Development and Evaluation of In-Flight Icing Index Forecast for Aviation

CYRIL MORCRETTE, KATIE BROWN, REBECCA BOWYER, PHILIP GILL, AND DAN SURI

Met Office, Exeter, United Kingdom

(Manuscript received 22 October 2018, in final form 1 April 2019)

ABSTRACT

An in-flight icing index from the literature is implemented in the Met Office forecasting system. Comparison of hindcasts of cloud fraction with ground-based remote sensing observations of liquid and ice cloud fraction are used to inform a reformulation of part of the index. Satellite-retrieved icing potential is then used to quantitatively assess the reliability and skill of the new index and compare its performance to the current operational one. Having shown that the new index has substantially better reliability, ways of presenting separate icing likelihood and severity are explored using a case study.

1. Introduction

Ice nuclei must be present in order for each liquid water droplet in a cloud to freeze as the temperature falls below 0°C. In the absence of sufficient ice nuclei, liquid cloud droplets can still be in a liquid state at temperatures between 0° and around −40°C (Rogers and Yau 1989). If an aircraft flies through such a cloud, the supercooled droplets can quickly freeze onto the surface of the wings and airframe (Bruford 1997). This accumulation of ice can seriously affect the handling of the aircraft, which can ultimately lead to a disastrous loss of control (Cole and Sand 1991; Lankford 2000, 2001; Politovich 2003).

Meteorological factors, such as air temperature, humidity, and the presence and size of supercooled cloud particles, are important in determining the probability and severity of an in-flight icing event. In addition to some automated forecasts derived from numerical weather prediction (NWP) model output, operational meteorologists at the Met Office create a number of products which consider the threat of in-cloud icing. The main group of products consists of significant weather charts for aviation purposes, which are issued at 6-hourly intervals. One of the main significant weather charts covers a large part of northwestern

Europe and considers forecast weather below 10 000 ft (3050 m). These charts are produced as part of the Met Office's contractual obligation to the Civil Aviation Authority in the United Kingdom and are known as "LOLAN" charts.

Meanwhile, the Met Office is one of only two World Area Forecast Centers (WAFc) that provide weather forecasting services to meet Annex 3 to the International Civil Aviation Organization (ICAO) Convention on Civil Aviation (ICAO 2018). As part of this commitment, significant weather charts are produced globally at 6-hourly intervals for forecast weather above 10 000 ft (3050 m, Fig. 1a). These also include a consideration of the threat of in-cloud icing.

More specific and detailed guidance on the threat of in-cloud icing is provided by Met Office operational meteorologists, mainly at British military bases. This output is often focused on specific airframes or more localized areas than the more generic LOLAN chart. Meanwhile, significant meteorological advisories (SIGMETs) for U.K. airspace, which comprises the London, Scottish, and Shanwick Flight Information Regions (Fig. 1b) and extends over part of the eastern Atlantic Ocean, are issued for a number of weather hazards, including severe in-cloud icing. Best practice at the Met Office tends to be to issue SIGMETs with lead times of approximately 1 hour. Valid for no more than 4 h at a time, the issuing of SIGMETs is based on a combination of NWP model output, operational meteorologist input, and pilot reports.

Techniques used by Met Office operational meteorologists to identify areas conducive to in-cloud icing are

Denotes content that is immediately available upon publication as open access.

Corresponding author: Cyril Morcrette, cyril.morcrette@metoffice.gov.uk

FIG. 1. Regions where the Met Office has a responsibility for providing in-flight icing forecasts for civil aviation (gray shading): (a) above and (b) below 10 000 ft (3050 m).

heavily weighted toward empirical thinking. Meanwhile, the current Met Office operational icing index, which is calculated automatically from NWP model output, is rarely used by forecasters due to its tendency to over-predict. The empirical considerations used by forecasters include an appreciation of synoptic situations known to be conducive to the potential for in-cloud icing. These include marine stratiform cloud, moist unstable airmasses, and areas on and ahead of warm and occluded fronts. Cold frontal precipitation bands, particularly when line convection is present, and regions of nimbostratus and cumulonimbus are all also associated with the risk of severe in-cloud icing. Otherwise, in-cloud icing is considered concentrated in cloudy environments, mainly with air temperatures between 0° and −20°C. Other considerations which have been identified as present in known in-cloud icing events, such as isothermal cloud layers and presence of some element of wind shear in cloud, are also borne in mind.

Human forecasters can undoubtedly add local knowledge and compensate for known model biases, before issuing an aviation forecast. As a result, a purely model-derived icing-risk product is not our goal. However, a skillful and trusted automatically generated icing index would be useful for operational meteorologists at the Met Office to consult as they prepare their forecasts.

When considering an icing threat, both currently and in the next few hours, NWP-derived indices can

be combined with satellite or ground-based radar observations in order to provide a consistent estimate of the locations of icing risks in the “nowcast” range (e.g., Tafferner et al. 2003; Le Bot 2004; Bernstein et al. 2005; Kalinka et al. 2017).

The focus here is on longer lead times (from $T + 12$ to $T + 36$), where the NWP output is not combined with observations. To implement such a model-derived icing product its quality needs to be measured, and choices relating to how it is calculated need to be assessed. Not only is this quantitative assessment necessary to show that the new index is skillful and worth looking at while preparing a forecast, it is essential in order to justify subsequent changes and improvements to the index in the future.

One way of verifying icing forecasts is using pilot report (PIREPs). However, there are a number of issues relating to location errors, sparsity of null reports, and avoidance of at-risk areas, which mean that statistical evaluations of icing algorithms using PIREPS must include a careful consideration of these sampling biases (Brown et al. 1997; Kalinka et al. 2017). Additionally, there are nonmeteorological factors that can modulate the severity of an icing event. These include the type of aircraft, whether it has de-icing equipment, the time spent in cloud, and its flight history. The latter is important as the same aircraft, flying through the same cloud, would be more vulnerable to icing if it is descending from altitude and has a cold

fuselage, rather than if it is ascending from the relatively warmer surface.

Minnis et al. (2005) use satellite remote sensing to infer the presence of conditions conducive to icing. Their methods have been implemented at the Met Office using data from the Meteosat Second Generation (MSG; Schmetz et al. 2002) satellite over the Greenwich Meridian (Francis 2007). As a result, an alternative way of evaluating forecast icing indices is to verify against satellite-based icing potential (Bowyer and Gill 2019).

In this paper, an icing index adapted from Belo-Pereira (2015) is implemented in the Met Office forecast system and used to produce hindcasts of icing risk based on NWP model output out to $T + 36$. Satellite-based verification on a 0.25° , 0.5° , and 1.25° grid shows that the new index is more reliable than the one currently produced operationally. To complement the passive top-down evaluation from satellite, ground-based active remote sensing is used to evaluate one of the elements that make up the icing index and address an apparent frequency bias. This leads to a further refinement of how the new index is calculated, which, when compared to the current operational index, still leads to both improved reliability and improved ability to distinguish between icing and nonicing events, when evaluated using the satellite data.

Following discussion with aviation forecasters, it was found that for automated icing products to be useful as initial guidance, they need to include a separate quantification of likelihood and severity. Having shown that the new icing index is statistically more reliable than the current operational one, the new index is used to quantify likelihood. The in-cloud condensate amount in the supercooled liquid water temperature range is then used to quantify severity. A simple 3×3 color matrix is introduced, which quickly summarizes the probability and severity elements of the NWP-derived index.

The model simulations, icing index calculations, satellite-based evaluation techniques, and ground-based data used for comparison are described in section 2. Results of the satellite-based and ground-based evaluations are presented in sections 3a and 3b, while a case study from some recent icing events in section 3c illustrates some of the key points about the performance of the control and new schemes. A discussion appears in section 4, where the 3×3 color matrix is also introduced, and conclusions are drawn in section 5.

2. Methods

a. Model simulations

All hindcasts were performed with the Met Office Unified Model (MetUM), using the Global Atmosphere

6 configuration (GA6; Walters et al. 2017). All simulations were initialized from Met Office global operational analyses and were run using a resolution known as N768, which consists of a regular latitude–longitude grid with $2 \times N$ points of longitude and $1.5 \times N$ points of latitude, leading to a longitude spacing of 0.23° and a grid size of 26 km on the equator. All simulations were initialized at 0000 UTC and run for 36 h.

For evaluation against satellite observations, 365 simulations were carried out covering 1 March 2016–28 February 2017, and the comparisons were done at 1200 UTC using model data valid at $T + 36$. For comparison against ground-based radar and lidar observations, 30 simulations were initialized at 0000 UTC each day from 31 March to the end of April 2007 and hourly model data with a lead time from $T + 12$ to $T + 36$ were concatenated together. Finally, for the icing event case study occurring around noon on 26 January 2017, the simulation was started a day and a half earlier, at 0000 UTC 25 January 2017.

b. Icing indices

The Met Office icing index that is currently provided operationally to the World Area Forecast System (WAFS) is based on Schultz and Politovich (1992). If the model temperature is between 0° and -20°C , and there is some cloud present (i.e., cloud fraction C is greater than zero), then the icing index is set to the value of the relative humidity. This index is the control.

The new icing indices being tested in the Met Office system are based on the simplified forecast icing potential (SFIP) described by Belo-Pereira (2015), itself based on the work of McDonough et al. (2004), Bernstein et al. (2005), and Wolff et al. (2009). The SFIP is defined as

$$\text{SFIP} = M_T(aM_{\text{RH}} + bM_w + cM_{\text{LWC}}). \quad (1)$$

The SFIP is based on a combination of membership functions M_ϕ derived by Bernstein et al. (2005), which quantify how occurrences of icing in pilot reports vary depending on meteorological conditions ϕ at the same time and place, as estimated from a database of NWP model output. The temperature membership function M_T used here differs slightly from the one used by Bernstein et al. (2005) and Wolff et al. (2009); for simplicity it does not differentiate between convective and stratiform cloud conditions. Additionally, in contrast to Belo-Pereira (2015), a different temperature membership function is not used for the lowest flight levels. Here, M_T is defined as

$$M_T = \begin{cases} 0.0, & \text{if } T \leq T_1, \\ (T - T_1)/(T_2 - T_1), & \text{if } T_1 < T \leq T_2, \\ 1.0, & \text{if } T_2 < T \leq T_3, \\ 1.0 - [(T - T_3)/(T_4 - T_3)], & \text{if } T_3 < T \leq T_4, \\ 0.0, & \text{if } T > T_4, \end{cases} \quad (2)$$

with $T_1 = -28^\circ\text{C}$, $T_2 = -12^\circ\text{C}$, $T_3 = -1^\circ\text{C}$, and $T_4 = +1^\circ\text{C}$. This is shown in Fig. 2a.

The relative humidity (RH) membership function, which unlike Belo-Pereira (2015) is again the same irrespective of the flight level, is qualitatively similar to that used by the other authors. Here it is specifically defined as

$$M_{RH} = \begin{cases} 0, & \text{if } RH \leq RH_1, \\ \left(\frac{RH - RH_1}{RH_2 - RH_1} \right)^2, & \text{if } RH_1 < RH \leq RH_2, \\ 1, & \text{if } RH > RH_2, \end{cases} \quad (3)$$

with $RH_1 = 0.6$ and $RH_2 = 0.95$ (Fig. 2b). Next, the vertical velocity w membership function is defined as

$$M_w = \begin{cases} -0.4, & \text{if } w \leq -0.1 \text{ m s}^{-1}, \\ -0.4 + 0.4 \left(\frac{w - w_1}{w_2 - w_1} \right), & \text{if } -0.1 < w \leq 0.0 \text{ m s}^{-1}, \\ \frac{w - w_2}{w_3 - w_2}, & \text{if } 0.0 < w \leq 0.5 \text{ m s}^{-1}, \\ 1, & \text{if } w > 0.5 \text{ m s}^{-1}, \end{cases} \quad (4)$$

with $w_1 = -0.1$, $w_2 = 0.0$, and $w_3 = 0.05 \text{ m s}^{-1}$ (Fig. 2c). Finally, the liquid water content (LWC) membership function is defined as

$$M_{LWC} = \begin{cases} LWC/0.4, & \text{if } LWC \leq 0.4 \text{ g kg}^{-1}, \\ 1.0, & \text{if } LWC > 0.4 \text{ g kg}^{-1}. \end{cases} \quad (5)$$

The last thing to define is the relative weights given to each of the membership functions. Following Belo-Pereira (2015), $a = 0.35$, $b = 0.2$, and $c = 0.45$, and any negative values of SFIP are reset to zero. Both the control and SFIP indices can be calculated on a three-dimensional grid, using data from a numerical weather prediction model.

There are questions regarding how well NWP models capture the physical processes that determine the

partitioning of condensate into the liquid or ice phase (e.g., Kalesse et al. 2016; Furtado and Field 2017). To avoid a SFIP index performing badly simply as a result of an incorrect phase partition; despite getting the temperature, humidity, and total condensate correct; a slightly different SFIP index was considered, where the condensate membership function M_{LWC} took the sum of the LWC and ice water content (IWC) as its input, rather than just the LWC.

It is worth noting that the RH membership function, which rises from zero to one, over an RH range of 0.6–0.95, looks very similar to the parameterization of cloud cover developed by Sundqvist et al. (1989). This prompts one to ask, given the efforts to improve the cloud cover forecasts in NWP models (e.g., Tiedtke 1993; Wilson et al. 2008; Forbes and Ahlgrimm 2014; Furtado et al. 2016), whether using a model's cloud fraction C would be a viable alternative to using M_{RH} when calculating SFIP. Indeed, the model's cloud fraction quantifies the probability of encountering cloud within a grid box; and is used by the model for doing microphysical and radiative transfer calculations. So it would make sense to test the inclusion of cloud fraction as part of an icing index.

This leads to several permutations of SFIP indices. Either M_{LWC} can be calculated from just the liquid (L) or from the liquid and ice (LI). Additionally, the relative humidity membership function (RH) can be replaced with either the liquid cloud fraction (LCF) or the bulk cloud fraction (BCF, which is the cloud fraction irrespective of whether the condensate is in the liquid or ice phase). These will be referred to as SFIP-L-RH, SFIP-L-LCF and SFIP-LI-BCF, and scores quantifying their performance are summarized in Tables 1 and 2.

c. Satellite-based evaluation

Minnis et al. (2005) describe how an icing potential can be calculated from satellite-retrieved cloud-top temperature, liquid water path, and effective radius. Their algorithm has been implemented to work on MSG data (Francis 2007), which cover the region 60°N – 60°S , 60°W – 60°E . The comparison presented here uses data valid at 1200 UTC each day for 12 months from 1 March 2016.

The satellite icing potential is available on a 0.25° grid, while the model forecasts are on a similar ($N768$, $\approx 26\text{-km}$ grid). To reduce double-penalty effects (e.g., Mason 2003), both datasets have been regridded to 1.25° . The impact of the regridding on the apparent skill of the icing indices is shown in Tables 1 and 2 and discussed in section 4.

The satellite retrieval will not be able to determine the icing potential if there is optically thick ice cloud located above lower-level clouds. In this case, the satellite icing potential reports missing data, and those

FIG. 2. Membership functions used for calculating the SFIP in Eq. (1).

locations are ignored from the model fields prior to carrying out any comparisons. The 5×5 regridding is calculated as a mean of the pixels that do have valid data, and the masking of the model fields to account for missing observations is only carried out after the regridding.

The satellite icing potential is a two-dimensional product (2D). The model-derived icing indices are three-dimensional (3D). In order for the two to be compared, the model-derived indices need to be collapsed in the vertical in some way.

One option is simply to take the vertical maximum of the model icing index, and this is done here. It could be argued however, that when passive satellite sensors detect a moderate icing risk near cloud top, there could be a region of higher risk located lower down, which may be undetected. Therefore, taking the maximum value in each model column would not be representative of what the satellite is seeing.

Here a pragmatic filtering technique is used, similar to the one developed by [Bowyer and Gill \(2019\)](#), for turning the 3D model icing indices into a 2D field ready for comparison against the satellite icing potential. For each model column, one starts at the top

and searches down through the icing index field until a certain threshold value is found. Any values smaller than the threshold are ignored. If the threshold is encountered, that becomes the value of the 2D field at that location and any icing values occurring below, whether smaller or larger, are ignored. If the threshold is never encountered the value of the 2D field is set to zero. As the goal of this study is to compare several icing indices against each other, one must ensure that the choice of threshold does not favor one index. For each of the indices studied here, the evaluation is performed four times; once with the vertical maximum and once with the filtering threshold set to 0.05, 0.1, and 0.2.

Additionally, when using the satellite icing potential to evaluate the model products, one needs to decide on the minimum value of satellite icing potential that is assumed to indicate that an event occurred. Because of uncertainty as to the best choice of threshold, three different thresholds are used in this work. The first is at 0.05 (as inspection of satellite icing potential histograms—not shown—indicates that this discriminates between all the zero and nonzero values). The other two thresholds are set at 0.4 and 0.7, which [Minnis et al. \(2005\)](#) state are the thresholds between low-and-medium and

TABLE 1. ROC skill score for different icing indices and different methods of converting the 3D model fields to a 2D field for comparison with the satellite observations. The RSS is calculated from the AUC as $RSS = (AUC - 0.5) \times 2$. All have been calculated assuming the middle threshold of 0.4 for deciding that the satellite icing potential has recorded an event. The first four rows are calculated with the data having been averaged to a 1.25° grid, the next four rows have been averaged to a 0.5° grid, while the bottom four rows use a 0.25° grid. The data used are valid at 1200 UTC for 12 months from 1 Mar 2016, covering the region 60°N – 60°S , 60°W – 60°E .

	Resolution	Vertical			
		max	$f = 0.05$	$f = 0.1$	$f = 0.2$
Control	1.25°	0.71	0.52	0.54	0.56
SFIP-L-RH	1.25°	0.72	0.51	0.59	0.62
SFIP-L-LCF	1.25°	0.63	0.52	0.52	0.32
SFIP-LI-BCF	1.25°	0.71	0.56	0.62	0.56
Control	0.50°	0.65	0.52	0.52	0.53
SFIP-L-RH	0.50°	0.66	0.45	0.54	0.56
SFIP-L-LCF	0.50°	0.55	0.43	0.45	0.32
SFIP-LI-BCF	0.50°	0.64	0.49	0.55	0.51
Control	0.25°	0.59	0.50	0.50	0.50
SFIP-L-RH	0.25°	0.61	0.42	0.50	0.51
SFIP-L-LCF	0.25°	0.48	0.37	0.39	0.30
SFIP-LI-BCF	0.25°	0.58	0.44	0.49	0.46

medium-and-high probabilities for the satellite icing potential. Combined with the 4 ways of converting from 3D to 2D fields, this leads to 12 different ways of evaluating the model icing index using the satellite observations.

Although all the icing indices and the satellite icing potential are dimensionless quantities with a range from zero to one, this does not mean that they are numerically equivalent. So as well as exploring three different choices for the definition of an event in the observations, comparisons are also made by considering a range of thresholds (from 0 to 1 in 0.01 steps) for defining an event in the model.

Hit rate and false-alarm rate are calculated at each of these thresholds, allowing the creation of receiver operating characteristic (ROC) curves (e.g., Wilks 2006). The area under the curve (AUC) can be used as a measure of the discriminatory skill of the forecast, which can further be summarized using the ROC skill score, $RSS = (AUC - 0.5) \times 2$.

To check whether the model icing indices can usefully be interpreted as probabilities, reliability diagrams are used (e.g., Wilks 2006). These show, for any model predicted value, the observed frequency of an event. If the index is a reliable predictor of probability then the data should fall along the diagonal. When interpreting reliability diagrams there is a question as to how far data can be off the diagonal and yet still be consistent with a

TABLE 2. BSS for different icing indices and different methods of converting the 3D model fields to a 2D field for comparison with the satellite observations. All have been calculated assuming the middle threshold of 0.4 for deciding that the satellite icing potential has recorded an event. The first four rows are calculated with the data having been averaged to a 1.25° grid, the next four rows have been averaged to a 0.5° grid, while the bottom four rows use a 0.25° grid. The data used are valid at 1200 UTC for 12 months from 1 Mar 2016, covering the region 60°N – 60°S , 60°W – 60°E .

	Resolution	Vertical			
		max	$f = 0.05$	$f = 0.1$	$f = 0.2$
Control	1.25°	−13	−5.9	−5.2	−4.7
SFIP-L-RH	1.25°	0.68	0.79	0.84	0.87
SFIP-L-LCF	1.25°	0.74	0.60	0.70	0.74
SFIP-LI-BCF	1.25°	0.69	0.79	0.82	0.85
Control	0.50°	−22	−4.7	−4.5	−4.4
SFIP-L-RH	0.50°	0.67	0.62	0.66	0.41
SFIP-L-LCF	0.50°	0.73	0.78	0.71	0.71
SFIP-LI-BCF	0.50°	0.68	0.39	0.69	0.46
Control	0.25°	−26	−3.2	−3.1	−3.2
SFIP-L-RH	0.25°	0.69	0.35	0.28	0.33
SFIP-L-LCF	0.25°	0.51	0.30	0.37	0.38
SFIP-LI-BCF	0.25°	0.70	0.35	0.38	0.41

reliable prediction. To assess this, consistency bars have been added around the diagonal following the bootstrap resampling method of Bröcker and Smith (2007).

d. Comparison against ground-based remote sensing observations

Although the satellite-based retrieval provides data over a large spatial domain, it is unable to detect any supercooled liquid cloud that may be located below thick ice cloud. Ground-based remote sensing observations from radars and lidars at three sites operated by the Atmospheric Radiation Measurement (ARM) program (Stokes and Schwartz 1994) are therefore used to provide a complementary evaluation of one element of the SFIP icing index. The data used are from the Southern Great Plains (SGP), United States (36.606°N , 262.515°E); Darwin, Australia (12.425°S , 130.892°E); and Murgtal, Germany (48.545°N , 8.405°E). Liquid and ice water contents are inferred from the radar and lidar data as described by Illingworth et al. (2007). For comparison with the model, an observed profile of cloud fraction is derived from the height-resolved high-frequency observations above each site. For the height range corresponding to each of the model levels, this is done by calculating the fraction of observations that contain liquid or ice cloud within a one hour period and, following Illingworth et al. (2007), assuming that the time-averaged cloud cover is a proxy for the spatially

FIG. 3. Maps of satellite-derived icing potential from Meteosat Second Generation (MSG) at 1200 UTC 1 Mar 2016: (a) data on 0.25° grid and (b) data on 1.25° grid.

averaged cloud cover. This provides a separate profile of observed liquid and ice cloud fractions each hour for the month of April 2007. The month of April 2007 was chosen for validation as it was a period when observational data were available from these three sites with very different weather conditions. Hogan et al. (2001) discussed the need to filter model fields to remove

the thin high-level cirrus clouds that the ground-based instruments would not be able to detect. This ensures a fairer comparison between model and observations. The filtering method, which is described by Brooks (2005) and Morcrette et al. (2012), consists in knowing the minimum detectable IWC as a function of height for each radar and

FIG. 4. Maps of two-dimensional model-derived icing index at 1200 UTC 1 Mar 2016, from $T + 36$ model fields. Data are regridded to $1.25^\circ \times 1.25^\circ$. (top) The control operational icing index based on temperature and relative humidity when cloud is present; (bottom) the SFIP-L-RH. Different ways of converting from a three-dimensional to a two-dimensional field are shown in each column: (a),(e) use the vertical maximum. Other columns use filtering using a threshold of (b),(f) 0.05, (c),(g) 0.1, and (d),(h) 0.2.

FIG. 5. Reliability diagrams for the control index (based on temperature and relative humidity when cloud is present), evaluated against satellite-derived icing potential. The four columns are for different ways of converting from 3D model indices to a 2D field for comparison with the observations. The three rows correspond to choices of threshold for deciding that an event has occurred in the observations ($s = 0.05$, $s = 0.4$, and $s = 0.7$). Each panel includes kite shading to indicate climatology and the no-skill line and consistency bars are shown around the diagonal. The BSS also appears in each panel. The data used are valid at 1200 UTC for 12 months from 1 Mar 2016, on a $1.25^\circ \times 1.25^\circ$ grid covering the region 60°N – 60°S , 60°W – 60°E .

assuming an analytical shape for the subgrids variation in IWC. The filtered model ice cloud fraction consists of that portion of the model IWC that would be expected to be detectable by the ground-based radar.

3. Results

a. Icing index evaluation using satellite observations

As an illustration of the data being used for evaluation, Fig. 3 shows the satellite icing potential at 1200 UTC 1 March 2016. The data are on a 0.25° grid in Fig. 3a and have been regridded to 1.25° in Fig. 3b.

Meanwhile, illustrative two-dimensional icing index fields are shown for the same time in Fig. 4. Figures 4a–d

are the control operational icing index based on temperature and relative humidity when cloud is present. Figures 4e–h are SFIP-L-RH (simplified forecast icing potential calculated using temperature, vertical velocity, liquid condensate amount, and relative humidity). Different ways of converting from a three-dimensional to a two-dimensional field are shown in each column: Figs. 4a and 4e use the vertical maximum. Other columns use filtering with a threshold of 0.05 (Figs. 4b,f), 0.1 (Figs. 4c,g) and 0.2 (Figs. 4d,h).

The effect of the search algorithm is to remove small values of the 2D icing index field which are present when using a vertical maximum. It also reduces the amount of large values, found when using the vertical maximum algorithm, as these are typically located below regions of

FIG. 6. As in Fig. 5 but for reliability diagrams for SFIP-L-RH.

moderate icing index which are numerically large enough to stop the search algorithm looking farther down.

Figure 5 shows 12 reliability diagrams for the control icing index based simply on relative humidity when cloud is present between 0° and -20°C . A total of 10 out of the 12 reliability diagrams have Brier skill scores (BSSs) that are negative, indicating that the index is worse than a prediction based simply on climatology. It is worth noting that, generally, the control icing index has poor “reliability” in the probabilistic forecast sense, which means that the data points are far from the diagonal. The control index also has poor “resolution” as the line through the data is rather flat. This means that events occur with very similar frequencies in reality irrespective of the numerical value of the model-produced index. These findings are consistent with the anecdotal feedback from forecasters, which states that they do not use the control index as part of any decision making process as it is not found to be useful.

Figure 6 shows reliability diagrams for SFIP-L-RH. The exact values of the BSS depend on the choice of threshold for defining an event in the observations and the manner in which the 3D model data are converted to a 2D field. It can be seen that using a threshold of $s = 0.4$ for deciding that an event occurred in the observations leads to the most reliable evaluation of the SFIP-L-RH index, while thresholds of $s = 0.05$ and $s = 0.7$ lead to an apparent under or overprediction by the model respectively. Additionally, there is a large range of model indices where the data fall within the range of the consistency bars, indicating that the data are not inconsistent with being reliable. This means that SFIP-L-RH could usefully be interpreted as a probability.

For brevity, as Fig. 5 shows that the control index has such poor reliability, its ROC curves are not shown. As a summary though, the RSS for the control using the satellite threshold of 0.4 appear in Table 1. A series of

FIG. 7. ROC curves for SFIP-L-RH evaluated against satellite-derived icing potential. The panels are arranged as in Figs. 5 and 6, and the RSS also appears in each panel.

ROC curves for SFIP-L-RH, which is notably more reliable than the control index, are shown in Fig. 7. The RSS for other ways of defining SFIP, evaluated using the satellite threshold of 0.4, also appear in Table 1. This table also includes results from evaluations carried out on 0.5° and 0.25° grids. Results of BSS are shown in Table 2.

b. Evaluation of cloud forecasts using ground-based remote sensing observations

An example of the data used for evaluation in this section, Fig. 8 shows time–height cross sections of cloud fraction over the SGP site in April 2007 for both the observations and the model.

Averaged over the month, profiles of cloud fraction in the observations and the model are shown in Fig. 9 for each of the three sites. These profiles are calculated considering liquid and frozen cloud fraction separately.

In the left-hand column (Figs. 9a–c), the cloud fraction profile is calculated as a simple mean over the month. In the right-hand column (Figs. 9d–f), the profiles have been calculated as the frequency of occurrence (FOO) of cloud fraction greater than 2%.

Figure 10 is similar to Fig. 9, but the mean cloud fraction and frequency of occurrence are calculated without taking account of the phase of the condensate.

By defining cloud occurrence events as when cloud fraction is greater than 2%, one can calculate the number of hits a , false alarms b , misses c , and correct negatives d . The frequency bias can then be found from $B = (a + b)/(a + c)$. Focusing on the temperature range from 0° to -30°C , the frequency biases are shown in Fig. 9. In this case a hit indicates that the model has correctly predicted the occurrence of liquid cloud. While still focusing on the same temperature range, the frequency

FIG. 8. Time–height cross section of bulk (liquid or ice) cloud fraction at SGP in April 2007: (a) as observed by vertically pointing radar/lidar and (b) as modeled by the Unified Model (GA6; N768: catenation of forecasts from $T + 12$ to $T + 36$ h). The modeled cloud has been filtered to account for clouds that the observational instruments would not be expected to detect. Lines show the 0° and -20°C isotherms (obtained from the model). A height of 50 000 ft corresponds to 15 240 m.

biases in Fig. 10 are calculated irrespective of cloud phase, which leads to a reduction in the frequency biases.

c. Case study: 26 January 2017

On 26 January 2017, a number of incidents of in-flight ice accretion were reported to aviation forecasters at the Met Office. The time and locations of these events are shown in Table 3.

An example of a LOLAN chart, for weather-related aviation hazards below 10 000 ft (3050 m), issued at 0300 UTC and valid from 0800 to 1700 UTC 26 January 2017 is shown in Fig. 11, while an explanation of the abbreviations used on the chart can be found in Lankford (2001). Three incidents occurred below 3000 ft (910 m) within Region C where, according to the LOLAN chart, the 0°C isotherm was forecast to be between the surface and 3500 ft (1070 m) over land. In this region the forecast was for between 5 and 8 oktas of stratus or stratocumulus, with bases between 1000 and 2000 ft (305–610 m) and cloud tops between 2500 and 3500 ft (760–1070 m). Moderate icing was forecast for these clouds.

A WAFC chart, for weather-related aviation hazards above 10 000 ft (3050 m), valid at 0600 UTC 26 January 2017, is shown in Fig. 12 (unfortunately the

1200 UTC chart was not archived). The large expanse of cloud from Iberia to the west of Ireland and extending east into Scandinavia is marked as having the potential for moderate icing from below 10 000 to 18 000 ft (3050–5490 m).

A map of the satellite icing potential at 1200 UTC 26 January 2017, zoomed in over northwest Europe and the United Kingdom, is shown in Fig. 13a. Regions where the icing potential retrieval was impossible due to overlying ice cloud are masked in white. The control and SFIP-LI-BCF model indices are shown in Fig. 13b and c, filtered using a threshold of 0.1. The vertical maximum of SFIP-LI-BCF is shown in Fig. 13d. No masking has been applied to the model fields shown in Fig. 13, although the quantitative evaluation of Brier and ROC skill scores shown earlier were calculated only where the satellite icing potential had a valid retrieval. In each panel, the location of the three reported icing incidents are shown by asterisks.

A different way of visualizing the model data from this day is to plot time–height cross sections at the locations where icing events occurred. These are shown in Fig. 14 for both the control index and SFIP-LI-BCF.

FIG. 9. Profiles of liquid and ice cloud fraction for April 2007. (left) Mean over the month and (right) FOO of cloud fraction greater than 2%. The frequency bias for the occurrence of liquid cloud in the temperature range from 0° to -30°C is shown. (a),(d) SGP; (b),(e) Darwin; and (c),(f) Murgtal. Horizontal lines show the mean height of the 0° , -10° , -20° , and -30°C isotherms. A height of 50 000 ft corresponds to 15 240 m.

4. Discussion

The work in this paper was motivated, in part, by the fact that the current Met Office operational icing index, which is calculated automatically from NWP model output, has such a tendency to overforecast that it is not seen as useful by many aviation forecasters and is hence rarely consulted. This impression is readily confirmed by Fig. 5, which showed that the control index has no reliability and has consistently very poor (often negative) Brier skill scores, irrespective of how the model index is compared to the satellite icing potential.

By comparison, the SFIP-L-RH index shows a large improvement in terms of reliability (Fig. 6). It does depend on exactly how the evaluation against satellite observations is carried out, but there are a number of permutations where SFIP-L-RH has positive Brier skill scores and where the data fall within the Bröcker and Smith (2007) 5th–95th percentile confidence interval, suggesting that the data are not inconsistent with being reliable.

It could be argued that the inclusion of the RH membership function in the calculation of SFIP is just a proxy for cloud fraction. After all, the shape of M_{RH} is reminiscent of the parameterization of cloud cover as a

function of relative humidity developed by Sundqvist et al. (1989). Unfortunately, comparison of SFIP-L-LCF to SFIP-L-RH, in Tables 1 and 2, shows that the replacement of the relative humidity membership function with the liquid cloud fraction does not generally improve the Brier or ROC skill scores.

However, comparison of model profiles of liquid and ice cloud fraction against ground-based remote sensing observations from some ARM sites shows why this is the case. Although the model mean model cloud fraction is quite well simulated in the temperature range from 0° to -30°C , it is the frequency of occurrence of nonzero cloud cover in this temperature range that is of most relevance to an icing index. In terms of liquid cloud fraction at subzero temperatures, notable frequency biases are seen over the three ARM sites (Fig. 9). Assessing the frequency of occurrence of model cloud, ignoring the phase obtained from the model, leads to reduced frequency biases in the temperature range from 0° to -30°C (Fig. 10). This information can be pulled through into the definition of SFIP. The SFIP is then calculated using the sum of LWC and IWC as the input to the M_{LWC} membership function and the relative humidity membership function M_{RH} is replaced with the model bulk cloud fraction

FIG. 10. As in Fig. 9, but without discriminating between the phase of the cloud condensate. A height of 50 000 ft corresponds to 15 240 m.

(that is the cloud fraction irrespective of whether it is liquid or ice). It is worth noting that the temperature membership function still ensures that SFIP will be zero outside of the range from -28° to $+1^{\circ}\text{C}$. Tables 1 and 2 shows that SFIP-LI-BCF can maintain the improved reliability of SFIP-L-RH relative to the control (in terms of BSS), while also beating the control in terms of RSS on a 1.25° grid.

The results discussed so far are exemplified by the case study of the icing events reported on 26 January 2017. Three occurrences of significant ice accretion were reported between 2000 and 2500 ft (around 610–760 m) over central England that day. The locations of these are shown overlaid on the satellite icing potential (Fig. 13a) and over the filtered control and SFIP-LI-BCF indices (Fig. 13b and c) and vertical maximum of SFIP-LI-BCF (Fig. 13d).

As is typical of the control index, vast spatial regions have large values (Fig. 13b), with limited areas having nonzero but small values (as would be expected from an index that replicates the numerical value of the relative humidity when cloud is present). This would be even worse if the vertical maximum of the control index was used (not shown). The new SFIP-LI-BCF index by contrast has more small values (Fig. 13c), (even when

using the vertical maximum: Fig. 13d) and subjectively seems to better replicate the extent and shape of the region of higher satellite icing potential seen over central England and the North Sea.

For the three locations where pilots reported inflight icing to aviation meteorologist on 26 January, time–height cross sections of icing index have been produced (Fig. 14). These start at 0000 UTC on the day before the reports, so these correspond to $T + 36$ h forecasts for the incidents concerned. The control index correctly flags that there is a risk of icing but as soon as nonzero values are produced, the value of the index is frequently quite high. This provides little indication of where more or less severe conditions are (Figs. 14a–c). The new SFIP-LI-BCF index by contrast shows a more focused shallow region of increased risk near cloud top,

TABLE 3. Latitude, longitude, altitude, and time of the three locations where in-flight icing was reported on 26 Jan 2017.

Location	1	2	3
Latitude ($^{\circ}\text{N}$)	53.09	53.03	51.75
Longitude ($^{\circ}\text{W}$)	0.17	0.49	1.58
Height above ground level (ft)	2000	2500	2500
Time (UTC)	1300	0850	1215

FIG. 11. Chart describing significant weather below 10 000 ft (3050 m) for aviation purposes valid from 0800 to 1700 UTC 26 Jan 2017.

with lower risk below. It is indeed near cloud top that all three reported events occurred. Feedback from operational aviation meteorologists indicates that figures such as Figs. 14d–f, if produced with sufficient lead time, can provide useful additional guidance when it comes to issuing a forecast.

With the SFIP index being able to take any real value between 0 and 1, an additional request from forecasters was for guidance on the quantitative interpretation of the index. Clearly, larger numbers meant more risk, but how should the numbers be interpreted and what were the critical values? In this regard, the possibility of SFIP indices being calculated to many decimal places, leading to a multitude of different shades when displayed on a color scale, was not conducive to conveying a clear message.

Discussions with operational aviation meteorologists have highlighted a need for any automated model-derived products to provide a quantitative indication of both likelihood of an icing event and the severity of the risk. Since the SFIP-LI-BCF has been shown to have some reliability in the statistical sense; that is observed events occurring with the frequency that the forecast

index suggests over a range of forecast indices; it would seem sensible to interpret SFIP-LI-BCF as the likelihood of icing conditions. Although the verification in section 3a showed the best results when the satellite icing potential was compared to the model index that had been collapsed to 2D by filtering, when displayed as a map for forecasting purposes, it is better for the 2D likelihood to be calculated from the vertical maximum of the SFIP-LI-BCF, as this highlights the maximum values that an aircraft may fly through as it goes from the surface to cruising altitude, or vice versa. For simplicity, it is decided to interpret SFIP-LI-BCF larger than zero but less than 0.33 as “low” likelihood. SFIP-LI-BCF between 0.33 and 0.66 is taken to be “medium” likelihood, and SFIP larger than 0.66 is “high” likelihood.

To estimate the severity of icing, full global 3D fields of liquid and ice condensate amount are considered. These are divided by the bulk cloud fraction to find the in-cloud condensate and then are multiplied by the temperature membership function from -28° to $+1^{\circ}\text{C}$ to focus on the temperature range of interest. Probability distribution functions (PDF)

FIG. 12. Chart describing significant weather between 10 000 and 45 000 ft (3050–13 700 m) for aviation purposes valid at 0600 UTC 26 Jan 2017.

of these in-cloud condensates are constructed and, somewhat arbitrarily, it is decided to define low severity as the lowest 70 centiles of this PDF. Meanwhile, medium severity is taken to be the 70th–90th centiles, while high severity is defined as the top 10 centiles of the PDF. For the UM using GA6, using global data for the whole of the months of March and August 2016, this leads to thresholds between low-and-medium and medium-and-high located at 1.3 and $6.2 \times 10^{-5} \text{ kg kg}^{-1}$, respectively. This calibration step, which is model specific and can be done using a climatology, should be revisited whenever the operational model is upgraded.

One advantage of defining severity in this way comes when icing indices produced by different NWP systems are compared. Even though different models may have different cloud fraction and cloud condensate biases, if the thresholds are calculated for each system independently, then high severity will have the same interpretation; it means the conditions are expected to be in the worse 10% that the model is capable of producing.

Having defined low, medium, and high severity in this way, there is no suggestion that these severities will match the definitions of “trace,” “light,” “moderate,” or “severe” that appear in icing PIREPS. However, it is well known that the severity of an icing event, although heavily influenced by the local meteorology, is also strongly dependent on the type of aircraft, flight history, and length of time spent in the supercooled conditions. So a direct mapping is not desirable.

The approach here is pragmatic and designed to highlight the regions where, meteorologically, the conditions are at their worst. One way of perhaps aligning our model-predicted severities with those that appear in PIREPS, would be to measure the relative abundance of light, moderate, and severe icing events that appear in PIREPS and then find the values in the model in-cloud condensate PDF that would split it into three categories that occur with the same relative frequency as the three PIREPS categories. In this way, the model severity would be phrased as low, medium, and high and the thresholds between them could be constructed such that low severity occurs with the same frequency as trace and

FIG. 13. Maps of icing index at 1200 UTC 26 Jan 2017: (a) satellite icing potential from Meteosat Second Generation (on a 0.25° grid), (b) control operational index, (c) SFIP-LI-BCF, and (d) vertical maximum of SFIP-LI-BCF. In (b) and (c) the 3D model data have been converted to 2D by filtering using a threshold of 0.1. Model indices are calculated from $T + 36$ model fields and are shown on the model's 0.23° longitude grid. The locations of three reported icing incidents are shown by asterisks.

light in the PIREPS, medium with the same frequency as moderate, and high with the same frequency as severe. Alternatively the thresholds could be chosen as a way to calibrate one NWP center's icing severity predictions against another center's predictions. Both of these are areas for future work.

With these caveats regarding the definition of low, medium, and high likelihood and low, medium, and high severity, attention is now focused on the aesthetics of how the information could be displayed in order to be most useful. Figure 15 shows a combination of icing likelihood and icing severity using a simple 3×3 color matrix. The map shows that the three reported incidents all occurred near regions of both high likelihood and high severity. Meanwhile, the time–height plots indicate regions of high likelihood of icing near the observed events. In terms of severity, the observed events occurred close to regions

forecast, with a lead-time of 30–36 h, to have in-cloud condensate amount in the top decile of the multi-month global climatology.

Although the performance of the different indices in this case study is rather subjective, it is consistent with what was assessed quantitatively in section 3a: the control index tends to overpredict the risk of icing and the index often produces high values leading to little scope for differentiating between levels of risk. The SFIP-LI-BCF by comparison produces numbers that span the range of values from 0 to 1, and high values are focused on smaller regions both horizontally and vertically.

A final point for discussion is the difference in the quantitative assessment of skill depending on whether the forecasts are evaluated on a 0.25° , 0.5° , or 1.25° grid. Table 1 shows that RSS is systematically lower when assessed on the finer grid, meaning that the

FIG. 14. Time–height cross sections of (a)–(c) control and (d)–(f) SFIP-LI-BCF for the three locations in Table 3 valid from 0000 UTC 25 Jan 2017. The time and height of the three reported icing incidents are shown by asterisks. The model data are taken from the model's 0.23° longitude grid. A height of 6000 ft corresponds to 1829 m.

forecast is less capable of predicting the correct location of events on a 0.25° scale than on a 1.25° scale. This is evidence of the double-penalty effect. When regridding to 1.25° the forecast fields are effectively not being penalized for displacement errors on scales up to 1.25° . Table 2 shows that when the control index is evaluated on a 0.25° or 0.5° grid, the reliability is very poor with negative BSS, as previously discussed for the 1.25° grid.

When comparing the performance of the different indices on the different grids using BSS, the SFIP-based indices all have some reliability, irrespective of the grid, whereas the control index never does. Meanwhile, although the discriminatory skill (based on RSS) is not as good as the control for the SFIP indices on the 0.25° grid, this reduction is less apparent at 0.5° and hardly present at 1.25° (where the RSS for the control and SFIP indices is similar). If the indices had only been evaluated on the 0.25° grid, one would conclude that an improvement in BSS was only

possible at the expense of RSS. However, on the 1.25° grid the SFIP indices maintain RSS while vastly improving BSS. Although this has required regridding to 1.25° , this is the grid that operational icing indices are currently shown on.

5. Summary and Conclusions

An icing index from the literature has been implemented in the Met Office forecasting system. Ground-based remote sensing observations are used to evaluate model hindcasts of cloud cover in the supercooled liquid water temperature range. The model is shown to overestimate the frequency of occurrence of liquid cloud. In the temperature range from 0° to -30°C , if one ignores the phase of the cloud predicted by the model, but just assesses the frequency of occurrence of cloud in general, the biases are reduced. Satellite-based retrievals of icing potential are used to quantitatively assess the skill of the model icing hindcasts. It is shown

FIG. 15. Icing index displayed using the 3×3 likelihood–severity color matrix discussed in the text: (a) map valid at 1200 UTC 26 Jan 2017, from $T + 36$ model fields (the two-dimensional likelihood is calculated from the vertical maximum SFIP-LI-BCF; the severity is calculated relative to the centiles of 2 months of global three-dimensional in-cloud condensate amount, with the two-dimensional field being maximum value in each column); (b)–(d) time–height cross section of forecasts from $T + 0$ to $T + 36$ h at three locations indicated by asterisks in (a) (the likelihood is calculated from SFIP-LI-BCF and the severity from the centiles of 2 months of global three-dimensional in-cloud condensate amount). The model data are taken from the model’s 0.23° longitude grid. A height of 6000 ft corresponds to 1829 m.

that the current icing index performs poorly, in line with anecdotal evidence that it is unreliable and not very useful. The new icing index, by comparison, shows evidence of reliability. When one element of the index, the relative humidity membership function, is replaced with the model-predicted bulk cloud fraction, it can improve reliability when assessed using passive satellite data on a 0.25° grid, while greatly reducing cloud occurrence frequency biases when compared to ground-based active sensing. Finally, to help with the rapid interpretation of the index in an operational context, a new color scale is proposed which quantifies both likelihood and severity in a traceable way. The SFIP-LI-BCF index is now being made available operationally in real time to aviation meteorologists at the Met Office.

There are several areas for potential future work: (i) the inclusion of information from an ensemble-based

prediction system, (ii) the calibration of the severity component against pilot reports or severity indices produced by other NWP centers, (iii) making use of more recent satellite icing products (e.g., [Smith et al. 2012](#)), (iv) the inclusion of freezing rain, and (v) the assessment of the index, currently only evaluated in the global forecasting system, in convection-permitting regional modeling configurations.

Acknowledgments. Thanks are due to the United Kingdom’s Civil Aviation Authority (CAA) for funding CM, KB, and RB. This work has benefitted from discussions with: Bob Lunnon, Helen Wells, Piers Buchanan, Teil Howard, Peter Francis, Katie O’Boyle, and Ruth Steele. Some of the observational data were obtained from the Atmospheric Radiation Measurement (ARM) Climate Research Facility, a U.S. Department

of Energy Office of Science user facility sponsored by the Office of Biological and Environmental Research.

REFERENCES

- Belo-Pereira, M., 2015: Comparison of in-flight aircraft icing algorithms based on ECMWF forecasts. *Meteor. Appl.*, **22**, 705–715, <https://doi.org/10.1002/met.1505>.
- Bernstein, B., F. McDonough, M. Politovich, B. Brown, T. Ratvasky, D. Miller, C. Wolff, and G. Cuning, 2005: Current icing potential: Algorithm description and comparison with aircraft observations. *J. Appl. Meteor.*, **44**, 969–986, <https://doi.org/10.1175/JAM2246.1>.
- Bowyer, R. L., and P. G. Gill, 2019: Objective verification of global in-flight icing forecasts using satellite observations: Verification of WAFS icing forecasts using satellite observations. *Meteor. Appl.*, <https://doi.org/10.1002/met.1788>, in press.
- Bröcker, J., and L. A. Smith, 2007: Increasing the reliability of reliability diagrams. *Wea. Forecasting*, **22**, 651–661, <https://doi.org/10.1175/WAF993.1>.
- Brooks, M. E., 2005: Clouds and radiation: Using radar/lidar synergy to evaluate and improve models. Ph.D. thesis, University of Reading, <http://www.met.reading.ac.uk/Library/phdsearch.html>.
- Brown, B. G., G. Thompson, R. T. Buintjes, R. Bullock, and T. Kane, 1997: Intercomparison of in-flight icing algorithms. Part II: Statistical verification results. *Wea. Forecasting*, **12**, 890–914, [https://doi.org/10.1175/1520-0434\(1997\)012<0890:IOIFIA>2.0.CO;2](https://doi.org/10.1175/1520-0434(1997)012<0890:IOIFIA>2.0.CO;2).
- Bruford, D., 1997: *Air Pilot's Practical and Theoretical Weather Manual*. Airlife, 158 pp.
- Cole, J., and W. Sand, 1991: Statistical study of aircraft icing accidents. *Proc. 29th Aerospace Science Meeting*, Reno, NV, American Institute of Aeronautics and Astronautics, AIAA 91-0558, <https://doi.org/10.2514/6.1991-558>.
- Forbes, R., and M. Ahlgrim, 2014: On the representation of high-latitude boundary layer mixed-phase cloud in the ECMWF global model. *Mon. Wea. Rev.*, **142**, 3425–3445, <https://doi.org/10.1175/MWR-D-13-00325.1>.
- Francis, P., 2007: Detection of aircraft icing conditions over Europe using SEVIRI data. *2007 EUMETSAT Meteorological Satellite Conf./15th AMS Satellite Meteorology and Oceanography Conf.*, Amsterdam, Netherlands, AMS/EUMETSAT, 8 pp., https://www.eumetsat.int/website/home/News/ConferencesandEvents/DAT_2042918.html.
- Furtado, K., and P. Field, 2017: The role of ice microphysics parametrizations in determining the prevalence of supercooled liquid water in high-resolution simulations of a Southern Ocean midlatitude cyclone. *J. Atmos. Sci.*, **74**, 2001–2021, <https://doi.org/10.1175/JAS-D-16-0165.1>.
- , P. R. Field, I. A. Boutle, C. J. Morcrette, and J. M. Wilkinson, 2016: A physically based subgrid parameterization for the production and maintenance of mixed-phase clouds in a general circulation model. *J. Atmos. Sci.*, **73**, 279–291, <https://doi.org/10.1175/JAS-D-15-0021.1>.
- Hogan, R. J., C. Jakob, and A. J. Illingworth, 2001: Comparison of ECMWF winter-season cloud fraction with radar-derived values. *J. Appl. Meteor.*, **40**, 513–525, [https://doi.org/10.1175/1520-0450\(2001\)040<0513:COEWSC>2.0.CO;2](https://doi.org/10.1175/1520-0450(2001)040<0513:COEWSC>2.0.CO;2).
- ICAO, 2018: Annex 3—Meteorological services for international air navigation. Tech. Rep., International Civil Aviation Organization, 218 pp., <https://store.icao.int/annex-3-meteorological-services-for-international-air-navigation-digital.html>.
- Illingworth, A. J., and Coauthors, 2007: CloudNet: Continuous evaluation of cloud profiles in seven operational models using ground-based observations. *Bull. Amer. Meteor. Soc.*, **88**, 883–898, <https://doi.org/10.1175/BAMS-88-6-883>.
- Kalesse, H., and Coauthors, 2016: Understanding rapid changes in phase partitioning between cloud liquid and ice in stratiform mixed-phase clouds: An Arctic case study. *Mon. Wea. Rev.*, **144**, 4805–4826, <https://doi.org/10.1175/MWR-D-16-0155.1>.
- Kalinka, F., K. Roloff, J. Tendel, and T. Hauf, 2017: The in-flight icing warning system ADWICE for European airspace—Current structure, recent improvements and verification results. *Meteor. Z.*, **26**, 441–455, <https://doi.org/10.1127/metz/2017/0756>.
- Lankford, T. T., 2000: *Aviation Icing*. McGraw-Hill, 352 pp.
- , 2001: *Aviation Weather Handbook*. McGraw-Hill, 656 pp.
- Le Bot, C., 2004: SIGMA: System of Icing Geographic Identification in Meteorology for Aviation. Preprints, *11th Conf. on Aviation Range and Aerospace Meteorology*, Hyannis, MA, Amer. Meteor. Soc., P6.5, <https://ams.confex.com/ams/pdfpapers/81704.pdf>.
- Mason, I. B., 2003: Binary events. *Forecast Verification—A Practitioner's Guide in Atmospheric Science*, I. T. Jolliffe and D. B. Stephenson, Eds., Wiley, 37–76.
- McDonough, F., B. C. Bernstein, M. K. Politovich, and C. A. Wolff, 2004: The forecast icing potential algorithm. *42nd AIAA Aerospace Sciences Meeting and Exhibit*, Reno, NV, American Institute of Aeronautics and Astronautics, AIAA 2004-0231, <https://doi.org/10.2514/6.2004-231>.
- Minnis, P., and Coauthors, 2005: Near real-time satellite cloud products for nowcasting applications. *Proc. WWRP Symp. on Nowcasting and Very Short Range Forecasting*, Toulouse, France, WWRP, 4.19, <https://enso.larc.nasa.gov/icing/pub/conf/Minnis.4.19.WSN05.pdf>.
- Morcrette, C. J., E. J. O'Connor, and J. C. Petch, 2012: Evaluation of two cloud parametrization schemes using ARM and Cloud-Net observations. *Quart. J. Roy. Meteor. Soc.*, **138**, 964–979, <https://doi.org/10.1002/qj.969>.
- Politovich, M. K., 2003: Aircraft icing. *Encyclopedia of Atmospheric Science*, J. R. Holton, J. A. Curry, and J. A. Pyle, Eds., Academic Press, 68–75.
- Rogers, R. R., and M. K. Yau, 1989: *A Short Course in Cloud Physics*. 3rd ed. Butterworth-Heinemann, 304 pp.
- Schmetz, J., P. Pili, S. Tjemkes, D. Just, J. Kerkmann, S. Rota, and A. Ratier, 2002: An introduction to Meteosat Second Generation (MSG). *Bull. Amer. Meteor. Soc.*, **83**, 977–992, [https://doi.org/10.1175/1520-0477\(2002\)083<0977:AITMSG>2.3.CO;2](https://doi.org/10.1175/1520-0477(2002)083<0977:AITMSG>2.3.CO;2).
- Schultz, P., and M. K. Politovich, 1992: Towards the improvement of aircraft-icing forecasts for the continental United States. *Wea. Forecasting*, **7**, 491–500, [https://doi.org/10.1175/1520-0434\(1992\)007<0491:TTIOAI>2.0.CO;2](https://doi.org/10.1175/1520-0434(1992)007<0491:TTIOAI>2.0.CO;2).
- Smith, W. L. J., P. Minnis, C. Leeger, D. Spangenberg, R. Palikonda, and L. Nguyen, 2012: Determining the flight icing threat to aircraft with single-layer cloud parameters derived from operational satellite data. *J. Appl. Meteor. Climatol.*, **51**, 1794–1810, <https://doi.org/10.1175/JAMC-D-12-057.1>.
- Stokes, G. M., and S. E. Schwartz, 1994: The Atmospheric Radiation Measurement (ARM) program: Programmatic background and design of the cloud and radiation test bed. *Bull. Amer. Meteor. Soc.*, **75**, 1201–1221, [https://doi.org/10.1175/1520-0477\(1994\)075<1201:TARMPP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1994)075<1201:TARMPP>2.0.CO;2).
- Sundqvist, H., E. Berge, and J. E. Kristjánsson, 1989: Condensation and cloud parameterization studies with a mesoscale numerical weather prediction model. *Mon. Wea. Rev.*, **117**, 1641–1657, [https://doi.org/10.1175/1520-0493\(1989\)117<1641:CACPSW>2.0.CO;2](https://doi.org/10.1175/1520-0493(1989)117<1641:CACPSW>2.0.CO;2).

- Tafferer, A., T. Hauf, C. Leifeld, T. Hafner, H. Leykauf, and U. Voigt, 2003: ADWICE: Advanced diagnosis and warning system for aircraft icing environments. *Wea. Forecasting*, **18**, 184–203, [https://doi.org/10.1175/1520-0434\(2003\)018<0184:AADAWS>2.0.CO;2](https://doi.org/10.1175/1520-0434(2003)018<0184:AADAWS>2.0.CO;2).
- Tiedtke, M., 1993: Representation of clouds in large-scale models. *Mon. Wea. Rev.*, **121**, 3040–3061, [https://doi.org/10.1175/1520-0493\(1993\)121<3040:ROCILS>2.0.CO;2](https://doi.org/10.1175/1520-0493(1993)121<3040:ROCILS>2.0.CO;2).
- Walters, D., and Coauthors, 2017: The Met Office Unified Model Global Atmosphere 6.0/6.1 and JULES Global Land 6.0/6.1 configurations. *Geosci. Model Dev.*, **10**, 1487–1520, <https://doi.org/10.5194/gmd-10-1487-2017>.
- Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. Elsevier, 648 pp.
- Wilson, D. R., A. C. Bushell, A. M. Kerr-Munslow, J. D. Price, and C. J. Morcrette, 2008: PC2: A prognostic cloud fraction and condensation scheme. I: Scheme description. *Quart. J. Roy. Meteor. Soc.*, **134**, 2093–2107, <https://doi.org/10.1002/qj.333>.
- Wolff, C., F. McDonough, M. K. Politovich, and G. Cuning, 2009: The forecast icing product: Recent upgrades and improvements. *First AIAA Atmospheric and Space Environments Conf.*, San Antonio, TX, American Institute of Aeronautics and Astronautics, AIAA 2009-3531, <https://doi.org/10.2514/6.2009-3531>.