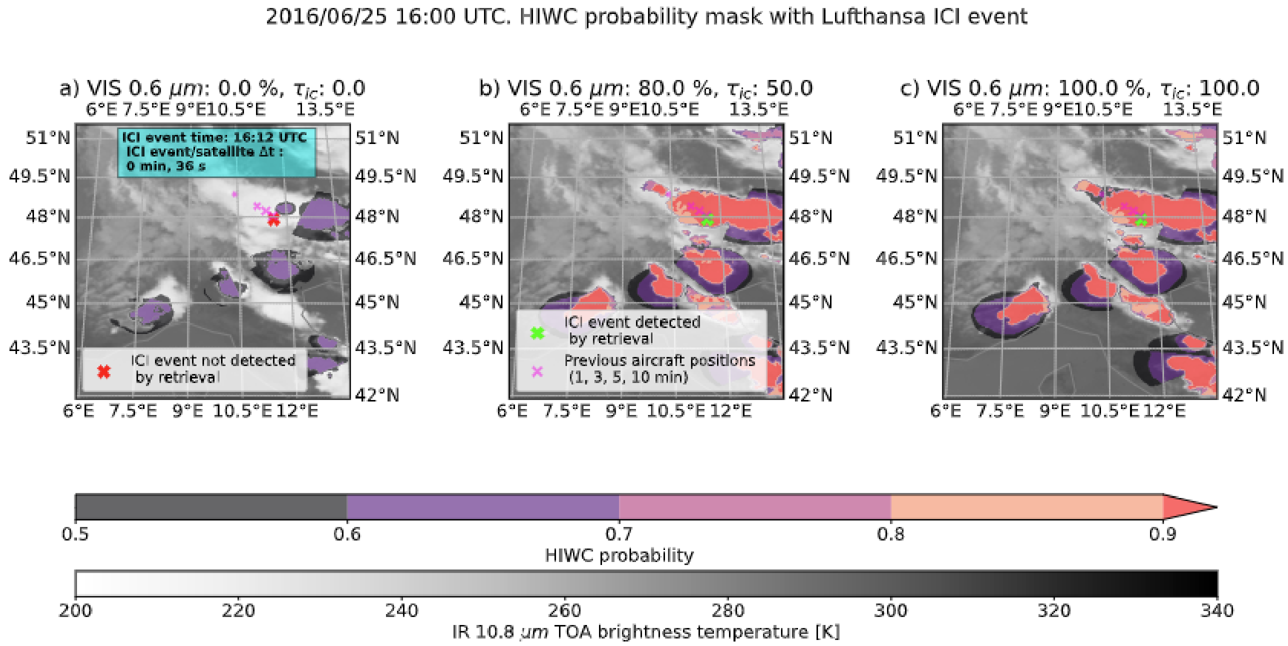

Figure A7. As Fig. 12, where the nighttime instrumental value approach was applied to a daytime scene, to verify its effect where cloud optical properties would be otherwise available.

Comment 9: All Figures in Appendix, what is the purpose of plotting the wind information on the maps? It seems like an unnecessary detail that adds clutter to the map.

Author reply:

We think that wind information is an important information to include because our ICI retrieval did not use it as input explicitly. We expect that ice particles are dispersed with according to the wind field, and one can see that the HIWC masks follow the wind field lines downstream of the most active convective cells, thus behaving physically without enforcing this information explicitly in the machine learning model. The discussion about its significance can be found in Sect. 4.2, Fig. 10.

Summary statement – Anonymous reviewer 2

In this paper, Aricò et al. present a method for detecting high ice water content (HIWC) conditions associated with aviation ice crystal icing (ICI) events using geostationary satellite data. They apply a machine learning approach to derive a set of input variables from Meteosat Second Generation (MSG) SEVIRI derived products. Training of a random forest is accomplished using CloudSat radar and lidar (DARDAR) products. Finally, the authors obtained a database comprised of in-service ICI events which they have used to asses performance of their detection method.

The paper is clearly written and well-organized. It begins with a comprehensive description of the ICI threat to aviation and a review of previously published HIWC detection methods. It provides thorough descriptions of the various data sets used. Explanations of the random forest approach, why it was chosen, and how it was applied to this problem are detailed and clear. Results presented are meaningful, and conclusions are well-justified. Significantly, the authors have demonstrated the need for a dedicated HIWC detection product by showing displacement between their HIWC Mask and detected convective cells. Overall, this is a very good paper that needs only minor revision.

Specific points:

1. If I understand correctly, the method is limited to cruise altitudes, but it’s not clear why. ICI engine events have occurred during ascents and descents, so the hazard is not limited to cruise altitudes.
2. While I understand why the method is limited to daytime, given its reliance on products derived from visible channels, I think it would be appropriate to at least discuss how you might develop a corresponding nighttime method.
3. It’s difficult to see the symbols in certain figures (e.g., Figs 1, 7, 9, 10). The images and symbols could be enlarged and/or the symbol color could have better contrast with the background image.
4. Verification against in-service ICI events is very important, but the Lufthansa database apparently only includes air data system (ADS) events (e.g., TAT anomalies), not engine events. I assume the authors did not have access to the latter. Some discussion of the relationship between ADS events and engine events would bolster the significance of your results.
5. In the box and whisker plots shown in Fig. 6, some of the variables on the vertical axis only have outlier points, i.e., no box and whiskers. Could you explain how this should be interpreted?

Comment 1: If I understand correctly, the method is limited to cruise altitudes, but it’s not clear why. ICI engine events have occurred during ascents and descents, so the hazard is not limited to cruise altitudes.

Author reply:

Although ICI events were observed during other flight phases than cruise, we tried to convey the message that geostationary satellites are not able to detect HIWC conditions in-clouds, thus making this scenario challenging to detect. This can be found in:
Page 9, Line 197: *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). Continues in Line 196: *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.*
Page 20, Line 347: *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.*

Manuscript changes:

We tried to emphasize further this concept:
Page 9, Line 201: *We speculate that this is due to passive sensors measuring emitted and reflected radiation in proximity to cloud tops, thus inherently limiting the in-cloud HIWC detection.*
Page 25, Line 424: *Cruise levels are considered because, even if ICI events are possible during the ascent and descent of an aircraft, passive remote sensing platforms are more sensitive to cloud tops and column-integrated quantities*.

Comment 2: While I understand why the method is limited to daytime, given its reliance on products derived from visible channels, I think it would be appropriate to at least discuss how you might develop a corresponding nighttime method.

Author reply:

It is a very good point, and we have added a night-time product demonstration for few Lufthansa ICI cases. Although a dedicated night-time product was not developed, it gave us the chance to test the current algorithm version during night-time.

Manuscript changes:

Section 4.3, and additional Lufthansa ICI case in the Appendix.
Page 23, Line 389: *The retrieval is here tested during nighttime. In this scenario, the random forest model does not have access to visible channel information and cloud optical thickness. Furthermore, it has been trained exclusively with day-time samples. Nevertheless, it can access infrared channels and convection related variables. In night-time mode, we decided to use instrumental values to fill the missing optical information required by the random forest approach.*

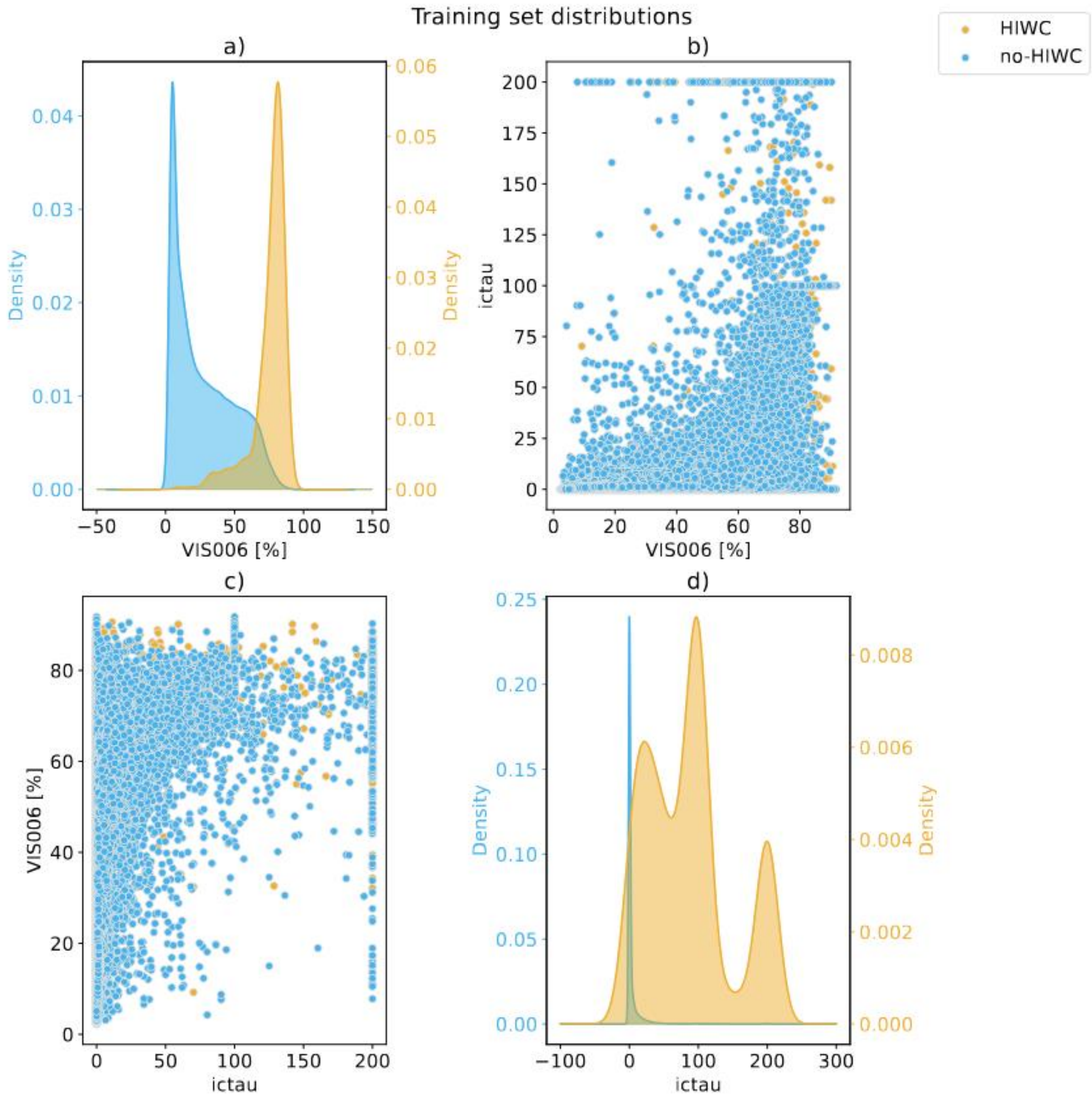

Figure 11. a) *VIS006* and d) *ictau* samples distributions in the training dataset, for HIWC (orange) and no-HIWC (blue). Negative values in both distributions arise because of the curve smoothing for plot purposes. The real sample locations are visible in panels b), and d).

In Fig. 11, the distribution in the training dataset of VIS006 and ictau for HIWC and no-HIWC is shown. These distributions allowed us to select a bias-free value with which we filled the missing information in nighttime mode. In particular, this bias-free value is selected such that it favors neither HIWC prediction, nor non-HIWC, i.e. the instrumental value should be in a range where HIWC and no-HIWC training samples distributions overlap. The values are set to VIS006= 80% and ictau= 50. The significance of this choice is shown in Fig. 12.

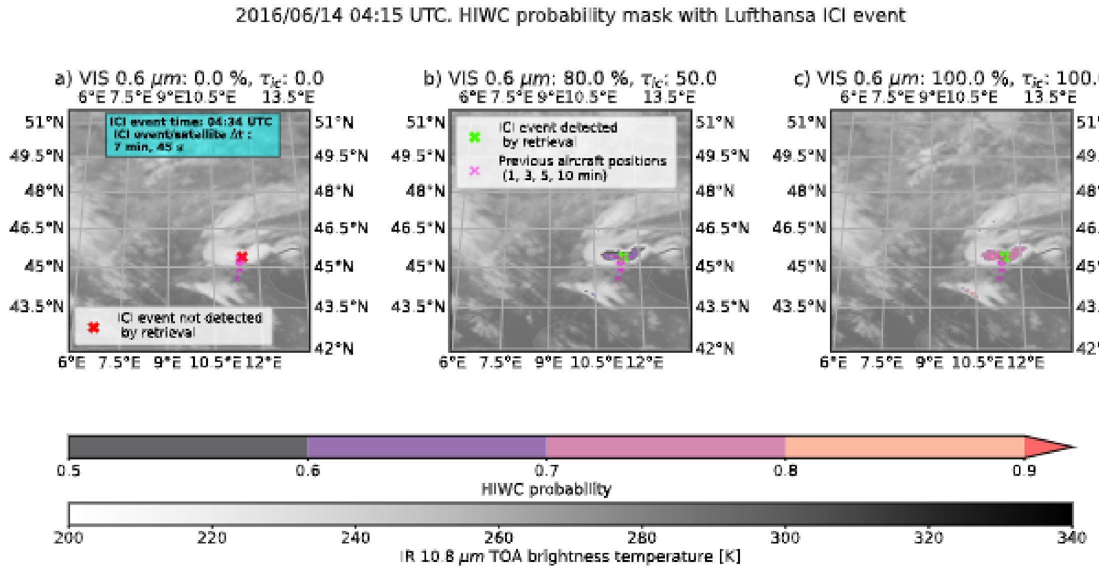

Figure 12. ICI retrieval nighttime mode demonstration example for a Lufthansa ICI event. Panel a) shows the HIWC mask setting instrumental values that favor no-HIWC prediction (see Fig. 11. Panel b) depicts the HIWC mask with bias free instrumental values. Panel c) shows the HIWC mask with instrumental values favoring HIWC predictions.

The mask in panel a), where we set VIS006= 0% and ictau= 0, is absent because the HIWC probability never exceeds 0.5. In panel b), the bias-free choice of VIS006= 80% and ictau= 50 leads to a smooth transition of HIWC probability between areas without detected HIWC and areas where HIWC is detected. Panel c) displays instead a sharp transition to high HIWC probabilities, as soon as this is detected by overcoming the 0.5 probability threshold. We observe that the constant instrumental values with which we fill missing information modulate the HIWC probability mask significantly. The choice made for panel b) is the best to achieve realistic results even with missing solar information. This demonstrates the good performance of the model even during nighttime.

Comment 3: It's difficult to see the symbols in certain figures (e.g., Figs 1, 7, 9, 10). The images and symbols could be enlarged and/or the symbol color could have better contrast with the background image.

Author reply:

Thanks for pointing that out and for providing a suggestion for a new layout. They have been modified in the manuscript accordingly.

Manuscript changes:

Figure 1, 7, 9, 10 were adjusted as suggested.

Comment 4: Verification against in-service ICI events is very important, but the Lufthansa database apparently only includes air data system (ADS) events (e.g., TAT anomalies), not engine events. I assume the authors did not have access to the latter. Some discussion of the relationship between ADS events and engine events would bolster the significance of your results.

Author reply:

Yes, that is correct, we do not have access to the engine events database. However, both TAT anomaly and engine failures can happen during the flight in HIWC regions. In conditions where TAT anomalies occur, the ice concentration is high, hence increasing the probability to have an engine failure. For this reason, TAT anomalies can be used as engine failure precursors. I do not have any information about the extent to which they correlate. We used TAT anomalies because those are the data available to us, and TAT anomalies are well correlated with HAIC regions.

Eventually the goal is to provide a prototype detection for those regions.
Explained in:
Page 2, Line 25: *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).*

Manuscript changes:

Added to emphasize the concept:
Page 2, Line 28: *Because those failures can happen in high ice concentration regions, on-board sensor anomalies, as for example the total air temperature (TAT) anomalies, are often used as precursors for engine failures (Haggerty, 2016; Rodríguez-Sanz et al., 2018).*

References:

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.

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. A.: High Ice Water Content and Airborne Temperature Measurement Anomalies in Tropical Convection, 32nd Conference on Environmental Information Processing Technologies, p. 1, 2016.

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.

Rodríguez-Sanz, Á., Arnaldo, R. M., Sánchez Ayra, E., and Gómez Comendador, F.: Detecting HAIC icing events from TAT anomalies vs8, 31st Congress of the International Council of the Aeronautical Sciences, 2018.

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.

Comment 5: In the box and whisker plots shown in Fig. 6, some of the variables on the vertical axis only have outlier points, i.e., no box and whiskers. Could you explain how this should be interpreted?

Author reply:

Thank you for bringing out this point, the explanation is now added.

Manuscript changes:

Page 16, Line 288: *This is denoted by the boxplot collapsing into a single line, which indicates that all the simulations carried out led to the same decrease in performance score, thus producing no distributions. Few outliers present for some variables, as for example D_A-1_2, D_A-1_3, IR_016, and ictau, indicates that only a minority of simulations led to a change in performance score.*