

The Ability of the ICE-T Microphysics Scheme in HARMONIE-AROME to Predict Aircraft Icing

BJØRG JENNY KOKKVOLL ENGDAHL,^{a,b} TIM CARLSEN,^b MORTEN KØLTZOW,^a AND TRUDE STORELVMØ^b

^a Norwegian Meteorological Institute, Oslo, Norway

^b Department of Geosciences, University of Oslo, Oslo, Norway

(Manuscript received 21 June 2021, in final form 3 December 2021)

ABSTRACT: In-cloud icing is a major hazard for aviation traffic and forecasting of these events is an important task for weather agencies worldwide. A common tool utilized by aviation forecasters is an icing intensity index based on supercooled liquid water from numerical weather prediction models. We seek to validate the modified microphysics scheme, ICE-T, in the HARMONIE-AROME numerical weather prediction model with respect to aircraft icing. Icing intensities and supercooled liquid water derived from two 3-month winter season simulations with the original microphysics code, CTRL, and ICE-T are compared with pilot reports of icing and satellite retrieved values of liquid and ice water content from *CloudSat-CALIPSO* and liquid water path from AMSR-2. The results show increased supercooled liquid water and higher icing indices in ICE-T. Several different thresholds and sizes of neighborhood areas for icing forecasts were tested out, and ICE-T captures more of the reported icing events for all thresholds and nearly all neighborhood areas. With a higher frequency of forecasted icing, a higher false alarm ratio cannot be ruled out, but is not possible to quantify due to the lack of no-icing observations. The increased liquid water content in ICE-T shows a better match with the retrieved satellite observations, yet the values are still greatly underestimated at lower levels. Future studies should investigate this issue further, as liquid water content also has implications for downstream processes such as the cloud radiative effect, latent heat release, and precipitation.

KEYWORDS: Cloud microphysics; Satellite observations; Numerical weather prediction/forecasting; Operational forecasting; Cloud parameterizations

1. Introduction


In-cloud icing is a major hazard for aviation traffic. When supercooled liquid water (SLW) comes into contact with a solid object such as an aircraft wing, the water will freeze and ice can accumulate upon it. The ice can alter the shape of the wings or airframe and lead to a loss of lift, and subsequently loss of control (Isaac et al. 2001; Serke et al. 2014). Most larger aircraft have deicing equipment, such as heated surfaces on the wings, yet atmospheric icing is an ongoing threat to aviation safety (Cao et al. 2018). Therefore, forecasts of icing conditions are one of the most important tasks for weather agencies worldwide (Politovich 2003).

Over the years, numerous studies have been conducted in order to improve the forecasts and verify current systems of atmospheric icing detections. Bernstein et al. (2005) designed the current icing potential (CIP) system based on a combination of available observations such as satellite, radar, surface, pilot reports, and numerical weather prediction (NWP) products, which is still the basis for the Forecast Icing Potential (FIP) issued by the National Oceanic and Atmospheric Administration (NOAA). Ellrod and Bailey (2007) tested out cloud top heights derived from geostationary satellites as a proxy for aircraft icing potential, but the system was

outperformed by the CIP based forecasts. Belo-Pereira (2015) evaluated different icing algorithms with input from the European Centre for Medium-Range Weather Forecasts (ECMWF) forecasts, among others a simplified version of the FIP, and found that inclusion of cloud liquid water was essential to avoid too many false alarms. Thompson et al. (2017) investigated the performance of a purely NWP based system of predicting aircraft icing and found that the NWP model was able to capture these events to a large extent.

Similar to the CIP/FIP method, common ways for aviation forecasters to identify icing conditions are to utilize observations such as reports from aircraft (AIREP), satellite images, radar, and radiosonde profiles, in addition to NWP model products such as pre-calculated icing indices, atmospheric profiles, and cloud characteristics. The icing indices derived from the NWP model output are often based on SLW (Thompson et al. 2017; Boudala et al. 2019). Unfortunately, the microphysical processes responsible for the generation and depletion of SLW are complex, and accurate representation of SLW in NWP models is difficult. Many NWP models tend to produce too much ice at the expense of SLW (Nygaard et al. 2011; Liu et al. 2011). This can lead to too few forecasts of icing conditions, especially of the most severe categories. Morcrette et al. (2019) accounted for this deficit by considering both the ice water content (IWC) and liquid water content (LWC) in their final version of the simplified forecast icing potential from Belo-Pereira (2015).

At the Norwegian Meteorological Institute (MET-Norway), the NWP model HARMONIE-AROME is used for operational short-term forecasts (Müller et al. 2017; Frogner

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

Corresponding author: Bjørg Jenny Kokkvoll Engdahl, bjorgjike@met.no

DOI: 10.1175/WAF-D-21-0104.1

© 2022 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

et al. 2019) and as the basis for an icing intensity index. Engdahl et al. (2020a) modified the cloud microphysics scheme in HARMONIE-AROME in order to improve the representation of SLW. The new scheme was given the name ICE-T. In a follow-up study, Engdahl et al. (2020b) validated ICE-T in simulations over a 3-month period for the winter season 2016/17, and showed how the new microphysics scheme led to an improvement of the estimates of ice loads on transmission lines caused by atmospheric icing.

In the current study, we shift the focus from the near-surface to the atmosphere. We want to investigate the general vertical distribution of liquid and ice water content, but also more specifically we want to evaluate ICE-T's ability to forecast winter season aircraft icing events. It is important to note that we are only validating the icing intensities and SLW from the NWP model, and not the icing forecasts issued by the aviation forecasters themselves (airmets and sigmets).

We utilize the same model output dataset generated in Engdahl et al. (2020b), and run this through the icing algorithm for operational aircraft icing forecasts, before we validate the results against pilot reports of aircraft icing (AIREPs). However, AIREPs are quite subjective and biased (e.g., Brown et al. 1997; Kalinka et al. 2017). Bowyer and Gill (2019) used geostationary satellites for objective verification of the icing potential forecast from the World Area Forecast System (WAFS). In the current study, we validate the vertical distribution of liquid water content (LWC) and ice water content, and the liquid water path from the simulations, against satellite retrieved data from *CloudSat-CALIPSO* and *AMS-R2*, for an objective validation of the distribution of SLW.

The remainder of the paper is organized as follows: the model and simulation configurations, icing algorithm, and observations are described in section 2. The results are presented in section 3, while the last section includes a brief summary and conclusions.

2. Methodology

a. HARMONIE-AROME

High Resolution Limited Area Model–Aire Limitée Adaptation Dynamique Développement International (HIRLAM-ALADIN) Research on Mesoscale Operational NWP in Euromed–Applications of Research to Operations at Mesoscale (HARMONIE-AROME) is the operational NWP model applied at MET-Norway and many other European countries (Müller et al. 2017; Bengtsson et al. 2017). HARMONIE-AROME is based on the AROME model developed at Météo-France (Seity et al. 2011; Brousseau et al. 2016) and is a nonhydrostatic and convection permitting model. Deep convection is resolved explicitly, while shallow convection is parameterized by the EDMFm scheme (Neggers et al. 2009), subgrid condensation and cloud fraction is determined by a statistical cloud scheme following Chaboureaud and Bechtold (2002) with input from the shallow convection and the HARTU turbulence schemes (Bengtsson et al. 2017). Surface parameterization is dealt with by the coupled independent surface model SURFEX (Masson et al. 2013), and the

radiation follows Morcrette et al. (2008) for short wave radiation and uses the rapid radiative transfer model by Mlawer et al. (1997) for long wave. More details on the HARMONIE-AROME version 40h1.1 physics parameterizations can be found in Bengtsson et al. (2017).

The cloud microphysics scheme in HARMONIE-AROME is a single-moment bulk microphysics scheme called ICE3 (Pinty and Jabouille 1998; Mascart and Bougeault 2011), with prognostic calculations of mass of cloud water, rain, cloud ice, snow, and graupel, and with physics essentially based on Lin et al. (1983). Studies by Liu et al. (2011) have shown that microphysics schemes with Lin-based physics tend to glaciate clouds prematurely and produce too little SLW as a result. Several changes to the ICE3 scheme were introduced in an option called OCND2 (Müller et al. 2017; Bengtsson et al. 2017). The changes were motivated by an excess of ice clouds and too little liquid water.

Engdahl et al. (2020a) showed in a series of idealized experiments that even with the changes introduced in OCND2, the microphysics scheme in HARMONIE-AROME had a tendency to prematurely initiate ice and consequently quickly deplete SLW. Therefore, they introduced further changes to the scheme based on the Thompson scheme (Thompson et al. 2004, 2008). The changes included stricter conditions for ice initiation, reduced efficiencies for snow and graupel to collect cloud water, and also a variable intercept parameter to allow for variable size distributions for rain. A detailed description of the changes can be found in Engdahl et al. (2020a, their Table 1). The final suite of changes showed a prolonged existence and higher amounts of SLW.

In the current study we will use the same 3-month dataset generated in Engdahl et al. (2020b). They ran two parallel simulations, one with the original microphysics scheme (CTRL) and one with the modified scheme (ICE-T) over the winter season 2016/17. Here, we will investigate the modified microphysics scheme's ability to forecast the conditions necessary for aviation icing, compared with the original scheme. HARMONIE-AROME (version 40h1.1) is the base for both simulations. CTRL is the original code with the OCND2 option active, including a bug fix from Engdahl et al. (2020a), while ICE-T contains all the modifications from the same study.

The resolution is similar to the operational setup at MET-Norway with 2.5-km distance between the grid points and 65 vertical levels. The domain consists of 739×949 grid points in the horizontal and covers Norway and the Norwegian Sea, Sweden, and parts of Finland, and can be seen in Fig. 1. There is no local data assimilation, and the initial and boundary conditions are taken from the ECMWF Integrated Forecast System–High Resolution model forecasts every hour, with a horizontal grid spacing of approximately 9 km. Forecasts are produced every day at 0000 UTC and are run for 36 h, with output every hour. To allow for spinup of clouds and precipitation we only use hours 13–36 from each forecast.

b. Icing algorithm

To calculate the icing intensities from the model simulations, we used the icing algorithm currently operational at MET-Norway. The routine is based on Makkonen (2000)

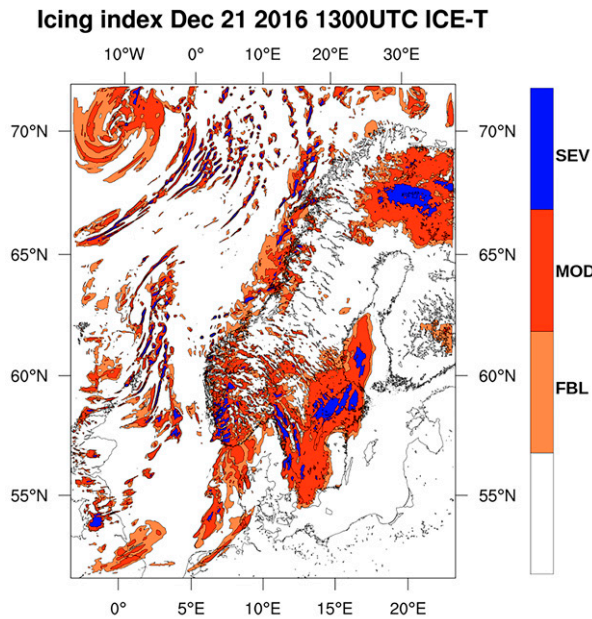


FIG. 1. Map of column maximum icing intensities from ICE-T at 1300 UTC 21 Dec 2016.

which is widely used in other icing studies (Hämäläinen and Niemelä 2017; Thompson et al. 2017; Engdahl et al. 2020b). Despite many uncertain assumptions that could be improved, we will make only minor changes to the MET-Norway icing algorithm since the purpose of this study is to investigate the relative differences between two model simulations. The icing algorithm calculates ice accumulation on a cylinder with the mass accretion rate dM/dt given by

$$\frac{dM}{dt} = \alpha_1 \alpha_2 \alpha_3 w v A. \quad (1)$$

Here, w is the mass concentration of the hydrometeors, v is the velocity relative to the object, A is the cross-sectional area of the object, α_1 represents the collision efficiency of the particles, α_2 is the sticking efficiency, and α_3 is the accretion efficiency.

The original icing algorithm model at MET-Norway uses all hydrometeors, humidity, and temperature from the NWP model as input. Airspeed relative to the ice accretion object is set to 92 m s^{-1} , which is assumed to be a typical aircraft speed during takeoff and landing when icing normally occurs. Wind speed is assumed to be negligible compared to the speed of the aircraft and is not used. The diameter of the ice-free aircraft wing is set to 15 cm.

The common way to use Eq. (1) is to only use liquid hydrometeors as input for w . However, the operational algorithm also includes the solid species snow, graupel, and cloud ice, whenever liquid water is also present, yet with a low impact. As α_2 is set to 0.01 for solid hydrometeors, only 1% of the mass contributes to the ice accumulation. The idea behind this proxy for SLW is that cloud water and rain are under predicted in mixed-phase clouds. Tests have shown little

TABLE 1. Thresholds based on the ice accumulation rates used to determine icing indices in the operational icing routine at MET-Norway.

Index	Ice accretion rate (cm h^{-1})	Intensity
1	0.1–0.625	Trace
2	0.625–2.5	Light
3	2.5–7.5	Moderate
4	>7.5	Severe

difference in the resulting icing when including ice species. Thus, to simplify our calculations, only liquid hydrometeors are used in this study.

The parameter α_1 is dependent on the median volume diameter, which is diagnosed based on the mass of rain and cloud water. The number concentration, N_d , of rain is also diagnosed from the mass mixing ratio, while for cloud water it is assumed to be $100 \text{ particles cm}^{-3}$, which is a common assumption when observations are lacking. α_2 is set to unity (all liquid particles stick to the object), and α_3 is calculated based on the temperature, airspeed, mass of SLW, and pressure [see Eq. (3.14) and detailed explanation in Makkonen (2000)].

As we are interested in the instantaneous icing intensity, the icing indices are calculated for each hour, and reset for the next hour, so ice does not accumulate longer than 1 h. The ice accumulation rate is converted into icing indices, based on simple thresholds displayed in Table 1. These thresholds are based on Politovich (2009) and are also used in Boudala et al. (2019). The idea of using icing indices based on the accumulated amount of ice on an airfoil can be traced back to the 1940s, when a cylinder placed on top of Mt. Washington, New Hampshire, was used as a reference for an aircraft wing. The rate of ice accretion on the cylinder was then translated into icing intensities (Jeck 2001).

However, for an individual aircraft, the accumulated rates depend on several variable factors such as aircraft speed, angle of attack, ice density, and type of aircraft, which the thresholds given in Table 1 do not account for. Therefore, these thresholds merely represent a guideline for icing intensities and cannot be directly translated into the icing severities reported by the pilots. The final output from the MET-Norway icing algorithm is an icing intensity index based on the available SLW. In Fig. 1 the column maximum icing indices are shown for ICE-T at 1300 UTC 21 December 2016, a date when icing was reported at the northern, western, and eastern parts of Norway.

c. Observations

1) AIREPS

Whenever an aircraft experiences atmospheric icing the pilots can issue a report in which they give information about the location, altitude, time, and severity of the icing. In Norwegian air space the reports will be received by the aviation forecasters on duty and disseminated to all other pilots in an AIREP. The term AIREP is used differently in other parts of

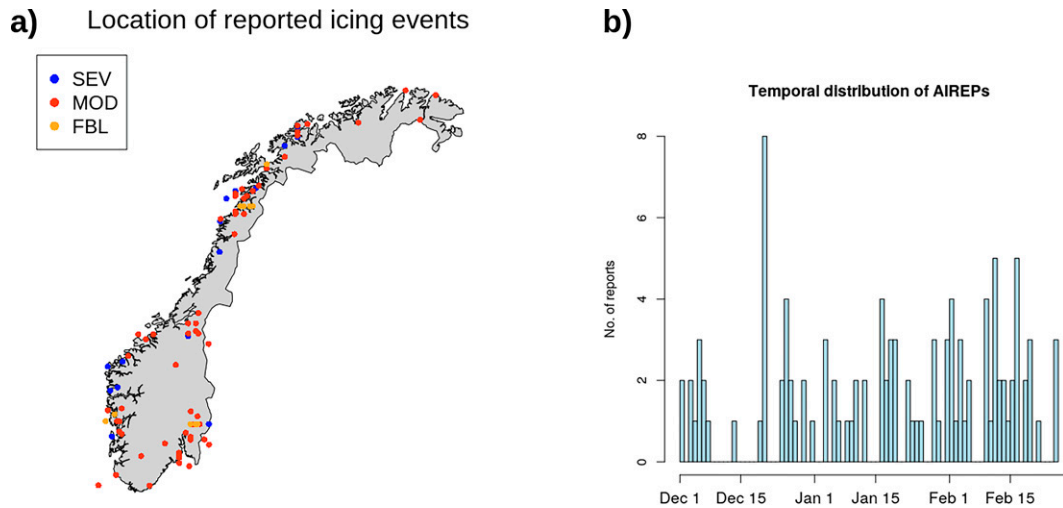


FIG. 2. (a) Location of reported icing events from AIREPs during 1 Dec 2016–28 Feb 2017. Blue dots represent severe icing (SEV), red dots represent moderate icing (MOD), and orange dots represent light icing (FBL). (b) Temporal distribution of the AIREPs.

the world, and the AIREPs used in this study are comparable to the more common PIREPs. AIREPs provide valuable in situ information directly from the users of the forecasts, and can be used both as warnings to other aircraft and as general feedback to the aviation forecasters.

AIREPs and PIREPs are also used for validation purposes in several studies (e.g., [Belo-Pereira 2015](#); [Thompson et al. 2017](#); [Kalinka et al. 2017](#)). Unfortunately, a number of problematic properties with AIREPs need consideration. For instance, we only have reports of experienced icing and no null cases (i.e., cases with observations of no icing), making the evaluation of false alarms difficult. Experience also suggests that pilots are more likely to report icing when it is not forecasted (missed), rather than when it is forecasted correctly. Experienced icing will also vary with type of aircraft, the pilot's subjective intuition and measures taken to avoid icing, i.e., by applying de-icing equipment ([Jeck 2001](#)). Furthermore, icing reports can sometimes be delayed, so that location and height may deviate from the actual icing area. These limitations are important to keep in mind when analyzing the results, yet the AIREPs still provide valuable information from the actual users of the forecasts and are the only existing in situ observations of aircraft icing ([Brown et al. 1997](#)).

There were 111 reported icing cases over Norway during the 3-month simulation period, for which 21 were reported as severe (SEV), 78 moderate (MOD), and 12 light (FBL). The reported locations and severity of each event can be seen in [Fig. 2a](#). The events are located near the main airports in Oslo and Trondheim and airports along the entire coast of Norway. [Figure 2b](#) shows the distribution of the AIREPs per day. The AIREPs are distributed quite evenly over the winter months, with the highest occurrence in a day happening on 21 December 2016 with 8 reports from several locations in Norway.

2) SATELLITE DATA

We compare the simulated vertical profiles of LWC and IWC with satellite retrievals that are based on observations with the cloud-profiling radar (CPR) on *CloudSat* ([Stephens et al. 2002](#)) and the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP; [Winker et al. 2010](#)) on *CALIPSO* for the 3-month simulation period. The LWC profiles from the *CloudSat* Radar-Only Cloud Water Product (2B-CWC-RO version P1_R05; [Austin et al. 2009](#)) and the IWC profiles from the *CloudSat* and *CALIPSO* Ice Cloud Property Product (2C-ICE version P1_R05; [Deng et al. 2010, 2015](#)) are regridded on a common pressure grid using pressure profiles from the ancillary ECMWF-AUX dataset (version P_R05, [Cronk and Partain 2017](#)) and nearest-neighbor interpolation. Subsequently, the profiles are averaged over the model domain for the 3-month period and converted to mixing ratios (g kg^{-1}) by means of the air density that is calculated from the ancillary temperature and pressure data for each radar bin. The LWC and IWC profiles were screened using the provided quality flags. The retrieval uncertainty for the individual profile varies with altitude and atmospheric conditions, but lies in the order of 20%–40%. As the radar reflectivity depends heavily on the particle sizes of the hydrometeors, the radar retrieval is strongly influenced by precipitation. The LWC profiles are derived from the composite dataset: the part of the profile warmer than 0°C is considered pure liquid, whereas in the mixed liquid and ice regime (between 0° and -20°C) the liquid and ice solutions are scaled linearly with temperature, with all ice below -20°C ([Austin et al. 2009](#)). The temperature profiles stem from the ECMWF-AUX product and were obtained by interpolating the ECMWF reanalysis product to the *CloudSat* CPR bins.

For LWP, we employ observations with the Advanced Microwave Scanning Radiometer (AMSR-2) on JAXA's *GCOM-W1* satellite ([Wentz et al. 2014](#)). The total cloud

TABLE 2. Maximum icing intensity in the vertical column in CTRL forecasts compared with AIREPs reported icing severity. Numbers in parentheses in the diagonal show the number with correct elevation of the forecasted icing. The boldface text indicates how well the maximum icing intensities in CTRL and ICE-T match the reported icing severities.

CTRL vs AIREPs	Obs light	Obs mod	Obs sev	Total CTRL
CTRL no icing	7	23	8	38
CTRL light	1 (3)	15	5	21
CTRL mod	4	36 (18)	7	47
CTRL sev	0	4	1 (0)	5
Total obs	12	78	21	111

liquid water in a vertical column of the atmosphere is retrieved for both precipitating and non-precipitating conditions and is provided on a 0.25° resolution. For comparison with HARMONIE-AROME, we compute the 3-month mean from the daily datasets for the model domain. Note that LWP data are only available over the ocean (closest distance to land of greater than around 25 km).

3. Results

In the following we compare the icing forecasts with AIREPs observations, before we validate the vertical distribution and integrated amounts of liquid water and ice against satellite observations.

a. Comparison against AIREPs

In this section we compare how well CTRL and ICE-T are able to 1) identify the reported icing events, 2) predict the reported icing severities, and 3) simulate the vertical distribution of reported icing.

For all 111 icing events reported in the AIREPs, we extracted information from the atmospheric column closest to the reported location and time (i.e., if AIREP time = 1340 UTC, it is matched to the model output from 1400 UTC), and calculated the vertical profile of icing. The maximum icing intensity in the column from CTRL and ICE-T is compared with AIREPs in Tables 2 and 3, respectively.

The icing intensity distributions are clearly shifted toward higher icing intensities in ICE-T compared with CTRL. For the reported icing cases ICE-T (CTRL) has no icing in 28 (38) cases, light icing in 17 (21) cases, moderate icing in 45 (47) cases and severe icing in 21 (5) cases. ICE-T had a hit rate of 75% for detecting the light or greater icing cases and captured 11 events not detected by CTRL, which had a hit rate of 66% and only captured one event ICE-T did not. No geographical or temporal patterns for the events captured by ICE-T but not by CTRL were found, i.e., the improved hit rate of ICE-T cannot be related to particular weather situations or locations. It is common to also calculate the false alarm ratio (Guan et al. 2001) or the equitable threat score when the hit rate is presented. Unfortunately, the lack of null cases (reports of no icing) makes it difficult to address this.

The entries in bold in Tables 2 and 3 show how well the maximum icing intensities in CTRL and ICE-T match the

TABLE 3. As in Table 2, but for ICE-T.

ICE-T vs AIREPs	Obs light	Obs mod	Obs sev	Total ICE-T
ICE-T no icing	5	17	6	28
ICE-T light	2 (2)	12	3	17
ICE-T mod	3	34 (23)	8	45
ICE-T sev	2	15	4 (3)	21
Total obs	12	78	21	111

reported icing severities. CTRL and ICE-T only matched the observed severity in 38 and 40 of the 111 cases, respectively. Only 4 of the 21 severe icing intensities in ICE-T corresponded with SEV AIREPs, while for CTRL 1 of 5 matched. Similar numbers for moderate events are 36 out of 47 for CTRL and 34 out of 45 for ICE-T. It should be noted that these are the maximum forecasted icing intensities inside the columns, and some of the simulations had lower intensities present for observed moderate and light events at the heights of the observations. Therefore, we can only be confident that the parts of Tables 2 and 3 summing up how the forecasts underestimate the intensity (i.e., the parts above the diagonals outlined in bold), 65 for CTRL and 51 for ICE-T, are in fact erroneous forecasts with respect to icing intensity. Furthermore, moderate and severe icing events are not always separated in previous studies, but verified together since the experienced icing severity is subjective (e.g., Kalinka et al. 2017). Taking a similar approach here would give a hit rate of 48% for CTRL and 62% for ICE-T for identification of moderate/severe icing events.

The AIREPs also contain information on the height of the experienced icing. We compared the reported icing severities from the AIREPs with the maximum forecasted icing in the corresponding model heights. Some AIREPs only include one specific height, others span ranges up to 4250 m from the lowest to the highest level. When only a single level was reported, we used the closest model level, otherwise we checked the entire interval for matching icing intensities. The results are written in parentheses in Tables 2 and 3. Once again we are comparing the maximum icing intensities, which means that for categories “light” and “moderate” the maximum icing intensity inside the column could be higher, while inside the interval the maximum intensity could be lower and therefore be a match. This way the hit rate could be higher for the height intervals compared with the columns in their entirety. Figure 3 shows an illustration of how this can happen. If a pilot reports “moderate” icing at a given height interval, the maximum column intensity could be higher outside the interval, and for the column as a whole counted as a miss, while inside the interval the maximum intensity could be “moderate” and then considered as a hit.

Where the column maximum intensity matched the severe icing severities reported by the pilots, ICE-T had 3 out of 4 inside the reported intervals, and CTRL none. For AIREPs of moderate severity, CTRL matched 18 and ICE-T 23. For light icing, CTRL had 3 at the correct height, and ICE-T 2. In total, for all icing events, out of the 73 columns with forecasted icing in CTRL, 45 were inside the reported height



FIG. 3. Illustration of an atmospheric column of computed icing indices from the simulations, and comparisons with the height interval reported in the AIREPs.

intervals. Similarly, ICE-T matched in 60 out of the forecasted 83 cases.

In summary, ICE-T better matches reported icing from AIREPs in terms of detection of icing events, the icing severity and height. However, the better detection of the observed icing events can in theory be achieved by simply forecasting icing conditions at all times. To complete this comparison it is therefore necessary to also evaluate the false alarm ratio in some form.

In addition, we have so far only evaluated the forecasts at the nearest grid point which introduces an uncertainty because of the lack of accuracy in the reported locations (discussed briefly above) and the general predictability limits in NWP models. The latter might be a particular issue in the experiments performed with no data assimilation in the regional model, and since we perform the evaluation for relatively long lead times (from +13 to +36 h) where it is well known that the predictability is reduced compared to shorter lead times.

Finally, operational forecasters issue warnings based on spatial patterns of certain extents in NWP output and not based on output from single grid points. To address these issues we extend the analysis to also include calculations of hit rate and forecast frequency as functions of neighborhood areas, a method also utilized by Xu et al. (2019).

We defined a set of neighborhood areas around the location for each icing event, stretching 0 (single point), 1, 3, 5, 10, 15, and 20 grid points in each direction from the location. The

largest neighborhood then covers 41×41 grid points ($102.5 \text{ km} \times 102.5 \text{ km}$ or 10506 km^2). Since the reported icing severities in AIREPs are subjective and uncertain we focus on the binary forecasts for icing/no icing in the following. To identify icing conditions in the forecasts we use different thresholds, and icing conditions are forecasted if a minimum of one grid cell, or more than 5%, 10% or 15% of the grid cells in a neighborhood area have icing intensities of “light” or higher. It should be noted that with larger neighborhoods it becomes more likely that the simulations will have at least one grid point with icing, ultimately making the forecasts converge toward always forecasting icing conditions, which makes them useless. The method of exploring icing in a larger area and using different thresholds for icing forecasts has also been used by Guan et al. (2001).

Figure 4a shows the hit rates for all thresholds averaged over all 111 AIREPs as a function of neighborhood size for both CTRL and ICE-T. ICE-T has a higher hit rate for all neighborhoods and all thresholds except for the lowest one (any icing inside the neighborhood) for the largest neighborhoods, where both simulations have a near perfect hit rate. Furthermore, since the confidence intervals do not overlap for many of the thresholds and neighborhoods, these differences are robust and show that ICE-T has a significantly higher hit rate than CTRL.

Figure 4b displays the icing forecast frequency (how often an icing forecast would be issued) as a function of neighborhood size for each threshold, for the same locations, but calculated over all time steps in the evaluated period. A higher forecasted frequency would make it possible to improve the hit rates in ICE-T over CTRL, shown in Fig. 4a, by chance, and not necessarily because ICE-T has a better representation of icing conditions. For the lowest threshold > 0 (any icing inside the neighborhoods) CTRL and ICE-T have almost the same icing forecast frequency, and the confidence intervals are completely overlapping. There is also some overlap on the second lowest threshold $> 5\%$ for the smallest neighborhoods. This indicates that ICE-T can give a higher hit rate without increasing the false alarm ratio significantly for these neighborhoods and thresholds. For the two highest thresholds ICE-T has significantly higher icing forecast frequencies in addition to the higher hit rates. This means that ICE-T can also increase the false alarm ratio, so the potential improved quality depends on the cost/loss ratio for the end user.

b. Icing frequency and vertical distribution

To check the general icing frequency and vertical distribution of the two simulations, we analyzed all grid points within the largest neighborhoods around the AIREPs locations (41×41 grid points) for every hour of the entire season. ICE-T has a higher icing frequency in total (5.2%) compared to CTRL (3.4%). The temperature profiles of the two simulations are almost entirely overlapping (not shown), so the increased icing frequency in ICE-T comes from increased SLW.

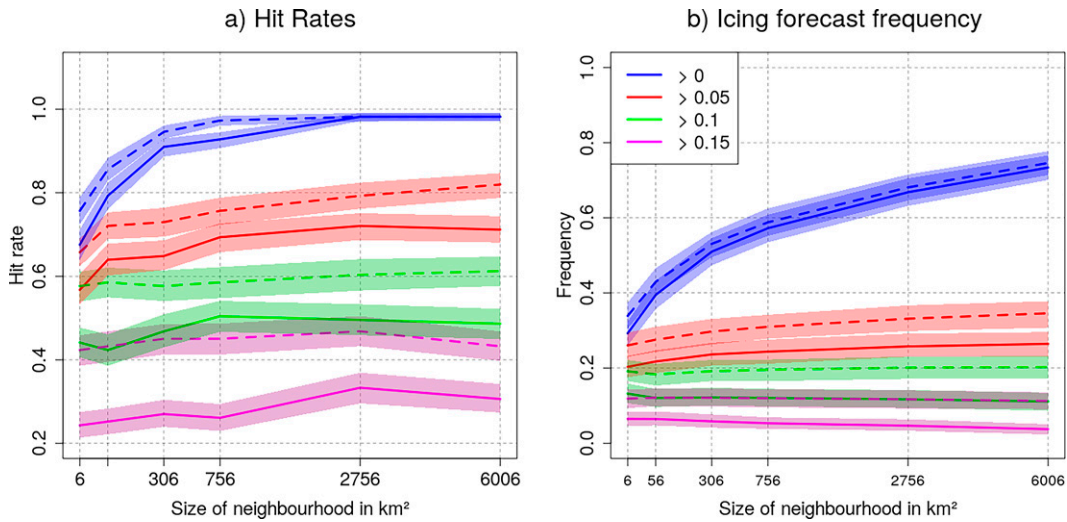


FIG. 4. (a) Hit rates and (b) icing forecast frequencies for all thresholds (see legend) as a function of neighborhood sizes for CTRL (solid lines) and ICE-T (dashed lines). Shaded areas indicate the 98% confidence intervals calculated using a bootstrap method with 1000 replicas sampled from the forecasts and observation data with replacement.

Figure 5 shows the vertical distribution of icing frequencies for CTRL and ICE-T, and the difference between the two simulations (ICE-T – CTRL). The increased icing frequency in ICE-T is evident for most heights, and especially for moderate and severe events, as ICE-T has 2 and 6 times higher frequencies than CTRL, respectively. Both simulations have their highest icing frequency between 1.5- and 2.5-km altitude, and the largest increase in frequencies by ICE-T occurs from 1 to 3.5 km. This could lead to an increased capture of icing events at these altitudes with ICE-T. Furthermore, for the reported AIREPs, it was found that both CTRL and ICE-T placed the icing slightly too low (not shown). Above 5 km, CTRL has a higher icing frequency than ICE-T, but these events are rare in both simulations.

c. Distribution of liquid water and ice

1) VERTICAL PROFILES

We averaged the vertical profiles of LWC (cloud water and rain) and IWC (cloud ice, snow, and graupel) over the entire season over almost the entire domain (the domain was slightly cropped due to technical issues). The profiles were compared to the corresponding satellite retrieved profiles over the same area, and the results can be seen in Fig. 6a.

It is clear that ICE-T has both more ice and liquid water in the atmosphere than CTRL. The corresponding temperature profiles (not shown) are so similar that the difference in amount of LWC can only be due to the microphysics. Separating the profiles into the different hydrometeors (Fig. 6c) reveals that the increased LWC and IWC in ICE-T is mainly

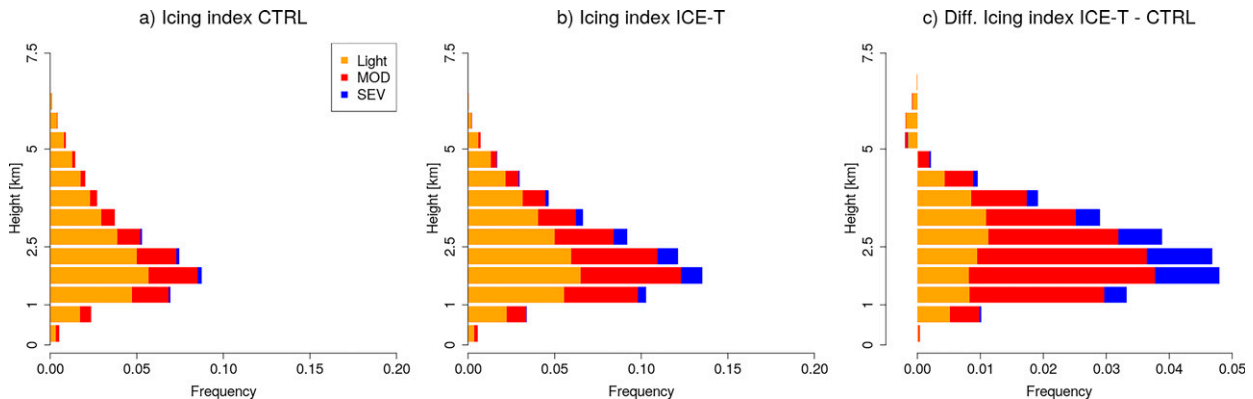


FIG. 5. Icing frequency distribution with height for (a) CTRL, (b) ICE-T, and (c) the difference in icing frequencies (ICE-T – CTRL) for a neighborhood area of 41×41 grid boxes for the 111 locations of reported aircraft icing. Each bar represents an interval of 500 m. The colored sections represent light (orange), moderate (red), and severe (blue) icing intensities.

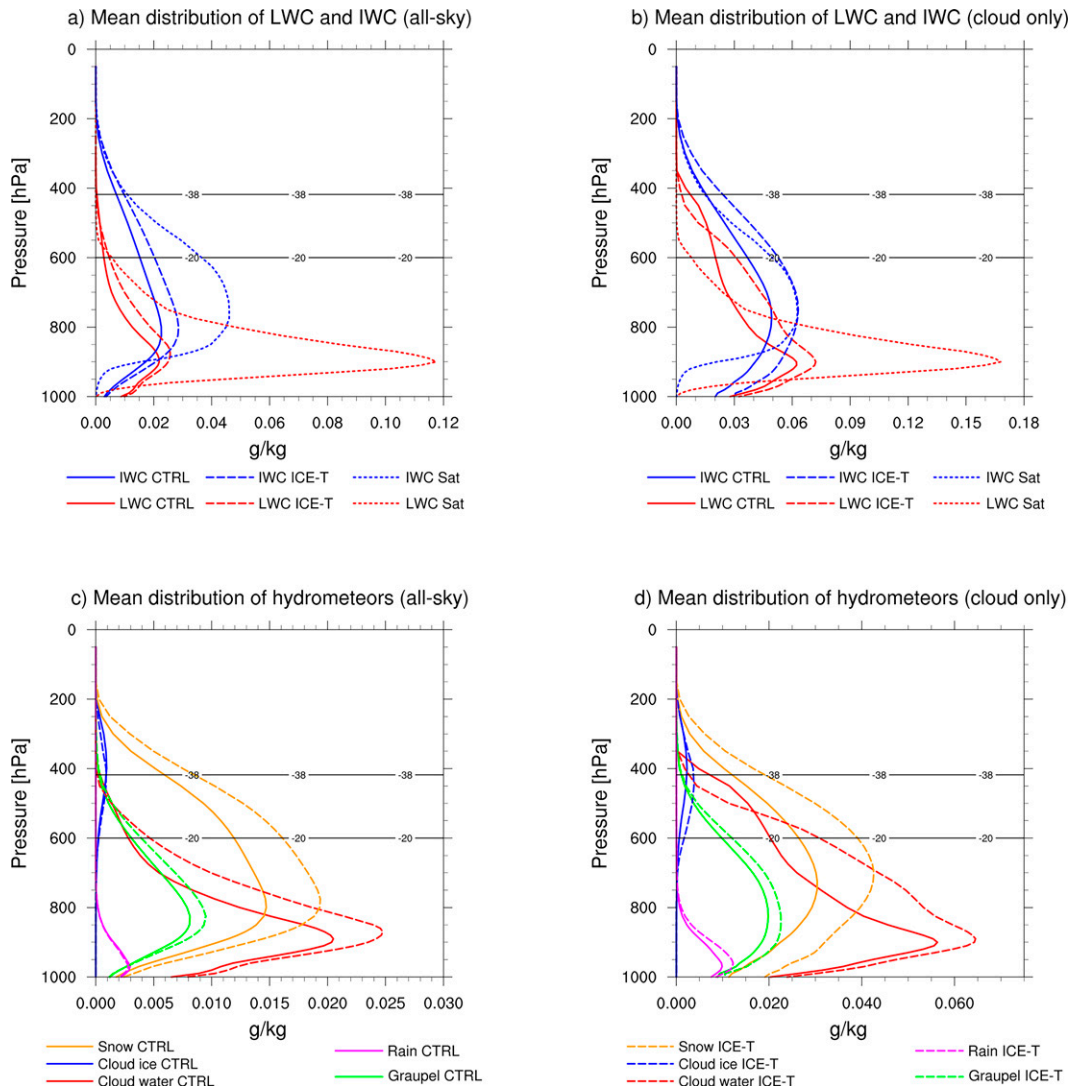


FIG. 6. Vertical distribution of mean LWC (red) and IWC (blue) in the atmosphere for CTRL (solid lines) and ICE-T (long dashed lines) and satellite-retrieved data (short dashed lines), averaged from 1 Dec 2016 to 28 Feb 2017 for (a) all-sky and (b) cloud-only. Vertical distribution of hydrometeors for CTRL (solid lines) and ICE-T (dashed lines) for (a) all-sky and (b) cloud-only. Horizontal lines represent the mean heights of -20° and -38° C. The 0° C isotherm is approximately at 1000 hPa.

due to increased content of snow and cloud water. This is in accordance with Hellmuth et al. (2021), which validated snow water content for both CTRL and ICE-T at Haukelisetter test site in southern Norway, and found that ICE-T generally had higher values than CTRL.

The increased snow water content can be traced back to several of the changes in the modified scheme described in Engdahl et al. (2020a), for example, snow riming, where rimed snow would readily be regarded as graupel in CTRL, while a greater portion would remain snow in ICE-T. Also, the mass-diameter relation for snow was changed, which results in smaller particles and subsequently slower fall speeds in ICE-T. An interesting feature is that atmospheric graupel is increased in ICE-T (Fig. 6d), even though Engdahl et al.

(2020b) showed a decrease in precipitation as graupel. This could be due to a lower graupel fall speed in ICE-T, compared with CTRL. Engdahl et al. (2020a) changed several physical parameterizations leading to more cloud water, i.e., stricter conditions for ice nucleation, and lower efficiency for snow and graupel collecting cloud water.

Both the increased snow and cloud water content are in accordance with the findings in Engdahl et al. (2020b). Liu et al. (2011) showed that the Lin-based microphysics scheme had the lowest content of hydrometeors in the atmosphere and the highest precipitation rates, which suggests high precipitation efficiency. A similar feature is seen in this study, as the CTRL simulation has a lower total condensate in the atmosphere. Combined with the results from

Engdahl et al. (2020b), which showed that CTRL had more total surface precipitation than ICE-T and a tendency to place the precipitation on the upslope of the terrain, as opposed to the lee side in ICE-T, it appears that CTRL has a higher precipitation efficiency than ICE-T.

Although ICE-T has increased IWC and LWC, when compared to the satellite retrieved profiles (Figs. 6a,b), it is evidently not enough. It appears that both simulations are lacking a substantial amount of liquid water at almost all heights. The shape of the simulated and observed profiles look similar, yet the peak simulated LWC is approximately 0.026 g kg^{-1} (ICE-T), while the observed peak is around 0.12 g kg^{-1} , almost 5 times higher.

Observed values of IWC are also considerably higher than the corresponding simulated values. The peak is around 800 hPa with maximum of 0.023 and 0.029 g kg^{-1} for CTRL and ICE-T, respectively. The highest values of IWC from the satellite data are found between 800 and 650 hPa and are approximately 0.046 g kg^{-1} . The simulated temperature profiles are a few degrees warmer than the corresponding values from the reanalysis data used for the satellite retrieval (not shown), so the lack of LWC cannot be due to a colder atmosphere in the simulations. Hellmuth et al. (2021) also found a higher IWC in their retrieved values based on remote sensing, compared with the simulations. Below 900 hPa there is an apparent overestimation of IWC by the simulations, but this is likely due to precipitating particles not captured by the satellite observations.

The satellites of the A-Train constellation pass over the domain during similar hours every day (0900–1400 UTC for ascending orbits, 0000–0400 UTC for descending orbits). This could lead to a temporal bias when compared with the simulated values. Yet, when we computed the vertical profiles from the simulations based only on output from these time intervals, only small differences were found (not shown).

To determine if the missing IWC and LWC is a sign of too few clouds or an underestimation of water content inside the clouds, we did the same averaging with the grid boxes containing any mass of hydrometeors, omitting model boxes with no hydrometeors present (Fig. 6b). The figure shows a good match between ICE-T and the observed values of IWC, and an improvement also for CTRL, suggesting that if it is too low overall, but it is reasonable within cloud, then the simulations have fewer ice clouds compared to the observations. It should be noted, however, that the simulations were run with the subgrid condensation scheme activated, which means that some of the grid boxes may have a cloud fraction less than 1. The hydrometeor mass mixing ratio output are grid box averages, and should be divided by the cloud fraction in order to retrieve the correct in-cloud condensate. Unfortunately, the cloud fraction field is not in the simulation output, so the calculation could not be made. However, the cloud fraction is mostly 1 or 0, so we do not believe this would change the results considerably.

The simulated LWC peak values are still far too low compared with the satellite data, which indicates that the lack of LWC is mainly an in-cloud liquid water content deviation and not related to the liquid cloud cover. Conversely, at higher

levels, there appears to be too much liquid water in the simulations. However, the observed satellite profiles cannot be seen as exact reference values, as they are subject to uncertainties especially in the mixed-phase temperature regime. As mentioned in the methodology section, the observed LWC is determined by linearly scaling the liquid cloud water with temperature between 0° and -20°C . According to the atmospheric reanalysis, which forms the input for the LWC retrieval algorithm, the mean height of the -20°C isotherm is around 600 hPa. Consequently, the retrieved LWC above this height is zero. However, theoretically liquid water can exist in the supercooled state at temperatures down to -38°C , which is why the simulated LWC profiles overestimate the satellite profiles at higher altitudes. This is further amplified by the somewhat warmer simulated temperature profiles as compared to the reanalysis values (not shown).

If we were to reassign the model total liquid and ice into separate liquid and ice using the same temperature-based function as the *CloudSat* retrievals, the apparent overestimation of the simulations LWC above 600 hPa and IWC below 900 hPa would be improved, yet the peak in LWC at 900 hPa would still be underestimated. Nevertheless, the modifications in the ICE-T scheme lead to a better agreement with observed condensate profiles.

2) LIQUID AND ICE WATER PATH

We also computed the integrated liquid and ice water content in the atmosphere, known as the liquid (LWP) and ice water paths (IWP). Figure 7 shows the spatial distribution of the mean LWP and IWP over the domain for the entire season for both simulations and AMSR-2 (only LWP), including the difference in LWP between the simulations and AMSR-2. As with the vertical profiles of LWC and IWC, ICE-T has higher values of LWP and IWP than CTRL. The domain averages for LWP and IWP are 50 and 94 g m^{-2} for CTRL, and 66 and 122 g m^{-2} for ICE-T.

We compared the simulated LWP with satellite retrieved values from AMSR-2 (Fig. 7c). The passive microwave radiometer observations with AMSR-2 only cover ocean areas, but are quite robust: the use of low-frequency microwave observations reduces the sensitivity to overlying ice clouds, and allows retrievals during day- and nighttime, thus reducing diurnal artifacts (Elsaesser et al. 2017). Due to difficulties in retrieving vertical information of LWC from the satellite as discussed above, the LWP as an integrated quantity offers a more rigorous comparison with the simulations. In addition, more cloud scenes are covered by the scanning instrument AMSR-2 as compared to the nadir-looking CPR on *CloudSat*.

The spatial distribution of the simulated LWP over the ocean areas corresponds very well to AMSR-2 retrievals, especially for ICE-T. For example, both simulations reproduce the high-LWP features visible in the AMSR-2 data off the southwestern tip of Norway and in the northwestern part of the domain. CTRL underestimates the LWP for most of the domain (Fig. 7d), while ICE-T also generally underestimates but also has some spots with overestimation (Fig. 7e). The domain average of LWP for AMSR-2 is 78 g m^{-2} .

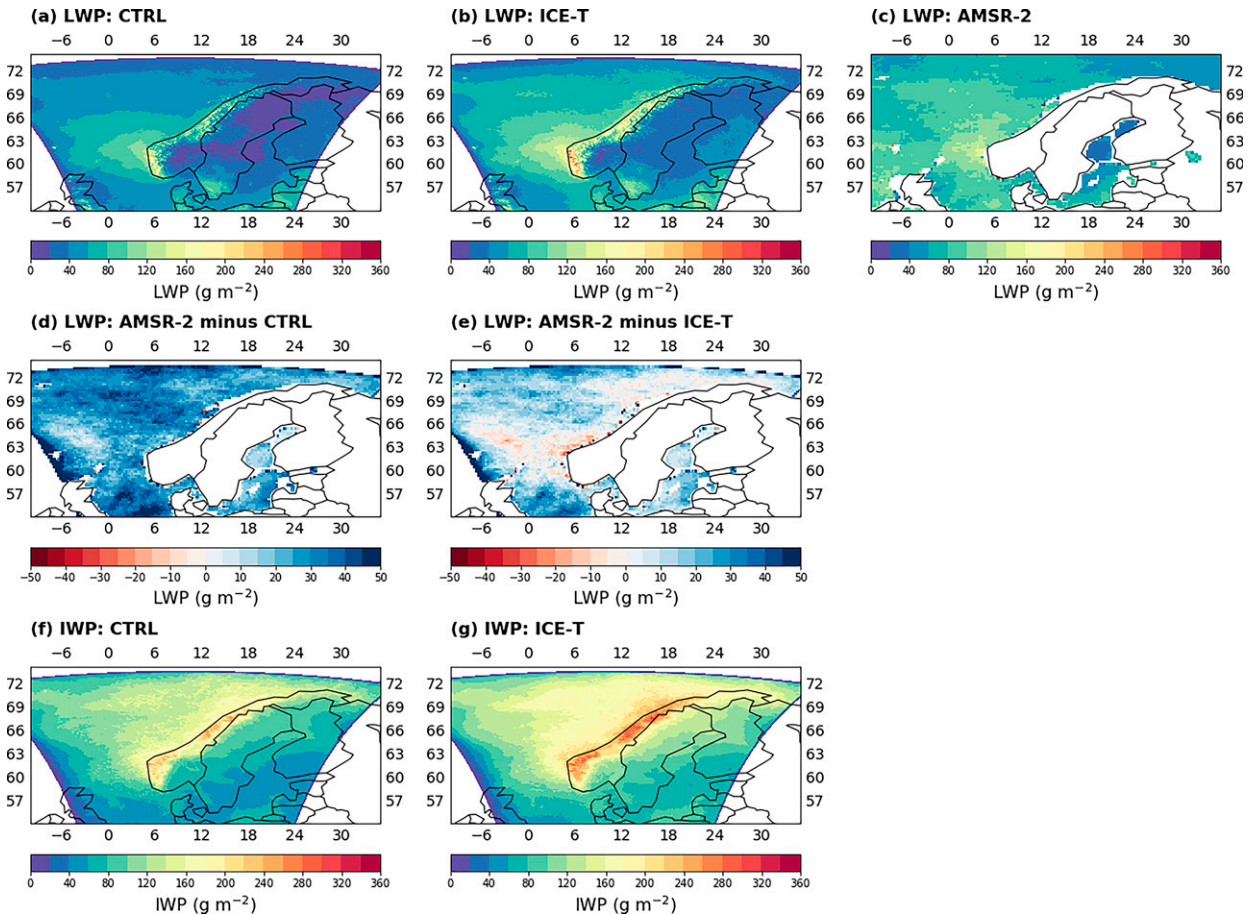


FIG. 7. Comparison of the spatial distribution of mean LWP from (a) CTRL, (b) ICE-T, and (c) AMSR-2. Difference in LWP between AMSR-2 and (d) CTRL and (e) ICE-T. IWP from (f) CTRL and (g) ICE-T. All values are averaged over the entire period 1 Dec 2016–28 Feb 2017.

The corresponding ocean-only domain averages are 52 and 69 g m^{-2} for CTRL and ICE-T, respectively. CTRL is considerably underestimating the LWP, while ICE-T shows an improved LWP compared to the satellite observations.

From Figs. 7f and 7g, it is clear that the total amount of ice condensate in the atmosphere is also increased with ICE-T. The spatial distribution remains similar, yet the amount is increased throughout most of the domain. This might seem a bit contradictory, as we wanted to make the microphysics scheme more “SLW-friendly” by slowing down the processes that convert liquid to ice. Yet, even with increased atmospheric ice, the total solid surface precipitation over the 3-month period is 4.4% larger for CTRL than ICE-T, which further strengthens our abovementioned argument that snow and graupel fall slower in ICE-T, leading to a reduced precipitation efficiency. Unfortunately, we do not have robust measurements of the IWP, yet Field et al. (2017) showed that AROME underpredicted the IWP in their study, so the increased ice in the atmosphere is not necessarily an unwanted feature.

Lenaerts et al. (2017) validated reanalysis and climate models in polar regions against satellite retrieved values of LWP

and IWP from *CloudSat-CALIPSO* over 3 years (2007–10). Their Arctic Ocean domain partly overlaps with our domain, and the retrieved values are similar to the values presented here. They also showed that the LWP and IWP in the reanalyses and climate models varied greatly, illustrating that accurate simulations of LWC and IWC remain challenging for models.

With the combined observations from *CloudSat-CALIPSO* and AMSR-2, it seems clear that the LWC and LWP are improved with ICE-T, yet both CTRL and ICE-T still underestimate the amount of liquid water.

4. Conclusions

In this study we have validated the new modified microphysics scheme in HARMONIE-AROME, ICE-T, with respect to aircraft icing observations and distribution of LWC and IWC over a 3-month winter period. The results show an increase of both LWC and IWC in most of the atmosphere, and consequently higher icing frequency and more severe icing with the modified scheme. ICE-T has, compared with

reported icing from pilots, an overall higher hit rate than CTRL in detecting icing events, their height intervals, and intensities. The hit rates in both ICE-T and CTRL depend both on how icing events are defined and the spatial scale (size of neighborhood) investigated. For the lowest thresholds for defining an icing event (at least one grid point or more than 5% of grid points) and the smallest neighborhoods, there is no significant increase in icing forecast frequency connected to the higher hit rate, which suggests that the improved hit rate does not come with a significantly higher false alarm ratio. For the larger neighborhood areas and higher thresholds, we also see an increased icing forecast frequency along with the increased hit rate, which suggests that higher false alarm rates are possible. However, the lack of “null cases” makes a comprehensive verification difficult, as a direct false alarm ratio cannot be calculated. Therefore it is hard to verify if ICE-T really gives a better icing forecast based on the information from the available AIREPs, though the results indicate that this is likely. Ultimately, the improved quality of the forecasts depends on the cost/loss ratio of the end users. Brown et al. (1997) discuss the trade-off between the algorithms ability to detect icing and over-forecasting. They suggest that the end users themselves can provide input on minimum standards for the verification statistics.

The forecasted icing intensities rely heavily on the thresholds chosen in Table 1. A retuning of the applied icing algorithm and thresholds used for icing intensities to better fit ICE-T could therefore further improve the usefulness.

Judging by the satellite retrieved profiles of LWC and IWC from *CloudSat*–*CALIPSO*, ICE-T represents a step in the right direction compared to CTRL, but still has too little LWC. This is also confirmed by comparing with the LWP retrieved from AMSR-2. The underestimation is present also in the cloud-only calculations, which means that this is mainly not cloud cover related, but an in-cloud LWC deviation.

The underestimation of LWC in the atmosphere could have downstream implications for the forecasts, especially for the radiative energy budget as the radiative fluxes strongly depend on the phase and amount of the hydrometeors. The processes of condensation and freezing release heat to the ambient air, and failure of the NWP model to represent this, may result in errors in the thermodynamic profile. The amount of LWC is also important for precipitation formation and can lead to errors in the timing, location, phase, and amount of precipitation. To find the source of the lack of LWC is therefore an important task to follow up in future studies.

In our study we have only checked the average statistics over the entire period for LWC and IWC. It is possible that the LWC is underestimated in certain weather conditions. For instance, shallow stratocumulus layers following cold air outbreaks have proved to be a challenging task in terms of simulating the LWP correctly (Forbes et al. 2016; Field et al. 2017). A follow-up study should try to investigate specific weather cases more closely.

Acknowledgments. We thank Tor Ivar Mathisen at MET-Norway and Gregory Thompson at NCAR for valuable discussions and insights during the progress of this study. We

would also like to thank Johanne Mehren for preparations of the AIREPs data and Esbjörn Olsson at SMHI for information regarding the icing algorithm. This work is a part of the WISLINE project funded by the Norwegian Research Council, Grant 244106/E10. The contribution from M. Køltzow was supported by the Norwegian Research Council Project 280573 “Advanced models and weather prediction in the Arctic: Enhanced capacity from observations and polar process representations (ALERTNESS).” The contributions from T. Carlsen and T. Storelvmo were supported by the European Research Council (ERC) through Grant StG 758005.

Data availability statement. The 3-month HARMONIE-AROME dataset for CTRL and ICE-T can be found here: https://thredds.met.no/thredds/catalog/metusers/bjorgjke-3mnd_ws/catalog.html. The standard *CloudSat* and *CALIPSO* data products (version R05) used in this study (2B-CWC-RO, 2C-ICE, ECMWF-AUX) were downloaded from the *CloudSat* Data Processing Center’s (at Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins) website (<http://www.cloudsat.cira.colostate.edu/>). AMSR data are produced by Remote Sensing Systems and were sponsored by the NASA AMSR-E Science Team and the NASA Earth Science MEaSUREs Program. Data are available at www.remss.com.

REFERENCES

- Austin, R. T., A. J. Heymsfield, and G. L. Stephens, 2009: Retrieval of ice cloud microphysical parameters using the CloudSat millimeter-wave radar and temperature. *J. Geophys. Res.*, **114**, D00A23, <https://doi.org/10.1029/2008JD010049>.
- 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>.
- Bengtsson, L., and Coauthors, 2017: The HARMONIE-AROME model configuration in the ALADIN-HIRLAM NWP system. *Mon. Wea. Rev.*, **145**, 1919–1935, <https://doi.org/10.1175/MWR-D-16-0417.1>.
- Bernstein, B. C., F. McDonough, M. K. Politovich, B. G. Brown, T. P. Ratvasky, D. R. Miller, C. A. Wolff, and G. Cunning, 2005: Current icing potential: Algorithm description and comparison with aircraft observations. *J. Appl. Meteor.*, **44**, 969–986, <https://doi.org/10.1175/JAM2246.1>.
- Boudala, F., G. A. Isaac, and D. Wu, 2019: Aircraft icing study using integrated observations and model data. *Wea. Forecasting*, **34**, 485–506, <https://doi.org/10.1175/WAF-D-18-0037.1>.
- 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.*, **26**, 610–619, <https://doi.org/10.1002/met.1788>.
- Brousseau, P., Y. Seity, D. Ricard, and J. Léger, 2016: Improvement of the forecast of convective activity from the AROME-France system. *Quart. J. Roy. Meteor. Soc.*, **142**, 2231–2243, <https://doi.org/10.1002/qj.2822>.
- Brown, B. G., G. Thompson, R. T. Bruintjes, 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).
- Cao, Y., W. Tan, and Z. Wu, 2018: Aircraft icing: An ongoing threat to aviation safety. *Aerosp. Sci. Technol.*, **75**, 353–385, <https://doi.org/10.1016/j.ast.2017.12.028>.
- Chaboureau, J.-P., and P. Bechtold, 2002: A simple cloud parameterization derived from cloud resolving model data: Diagnostic and prognostic applications. *J. Atmos. Sci.*, **59**, 2362–2372, [https://doi.org/10.1175/1520-0469\(2002\)059<2362:ASCPDF>2.0.CO;2](https://doi.org/10.1175/1520-0469(2002)059<2362:ASCPDF>2.0.CO;2).
- Cronk, H., and P. Partain, 2017: CloudSat ECMWF-AUX auxiliary data product process description and interface control document. CIRA, Colorado State University, 15 pp., https://www.cloudsat.cira.colostate.edu/cloudsat-static/info/dl/ecmwf-aux/ECMWF-AUX_PDICD.P_R05.rev0_.pdf.
- Deng, M., G. G. Mace, Z. Wang, and H. Okamoto, 2010: Tropical composition, cloud and climate coupling experiment validation for cirrus cloud profiling retrieval using CloudSat radar and CALIPSO lidar. *J. Geophys. Res.*, **115**, D00J15, <https://doi.org/10.1029/2009JD013104>.
- , —, —, and E. Berry, 2015: CloudSat 2C-ICE product update with a new *ze* parameterization in lidar-only region. *J. Geophys. Res. Atmos.*, **120**, 12198–12208, <https://doi.org/10.1002/2015JD023600>.
- Ellrod, G. P., and A. A. Bailey, 2007: Assessment of aircraft icing potential and maximum icing altitude from geostationary meteorological satellite data. *Wea. Forecasting*, **22**, 160–174, <https://doi.org/10.1175/WAF984.1>.
- Elsaesser, G. S., C. W. O'Dell, M. D. Lebsock, R. Bennartz, T. J. Greenwald, and F. J. Wentz, 2017: The Multi-Sensor Advanced Climatology of Liquid Water Path (MAC-LWP). *J. Climate*, **30**, 10193–10210, <https://doi.org/10.1175/JCLI-D-16-0902.1>.
- Engdahl, B. J. K., G. Thompson, and L. Bengtsson, 2020a: Improving the representation of supercooled liquid water in the HARMONIE-AROME weather forecast model. *Tellus*, **72A**, 1–18, <https://doi.org/10.1080/16000870.2019.1697603>.
- , B. E. K. Nygaard, V. Losnedal, G. Thompson, and L. Bengtsson, 2020b: Effects of the ICE-T microphysics scheme in HARMONIE-AROME on estimated ice loads on transmission lines. *Cold Reg. Sci. Technol.*, **179**, 103139, <https://doi.org/10.1016/j.coldregions.2020.103139>.
- Field, P. R., and Coauthors, 2017: Exploring the convective grey zone with regional simulations of a cold air outbreak. *Quart. J. Roy. Meteor. Soc.*, **143**, 2537–2555, <https://doi.org/10.1002/qj.3105>.
- Forbes, R., A. Geer, K. Lonitz, and M. Ahlgrim, 2016: Reducing systematic errors in cold air outbreaks. *ECMWF Newsletter*, No. 146, ECMWF, Reading, United Kingdom, 17–22.
- Frogner, I.-L., and Coauthors, 2019: HarmonEPS—The harmonie ensemble prediction system. *Wea. Forecasting*, **34**, 1909–1937, <https://doi.org/10.1175/WAF-D-19-0030.1>.
- Guan, H., S. G. Cober, and G. A. Isaac, 2001: Verification of supercooled cloud water forecasts with in situ aircraft measurements. *Wea. Forecasting*, **16**, 145–155, [https://doi.org/10.1175/1520-0434\(2001\)016<0145:VOSCWF>2.0.CO;2](https://doi.org/10.1175/1520-0434(2001)016<0145:VOSCWF>2.0.CO;2).
- Hämäläinen, K., and S. Niemelä, 2017: Production of a numerical icing atlas for Finland. *Wind Energy*, **20**, 171–189, <https://doi.org/10.1002/we.1998>.
- Hellmuth, F., B. J. K. Engdahl, T. Storelvmo, R. O. David, and S. J. Cooper, 2021: Snowfall model validation using surface observations and an optimal estimation snowfall retrieval. *Wea. Forecasting*, **36**, 1827–1842, <https://doi.org/10.1175/WAF-D-20-0220.1>.
- Isaac, G. A., S. G. Cober, J. W. Strapp, A. V. Korolev, A. Tremblay, and D. L. Marcotte, 2001: Recent Canadian research on aircraft in-flight icing. *Can. Aero. Space J.*, **47** (3), 1–9, <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.471.2998&rep=rep1&type=pdf>.
- Jeck, R. K., 2001: A history and interpretation of aircraft icing intensity definitions and FAA rules for operating in icing conditions. Tech. Rep. DOT/FAA/AR-01/91, Office of Aviation Research, 43 pp., <http://www.tc.faa.gov/its/worldpac/techrpt/ar01-91.pdf>.
- 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>.
- Lenaerts, J. T. M., K. Van Tricht, S. Lhermitte, and T. S. L'Ecuyer, 2017: Polar clouds and radiation in satellite observations, reanalyses, and climate models: Polar clouds and radiation. *Geophys. Res. Lett.*, **44**, 3355–3364, <https://doi.org/10.1002/2016GL072242>.
- Lin, Y.-L., R. D. Farley, and H. D. Orville, 1983: Bulk parameterization of the snow field in a cloud model. *J. Climate Appl. Meteor.*, **22**, 1065–1092, [https://doi.org/10.1175/1520-0450\(1983\)022<1065:BPOTSF>2.0.CO;2](https://doi.org/10.1175/1520-0450(1983)022<1065:BPOTSF>2.0.CO;2).
- Liu, C., K. Ikeda, G. Thompson, R. Rasmussen, and J. Dudhia, 2011: High-resolution simulations of wintertime precipitation in the Colorado headwaters region: Sensitivity to physics parameterizations. *Mon. Wea. Rev.*, **139**, 3533–3553, <https://doi.org/10.1175/MWR-D-11-00009.1>.
- Makkonen, L., 2000: Models for the growth of rime, glaze, icicles and wet snow on structures. *Philos. Trans. Roy. Soc. London*, **358A**, 2913–2939, <https://doi.org/10.1098/rsta.2000.0690>.
- Mascart, J., and P. Bougeault, 2011: The meso-NH atmospheric simulation system: Scientific documentation. Part III: Physics. Météo-France Tech. Rep., 153 pp.
- Masson, V., and Coauthors, 2013: The SURFEXv7.2 land and ocean surface platform for coupled or offline simulation of Earth surface variables and fluxes. *Geosci. Model Dev.*, **6**, 929–960, <https://doi.org/10.5194/gmd-6-929-2013>.
- Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.*, **102**, 16663–16682, <https://doi.org/10.1029/97JD00237>.
- Morcrette, C., K. Brown, R. Bowyer, P. Gill, and D. Suri, 2019: Development and evaluation of in-flight icing index forecast for aviation. *Wea. Forecasting*, **34**, 731–750, <https://doi.org/10.1175/WAF-D-18-0177.1>.
- Morcrette, J.-J., H. W. Barker, J. N. S. Cole, M. J. Iacono, and R. Pincus, 2008: Impact of a new radiation package, McRad, in the ECMWF integrated forecasting system. *Mon. Wea. Rev.*, **136**, 4773–4798, <https://doi.org/10.1175/2008MWR2363.1>.
- Müller, M., and Coauthors, 2017: AROME-MetCoOp: A Nordic convective-scale operational weather prediction model. *Wea. Forecasting*, **32**, 609–627, <https://doi.org/10.1175/WAF-D-16-0099.1>.
- Neggers, R. A. J., M. Köhler, and A. C. M. Beljaars, 2009: A dual mass flux framework for boundary layer convection. Part I: Transport. *J. Atmos. Sci.*, **66**, 1465–1487, <https://doi.org/10.1175/2008JAS2635.1>.
- Nygaard, B. E. K., J. E. Kristjánsson, and L. Makkonen, 2011: Prediction of in-cloud icing conditions at ground level using

- the WRF Model. *J. Appl. Meteor. Climatol.*, **50**, 2445–2459, <https://doi.org/10.1175/JAMC-D-11-054.1>.
- Pinty, J. P., and P. Jabouille, 1998: A mixed-phase cloud parameterization for use in mesoscale non-hydrostatic model: Simulations of a squall line and of orographic precipitations. *Conf. on Cloud Physics*, Everett, WA, Amer. Meteor. Soc., 217–220.
- Politovich, M., 2003: Aircraft icing. *Encyclopedia of Atmospheric Sciences*, J. R. Holton, Ed., Academic Press, 68–75, <https://doi.org/10.1016/B0-12-227090-8/00055-5>.
- , 2009: Predicting inflight aircraft icing intensity—A surrogate for severity. FAA Tech. Doc., 21 pp., https://aviationweather.gov/docs/icing/How_to_properly_use_an_Icing_Forecast.pdf.
- Seity, Y., P. Brousseau, S. Malardel, G. Hello, P. Bénard, F. Bouttier, C. Lac, and V. Masson, 2011: The AROME-France convective-scale operational model. *Mon. Wea. Rev.*, **139**, 976–991, <https://doi.org/10.1175/2010MWR3425.1>.
- Serke, D., and Coauthors, 2014: Supercooled liquid water content profiling case studies with a new vibrating wire sonde compared to a ground-based microwave radiometer. *Atmos. Res.*, **149**, 77–87, <https://doi.org/10.1016/j.atmosres.2014.05.026>.
- Stephens, G. L., and Coauthors, 2002: The *CloudSat* mission and the A-Train: A new dimension of space-based observations of clouds and precipitation. *Bull. Amer. Meteor. Soc.*, **83**, 1771–1790, <https://doi.org/10.1175/BAMS-83-12-1771>.
- Thompson, G., R. M. Rasmussen, and K. Manning, 2004: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part I: Description and sensitivity analysis. *Mon. Wea. Rev.*, **132**, 519–542, [https://doi.org/10.1175/1520-0493\(2004\)132<0519:EFOWPU>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<0519:EFOWPU>2.0.CO;2).
- , P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev.*, **136**, 5095–5115, <https://doi.org/10.1175/2008MWR2387.1>.
- , M. K. Politovich, and R. M. Rasmussen, 2017: A numerical weather model's ability to predict characteristics of aircraft icing environments. *Wea. Forecasting*, **32**, 207–221, <https://doi.org/10.1175/WAF-D-16-0125.1>.
- Wentz, F., T. Meissner, C. Gentemann, K. Hilburn, and J. Scott, 2014: Remote Sensing Systems GCOM-W1 AMSR2 Daily Environmental Suite on 0.25 deg grid, version 8.2. Remote Sensing Systems, accessed 26 May 2021, www.remss.com/missions/amr.
- Winker, D. M., and Coauthors, 2010: The CALIPSO mission: A global 3D view of aerosols and clouds. *Bull. Amer. Meteor. Soc.*, **91**, 1211–1230, <https://doi.org/10.1175/2010BAMS3009.1>.
- Xu, M., G. Thompson, D. R. Adriaansen, and S. D. Landolt, 2019: On the value of time-lag-ensemble averaging to improve numerical model predictions of aircraft icing conditions. *Wea. Forecasting*, **34**, 507–519, <https://doi.org/10.1175/WAF-D-18-0087.1>.