Harmonization and Verification of Three National European Icing Forecast Models Using Pilot Reports

Christoph Knigge,1 Katie Bennett,2 Christine Le Bot,3 Mara Gehlen-Zeller,1 and Svenja Koos1

1Deutscher Wetterdienst, Offenbach Am Main, Germany
2UK Met Office, Exeter, UK
3Méteo-France, Toulouse, France

Correspondence should be addressed to Christoph Knigge; christoph.knigge@dwd.de

Received 5 November 2021; Revised 22 March 2022; Accepted 20 April 2022; Published 26 May 2022

1. Introduction

The European airspace has a strongly fragmented character. At the continental scale, there are multiple air navigation service providers each responsible for a portion of the European airspace. Furthermore, air traffic control in Europe is very complex because the airspace is one of the busiest in the world with over 30,000 flights daily on average (2018, before COVID-19 pandemic) and a very high airport density. In the year 2000, the Single European Sky initiative was created to structure airspace and air navigation services in Europe to overcome the fragmentation and increase flight capacity and air traffic management efficiency. The Single European Sky Air Traffic Management Research (SESAR) program aims at modernizing and harmonizing the European airspace through the development and implementation of a new set of operational procedures and systems among European aviation stakeholders [1, 2].

In order to achieve the goals of SESAR, aeronautical and meteorological information need to be harmonized and consistent over Europe. Several national meteorological service providers operate their own Numerical Weather Prediction (NWP) systems on various grid scales (global to regional) and time scales resulting in several weather forecasts available over Europe at any given time. Additionally, harmonized global aviation weather forecasts provided by the World Area Forecast Centre’s are available. However, for the highly fragmented and high-capacity European airspace, there is a need for more detailed weather forecast information (high-resolution aeronautical
meteological (MET) information for short-distance flights, terminal areas, and airports. The SESAR Deployment project [3] 2015_068_AF5 "European Harmonised Forecasts of Adverse Weather (Icing, Turbulence, Convection and Winter Weather)" aims at fulfilling this need by providing European harmonized MET products of aviation weather hazards in high resolution and of high accuracy.

Alongside turbulence and convection, in-flight icing is seen as one of the most important meteorological phenomena, which may lead to hazardous flight conditions for aircraft. In-flight icing typically occurs in clouds with supercooled water droplets. The temperature within these clouds is usually between 0 and -20°C [4] and can be, depending on the surrounding conditions (such as the absence of ice nuclei), as cold as -40°C [5]. An aircraft flying through such a cloud may collect some of the supercooled droplets on its wings, propellers, or turbines, measuring instruments and other structures of the aircraft. The droplets freeze instantaneously in contact with the aircraft (large droplets may remain in a semi-liquid state for up to one second [5]) and can accrete to form an ice layer of up to several centimeters [6]. These ice layers can have a strong impact on the maneuverability of the aircraft and may lead to hazardous flight conditions in such a way that measuring instruments and the aircraft aerodynamics are compromised. In the latter case, icing causes an increased drag and a decreased lift, which results in control problems [7]. Green [8] gives a long-term overview (1978–2002) of accidents and incidents related to in-flight icing. Further descriptions of some individual incidents are discussed by Bernstein et al. [9].

Due to the hazardous nature of in-flight icing, aircraft are equipped with anti-icing and de-icing systems. The two most commonly used systems are heated surfaces and pneumatic rubber boots on the leading edge of the wings and at other locations where ice accretion may lead to flight problems. However, their utilization leads to a significant increase in fuel consumption, and therefore, avoidance of icing conditions is preferred. Consequently, there is a demand for more reliable icing forecasts to reduce costs as well as increase flight safety.

In this study, a harmonized icing forecast over Europe is produced using forecast data from Deutscher Wetterdienst (DWD), Météo-France (MF), and Met Office (MO). The three icing forecast products, each based upon a different NWP model, are ADWICE from DWD [5] (postprocessing based on model ICON-EU [10]), ARPEGE from MF [11], and Unified Model (UM) from MO [12]. The harmonized product, hereafter called by its working title “HarmonICE,” is produced by combining the different icing indices created by each of the individual NWP models.

The primary goal of this study is to analyze different methods combining the icing indices, such as the minimum, maximum, or mean of all data and, as a more complex approach, the method suggested by Pepe and Thompson [13] described in detail below. The approach with the highest forecast skill will be deployed operationally to provide a harmonized European icing forecast with a high accuracy and consistency across national borders and among airspace users and aviation stakeholders.

Pilot reports (PIREPs) are used to verify the forecasts as supported by previous studies, for example, Brown et al. [14] who describe in detail the advantages and disadvantages of using PIREPs for NWP verification. PIREPs provide information about hazardous weather and are currently the only direct regular measure of in-flight icing events. An overview of the PIREPs used in this study and the methods for the combination of the individual icing forecasts are described in the next section, followed by a detailed description of the results and an outlook.

2. Data and Methods

PIREPs from the winter season 2017/18 (October 2017 to April 2018) are used in order to determine the best approach for combining the three icing forecasts based on single NWP models and to verify the final harmonized icing forecast, HarmonICE. Icing events over Europe typically occur more frequently during this period of the year than in other months. However, compared to the large amount of forecast data available in the time period, there are few observations of the icing phenomenon.

2.1. Observation Data. An icing event reported by a PIREP provides a point observation of the existence and severity of icing (categorized as none, trace, light, moderate, or severe) at the location of the aircraft. Alongside the five-level icing intensity, the geographic location, altitude, and observation time are reported. PIREPs have been used as icing observation data in several other studies such as Brown et al. [14], Taffner et al. [15], and Kalinka et al [5].

In total, 4268 PIREPs were collected during the winter season. The geographical distribution of “no icing” and “icing” events is shown in Figure 1. The majority of the positive icing events were reported over land in the middle of Europe and close to large cities. Aircraft sending these PIREPs were often located at middle flight levels during approach or landing. This can be seen in the altitude distribution of the PIREPs in Figure 2. In contrast, negative reports were mostly sent from over the sea in the western part of the analyzed area (Figure 1) and from higher flight levels (Figure 2). As a result, a large number of ”no icing” reports are at heights that are above the maximum forecast level (FL 360) where supercooled liquid water rarely exists due to very low temperatures.

The frequency distributions of the PIREPs separated into four icing intensities can be seen in Figure 3. Trace icing observations are included in the light icing intensity category as there are few trace icing reports compared to light icing reports. By far, most PIREPs report moderate icing. A relatively small number of negative reports are available (662) compared to the number of positive icing events (3606). This distribution can be explained by the fact that pilots are expected to avoid areas with severe icing intensities and that light icing intensities may not give rise to a level of
concern for the pilot to consider making a report. Furthermore, icing intensity is subjective since no quantitative definition of severity exists. A detailed discussion about the low quality and quantity of PIREPs is given by Brown et al. [14]. They conclude that PIREPs are challenging to use in an absolute sense for NWP verification. However, in a relative sense, they can be useful, for example, for comparing the detection rates of different icing forecasts or the combination of them as presented in this study.

2.2. Forecast Data. The three icing forecasts used for the harmonized product from DWD, MF, and MO are based on differed NWP systems. ADWICE from DWD provides hourly data output with a horizontal grid spacing of 0.0625° and 32 vertical layers up to 225 hPa. The icing intensities provided by ADWICE match the four categories used for the PIREP values in the study. The icing index based on ARPEGE from MF has a lower horizontal grid resolution of 0.1° and 14 layers up to 400 hPa vertically. The data output is hourly up to forecast hour 12 and then every three hours with an icing intensity range between 1 and 10 in steps of one. The UM icing index from MO provides hourly data with 0.140626° horizontal grid spacing in east-west direction and 0.09375° in north-south direction with 16 vertical grid layers extending up to 250 hPa. The icing intensity is continuous ranging between 0 and 1. All models are run every six hours (0, 6, 12, and 18 UTC) and simulate up to 36 hours or longer.

The different grid spacings and time resolutions have to be considered in the final harmonized product. The approach used for HarmonICE uses the finest of the aforementioned horizontal grid resolutions of 0.0625° and 29 vertical grid layers (up to 225 hPa or flight level 360). This is done so that in the final product, there is no loss of icing intensity information resulting from extrapolation to a

Figure 1: Geographical distribution of the PIREPs used in this study. The area shown represents the geographical coverage of HarmonICE.

Figure 2: Distribution of the PIREPs by altitude (left: positive icing, right: negative icing). Note that different ranges on the abscissa are shown due to the relatively small number of no icing reports. One report may be counted multiple times because the report identifies multiple levels.
coarser grid. Furthermore, customers who currently use the finer resolved single product could have reservations against the harmonized product if it contained less information. Any reduction of information in the vertical dimension would result in less flight levels available in the forecasts, reducing the icing intensity information available when compared to existing single model forecasts, which is undesirable for customers. This may give the impression that the harmonized product is not as good as the higher resolved single product and result in reluctance to use HarmonICE.

Data output of HarmonICE is hourly up to forecast hour 36. The final output is the categorical icing intensity forecast utilizing the categories recommended in ICAO Annex 3 [16], which are the same categories used for PIREPs. The domain of HarmonICE goes from 23.5° West to 62.5° East and from 29.5° to 70.5° North (see Figure 1).

In order to combine the three different centers’ forecasts, each individual icing forecast is re-gridded to the HarmonICE grid using linear interpolation. Time interpolation is performed with the ARPEGE data, and the icing indices of ARPEGE and UM are transformed to the four-step categorical range of HarmonICE by applying the thresholds listed in Table 1. This transformation is necessary because unlike ADWICE, which produces categorical icing intensity forecasts, there are no internal calculations to a categorical icing intensity within ARPEGE and UM. The thresholds listed in Table 1 were calculated during an earlier phase of the project using PIREPs from another winter season. They are currently fixed at the given values but will be verified and if necessary recalculated in future.

Due to storage problems, some gaps in the forecast data exist for the analyzed winter season 2017/18. All three datasets cover approx. 70% of the complete winter time period so not all PIREPs gathered could be used for comparison with the forecast data. In total, 3010 PIREP forecast data comparisons could be performed.

| Table 1: Thresholds used for the transformation of the original icing index potential based on ARPEGE and UM to the four-step range used in HarmonICE. |
| HarmonICE | ARPEGE | UM |
| No = 0    | 0      | 0–0.59 |
| Light = 1 | 0.1–3.9| 0.6–0.79 |
| Moderate = 2 | 4.0–7.9 | 0.8–0.89 |
| Severe = 3 | 8.0–10.0 | 0.9–1.0 |

2.3. Analyzing Methods. This section describes the data processing required to enable evaluation and verification of the forecasts. First, the forecast data corresponding to each PIREP are extracted from the four-dimensional data files. Second, the individual icing forecasts are combined in order to produce the harmonized icing intensity.

2.4. Pairing Data. For each PIREP of the examined winter season, the position, observation time, and icing intensity were compared with the forecast data. The point in time for each PIREP was shifted to the nearest full hour so that a maximum time difference between an icing forecast validity time and the PIREP time was 30 minutes. Within this half-hour window, the detected icing area (cloud) could be advected horizontally some distance according to the mean wind at this altitude leading to some spatial deviation in the location of the icing forecast. Assuming typical mean wind speed of around 15 m/s in a middle altitude of the troposphere, the advection would lead to a distance of around 30 km between the measuring point and the potential new position of the measured cloud half an hour earlier or later. Another source of uncertainty lies in the reported information in the PIREP (e.g., rounding errors of the longitude and latitude values or a time offset during the observation resulting from aircraft travel). In order to account for these uncertainties in time and space, a box was placed around each PIREP and the maximum forecast icing intensity in the box was used to compare to the observed icing intensity (similar to other studies, e.g., Kalinka et al. [5]). The sensitivity of the results to the size of such a box was tested by using four different box sizes. The smallest box is simply a single grid length in all dimensions, which results in a point-to-point comparison. All other boxes are increased vertically by one grid point up and down (totaling three grid points in height), and by one, two, and three grid points, respectively, in all four horizontal directions. Thus, the largest box has dimensions $3 \times 7 \times 7$, totaling 147 grid points, that is, three vertical grid points and seven horizontal grid points in North-South and East-West directions (denoted as 3v7h), corresponding to approximately 600 m vertically and 50 km in both horizontal directions.

2.5. Merging Data. For merging the individual icing forecasts, different approaches were tested. Besides the classical method to calculate the mean of the three forecasts, the combination of the maximum and minimum was calculated. Additionally, the method proposed by Pepe and Thompson [13] is used. This method combines the forecast values $p_1,$
p2, p3 by choosing a weighting factor $\lambda \in [-1,1]$ in such a way that the combinations $\lambda p_i + p_j$ (with $i, j \in \{1,2,3\}$) result in an optimal value for the chosen skill score for each pair of possible combinations ($p_i, p_j$). The pair with the highest resulting skill score is then combined into a new intermediate combined icing forecast value $p_{ij}$. This process is then repeated for the remaining pair of icing forecasts $p_k, p_{ij}$ (with $k \in \{1,2,3\}$) in order to calculate a second weighting factor for the final product, which has the following form: $p_{ijk} = p_i + \lambda_1 p_j + \lambda_2 p_k$.

2.6. Evaluating Data. Former studies often used a binary yes/no icing analyzing method, for example, Brown et al. [14] and Kalinka et al. [5]. A contingency table enables the calculation of POD (probability of detection) for yes and no icing as well as calculation of the FAR (false alarm rate). The AUC (area under the curve) can also be calculated. These skill scores provide an overview of the forecast quality and enable comparisons of different NWP model performance to be made. However, this binary method distinguishes only between an icing and a no icing event with no consideration to the intensity. For this reason and the fact that only a small number of negative icing reports are available (they are needed for the FAR calculation), a more complex multcategory contingency table is used to analyze the data (e.g., Murphy [17], Brooks [18]). A detailed description of the method applied in this study for calculating skill scores from the multcategory contingency tables is given in the next section alongside an example.

3. Results and Discussion

In this section, the results are presented for combining individual in-flight icing forecasts using simple and more sophisticated methods. The skill of each method for combining the individual in-flight icing forecasts as well as of each individual in-flight icing forecast is shown. Based on these results, the recommended approach for combining the in-flight icing forecasts for HarmonICE is presented.

3.1. Example of a Combination of Individual Forecasts. One of the simplest ways of combining the icing forecasts is by calculating the average icing intensity (mean method). An example of such an icing intensity field is shown in Figure 4. A pressure level of 600 hPa on 11 December 2017 at 6 UTC (model run at 00 UTC) is chosen. At this time, the single forecasts show some significant differences in the icing intensities (ADWICE has only very small areas of severe
icing) and in the areas where icing occurs (ARPEGE has no icing in the northeastern and northwestern part of the domain). However, the position and spatial extension of the frontal zone agree well between all models. In the HarmonICE field, the mean method is shown to remove icing areas with light intensities predicted by only one of the three forecast products and typically reduces the higher icing intensity areas since the highest intensity predicted by all products appears only in few areas.

The two PIREPs available at this flight level and time, shown as colored dots in Figure 4, agree well with the forecast icing intensities of HarmonICE. In contrast, not all of the individual forecasts are a good fit to the PIREPs, further highlighting the benefit of intelligently combining the individual in-flight icing forecasts into HarmonICE. Analysis of all data during the winter season for this mean-method case is shown in Figure 5. The multicategory contingency table belongs to the point comparison (smallest model box size around each PIREP) and can be read as follows: on the green diagonal, the forecast values agree perfectly with the observed data. An optimal forecast would generate nonzero values only along this diagonal; all other entries would be zero. The blue cells show under-forecast cases where the predicted icing intensity is smaller than the corresponding PIREP icing intensity. The red cells show the over-forecast cases. The darker the red and blue colors, the larger the differences between the forecast icing intensities and each corresponding PIREP. The percentage of the three categories (perfect forecast, under-forecast, and over-forecast) is mentioned below the table. The gray cells give the sum of the corresponding rows and columns. As mentioned previously, 3010 PIREPs were compared with a forecast value. This is less than the total amount of PIREPs available in the winter season (see section “Data and Methods”) since not all forecasts were not available over the entire time period and some PIREPs were reported from an altitude above the model domain height.

In this example of averaging the individual model forecasts, the results show a tendency to underestimate the icing intensities. Nearly half of the compared values of the mean-method data are smaller than the corresponding PIREP. A perfect forecast, where the predicted icing intensity and the corresponding observed icing intensity are identical, is observed in 41.7% of all cases.

The multicategory contingency table provides a variety of different information about the analyzed forecast product. However, it does not provide one single skill score, which would allow an overall evaluation of the product and a simple comparison with multicategory contingency tables derived from other forecast products. In order to create a single skill score to use to compare the individual icing forecasts, the PODs of each icing intensity, as well as the total hit rate, the FARs of each icing intensity as well as the total false-forecast rate and under- and over-forecasting rates were calculated. For the latter four values, the total amount of observed PIREPs (3010) was used as the divisor. To calculate the icing intensity hit rates (no, light, moderate, and severe), the number of correctly forecast icing intensities (green values on the diagonal of the table in Figure 5) was divided by the amount of PIREPs with the corresponding icing intensity (gray values in the last line of the table in Figure 5). For example, the hit rate of no icing was calculated by dividing 236 by 247, which results in 0.96. Figure 6 shows the above-listed skill scores calculated from the multicategory contingency table (Figure 5). The very high hit rate for no icing reports (0.96) highlights the problems mentioned in section “Observation Data” section with PIREPs of no icing. Most of the no icing PIREPs are at altitudes where no icing clouds occur and thus these statistics look very good. The reason for taking both the single hit rates of each icing intensity (hit no, hit light, hit moderate, hit severe) and the total hit rate (hit all) into account is that in the total hit rate, each value has the same influence on the probability, while the single hit rates are differently influenced by one value because their total numbers are different.

In order to determine one single skill score from the values shown in Figure 6, they were merged into three different groups of skill scores: the hit rates for the single icing intensities (hit no, hit light, hit moderate, and hit severe), the total hit rate, which is equal to 1-false all and the under- and over-forecast rates. These three groups of scores contain all relevant information in the multicategory contingency table. By averaging the single scores within the groups and then averaging the three groups, one multicategory average skill score (MCASS) is obtained, which contains all relevant characteristics of the multicategory contingency table. The calculation described above can be mathematically formulated as

\[
MCASS = \frac{\sum_{i=1}^{4} hits_i}{4} + hits_{all} + \left( 1 - \frac{underf. + overf.}{2} \right),
\]

(1)

with the first term representing the average of the single hits, the second term representing all hits, and the third term giving consideration to under- and over-forecast events. The under-forecast and over-forecast values are subtracted from one in order to get a value that is best when it is equal to 1. The authors are aware that there are other scores or combinations of scores that can be derived from the multicategory contingency table such as the Heidke skill score (HSS) [19]. However, the MCASS (equation (1)) provides a good and sufficient indication of the icing forecast skills investigated in this study. This was also manually checked by
comparing a set of calculated MCASS with the corresponding multicategory contingency tables.

One additional advantage to the formulation of the MCASS over skill scores such as the HSS is that it can be modified by different weighting factors for each term in equation (1) in order to consider certain single terms more than others. This might be the case if customer requirements are modified in a way that, for example, over-forecasting must have a stronger impact than under-forecasting or the hit rate of severe must be considered stronger than the other hit rates. In the current study, however, the single terms of equation (1) are weighted homogeneously. More detailed analysis of the differences as well as the advantages and disadvantages of the MCASS in comparison with other skill scores will be investigated in future studies.

The MCASS was computed for each individual icing forecast and each individual combination method discussed in the “Merging Data” section (results for the mean method presented in Figures 5 and 6).

3.2. Overview of Various Combination Methods. Following the example for the multicategory contingency table in the previous section, the performance of each individual icing forecast and combination method is now assessed. A final value of MCASS is shown for the recommended combination method for producing HarmonICE.

For the Pepe and Thompson (PT) method, the best values for $\lambda_1$ and $\lambda_2$ were iteratively calculated as 0.6 and -0.2, respectively. The final optimized combination results in ADWICE + $\lambda_1$UM + $\lambda_2$ARPEGE. In Figure 7, all results are summarized for the three forecast products as well as for the four combinations (min, mean, max, and PT). The box sizes used around the PIREPs are indicated as different colors of the bars. Figure 7 shows that when considering the individual forecasts, the box size only has a significant influence on ADWICE. The larger the box size, the better the results indicated by the MCASS. This can be explained by the relatively high grid resolution of ADWICE compared to the other two forecasts. The higher grid resolution leads to higher spatial fluctuations of the icing intensity within small areas, and hence, larger icing intensity values are more likely to be encountered by increasing the box size. Furthermore, ADWICE tends to underestimate the icing intensity more than the other models for the smallest box size (not directly shown here but reflected by the relatively small blue bar of ADWICE in Figure 7), which also leads to better results for an increasing box size. The box size dependency of ADWICE can also be seen in the combined PT method since ADWICE is included with the largest factor of one. Overall, the PT method leads to the best results of all forecast combinations and is better than each individual forecast product. This statement is independent of the box size around the PIREPs. For each box size (bar color), the PT method provides the largest MCASS value.

In order to give a better insight into the improvement of HarmonICE provided by the PT method, Figure 8 shows the multicategory contingency table, which belongs to the largest bar in Figure 7 box 3v7h of PT method red bar). The table in Figure 8 can be compared with that of Figure 5, which was obtained by using the mean method with the smallest box size (box 1v1h of the mean method in Figure 7—blue bar).

The MCASS difference between these two examples is approximately 0.12. This is less than the difference between the PT method and most of the individual forecasts. The perfect forecast using the PT method is 63.1%, more than 20 percentage points larger than with the mean method (41.3%,
see Figure 5), and the under-forecasting is reduced to 29.6%. The main reason for the higher skill using the PT method is that a large number of the light icing forecasts in the other methods are increased to moderate icing. This leads to a high hit rate of moderate icing, which is the group with the most icing reports and consequently has a large impact on the results. Despite this increase, the hit rate of the no icing events is still very good with the PT method (234 vs. 236 with the mean method). This shows that the influence on the hit rate of no icing reports is small.

4. Conclusion and Outlook

In this study, a method for harmonizing in-flight icing forecasts over Europe is introduced. Different approaches were tested in order to find the best combination of three in-flight icing forecast indices based on three different NWP models. PIREPs are used as verification data, although they have some disadvantages for this purpose. An example on a single day shows the differences in the individual forecast products and the combined forecast product using the mean method. The PIREPs reported during the single day event fit reasonably well to the combined data set.

In order to be able to compare the different combination methods, the MCASS was derived from each multicategory contingency table generated for various forecast combinations and the individual forecast products. The MCASS contains all relevant information given by the multicategory contingency table, that is, the hit rates of the four icing intensities, over- and under-forecasting, and the total hit rate. It allows quantitative comparisons of the forecast skill from each icing forecast product as well as from the combination methods presented.

The final comparison of the MCASS of the different forecast combinations and the individual forecast products shows that the PT method provides the best results including showing an improvement compared to each single icing intensity forecast. The sensitivity study of box sizes around the PIREP positions in which the forecast data were extracted shows that the results are independent of the chosen box size; however, it does have an influence on the McASS. In most cases, an increased box size leads to a larger MCASS, but for certain forecast products or combinations, the results were unaffected by box size or showed a decrease in MCASS.

Within the SESAR Deployment project, initial tests of the preoperational version of HarmonICE have run since the middle of 2019. Since 2020, the data have been provided as a six-hourly icing forecast over Europe. Verification to inform the revision of the best forecast combination method—and if
necessary, a new calculation of the weighting factors—is and will be performed annually with the PIREPs from the previous winter season. Within that recalculation, it will also be possible to adapt MCASS to user requests. The calculation method allows for a different weighting of the single skill score terms such as over-forecasting or a certain intensity hit rate (e.g., severe icing intensity). If one of these single skill score terms must be considered more strongly than the others, it can be implemented in the average skill score calculation.

PIREPs are used for icing verification since better alternatives have previously been unavailable. One alternative for determining observed icing intensities, or at least whether icing occurs or not, is the use of a satellite inferred icing potential [20]. Stretton et al. [21] showed that satellite data can be used to detect the “no icing” regions in cloud-free areas. Furthermore, cloud top heights and the corresponding icing potential at this height level can be derived from the satellite. Overall, satellite data cannot provide a three-dimensional dataset of icing intensities, but it can help to improve the forecast of the areas of icing conditions and accuracy of the verification dataset together with PIREPs. The use of satellite data as a verification tool alongside PIREPs will be explored in future work.

Data Availability

The underlying data have been produced for aviation purposes. They can be provided for aviation use and for research activities. Contact the author or VL.Kunden-service@dwd.de.

Disclosure

The contents of this publication are the sole responsibility of Deutscher Wetterdienst, Météo-France, and Met Office and do not necessarily reflect the opinion of the European Union. An interim status of the study was presented at the “International Conference On Icing Of Aircraft Engines and Structures JUNE 17–21 2019 MINNEAPOLIS MN.”

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

The authors thank all who have provided support and contributed to the European Harmonization In-flight Icing project, especially the project managers of the involved partners Stephanie Jameson (MO), Lauren Donohue (EUMETNET), and Stéphanie Wigniolle (MF), and acknowledge the support from user’s perspective by Rosalind Lapsley (EUROCONTROL). The activity as part of the Action “SESAR Deployment Programme implementation 2015–Cluster 2” N° 2015-EU-TM-0196-M is co-founded by the European Union (482k EUR).

References

[1] Eurocontrol, https://www.eurocontrol.int/news/new-traffic-record-set-37228-flights-one-day, 2019.
[2] “SESAR Joint Undertaking,” 2021, https://www.sesarju.eu.
[3] “SESAR Deployment Manager,” 2021, https://www.sesardeploymentmanager.eu.
[4] M. K. Politovich, “Aircraft icing caused by large supercooled droplets,” Journal of Applied Meteorology, vol. 28, no. 9, pp. 856–868, 1989.
[5] P. Kalinka, K. Roloff, J. Tendel, and T. Hauf, “The In-flight icing warning system ADWICE for European airspace – current structure, recent improvements and verification results,” Meteorologísche Zeitschrift, vol. 26, no. 4, pp. 441–455, 2017.
[6] W. R. Sand, W. A. Cooper, M. K. Politovich, and D. L. Veal, “Icing conditions encountered by a research aircraft,” Journal of Climate and Applied Meteorology, vol. 23, no. 10, pp. 1427–1440, 1984.
[7] M. K. Politovich, “Response of a research aircraft to icing and evaluation of severity indices,” Journal of Aircraft, vol. 33, no. 2, pp. 291–297, 1996.
[8] S. D. Green, “A study of U.S. Inflight icing accidents, 1978 to 2002,” in Proceedings of the 44th AIAA AerospaceSciences Meeting and Exhibit, Reno, Nevada, 2006.
[9] B. C. Bernstein, F. McDonough, M. K. Politovich et al., “Current icing potential: algorithm description and comparison with aircraft observations,” Journal of Applied Meteorology, vol. 44, no. 7, pp. 969–986, 2005.
[10] G. Zängl, D. Reinert, P. Ripodas, and M. Baldauf, “The ICON (ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI-M: description of the non-hydrostatic dynamical core,” Quarterly Journal of the Royal Meteorological Society, vol. 141, no. 687, pp. 563–579, 2015.
[11] J. Pailleux, J.-F. Geleyn, M. Hamrud, and P. Courrier, “Twenty-five years of IFS/ARPEGE,” ECMWF Newsletter, vol. 141, pp. 22–30, 2014.
[12] D. Walters, I. Boutle, M. Brooks et al., “The Met Office unified model global atmosphere 6.0/6.1 and JULES global land 6.0/6.1 configurations,” Geoscientific Model Development, vol. 10, no. 4, pp. 1487–1520, 2017.
[13] M. S. Pepe and M. L. Thompson, “Combining diagnostic test results to increase accuracy,” Biostatistics, vol. 1, no. 2, pp. 123–140, 2000.
[14] B. G. Brown, G. Thompson, R. T. Bruintjes, R. Bullock, and T. Kane, “Intercomparison of in-flight icing algorithms. Part II: statistical verification results,” Weather and Forecasting, vol. 12, no. 4, pp. 890–914, 1997.
[15] A. Tafferner, T. Hauf, C. Leifeld, T. Hafner, H. Leykauf, and U. Voigt, “ADWICE: advanced diagnosis and warning system for aircraft icing environments,” Weather and Forecasting, vol. 18, no. 2, pp. 184–203, 2003.
[16] ICAO, Annex 3 to the Convention on International Civil Aviation, Meteorological Service for International Air Navigation, Montreal, Canada, 20th edition, 2018.
[17] A. H. Murphy and R. L. Winkler, “A general framework for forecast verification,” Monthly Weather Review, vol. 115, no. 7, pp. 1330–1338, 1987.
[18] H. E. Brooks and C. A. Doswell III, “A comparison of measures-oriented and distributions-oriented approaches to forecast verification,” Weather and Forecasting, vol. 11, no. 3, pp. 288–303, 1996.
[19] P. Heidke, “Berechnung des Erfolges und der Güte der Windstärkevorhersagen im Sturmwarnungsdienst,” *Geografiska Annaler*, vol. 8, pp. 301–349, 1926.

[20] P. N. Francis, “Detection of aircraft icing conditions over Europe using SEVIRI data,” *AMS/EUMETSAT Meteorological Satellite Conference Session*, vol. 3, pp. 24–28, 2007.

[21] R. Stretton, P. Buchanan, and P. Francis, *Verification of Met Office Global Icing Potential Forecasts Using Satellite Observations,* *AMS 18th Conference on Aviation, Range, and Aerospace Meteorology*, pp. 23–26, Seattle, Washington, 2017.