Ship-icing prediction methods applied in operational weather forecasting

Eirik Mikal Samuelsen,[a,b]*

[a]*UiT - The Arctic University of Norway, Tromsø, Norway.* [b]*MET Norway - Norwegian Meteorological Institute, Tromsø, Norway*

*Correspondence to: UiT: P.O. Box 6050 Langnes, NO-9037 Tromsø, Norway; MET Norway: P.O. Box 6314 Langnes, NO-9293 Tromsø, Norway. E-mail: eirik.m.samuelsen@uit.no;eiriks@met.no

Sea-spray wetting of ships operating in cold environments imposes a great safety risk due to icing. For this reason marine-icing warnings have been a part of operational weather forecasting for the last five decades. Yet verification of such warnings has only been done sparingly. This paper evaluates different ship-icing methods applied in operational weather forecasting. The methods are tested against a unique data set from a single ship type from Arctic-Norwegian waters, and two screened data sets from several ship types from Alaska and the east coast of Canada. Missing and uncertain parameters in the latter data sets are supplemented by reanalysis data of different sources. Continuous icing-rate verification and sensitivity tests are presented for the physical icing models alongside categorical icing-rate verification which is applied in order to also evaluate icing nomograms that are still in use by several forecasting agencies. Furthermore, a newly proposed definition of the boundaries between the icing-rate severity categories is applied in the categorical-verification procedure. The overall best verification scores for the continuous and categorical icing rates are obtained by the Marine Icing model for the Norwegian COast Guard (MINCOG) and a physically-based Overland model updated from its initial version with a more realistic heat transfer. Finally, sensitivity tests highlight that very low air and sea-surface temperatures rarely occur over sea areas together with high waves due to fetch limitations, even for strong winds. For this reason models and nomograms that do not treat wind speed and wave height separately will provide inaccurate predictions of icing rate in such areas. Consequently it is preferable that methods applied in operational weather forecasting are replaced with methods capable of taking this effect into account.

Key Words: Polar meteorology; Wave-ship interaction; Extreme weather event; Arctic transportation; Icing observations; Freezing sea spray; Marine weather warning; Cold climate

Received . . .

1. Introduction

Ships operating in Arctic waters in sub-freezing temperatures are exposed to ice accretion. This phenomenon called icing appears when water from either the atmosphere or the ocean freeze onto the ship. Water from the atmosphere is provided by precipitation, fog or even vapour deposited directly (Shekhtman 1971). Water from the ocean comes from sea-spray droplets created through the interaction between a ship and waves, or droplets ripped off from the top of the wave crests by the wind in breaking-wave situations. Large portions of sea water can also flood the deck and some of the remaining water may freeze and contribute to icing (Shellard 1974; Løset *et al.* 2006). The perils related to icing have been known for centuries. Panov (1978) cites Buchinskiy (1960) claiming that there have been reports of icing since the 15th-16th centuries from inhabitants in the White Sea sailing towards Arctic waters in the late autumn. However, until the 1920s icing had virtually not been studied (Panov 1978). Especially after the 1930s closer attention was drawn to this hazardous phenomenon when more ships, particularly fishing trawlers, started operating in Arctic waters (Mertins 1968). The greatest risk is associated with rapid ice accretion above deck level raising the centre of gravity of the vessel, which may lead to destabilization and eventually capsizing (Shellard 1974). Thus, uneven ice build-up on one of the sides of the hull or superstructure is especially dangerous. Shellard (1974) documents 81 ships that were lost due to icing in the decades after the Second World War. Sawada (1968) addresses a list of 19 ships sinking in the northern seas of Japan costing the lives of 296 people in a 5-year period during the 1960s. The numerous loss of lives related to icing increased the necessity of including marine-icing warnings in operational weather forecasts. This lead to the development of the nomogram of Mertins (1968). This categorical icing nomogram is a statistical forecasting method for pure spray icing which is empirically derived from temperature, wind-speed, and sea-surface-temperature data based on 400 slow-moving fishing vessels in the sea areas around Iceland and Greenland, the Labrador Sea, the Barents Sea, and the Baltic Sea (Mertins 1968). The method is still applied by several weather forecasting agencies in Europe today (Ekeberg 2010). In addition, it is the only ship-icing method mentioned in the Forecasters' reference book of the UK Met Office (Meteorological Office College 1997). Even though the ships of today are more seaworthy than those in the 1950s and 1960s, there are still examples of fishing vessels capsizing due to icing in more recent decades. In February 2007 the fishing vessel "Lady of Graze" capsized in the Nantucket Sound at the East Coast of the USA costing the lives of 4 people (United States Coast Guard 2008). A recent inspection of the database of ship accidents from the Norwegian Maritime Authority also revealed two fishing vessels, Borgøygutt and Sjøheim, capsizing in January 1999 at the coast of Northern Norway, one of those costing the lives of 3 people (Norwegian Maritime Authority 2014). Icing may also be a concern for larger ships. Although capsizing is rare, icing can lead to slippery conditions on decks, ladders, and handrails. Iced-down rescue equipment is also a potentially dangerous outcome of icing. Finally, freezing antennas can hamper the radio communication (Løset *et al.* 2006).

Examinations of approximately 6,000 icing events have revealed that pure sea spray is the most frequent cause of icing (Table 5.1 in Zakrzewski and Lozowski (1989)). However, spray in combination with precipitation or fog is also prevalent especially in Arctic-Russian and Arctic-Norwegian sea areas. Brown and Roebber (1985) and Brown and Agnew (1985) have pinpointed that the WMO (World Meteorological Organization) Synop code (WMO 2015) does not include spray and snow in combination as a cause of icing, although snow was reported together with spray during an icing event in more than 60% of 960 icing reports in Canadian waters. Eide (1983) also discovered that snow occurred together with sea spray during 12 icing events on the weather ship AMI situated on Tromsøflaket outside the coast of Northern Norway in the late 1970s. This is in agreement with the data of Samuelsen *et al.*

(2017) in which most of the spray-icing events occurred in conjunction with snow showers or Arctic sea smoke. However, it is not clear whether these atmospheric water sources really contribute extensively to icing, or if they are just features related to a weather situation with cold winds blowing over relatively warm water where icing mainly occur due to sea spray (Samuelsen *et al.* 2017). In any case, all methods used in operational weather forecasting are focusing on pure spray icing only. In addition to the statistical method of Mertins (1968), Lundqvist and Udin (1977) made a statistical forecasting method for spray icing for the Baltic sea. This statistical method is based on data collection from merchant ships with 500 to 7000 deadweight tonnage (dwt) moving with normal speeds of 10-15 knots in the less saltier Baltic sea. These categorical icing curves in relationship with temperature and wind speed are for instance used today by some German forecasters (pers. comm., Fregattenkapitän Ingmar Behrendt, April 2016). According to Ekeberg (2010) the Swedish Meteorological and Hydrological Institute (SMHI) applies this method in icing forecasts. However, there are also indications that forecasters at SMHI applies the predictor from Overland (pers. comm., anonymous forecaster SMHI Enköping, September 2015). Another statistical forecasting method based on the nomogram from Sawada (1962), which is also cited in Shellard (1974) and Lundqvist and Udin (1977), has been applied by Japanese forecasters (Ekeberg 2010), and is still in use by some Canadian forecasters (pers. comm., Serge Desjardins, April 2016).

In addition to these statistical forecasting methods, several different physical forecasting models for wave-ship-interaction generated icing have been developed from the 1970s and onwards (Kachurin *et al.* 1974; Stallabrass 1980; Overland *et al.* 1986; Makkonen 1987; Henry 1995; Lozowski *et al.* 2000; Samuelsen *et al.* 2017). Wave-ship-interaction generated sea spray provides several orders of magnitude larger spray amounts than the sea spray blown off the white caps of the waves, and it is therefore reckoned to be the most important water source in dangerous icing events (Stallabrass 1980; Zakrzewski 1987; Lozowski *et al.* 2000). Most of these models calculate the icing rate as a fraction of the flux of the incoming spray generated in the collision process between the ship and the waves, and solve a heat-balance equation assuming thermal equilibrium between the heat released by freezing and heat fluxes from the atmosphere and the sea spray acting upon the brine. Normally this heat balance is assumed on a surface with simplified geometry at a specific position of a ship. In the recent years there have also been developed models based on Computational Fluid Dynamics (CFD) calculations to simulate the local air stream and icing rate taking into account the more complete geometry of a specific ship or structure (Hansen 2012; Kulyakhtin and Tsarau 2014; Hansen and Teigen 2015). Despite the fact that numerous physical spray-icing models have been developed in the research areas of Technology and Meteorology in the past decades, only a few of these models are incorporated in the operational forecasting environment of today (2017). The model of Stallabrass (1980) modified by Ross Brown (Henry 1995), and the model of Overland (Overland *et al.* 1986; Overland 1990) with or without adjustments (Desjardins 2013), are, to the knowledge of the author, the only physically-based spray-icing models applied in operational weather forecasting. In addition, a simplified version of a newly developed Marine-Icing model for the Norwegian COast Guard (MINCOG) has recently been implemented in the MET Norway forecasting tool (Samuelsen *et al.* 2017).

The predictor of Overland is divergent from the other approaches. This method is based on a physical model for wave-ship-interaction generated icing together with a simplified expression of the heat equation. However, instead of estimating the rate of icing for a specified spray flux, a constant freezing fraction is assumed. The freezing fraction is defined as the ratio between the icing flux and the spray flux. Overland (1990) applies the freezing fraction tuned in such a manner that fits best the observations of icing rates from Pease and Comiskey (1985) and Zakrzewski *et al.* (1989). Although being criticised for being unphysical and too sensitive to low sea-surface temperatures (Makkonen *et al.* 1991), it is probably the most widely used marine-icing method of today. The Overland predictor is applied by both Norwegian forecasters (Samuelsen *et al.* 2015), American forecasters (Nacional Oceanic and Atmospheric Administration 2015; Ryerson 2013), Canadian forecasters (with a slight adjustment, (Desjardins 2013)), Japanese forecasters (pers.comm., Satoshi Ogawa, Japan Meteorological Agency, November 2016), and Swedish forecasters. Additionally, it is widely used in several newer studies involving the topic of marine icing, e.g. climatology studies of icing in the Norwegian Sea,

the Barents Sea, and the Greenland Sea (Byrkjedal *et al.* 2011; Iden *et al.* 2012; Moore 2013), in several master theses in Science and Technology (Johansen 2013; Sollid 2013; Pedersen 2015; Gemynthe 2015), and in an American guide for ships operating in low temperatures (ABS - American Bureau of Shipping 2010).

Whereas icing has been included in marine weather forecasts for almost 50 years, the ship-icing-prediction methods have only been verified sparingly. Both Brown and Roebber (1985) and Roebber and Mitten (1987) compare the methods of Mertins (1968), Kachurin *et al.* (1974), Wise and Comiskey (1980), and Stallabrass (1980) both continuously and categorically, but only the method of Mertins (1968) is still applied today. Roebber and Mitten (1987) also evaluate a modified version of Overland *et al.* (1986) due to some calculation errors in the third degree polynomial fit for large values of the predictor, but this version is not applied by Overland (1990) or in later studies. Overland (1990) argues that the Overland predictor was an improvement over previous methods (e.g. Wise and Comiskey (1980) which is an extension of Mertins (1968)) referring to a verification study by Feit (1987). However, the latter study only compares 10 situations with icing and 82 situations without icing, and does not compare the continuous or categorical icing rate provided by Overland *et al.* (1986). Hansen (2012) compares ice-thickness accumulation on the stationary Weather ship AMI with calculations from Overland (1990). Samuelsen *et al.* (2015) evaluate Overland, ModStall, and some other methods against ship-icing data that are recorded during a voyage of the Norwegian Coast Guard vessel KV Nordkapp in the Barents Sea on the west side of a developing polar low. Prior to these studies there is no documented verification of icing on ships in Arctic-Norwegian waters. Samuelsen *et al.* (2017) distribute a unique data set of 37 icing events from the KV Nordkapp ships with a collection of necessary input parameters and icing rates. The problem of applying wave heights and periods as direct functions of the wind speed, a commonly-applied relationship in marine-icing modelling, is here emphasised. In the current paper this unique data set of Samuelsen *et al.* (2017) is used in a meticulous and thorough evaluation of the different icing methods applied in operational weather forecasting of today. Furthermore, the methods are tested against a screened and updated version of the data sets applied in Overland *et al.* (1986) and Roebber and Mitten (1987). The data sets are supplemented with reanalysis data. Several physical models are tested against icing-rate observations from these three data sets, and the results from four to six of these models are presented in more detail. The continuous icing-rate values from the physical models are divided into the four categories applied in operational weather forecasting: no icing, light icing, moderate icing, and severe icing, and tested along with icing-severity output from different nomograms applied operationally today. All methods are evaluated categorically applying a newly derived definition of the thresholds between the icing categories obtained from icing climatology.

2. Spray-icing forecasting methods

2.1. Marine icing theory

In the physical models for wave-ship-interaction icing the source of water is expressed in the term of a spray flux (R_w). This flux provides an upper boundary for the amount of ice that is accumulated pr. unit time. In order to derive the rate of ice accretion (R_i) on a certain position of the ship, a surface energy balance is assumed between the heat released from the freezing flux of water (Q_f), and the most important heat fluxes from the atmosphere, ocean, and the underlying surface. Furthermore, a freezing fraction (n) between R_i and R_w is calculated from this heat-balance equation. By taking into account the contribution from the turbulent heat fluxes from the atmosphere, the sensible or convective heat flux (Q_c), and the latent or evaporative heat flux (Q_e), the heat flux from the impinging water droplets or spray (Q_d), and the net radiative heat flux (Q_r), the following heat balance is obtained:

$$Q_f = Q_c + Q_e + Q_d + Q_r \tag{1}$$

A list of symbols explaining these and other terms is provided in the Appendix (Table A.I). A typical way of representing these heat fluxes is as follows:

$$Q_c = h_a \left(T_s - T_a\right) \tag{2}$$

$$Q_e = h_e \left(e_s \left(T_s\right) - R_H e_s \left(T_a\right)\right) \tag{3}$$

$$Q_d = R_w c_w \left(T_s - T_{\mathrm{sp}}\right) \tag{4}$$

$$\begin{aligned} Q_r =& \uparrow \mathrm{LW} - \downarrow \mathrm{LW} + \uparrow \mathrm{SW} - \downarrow \mathrm{SW} \\ =& \sigma (T_s + 273.15)^4 - \downarrow \mathrm{LW} - (1 - A) \downarrow \mathrm{SW} \end{aligned} \tag{5}$$

In addition to these fluxes the conductive heat flux may also be important in the initial phase of icing and in pulsed spray conditions (Kulyakhtin *et al.* 2016). However, it is convenient to neglect conduction, whereby consideration of the initial ice thickness is avoided, when dealing with prediction models providing information about future icing. In addition, pulsed spray conditions are difficult to model accurately since the spraying of a ship during wave-ship interactions is highly irregular and probably stochastic (Samuelsen *et al.* 2017). Conduction is also negligible when applying continuous spray-icing models with or without time-averaging the spray-flux term. A more thorough explanation of the applied heat-flux terms and the underlying methodology is provided by Samuelsen *et al.* (2017).

2.2. *Physical models*

Table I provides an overview of the icing models that are applied in operational weather forecasting today. A short description of the basic assumptions and differences between the physical models are elaborated in the following section:

Table I. Overview of icing models applied in operational weather forecasting (Ekeberg (2010); pers.comm., forecasters in Norway, Sweden, Germany, Canada, the USA, and Japan, 2014-2016).

Model/Method	Applied in
Physical models	
Overland	USA, Canada[†], Norway, Sweden, Japan
Modified Stallabrass	Canada, Norway
MINCOG	Norway[††]
Statistical methods/Nomograms	
Mertins	UK, Denmark, Norway
Lundqvist and Udin	Germany, Sweden
Sawada	Japan, Canada

[†] Applied with fetch factor (Section 2.2.3).
[††] A simplified version is applied from the winter season 2016 to 2017.

2.2.1. *Overland (Over)*

As mentioned in the introduction, the Overland predictor (Pr) has its basis in the heat equation similar to Eq.(1). Although Overland *et al.* (1986) initially take into account more terms in their heat-balance equation, they only use expressions from the convective heat flux (Q_c) and the heat flux from the impinging spray (Q_d) in the development of their commonly-applied predictor (Pr) which is defined

as follows:

$$\mathrm{Pr} = \frac{V(T_f - T_a)}{1 + \Phi\left(\mathrm{SST} - T_f\right)} \tag{6}$$

The basic principal in the derivation of Eq.(6) may be summarised as follows:

$$Q_f = Q_c + Q_d \tag{7}$$

using

$$Q_f = L_{\mathrm{fs}} R_i \tag{8}$$

$$Q_d = c_w R_w (T_f - \mathrm{SST}) = c_w \frac{R_i}{n}(T_f - \mathrm{SST}) \tag{9}$$

$$R_i = \rho_i \frac{dh}{dt} \tag{10}$$

$$Q_c = \rho_a c_p C_H V(T_f - T_a) = kV(T_f - T_a) \tag{11}$$

gives

$$L_{\mathrm{fs}} R_i - c_w \frac{R_i}{n}(T_f - \mathrm{SST}) = kV(T_f - T_a) \tag{12}$$

$$\begin{aligned} \frac{dh}{dt} &= \frac{\frac{k}{L_{\mathrm{fs}}\rho_i} V(T_f - T_a)}{1 + \frac{c_w}{nL_{\mathrm{fs}}}(\mathrm{SST} - T_f)} \\ \frac{dh}{dt} &\propto \frac{V(T_f - T_a)}{1 + \Phi(\mathrm{SST} - T_f)} = \mathrm{Pr}, \quad \Phi = \frac{c_w}{nL_{\mathrm{fs}}} \end{aligned} \tag{13}$$

A major difference between the Overland model and all the other sea-spray icing models, is the use of the so-called Bulk formulation for the sensible heat flux. By using this formulation Overland *et al.* (1986) indirectly assume that the turbulent fluxes governing the heat transfer between the atmosphere and the brine are based on applications of the Monin-Obukhov similarity theory to estimate the heat exchange between a surface on the earth and the atmosphere (Monin and Obukhov 1954). Details about how this method may be applied in climate or weather prediction models are found e.g. in Fairall *et al.* (2003). One important aspect of the theory is the assumption that the atmospheric flow is independent of the Reynolds number (Re) due to low viscosity. As a consequence the Q_c is proportional to V instead of $V^{0.8}$ applied in the models of Stallabrass (1980) and Samuelsen *et al.* (2017) where it is assumed that the heat-transfer coefficient is approximately proportional to $\mathrm{Re}^{0.8}$ both for a turbulent air flow along-flow and cross-flow relative to a vertical plate or cylinder. Nevertheless, instead of applying the physical relationship in Eq.(13) by calculating the heat-transfer coefficient directly for a constant freezing fraction, Overland *et al.* (1986) find the best empirical relationship between the observed icing rate (dh/dt) and the predictor Pr calculated from the input parameters in Eq.(6) by using a polynomial fit:

$$\frac{dh}{dt}(\mathrm{cm\ h^{-1}}) = A_1\mathrm{Pr} + B_1\mathrm{Pr}^2 + C_1\mathrm{Pr}^3 \tag{14}$$

where

$$A_1 = 2.73 \times 10^{-2},\ B_1 = 2.91 \times 10^{-4},\ C_1 = 1.84 \times 10^{-6}$$

However, when combining Eqs.(6), (11), (13), and (14) it is apparent that C_H also may be written as a function of Pr based on the coefficients from Eq.(14):

$$\frac{dh}{dt}(\text{cm h}^{-1}) = \frac{\rho_a c_p C_H \text{Pr}}{L_{\text{fs}}\rho_i} 3.6 \times 10^5 \tag{15}$$

$$C_H = (A_1 + B_1\text{Pr} + C_1\text{Pr}^2)\frac{L_{\text{fs}}\rho_i}{\rho_a c_p 3.6 \times 10^5} \tag{16}$$

The factor 3.6×10^5 is included in Eq.(16) in order to convert from cm h^{-1} to SI units. In other words, when assuming that the freezing fraction (n) is constant, the use of the polynomial fit from Eq.(14) is the same as applying a physical model of icing where the bulk-transfer coefficient for heat (C_H) is dependent on Pr which again is a function of V, T_a, and SST (Eq.(6)). Figure 1 illustrates how C_H increases for increasing Pr in the above-mentioned physical model. In effect it is apparent that C_H vary strongly over a typical range of Pr for the Overland model (Over). For a zero predictor C_H is around 12×10^{-3}. Overland *et al.* (1986) state that the unmodelled physics represent 30% of the sensible heat flux and hereby derive a value of $C_H = 9\times10^{-3}$ when only using the linear term in Eq.(14). However, this value is still one order of magnitude higher than typical values of C_H over open ocean (1-2$\times10^{-3}$, as e.g. seen in Figure 9.13 in Wallace and Hobbs (2006) or Figure 6 in Bourassa *et al.* (2013) for neutral conditions). In addition, when using only the linear term in Eq.(14), one obtains only the minimum value of C_H, i.e. when Pr = 0. For the onset of severe icing (Pr = 45.2) the value is around 20×10^{-3} using the full expression in Eq.(14).

While Overland *et al.* (1986) apply 58 of 85 observations from different ship types from March 1979 to January 1984 (Pease and Comiskey 1985), Overland (1990) includes 44 observations from a single ship recorded during a two-weeks cruise in February 1988 (Zakrzewski *et al.* 1989). This results in a higher freezing fraction in Overland (1990) compared to Overland *et al.* (1986), changing n from 0.04 (Φ = 0.4) in Overland *et al.* (1986) to 0.06 (Φ = 0.3) in Overland (1990) (Samuelsen *et al.* 2015). Overland (1990) also adjusts the predictor-value thresholds for the different icing-severity categories (Figure 1). However, Overland (1990) is not modifying the icing-rate (dh/dt) thresholds following the change in predictor-value thresholds. This may have been realised by applying Eq.(14). In order to exemplify: a change in the boundary between the light and moderate category, given by a Pr-value change from 20.4 to 22.4 m s^{-1} K, would change the icing-rate boundary value from 0.69 to 0.78 cm h^{-1}. Similarly would a change in the Pr boundary value between moderate and severe icing from Pr = 45.2 to 53.3 m s^{-1} K lead to a change in icing-rate boundary value from 2.00 to 2.56 cm h^{-1}. For categorical icing-rate verification in the current study both sets of thresholds are tested. It is also throughout the study decided to only test the Overland model of the version from Overland (1990) with n = 0.06 (Φ = 0.3) and not that from Overland *et al.* (1986) with n = 0.04 (Φ = 0.4), since the former is the version most applied.

2.2.2. *Overland modified polynomial (Over mod)*

Roebber and Mitten (1987) state that there is an error in the polynomial fit in Overland *et al.* (1986), and suggest using a second order polynomial of the form:

$$\frac{dh}{dt}(\text{cm h}^{-1}) = A_2 + B_2\text{Pr} + C_2\text{Pr}^2 \tag{17}$$

where

$$A_2 = 0.1982,\ B_2 = 3.07 \times 10^{-2},\ C_2 = 1.996 \times 10^{-4}$$

Although this polynomial fit gives some discrepancies for low values of the predictor, it greatly reduces the icing rates for large values of the predictor. The effect of this new polynomial fit is also noticeable when investigating the bulk-transfer coefficient for heat (C_H)

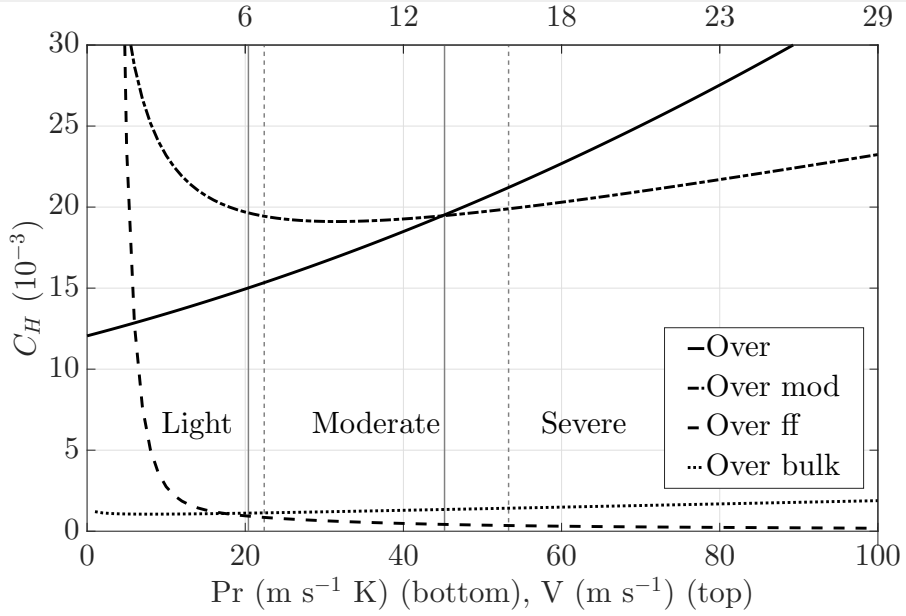

Figure 1. Illustration of how the bulk-transfer coefficient for heat (C_H) changes for different values of the Overland predictor (Pr) for the models Over, Over mod, Over ff, and Over bulk. The wind speed (V) presented on the top x-axis is calculated from the Predictor values by setting SST = 2.5 °C, T_f = −2 °C, and T_a = −10 °C using Eq.(6). The vertical solid lines visualise the boundaries between the icing categories from Pr in Overland *et al.* (1986), while the vertical dashed lines illustrate the boundaries from Overland (1990). Note that C_H is unitless.

in Figure 1. If neglecting the smallest values of the Pr where C_H goes to ∞, this modified model provides a more constant C_H with a mean value of 20×10^{-3} when Pr is in the range between 10 and 100 m s^{-1} K.

2.2.3. *Overland fetch factor (Over ff)*

Since there is no direct calculation of the spray flux in the Overland model, but an assumption of a constant ratio between the icing rate and spray flux in icing situations, there are inaccuracies when applying the model in fetch-limited areas. Thus, Desjardins (2013) introduces a fetch factor which is multiplied with the icing rate derived from the Overland model (Eq.(14)) in order to assign a wave-height dependency in the model and therefore reduce the magnitude of the predicted icing rate in such areas. This fetch factor is given as:

$$\mathrm{FF} = \frac{H_s}{\left(\frac{1.94V}{13}\right)^2} = \frac{H_s}{0.22\frac{V^2}{g}} \tag{18}$$

H_s is the significant wave height, and the denominator is an expression for a theoretical maximum wave height following the Pierson-Moscowitz spectrum (Pierson and Moskowitz 1964) based on a fully developed sea. This implies that it is assumed that the waves are in equilibrium with the local prevailing wind which is often not the case in icing situations (Samuelsen *et al.* 2017). Swell waves are for instance not taken into account. The expression in Eq.(18) is adjusted from the original version to be able to apply wind speed (V) in m s^{-1} instead of knots. The overall icing rate (dh/dt) is then calculated by multiplying this factor (FF) to the icing rate calculated from the polynomial fit of Overland (1990) (Eq.(14)). Notice that this new expression is not valid for low wind speeds. The effect of this adjustment on the C_H is visualised in Figure 1 *. It is apparent that the effect of Over ff is a reduction of the C_H for increasing values of Pr. Notice that for different values of SST, T_a and T_f, the values of V and H_s may be different, and the reduction of the C_H for the Over ff model compared to the original Overland model will during such circumstances also be somewhat different from what

*In Figure 1 it is assumed that Pr changes due to changes in wind speed only. The wind speed (V) is then calculated from the fixed Pr by applying constant values for the sea-surface temperature (SST = 2.5 °C), the freezing temperature of the brine set equal to the freezing temperature of the incoming sea water (T_f = −2 °C), and the air temperature (T_a = −10 °C). In order to obtain an increase in wave height with increasing wind speed, H_s is calculated from V assuming a constant fetch of 100 nautical miles (nm) using a polynomial fit based on data listed in a handbook of oceanographic tables (Bialek 1966; Zakrzewski 1987; Samuelsen *et al.* 2017). In the later calculations the observed wave heights are applied.

is illustrated in Figure 1. For small values of V, FF is greater than 1, which results in very large C_H. Note that the model is not defined for zero wind speed.

Since the effect of both Over mod and Over ff is to reduce the icing rate in the original Overland model, it is decided to only present one combined model in the main article where both modifications are implemented (Over mod ff). However, the verification scores when applying these two methods separately are presented in the supplementary material.

2.2.4. *Overland bulk calculation (Over bulk)*

As illustrated in Figure 1 the result of fitting the predictor towards the observations in the Overland data set is yielding a large C_H which might be unrealistic. For this reason a new model is implemented where it is assumed that the heat transfer between the atmosphere and the ocean is comparable to the heat transfer between the atmosphere and those ship surfaces containing brine water. The basic idea of this model is that the atmospheric eddies governing the former heat transfer also are important for the latter. As a consequence, turbulent eddies that are generated by other sources, e.g. flow disturbances induced by the roughness or the movement of the ship, are neglected. A typical formulation of C_H for the ocean-atmosphere exchange is expressed as follows (Eq.(11.17) Arya (2001)):

$$C_H = \frac{\kappa^2}{(\log \frac{z}{z_0} - \psi_m)(\log \frac{z}{z_{0T}} - \psi_h)} \tag{19}$$

κ is the von-Karman constant set to 0.4, z_0 is the roughness length, and z_{0T} is the roughness lengths for temperature set to 10% of z_0 (Reijmer *et al.* 2003). ψ_m and ψ_h are stability correction terms which may be defined according to Eq.(2.85)-(2.89) in Foken and Nappo (2008). These terms are dependent on the static stability which may be calculated through the Bulk-Richardson number (Eq.(11.18) in Arya (2001)). Since the air flow before reaching the ship is mostly affected by the roughness over the open ocean, it is assumed that the roughness length may be approximated following the relation of Charnock taking into the account the roughness of the waves: $z_0 = 0.015(u^*)^2 g^{-1}$ (Eq.(7.4.1 h) in Stull (1988)). The friction velocity u^* is then approximated according to a logarithmic wind profile: $u^* = \kappa^2 V^2 \left(\log\left(z/z_0\right)\right)^{-2} \phi_m^{-2}$, and z_0 is solved through an iterative process. ϕ_m is a non-dimensional wind shear which is defined according to the static stability (Eq.(11.9) of Arya (2001)). In Figure 1 the same values as in Over ff are applied for SST, T_f, T_a, and V. In addition, a reference height of z = 12 m is used since the observations of pressure and temperature on KV Nordkapp are recorded at this height (Samuelsen *et al.* 2015). In Figure 1 humidity is not taken into account, a constant density of 1.3 kg m^{-3} is applied, and the virtual temperature (T_v) is approximated to be equal to the dry temperature. In later calculations the density is calculated through the ideal gas law ($\rho_a = pR_d^{-1}T_v^{-1}$), and T_v is applied when calculating the Richardson number. As apparent in Figure 1, the C_H in this new model is considerably lower than in the original Overland with a mean value of 1.4×10^{-3} for Pr $\in$ [0,100] m s^{-1} K.

2.2.5. *Overland bulk calculation 2 (Over bulk2)*

When reducing the bulk-transfer coefficient for heat by one order of magnitude as is done in the Over bulk model, the convective heat flux is also reduced similarly compared to the original version of the model. Since it is assumed that only 6% of the spray flux is freezing, the heat from the sea water (Q_d) will greatly reduce the calculated icing rate than it will for higher freezing fractions, especially in relatively warm waters. For this reason the bulk model is further developed by including both the evaporative (Q_e) and the radiative heat flux (Q_r). The bulk-transfer coefficient for moisture (C_E) is then set equal to the one for heat (C_H = C_E). It is assumed that the freezing brine radiates as a black body with a temperature equal to the freezing temperature of the incoming sea water. Incoming longwave and shortwave radiation from reanalysis data are applied as in the MINCOG model (Samuelsen *et al.* 2017). Q_r is then the net effect of both the incoming and outgoing longwave and shortwave radiation. In addition, a constant freezing fraction

(n) of 0.5 is applied instead of 0.06. A value of n = 0.5 is more in agreement with the mean value of the calculated freezing fractions in the MINCOG model when applying the KV Nordkapp data and the Borisenkov spray-flux formulation (Table 3 in Samuelsen *et al.* (2017)).

2.2.6. Stallabrass (Stall)

The Stallabrass model is described in detail in Stallabrass (1979, 1980). The basic idea is to solve Eq.(1), but the radiative heat flux is neglected. Instead of assuming the same freezing temperature as the incoming sea water as in the Overland model, salt expulsion is incorporated in the freezing process. Thus a new freezing temperature of the brine is applied derived from experiments in Tabata *et al.* (1967):

$$T_s = (1+n)T_f \tag{20}$$

The droplet temperature of the spray (T_d) is solved through a droplet cooling equation, which provides a lower temperature than the SST used in the Overland model. One of the challenges is then to find a reasonable estimate of the droplet cooling time (Δt). Stallabrass (1980) assumes a $\Delta t = s/W_r$ where s is the distance travelled and W_r is the relative wind speed or the wind speed of the droplets in a coordinate system following the ship. The droplets are assumed to have a constant spherical size with a diameter of 2 mm which is used in the calculation of a terminal velocity in the droplet cooling equation. Originally s is set equal to a value of 20 m, which provided the best fit according to icing-rate observations in Stallabrass (1980), but such a constant s is not applied in the current study. Hence, s is calculated according to a mathematical expression of the distance from the gunwale to the freezing plate on KV Nordkapp (Eq.(12) in Samuelsen *et al.* (2017)) by applying the angle between the ship and wind (β) in a coordinate system following the ship. β is then defined to be in the interval [90,180] $^\circ$ from a direction where the wind and ship are moving normal to each other to a direction where these two are moving in opposite directions. An advantage with such an approach for calculation of Δt is that it takes into account the shorter distances that droplets have to travel when the direction between the wind and the ship is less than 180 $^\circ$. For the front part of KV Nordkapp the distances are in the range of approximately 6 to 20 m. It is then assumed that the droplets follow a strait line from the gunwale to the midpoint of the plate.

The spray-cloud liquid-water-content formulation (l_{wc}) in the Stallabrass model is initially (Stallabrass 1979) adopted from Kachurin *et al.* (1974) and is a function of wave height ($l_{\mathrm{wc}} = 10^{-3}H_s$), but is developed and adjusted according to observed icing-rate values in Stallabrass (1980) ($l_{\mathrm{wc}} = 1.7 \times 10^{-4}H_s$). Furthermore, Stallabrass (1980) applies a vertical cylinder with a diameter of 3 m as a reference for the icing forecasts. This provides a heat-transfer coefficient of $h_a = 5.17W_r^{0.8}$. The spray temperature (T_{sp}) is assumed to be equal to the droplet temperature (T_d) of a spray consisting of separated droplets only. This temperature is derived from a droplet cooling equation:

$$\frac{dT_d}{dt} = \frac{6\mathrm{Nu}_d k_a}{\rho_w c_w d_r^2}\left(T_a - T_d\right) X_t \tag{21}$$

where

$$X_t = \left(1 + \epsilon\frac{L_v}{c_p p}\right)\frac{R_H e_s\left(T_a\right) - e_s\left(T_d\right)}{T_a - T_d}$$

$$\mathrm{Nu}_d = 0.37\mathrm{Re}_d^{0.6}$$

Stallabrass (1980) solved this equation by assuming that X_t is a constant when integrating T_d. The resulting expression for T_d is then solved by an iterative process. However, this leads to erroneous results for short droplet cooling times for low values of relative humidity (not shown). While Stallabrass applied a constant R_H of 0.9, the observed R_H is applied in the current study. When solving the ordinary differential equation directly (Eq.(21)) through numerical methods (e.g. Dormand and Prince (1980)) this error is circumvented.

2.2.7. Modified Stallabrass (ModStall)

The Modified Stallabrass model is developed by Ross Brown during the 1980s and 1990s (Henry 1995). It has its basis in the Stallabrass model, but some important changes are made regarding different formulations applied in the model. First and foremost the l_{wc}-formulation from Zakrzewski (1986) is applied, but the constant in the beginning of the expression is slightly adjusted by Samuelsen *et al.* (2015) yielding the following equation for the water content of the sea spray:

$$l_{\mathrm{wc}} = 6.36 \times 10^{-5} H_s V_r^2 \exp\left(-0.55(z - 3.5)\right) \qquad (22)$$
$$z \geq 3.5 \text{ m}$$

This formulation is based on spray data collected on the medium-sized fishing vessel (MFV) Narva (Borisenkov *et al.* 1975; Zakrzewski and Lozowski 1989). z = 0 m is originally set at the deck level, but is here adjusted to a calm mean sea level by subtracting the free-board height of an MFV of 3.5 m. In addition, according to Funk (2012), Canadian forecasters had experienced that the Stallabrass model underestimated icing rates. For this reason Brown adjusted the droplet-cooling time expression from Stallabrass (1980) in the ModStall model. Thus the droplet-cooling time is assumed to be equal to the spray-cloud duration time extracted from Figure 14 in Zakrzewski (1986) for ships travelling with an angle (α) of 180 $^\circ$ relative to the waves, i.e. in the opposite direction of the waves. Zakrzewski (1986) is then assuming that the ship speed is a direct function of wind speed and heading (Table 6 in Zakrzewski (1986)). Moreover, wave height and wave period are in Zakrzewski (1986) calculated from the wind speed for an assumed fetch of 200 nautical miles (nm). The spray-cloud duration time, and hence the droplet-cooling time in ModStall, is therefore only a function of wind speed:

$$\tau = t_{\mathrm{dur}} = 11.25 - \frac{V}{4} \qquad (23)$$

Since Brown presumably wanted to create an algorithm that could be applied for several vessels for application in operational weather forecasting, the ship-dependent variables are neglected or set as constants in all expressions. Firstly, a constant relative speed between a ship and waves (V_r) of 10 m s^{-1} is applied in Eq.(22). Secondly, icing rates are calculated for every half meter 3.5 to 9 m above deck level, i.e. 7.0 to 12.5 m above sea level. This is the assumed height of the superstructure of an MFV. The final icing rate is determined by calculating the mean value of the icing rate derived from all these levels. Finally, the absolute wind speed is applied in the spray-flux expression instead of the normal component of the relative wind speed to a vertical cylinder.

Since all the ship-dependent variables are masked in the original ModStall algorithm, a model that includes observed or calculated vessel speeds and directions is evaluated. In addition, z-values 6.5 to 8.5 m above sea level are applied adjusted for the location of the freezing plate on KV Nordkapp. Icing rates are calculated for every 0.1 m vertically. In the spray-flux term the normal component of

the relative wind speed according to a vertical plate is applied instead of the absolute wind speed:

$$\begin{aligned} R_w &= W_{\mathrm{rn}} l_{\mathrm{wc}} \\ &= (\mathbf{W_r} \cdot \mathbf{n_1}) \times 6.36 \times 10^{-5} H_s V_r^2 \exp\left(-0.55(z-3.5)\right) \\ z &\in [6.5, 8.5] \text{ m} \end{aligned} \tag{24}$$

However, the original droplet-cooling time expression (Eq.(23)) is kept, regardless of α, since the expression was not easily adjustable for different headings without changing it completely (Figure 14 in Zakrzewski (1986)). The results of the Modified Stallabrass with relative velocities (ModStall) is presented in the main article, whereas the supplementary material includes verification scores of the original Stallabrass model (Stall) and the original Modified Stallabrass model (ModStal org).

2.2.8. MINCOG

Samuelsen *et al.* (2017) developed a new model adjusted for the KV Nordkapp ship class. The spray flux derived from the Borisenkov *et al.* (1975) data is applied in the current study. For wind speeds below 5 m s^{-1}, spray-cloud duration time ($t_{\mathrm{dur}} = 0.1230 + 0.7008 V_r H_s V^{-1}$) is held as a constant equal to V = 5 m s^{-1} to avoid unrealistic large spray fluxes for very low wind speeds. The rest of the calculations are completed in the same manner as in the original MINCOG model.

2.3. Statistical methods/Nomograms

In the physical models icing rate is derived based on calculations of the spray and heat fluxes in the wave-ship interaction icing process. In the statistical methods the icing rate is estimated based on empirical relationships between icing and important input parameters collected from several icing events. While in most of the physical models a continuous icing rate is calculated and the icing is divided into categories based on this continuous icing-rate calculation, the icing output from the nomograms is only categorical. The advantage of the statistical methods relative to the physical models is the possibility of omitting the complex and somewhat uncertain calculation of the wave-ship-interaction spray flux and the turbulent heat transfer. The drawback of these methods is the challenge of applying enough important input parameters in the nomograms to yield accurate and physically reasonable results. This is in particular a problem when these methods are applied in areas and weather situations different from those that these empirically-based methods are derived from.

Three statistical methods are still applied in operational weather forecasting of today (Table I). These nomograms are reproduced and briefly described in the following section:

2.3.1. Mertins

The nomograms of Mertins (1968) are reproduced in Figure 2. The nomograms assume no icing for wind speeds below 10.8 m s^{-1} (Beaufort 6), $\mathrm{SST} > 8$ °C, or $T_a > -2$ °C. Vessel speed is not taken into account, and according to Mertins (1968) most of the ships, on which the nomograms are based, were only steaming at "low" speeds during the observations.

2.3.2. Lundqvist and Udin (LU), Sawada

Figure 3 is a reproduction of the nomograms of Lundqvist and Udin (1977) and Sawada (1962). LU is based on observations from merchant ships in the Baltic seas, and the nomogram of Sawada (1962) is based on icing events from Japanese patrol and fishing vessels. Lundqvist and Udin (1977) state that their nomogram is valid for ships of size 500-7000 dwt with speeds of 10-15 knots (5.2 to 7.7 m s^{-1}). The onset of icing for somewhat higher temperatures than those in for instance Mertins (1968) is in Lundqvist and Udin

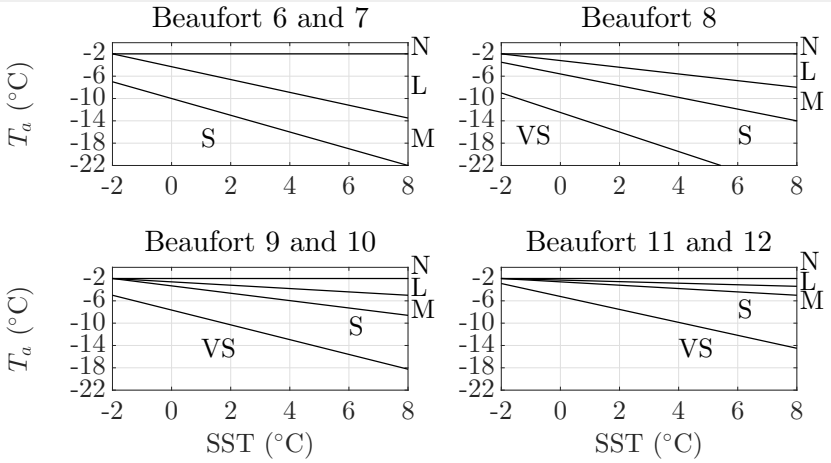

Figure 2. Reproduction of the nomograms of Mertins (1968). N, L, M, S, and VS are abbreviations applied for the icing categories "no icing", "light icing", "moderate icing", "severe icing", and "very severe icing". The VS category is not applied in the current study.

(1977) emphasised to be an effect of the less saltier water in the Baltic sea. Moreover, according to Figure 3 the observed icing in the Japanese fishing grounds also has a similar onset of icing relative to temperature for strong winds. An interesting signature in the LU graph is the reduction in icing severity for decreasing temperatures for low temperatures. No observations below -10°C were reported in conjunction with icing in the Baltic sea (Lundqvist and Udin 1977), and the graphs are here extrapolated linearly to -22 °C based on the extrapolation to -15 °C in the same paper. Lundqvist and Udin (1977) suggest that some of the reason for this decrease in icing for low temperatures is due to the fact that some of the droplets might freeze before hitting the ship in low temperatures. Lundqvist and Udin (1977) also discuss the effect of the sea-surface temperature on icing from their observations. Based on this discussion German forecasters have made a correction table to take into account the effect of sea-surface temperature on icing (pers. comm., Fregattenkapitän Ingmar Behrendt, April 2016). Table II is a reconstruction of this table which is based on Figure 15 in Lundqvist and Udin (1977).

Table II. Diagram to correct the icing-rate categories provided by LU according to the sea-surface temperature.

	New icing category based on x = SST (°C)			
LU Cat.†	$x \geq 6$	$6 > x \geq 4$	$4 > x \geq 2$	$x < 2$
N	N	N	N	N
L	N	L	L	L
M	N	L	M	M
S	N	L	M	S

† N: No icing, L: Light icing, M: Moderate icing, S: Severe icing.

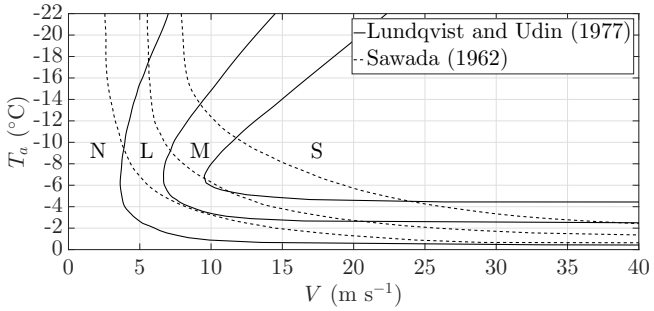

Figure 3. Reproduction of the nomograms of Lundqvist and Udin (1977) and Sawada (1962). N, L, M, and S are abbreviations applied for the icing categories "no icing", "light icing", "moderate icing", and "severe icing". For the LU nomogram the boundaries between the categories are linearly extrapolated to -22 °C.

3. Data sets applied

Three data sets are applied in the evaluation of the different models. The main data set is the thoroughly screened and quality checked data set from a single ship type presented in Samuelsen *et al.* (2017). However, the models are also evaluated against two other data sets from different ship types in order to test a more general application of the models. The dataset from MT Zandberg applied by Overland (1990) is not evaluated since the position data are missing in Zakrzewski *et al.* (1989) and it is therefore not possible to substitute some of the missing parameters like relative humidity and wave period from reanalysis data.

3.1. KV Nordkapp data (KVN)

All methods are tested against the unique data set with 37 observations from the Norwegian coast-guard-vessel class: KV Nordkapp (Figure 4). In addition, 30 no-icing events are also included. In Samuelsen *et al.* (2017) a total of 41 no-icing events are included. However, in 11 of these events the subjective ice-accretion rate indicated icing, although the accumulated ice is measured to be 0 cm. These 11 events are removed in order to reduce the possibility of applying erroneous no-icing events.

3.2. Pease and Comiskey data (P&C)

Overland *et al.* (1986) apply 58 of the 85 observations from Pease and Comiskey (1985) collected outside Alaska. However, an additional screening of these data is necessary. Following the argumentation of Overland *et al.* (1986) that all the data points should be open ocean, only the data points that are flagged as ”open ocean” (o) are kept. For instance are all the data points with suspiciously high icing rates removed (marked with ”s” or ”so” in Pease and Comiskey (1985)). It is also necessary to remove data with zero vessel speed and an angle between the wind and ship (β) less than 90 $^\circ$ in order to avoid downwind situations in which β is not defined for the distance calculation applied by some of the models. In contrast to the KVN data set, only one position point is provided for each event, although the duration of the event is from 1 to 24 h. Observations lasting longer than 12 h are removed to avoid excessive variations in the meteorological and oceanographic conditions during the accumulation period. The remaining 23 observations are visualised in Figure 5. Since no information is provided in Pease and Comiskey (1985) with regard to the representation of the numbers given for the position and meteorological variables for a certain event, it is assumed that the presented numbers are the middle values during the event. As suggested by Pease and Comiskey (1985) the National Weather Service temperature analysis is applied instead of the observed temperature, and a new saturation vapour pressure is calculated from this temperature value instead of applying the listed

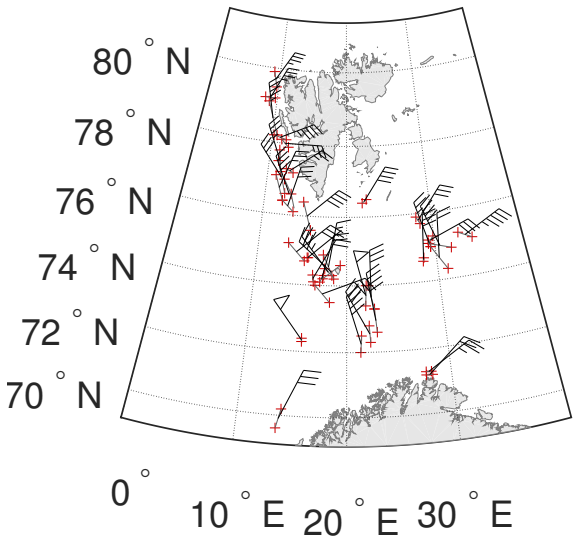

Figure 4. Position of the 37 icing events from Arctic-Norwegian waters presented in Samuelsen *et al.* (2017). Start and end positions are marked with crosses. The straight lines between the start and end positions illustrate the assumed paths. Wind barbs represent the mean observed wind speed (in knots) and direction between the start and end positions.

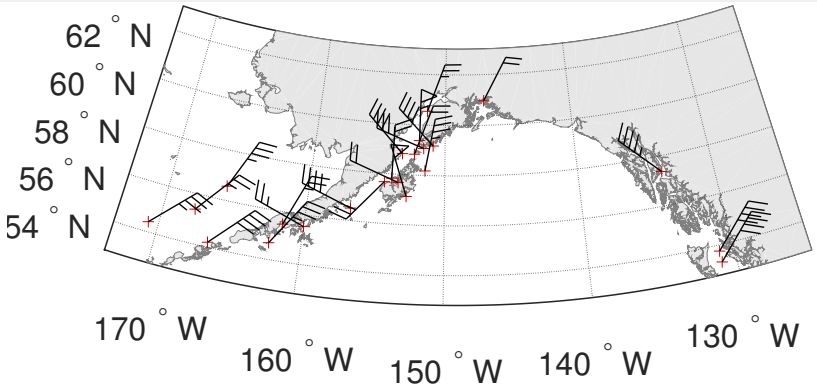

Figure 5. Positions of the 23 screened icing events outside Alaska in the Pease and Comiskey data set applied by Overland *et al.* (1986). The position of an event is marked with a cross. Wind barbs represent the applied wind speed (in knots) and direction.

saturation vapour pressure. Relative humidity is then calculated from the listed vapour pressure and this newly derived saturation vapour pressure. Also following Pease and Comiskey (1985) the wind speed is derived from the maximum value of that visually estimated and that provided by the National Weather Service analysis. Other parameters not provided by Pease and Comiskey (1985) are extracted as follows:

- Salinity (S_w) is collected from the annual climatology values (1900-1992) from the NODC_WOA94 data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at `http://www.esrl.noaa.gov/psd/` (Monterey and Levitus 1997).
- Mean sea-level pressure (p), incoming longwave radiation ($\downarrow$ LW), and incoming shortwave radiation ($\downarrow$ SW) are collected from ERA-Interim reanalysis data (Dee *et al.* 2011). Daily average values are derived from this data set and applied in icing calculations.
- Wave period (P_s) and direction (D_W) are collected from the wave-hindcast data set named Wavewatch III provided by NOAA (Tolman 2014). Wave-period values are collected every three hour from 0000 UTC to 2100 UTC, and the daily average values are applied. Deep-water approximation is assumed for calculation of the wave speed and wave length. For the wave-direction parameter the dominating direction is applied by comparing the direction from all available junctures of a day. Both data sets with 1/15-2/15 ° horizontal resolution and 1/6-1/4 ° resolution are applied in order to cover all data points. For two of the data points close to the shoreline the wave-hindcast-data values are missing, and the wave period is calculated from the wind speed and a minimum fetch of 100 nautical miles from the formula provided by Zakrzewski (1987). The wave direction is in these events assumed to be equal to the wind direction.

It is decided to not replace the visually-estimated wave heights provided by Pease and Comiskey (1985) with reanalysis data, since application of a visually-estimated value in MINCOG provides better verification scores of icing rates from the KV Nordkapp data compared to the application of significant wave height from reanalysis data with 10 km horizontal resolution (Samuelsen *et al.* 2017). It is also decided to apply as the observed rates the event icing rates instead of maximum icing rates, since it is uncertain how the latter ones are obtained (Section 4).

3.3. Roebber and Mitten data (R&M)

Roebber and Mitten (1987) present 307 icing events from several ship types. After selecting only those data with a given ship speed, wind direction, and longitude and latitude, only 27 observations of icing are remaining. Position data are necessary in order to supplement the data set with reanalysis data. In fact reanalysis temperature is applied from ERA-Interim, since it is believed that

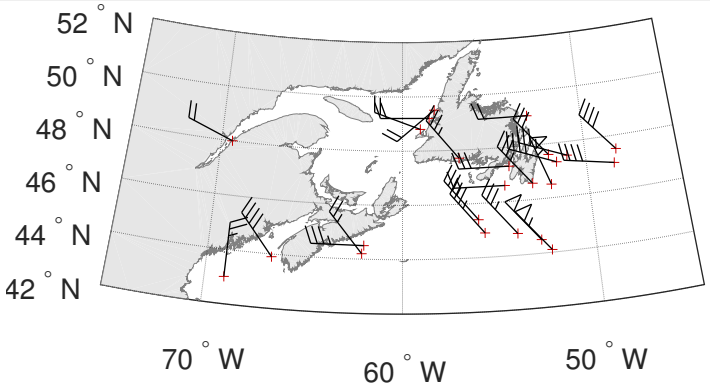

Figure 6. Positions of the 24 screened icing events outside the east coast of Canada in the data set presented by Roebber and Mitten (1987). The position of an event is marked with a cross. Wind barbs represent the observed wind speed (in knots) and direction.

there are some inaccuracies regarding the temperature measurements in the data set (e.g. -15.0 °C observed and -5.0 °C in the ERA-Interim data for a light icing case). When selecting only those events in which the ERA-Interim temperature is negative, a total of 24 observations are remaining (Figure 6). On the contrary of the P&C data, neither sea-surface temperature or humidity data are provided in the original data set. Relative humidity is derived from the daily average temperature and dew-point temperature in the ERA-Interim data, while sea-surface temperature is collected from NOAA High Resolution SST data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at `http://www.esrl.noaa.gov/psd/` (Reynolds *et al.* 2007). Salinity, mean sea-level pressure, incoming longwave and shortwave radiation, wave period, and wave direction are collected in the same manner as for the P&C data. However, for the wave parameters mostly the coarsest horizontal resolution of (1/2)×(1/2) ° of the Wavewatch III is available for these data points, but for two of the events the higher horizontal resolution of (1/6)×(1/4) ° is available and applied. The wave heights are stated to be estimated from the observed wind speed in Roebber and Mitten (1987), and by following Samuelsen *et al.* (2017) it would be preferable to not apply such empirically-based relationships between wind speed and wave height in icing modelling. However, due to the low wave heights obtained when applying the daily average value of the Wavewatch III with a course horizontal resolution of approximately 56 km (1/2 ° in north-south direction), it is decided to apply the estimated wave heights instead of the values obtained from reanalysis data. Only for the wave period and direction the Wavewatch III values are applied. As for the P&C data only one position point and a single date is presented for each event in the data set. These events may therefore last from 1 to 24 h like the P&C data. However, since no duration time is provided for these events, it is not possible to only select the observations lasting less than 12 h as is done for the P&C data.

4. Icing severity

4.1. Challenges regarding icing-severity definition

When forecasting icing severity there are several aspects that need to be considered. Firstly, when a ship moves spatially and temporally, atmosphere and ocean parameters vary as well as the rate of which ice can accumulate. For shorter time periods and shorter distances at open sea, the variations in these parameters are smaller. The total load of ice is dangerous but very ship-type dependent. On the other hand fast accumulation of ice makes the removal process of ice more difficult. In general prediction of icing is mainly based on the rate of ice accumulation and the ability to remove the ice rather than the overall load of ice. For shorter time periods the changes in the environmental parameters are less, and a calculated or predicted icing rate based on a certain set of input parameters are more reasonable to apply than the icing rate predicted for a longer time span. When applying nomograms like the ones in Figure 2 and 3 one must assume that the parameters remain constant or with little variation during the accumulation period or the period in consideration.

Furthermore, ice will accumulate differently on different positions of a ship and among different ship types. Observed ice accumulation is often given with a single value in cm without providing information regarding the location on a ship at which ice

Table III. Different definitions of icing-rate severity categories collected from the literature. The lower panel presents the boundaries between the categories at which the values are converted into units of cm h^{-1} and divided into the three main categories light, moderate, and severe.

Cat/Source	Mertins (1968)	LU[†]	WMO[††]	WC [†††]	BR [$]	Overland [$$]
Trace	-	-	-	0.25-0.64 cm $(3h)^{-1}$	< 0.20 cm h^{-1}	-
Light	1-3 cm $(24h)^{-1}$	0.5-2 cm $(12h)^{-1}$	1 cm $(3h)^{-1}$	0.64-1.27 cm $(3h)^{-1}$	0.20-0.40 cm h^{-1}	< 0.7 cm h^{-1}
Moderate	4-6 cm $(24h)^{-1}$	1-3 cm $(4h)^{-1}$	1-5 cm $(3h)^{-1}$	1.27-1.91 cm $(3h)^{-1}$	0.40-0.96 cm h^{-1}	0.7-2.0 cm h^{-1}
Severe	7-14 cm $(24h)^{-1}$	$>$4 cm $(4h)^{-1}$	6-12 cm $(3h)^{-1}$	1.91-3.18 cm $(3h)^{-1}$	> 0.96 cm h^{-1}	2.0-4.0 cm h^{-1}
Very Severe	≥ 15 cm $(24h)^{-1}$	-	$>$12cm $(3h)^{-1}$	> 3.18 cm $(3h)^{-1}$	-	$>$4.0 cm h^{-1}
			Icing-rate unit $\left(\text{cm h}^{-1}\right)$			
Light	≤ 0.17	≤ 0.25	≤ 0.33	≤ 0.42	≤ 0.40	≤ 0.70
Moderate	0.17-0.29	0.25-1.0	0.33-2.0	0.42-0.64	0.40-0.96	0.70-2.0
Severe	> 0.29	> 1.0	> 2.0	> 0.64	> 0.96	> 2.0

[†] Lundqvist and Udin (1977)
[††] WMO definition from 1975 according to Lundqvist and Udin (1977)
[†††] Converted from inches and derived from Fig. 6 in Wise and Comiskey (1980). The naming of the 5 categories are somewhat different in Wise and Comiskey (1980).
[$] Brown and Roebber (1985)
[$$] Overland *et al.* (1986), and very severe from Overland (1990).

was accumulated. In some data sets (Stallabrass 1980) the recorded icing rate is a weighted mean from measurements on various parts of the ship. Sometimes a total ice load accumulation rate (tonnes h^{-1}) is given (Zakrzewski and Lozowski 1989), but this parameter also varies among ship types according to the total area of the ship on which the ice is accumulated.

Due to these aforementioned circumstances it is not easy to define and divide icing-rate severity into general categories valid for all ship types. This is probably the reason why a large variety of definitions exists in the literature (Table III). In for instance Sawada (1962) no definition of the different icing-rate categories are provided at all. As seen in Table III there are also variations in the number of categories into which the icing-rate severity are divided. Most authors have at least divided the icing rates into the three categories: ”light”, ”moderate” and ”severe”, and others include ”trace” and ”very severe” icing. At the Norwegian Meteorological Institute icing severity is only included in the marine weather forecasts whenever moderate or severe icing is expected (MET Norway 2015), whereas for instance the Japan Meteorological Agency also includes forecasts with light icing (Japan Meteorological Agency 2017). However, there is no clear information provided along with the warning or forecast stating the definition of the icing category applied. Probably it is dependent on the method a specific forecaster has applied when constructing the forecast. But a given forecast may also have been derived without the forecaster considering or having knowledge of the meaning of the specific icing category. Nevertheless, the current study will base the icing-rate category division on the three most common categories that are applied operationally: light, moderate and severe.

4.2. New definitions of icing severity

In addition to the large variations among the different definitions applied, it is also unclear how a certain definition is obtained. Overland *et al.* (1986) state that their definition is based on the ”subjective nomenclature from Alaskan observers”, and that this definition divide their 58 icing observations from the Pease and Comiskey (1985) data set equally. This is the same as stating that the light icing category represents the lower one-third amount (33rd percentile) of a cumulative distribution function (CDF) of icing observations ordered from low to high, and that the boundary between moderate and severe icing should be at the two-third amount (67th percentile) of such a CDF. Furthermore, Overland *et al.* (1986) also propose a definition of a so-called ”potential icing rate” as an average of the maximum icing rate and the event icing rate. While the event icing rate is defined as the total ice accumulation during a certain icing period divided by the number of hours of duration, it is unclear how the maximum icing rate is obtained; whether the maximum icing rate is calculated from the maximum event icing rate over a shorter time span during the same incident, or if there has been some additional instantaneous measurement of icing.

When applying observations from the KV Nordkapp data set, it therefore seems necessary to develop new icing-rate category definitions which are based both on the severity and the frequency of an expected icing-rate category for this particular ship type. In addition, a certain severity category should provide information about the danger or discomfort related to a certain icing rate. Borisenkov and Pchelko (1975) provide the following icing-category definitions based on the crews ability to remove ice for fishing vessels of length 30-40 m with displacements equal to 300-500 tonnes:

- Gradual (light) icing: $dM_i/dt \leq 1.5$ tonnes h^{-1}
- Rapid (moderate) icing: $dM_i/dt \in \langle 1.5, 4.0]$ tonnes h^{-1}
- Very rapid (severe) icing: $dM_i/dt > 4.0$ tonnes h^{-1}

According to Borisenkov and Pchelko (1975) this definition was confirmed by the "seminar on the ice problem" in September 1970. For a large ship like KV Nordkapp which is 105 m long with a displacement of approximately 3300 tonnes, it is reasonable to apply somewhat higher thresholds than the thresholds applied for fishing vessels during the 1970s. In addition, the observations are in general only given in total accumulation of ice on a certain position of a ship and not as ice loads. Fortunately, during the well documented severe icing event reported in Samuelsen *et al.* (2015), 110 tonnes of ice accumulated in a 17-hour period estimated from the readings of the ballast water. In the same time period 20 cm of ice accumulated at a fixed position of the ship. If one assumes that there is a relationship between the ice load accumulated on the whole ship, and the ice thickness accumulated at this fixed position, one may assume that 110 tonnes/17 h = 6.47 tonnes h^{-1} is approximately equal to 20 cm/17 h = 1.17 cm h^{-1}. This implies that an ice-accumulation rate of 0.18 cm h^{-1} at this specific position may roughly be approximated as 1 tonnes h^{-1} on the whole ship. From the subjective estimation of icing rate from the crew at all junctures every 3 hour during the whole trip described with the Synop-code

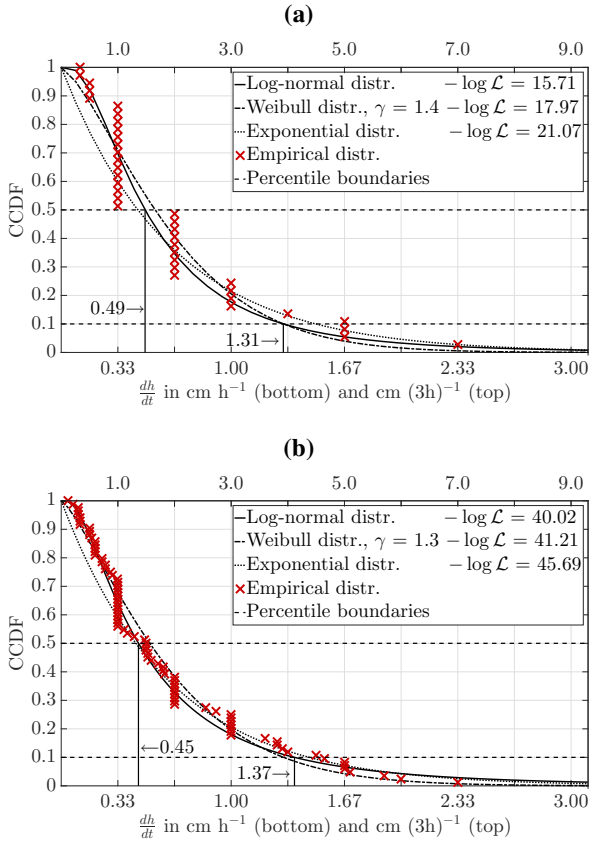

Figure 7. Observed icing rates from a) the KVN data set (37 events), and b) all three data sets merged together (84 events). Icing-rate observations are plotted as crosses together with the fitted lines of the observations following the three different distributions: log-normal, Weibull, and exponential. These distributions are organised in the legend from the closest to the least close fit according to the values of the negative log-likelihood function ($-\log \mathcal{L}$). Dashed horizontal lines represent the 50th and 90th percentile of the cumulative distribution function (CDF), i.e the 50th and the 10th percentile of the conjugative cumulative distribution function (CCDF). The vertical lines illustrate the intersection between these percentiles and the log-normal distribution which has the lowest negative log-likelihood-function value.

parameter $R_S = 2$, the visual inspection of photographs from the aftermath, and personal communication with participants of the voyage (pers.comm., Sveinung Løset, October 2014), it seems appropriate that this particular event is classified as a severe icing event (Samuelsen *et al.* 2015). This average accumulation thickness during the trip may therefore be a starting point for a possible limit for what may be considered as severe icing for this large ship type.

When investigating the observed icing rates of the 37 KV Nordkapp events, it is apparent that the ordered icing-rate observations follow a log-normal distribution most closely when investigating values of the negative log-likelihood function ($-\log \mathcal{L}$) of the different distributions presented in Figure 7 a). A compilation of all 84 icing-rate observations from a mixture of ship types from all three data sets presented in Figure 7 b) also fits a log-normal distribution the best. The value of the average icing rate from the severe icing event mentioned in the previous paragraph (1.17 cm h^{-1}), represents the boundary of the 87th percentile of the cumulative distribution function (CDF) of the log-normal distribution of the KVN data. When setting the boundary slightly higher at the 90th percentile, i.e. the 10th percentile of the conjugative cumulative distribution function (CCDF), the limit between moderate and severe icing would be at 1.31 cm h^{-1}. Interestingly the 90th percentile of the merged data set from several ship types, is close to this value (1.37 cm h^{-1}). For a 3-hour accumulation period the boundary would be around 4 cm. The question is then whether the 4 cm (3h)$^{-1}$ (8 cm in 6 hours) observations should be placed in the moderate or severe category for the KVN data set. However, since the visual estimated icing rate from the synop code (R_S) is only set equal to "ice building up rapidly" (R_S = 2) for the 3 of the 4 events with 5 cm (3h)$^{-1}$ or more in the KVN data set, severe icing is defined as an icing-rate value exceeding 4 cm (3h)$^{-1}$ (1.34 cm h^{-1}). The exponential and empirical distribution also show thresholds between 4 and 5 cm (3h)$^{-1}$ for the 90th percentile of the CDF of this data set. If one assumes a similar ratio for the KV Nordkapp ships for the boundary between moderate and severe icing, and the boundary between light and moderate icing, as the ratio of the boundaries applied for fishing vessels (4/1.5), the boundary between light and moderate icing would be around 0.44 cm h^{-1}. This value is only slightly lower than the 50th percentile value of 0.49 cm h^{-1} for the log-normal distribution of the KVN data, and 0.45 cm h^{-1} for the log-normal distribution of the merged data set (Figure 7). However, for observations in whole cm with a three-hour time interval, 1.5 cm (3h)$^{-1}$ seems like a more natural threshold than 1.3 cm (3h)$^{-1}$ or 1.4 cm (3h)$^{-1}$. For this reason 0.50 cm h^{-1} is selected as the threshold between light and moderate icing instead of 0.44 or 0.45 cm h^{-1}.

In summary the new definitions for light, moderate and severe icing for the KV Nordkapp ship class are as follows:

- Light icing (L): $\frac{dh}{dt} \leq 0.50$ cm h^{-1} (1.5 cm (3h)$^{-1}$)
- Moderate icing (M): $\frac{dh}{dt} \in \langle 0.50, 1.34]$ cm h^{-1} $\left(\langle 1.5, 4.0]\right.$ cm (3h)$^{-1}$)
- Severe icing (S): $\frac{dh}{dt} > 1.34$ cm h^{-1} (4.0 cm (3h)$^{-1}$)

Since these boundaries also approximately follow the 50th and 90th percentile of the merged data set from several ship types, it seems appropriate to apply these boundaries also when verifying categorical icing rates for other ship types than the KV Nordkapp data. Although this icing-rate category definition then is frequency based and based on observations, one must keep in mind that a single icing-category forecast of for instance moderate icing may be more dangerous for smaller ships than larger ships. At least for the KV Nordkapp ship type there is a similarity between this frequency definition applied for the severe events and the visually-estimated "rapid ice build-up" reports from the observers.

5. Results and discussion

In the following section results are presented that stem from the icing-rate calculations of the physical models applied operationally: Overland, Modified Stallabrass, and MINCOG. In addition, the combined effect of introducing a better polynomial fit from Roebber

Table IV. Verification scores for continuous icing rates for the KVN data, the two other data sets merged together, and all three data sets consolidated. Mean error (BIAS), mean-absolute error (MAE), and root-mean-square error (RMSE) have units cm h^{-1}, and the correlation coefficient (r) is unitless.

Name	KVN data (N = 37)				P&C and R&M data (N = 47)				All three data sets (N = 84)			
	BIAS	MAE	RMSE	r	BIAS	MAE	RMSE	r	BIAS	MAE	RMSE	r
	†	††	††	†††	†	††	††	†††	†	††	††	†††
Ref.		**0.44**	**0.62**	**0.27**		**0.48**	**0.64**	**0.24**		**0.48**	**0.66**	**0.18**
Over	7.66	7.76	18.33	−0.06	2.15	2.39	7.24	**0.29**	4.58	4.76	13.32	0.06
Over mod ff	1.73	1.79	2.89	0.06	0.78	1.13	2.92	0.21	1.20	1.42	2.91	0.16
Over bulk	−0.38	0.51	0.67	0.09	−0.49	0.49	0.68	**0.45**	−0.44	0.50	0.67	**0.23**
Over bulk2	**0.12**	0.48	**0.61**	**0.29**	−0.11	**0.33**	**0.43**	**0.59**	**−0.01**	**0.39**	**0.52**	**0.42**
ModStall	1.18	1.28	1.82	**0.55**	0.64	0.81	1.22	**0.40**	0.88	1.02	1.52	**0.47**
MINCOG	**0.06**	0.50	0.64	**0.38**	**−0.12**	**0.36**	**0.52**	**0.42**	**−0.04**	**0.42**	**0.58**	**0.39**

† Non-boldface is indicating that the mean error in the model is significantly (5% significance level) greater or lower than zero error (BIAS = 0). Boldface is indicating that there is not enough support from the data set to reject the null hypothesis that the BIAS in the model is greater or lower than zero (5% significance level).
†† The 95 % lower limit of the ordered MAE and RMSE when applying a Monte-Carlo simulation ($N = 10^4$) of a naïv reference error (Hyndman and Koehler 2006). Models with an MAE and an RMSE value below this reference value, have an MASE and an RMSSE significantly lower than 1 and these values are marked with boldface (see text for details).
††† The 95 % upper limit of no positive correlation ($r = 0$) assuming t-distribution. Models with an r-value marked with boldface are greater than this highlighted reference value.

and Mitten (1987) (Over mod), and the introduction of a fetch factor (Over ff) in the Overland model is presented and analysed in a model named Over mod ff. This combined model is included in the analysis since the Over ff model is applied operationally by Environment Canada (Table I), and the Over mod model is used in a former verification study of the Overland *et al.* (1986) model (Roebber and Mitten 1987). Moreover, the effect of applying a more realistic heat transfer in the Overland model is tested by analysing the results from the new model named Over bulk with the application of two additional heat flux terms and a higher freezing fraction in the model named Over bulk2. Verification of the continuous icing-rate values is presented for all these models. However, it is decided to only present plots and sensitivity analyses of important input variables for the three main models: Overland, Modified Stallabrass, and MINCOG, in addition to the Over bulk2 model that obtains the best verification scores among the other models. Furthermore, after having divided the computed icing rates into icing severity categories all these models are compared with each other, and compared with the three icing-severity nomograms named Mertins, LU, and Sawada. For completeness the verification scores of the Over mod, Over ff, the original Stallabrass model (Stall), and the original Modified Stallabrass model (ModStall org), and the variant of LU with a correction for sea-surface temperature (LUsst) (Table II) are presented in the supplementary material of this paper. For the multi-categorical verification scores only those results that apply the newly derived severity definition are presented in the main article. There is also an additional verification describing the scores when the following definitions are applied: the original definition of Overland *et al.* (1986), the original definition of the nomograms of Mertins (1968) and Lundqvist and Udin (1977) (Table III), and the predictor thresholds from Overland (1990). The tables describing the results from these latter analyses are all added to the supplementary material.

5.1. Continuous icing-rate verification

Table IV provides a summary of the verification scores for the predicted icing rates of the physical models for the KVN data set alone, the P&C and R&M data sets merged together, and for all three data sets merged together. In Figure 8 observed icing rates in cm h^{-1} are compared to the calculated ones for the main models for each of the data set separately. From both Table IV and Figure 8 it is apparent that the Overland model is heavily overestimating the icing rates. Although the overestimation is the smallest when applying the P&C data, some of the predicted rates are still one order of magnitude higher than the observed ones. The effect of changing the polynomial fit (Over mod) and introducing a fetch factor (Over ff) is clearly to reduce this overestimation. However, the model which combines both effects (Over mod ff) is still overestimating the icing rates (Table IV). The idea of applying a bulk-transfer coefficient for heat more in agreement with the coefficients applied in numerical weather prediction models for the heat exchange between the

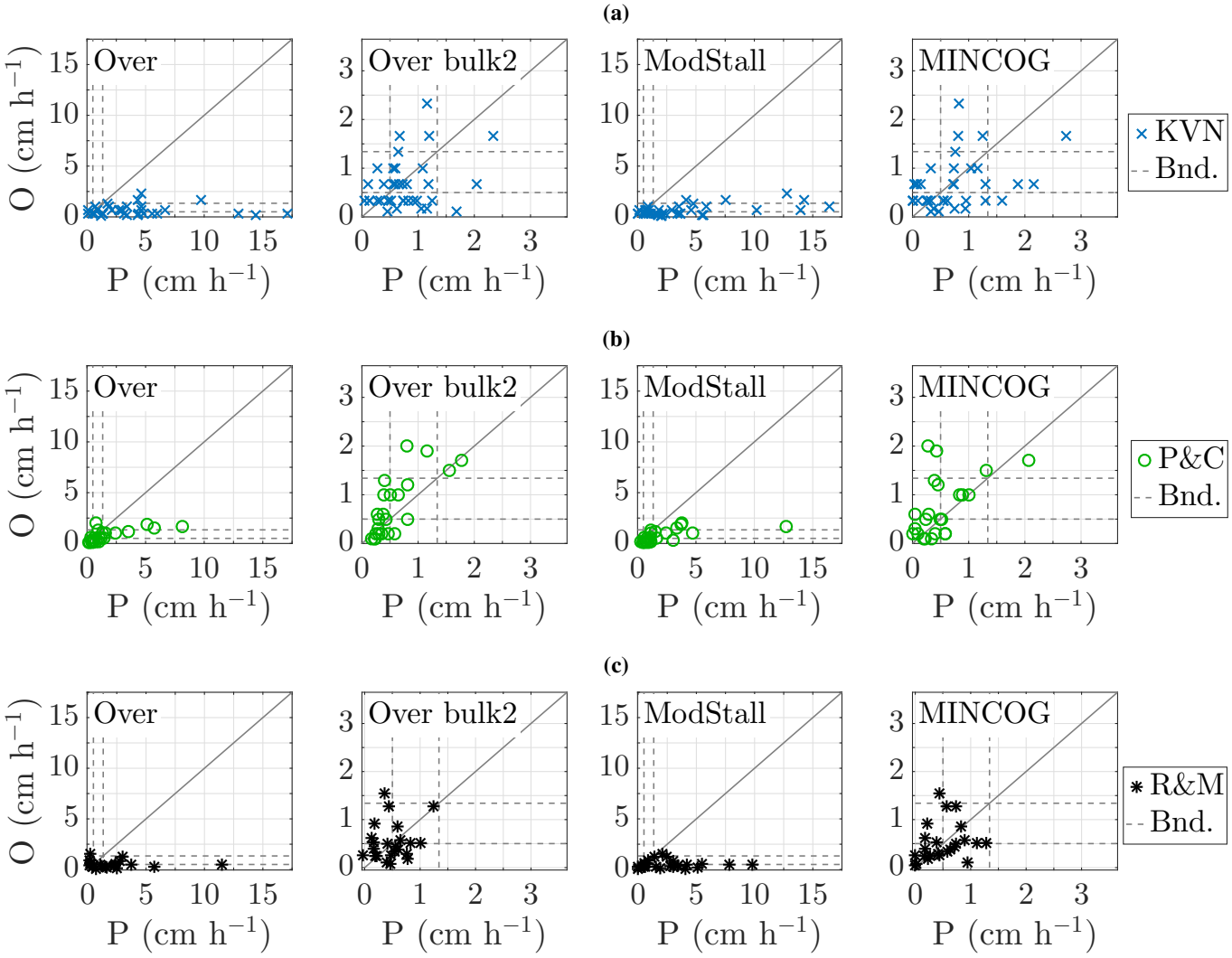

Figure 8. Predicted (P) against observed (O) icing rate for the main physical models applying input parameters from the following data sets: a) the KVN data, b) the P&C data, and c) the R&M data. Notice that the scaling on the axes are different for the models Over and ModStall since these two models are heavily overestimating the icing rates compared to the other two models presented. In addition, in order to enhance readability, four data points (3 in the KVN data set and 1 in the R&M data set) with predictions above 17.5 cm h^{-1} are left out of the plots for the Over model. The predicted values of these data points are 23, 73, and 76 cm h^{-1}, respectively, for the KVN data set, and 48 cm h^{-1} for the R&M data set. The observed values for these events are all below 2.0 cm h^{-1}, and three of them are below 1.34 cm h^{-1}, i.e. light or moderate icing is observed.

ocean and the atmosphere, is apparently a more effective method of reducing the icing rate as is apparent from the negative mean error (BIAS) in Table IV for the Over bulk model. However, since the freezing fraction (n) is assumed to be only 0.06 in the Over, Over mod ff, and the Over bulk model, the spray flux (R_w) is 17 times higher than the icing flux (R_i). This provides a large negative Q_d which contributes to melting (Figure 9). Hence, Over bulk has a distinct negative BIAS since the heat transfer coefficient is reduced by an order of magnitude compared to the original Overland algorithm (Figure 1). Reducing the spray flux to only the double of the icing flux (n = 0.5), and also taking into account the evaporative and radiative heat fluxes (Over bulk2), appear to provide more reasonable predictions compared to the predictions from the Over bulk model, and also compared to the predictions from the Over and the Over mod ff model. This change in freezing fraction is in effect a reduction of the spray flux by an order of magnitude from around hundreds of g m^{-2} s^{-1} to around tens of g m^{-2} s^{-1}. Figure 10 shows that both the spray flux and the icing flux are effectively reduced by several orders of magnitude in the Over bulk2 model compared to the Over model. It is apparent that the Over bulk2 fluxes are more realistic compared to the magnitude of the observed icing fluxes and the spray fluxes obtained by the other models, at least for the lower wave heights. In Figure 10 the mean values from the KVN data set are applied as standard conditions (Std)[†]. However, since the wave period is dependent on the wave height, it was decided to let these two variables vary together in this figure (Std2). The dependency

[†] T_a = −10.4 °C, V = 15.4 m s^{-1}, H_s = 3.9 m, $\beta = \alpha$ = 149 °, V_s = 4.0 m s^{-1}, P_s = 6.0 s (deep-water approximation applied), S_w = 34.9 ppt, R_H = 0.85, SST = 2.3 °C, ↓ LW = 236 W m^{-2}, ↓ SW = 0 W m^{-2}, p = 1002 hPa (Table 2 in Samuelsen *et al.* (2017))

between P_s and H_s is then calculated by applying a linear fit from the KVN data set ($P_s = 4.97 + 0.34H_s$, r = 0.74). Although there are other variables with possibly strong dependencies, like wind speed and wave height, it is not corrected for those since the correlation between these two variables is lower when an event occurred near land or sea ice with the wind blowing from the land or sea ice. This is specifically noticeable in the P&C data set (Figure 11). In the R&M data on the other hand the correlation is probably unrealistically high since the wave height is not observed, but estimated directly from the observed wind speed based on a linear dependency between those two parameters.

Table IV and Figure 8 also illustrate that the Modified Stallabrass model overestimates icing rates. The main reason for this overestimation is the long droplet cooling time, which alters the sign of the heat flux of the spray (Q_d), changing it to become a cooling term which is unique for this model compared to the other models (positive value in Figure 9). In fact applying a more realistic relative wind speed instead of the absolute speed, and applying the water contents from lower elevations in agreement with the elevations of the measurement site of KV Nordkapp increase this negative spray flux and hence increase the calculated icing rates. For this reason the overestimation in the ModStall model is even larger than in the original model (Table S.I). In addition, the ModStall model does not take into account the fact that the spray amounts are affecting the ship in pulses since a time-averaging term is not included in the calculation of the spray-flux, i.e. spray frequency and spray-cloud duration time are not taken into account which leads to an overall larger spray flux relative to that in the original Stallabrass (not shown) and the MINCOG model (Figure 10). In the original Stallabrass model the time-averaging and height effect are indirectly included in the R_w-term since the water amount is calibrated from icing observations on fishing trawlers. Moreover, the Stallabrass model and the MINCOG model have shorter droplet cooling times, and the Q_d-term therefore contributes to melting (Figure 9). Regarding the negative Q_d-term in the MINCOG model this is in part also due to the fact that the spray temperature is calculated as the average value of the sea-surface temperature and the temperature of the individual droplets. The reasoning behind this argument is that the spray may consist of both individual and consolidated droplets. Figure 10 also illustrates how high waves provide large spray fluxes and contribute to melting in the MINCOG model and freezing in

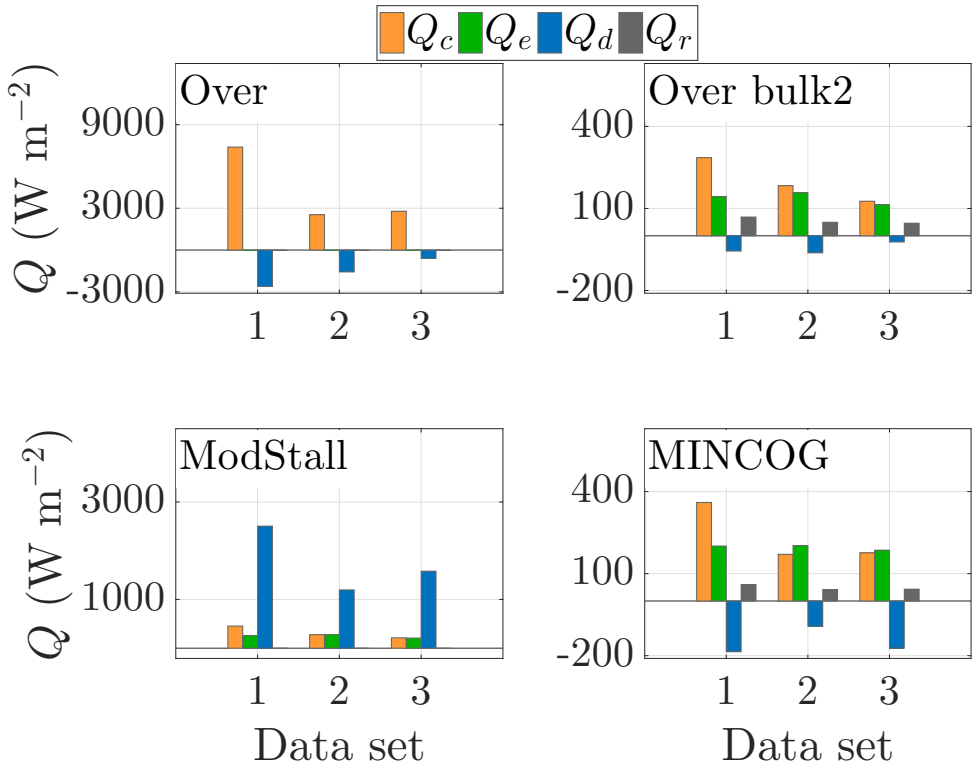

Figure 9. Mean values of the calculated heat fluxes for the main physical models for the three data sets applied. Data set 1 is the KVN data, data set 2 is the P&C data, and data set 3 is the R&M data. Notice that the scaling on the y-axes are different for Over and ModStall due to higher heat flux absolute values in these models compared to the other two models.

the ModStall model. This is evident from a considerable increase in the icing flux (R_i) for increasing wave heights in the ModStall, while the icing flux in the MINCOG model decreases for increasing wave height (Std2).

The MINCOG model has together with Over bulk2 the lowest errors for all three data sets, both when the data sets are considered separately or when these are merged together (Table IV). For all three data sets merged together, the differences between these two models in the mean-absolute error (MAE) and the root-mean-square error (RMSE) are only ± 0.09 cm h^{-1}. However, the correlation coefficient (r) is the highest in the MINCOG model compared to this other low-error model. For the P&C and R&M data merged together, the Over bulk2 has the lowest MAE and RMSE, and the highest r compared to these scores for the other models. It is however interesting that the MINCOG model, that is adjusted for the KV Nordkapp ship type, still provide quite accurate predictions for the mixed data set with several different ship types. This is particularly noticeable when comparing the scores of MINCOG to the other two operationally-applied models Overland and ModStall. When all three data sets are merged together, the Over bulk2 still have the lowest MAE and RMSE, and the highest r. A relatively high correlation coefficient, and relatively small errors for the Over bulk2 and the MINCOG model in each of the three data sets are also apparent from a visual inspection of Figure 8. It is namely noticeable that the points in these two models are closest to the diagonal compared to the other two models presented. Although the correlation in the ModStall model is high when investigating icing rate-predictions and observations from all three data sets separately (Figure 8), the marked overestimation in the model is also observable from this figure.

In Table IV reference values are applied in the top row to get an idea of the magnitude of the errors and correlation coefficient between the modelled and observed icing rates for the different data sets. For the BIAS and the correlation coefficient t-tests are performed to test whether the values are statistical greater or lower than zero (5% significance level). For the BIAS there is a different threshold for each of the models. For the models that are not significantly greater or lower than zero, the BIAS-values are marked with boldface in Table IV. For the correlation coefficient there is only one threshold value for each data set for all models, and only

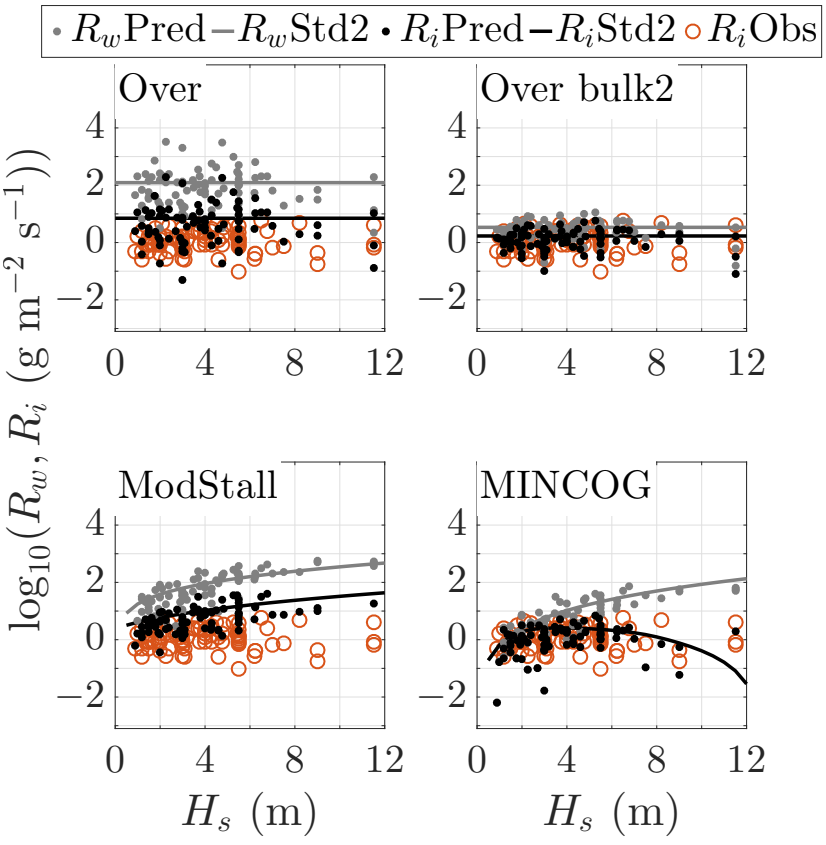

Figure 10. Predicted spray flux (R_w), and predicted and observed icing flux (R_i), plotted on a logarithmic scale for the main physical models versus wave height (H_s). Predicted (Pred) spray and icing fluxes for all 84 observations are marked with grey and black dots, respectively. For the KVN data the mean wave height between the start and the end position is applied. Observed icing flux converted from icing rate by applying an ice density (ρ_i) of 890 kg m^{-3} is plotted with circles (Obs). In addition, calculated spray and icing fluxes are plotted with lines, when only varying wave height, and applying constant standard conditions (Std2) for the other variables except wave period (details provided in text). Spray and icing fluxes are presented with units g m^{-2} s^{-1}.

the models with greater correlation than this given threshold are highlighted as statistically significant positive correlation (r >0). For the MAE and the RMSE it is initially difficult to state whether such icing-rate errors are large or small by just examining the values. However, Hyndman and Koehler (2006) suggest to introduce the scaled error scores mean-absolute-scaled error (MASE) and root-mean-square-scaled error (RMSSE) by dividing the MAE and RMSE with a "naïv reference error" (Appendix Table A.I). As a "naïv reference error" or "random walk error" a method is applied that calculates the mean-absolute difference and the root-mean-square difference between all observations organised in a certain order, e.g. chronologically. However, this method may be sensitive to this initially arranged order of the observations. Thus, a Monte-Carlo simulation is applied which redistributes the order of the observations randomly 10,000 times. A reference value is selected as being the 95% lower limit of these errors of random distributions organised in an order from low to high values. The reasoning behind the introduction of naïv reference errors in this context is to obtain an idea of the magnitude of the errors provided by the MAE and the RMSE for continuous icing rates relative to an error obtained from a forecast without skill. The models that have an error lower than the reference value are highlighted implying that these models also have an MASE or RMSSE significantly lower than one. A value below one indicates skill in these scaled error scores. It is observed that for the KVN data set, Over bulk2 is the only model with an RMSE significantly lower than this reference error, and none of the models have an MAE lower than the reference. On the contrary the MINCOG model has similar magnitudes for the MAE and RMSE compared to the Over bulk2, but the values are not low enough for this sample size (N = 37) to be statistically significantly lower than the values of the 95 % lower boundary of the naïv reference errors organised from low to high values. The mean values of the naïv errors are 0.53 cm h^{-1} and 0.73 cm h^{-1}, respectively, and the MAE and RMSE of MINCOG are at least lower than these values. Regarding the correlation coefficient between the predictions and the observations of icing rates for the KVN data both Over bulk2 and MINCOG have values above the reference of zero positive correlation. However, the correlation coefficient for MINCOG is the largest of these two models. Interestingly the ModStall model has a higher correlation coefficient than the MINCOG model. Nevertheless, as also visualised in Figure 8 a), the large positive BIAS, high MAE, and high RMSE in the ModStall model, in general make the predictions of the icing rates in this model less accurate than in the MINCOG model (Table IV).

For the P&C and R&M data merged together only the two low-error models have better MAE, RMSE, and r than the reference value. However, there is a significant underestimation in icing rate for the Over bulk2 model. In the MINCOG model there is also an underestimation that is not reckoned as significant. This is the case since the spread in the individual errors between the forecast and observation is larger in the MINCOG model than the Over bulk2 model, and there is therefore not enough support from the data to reject a null hypothesis of greater or equal to zero BIAS for the MINCOG model. One of the reasons for the underestimation in the MINCOG model might be that the P&C and R&M data include observations from other ship types than KV Nordkapp in which the estimations of the spray fluxes adjusted for KV Nordkapp are less accurate. In the screened P&C data the lengths of the ships are from 20 to 89 m, with a mean value of 44 m. Although the MINCOG model applies a water-content formulation from a 39 m long ship, the

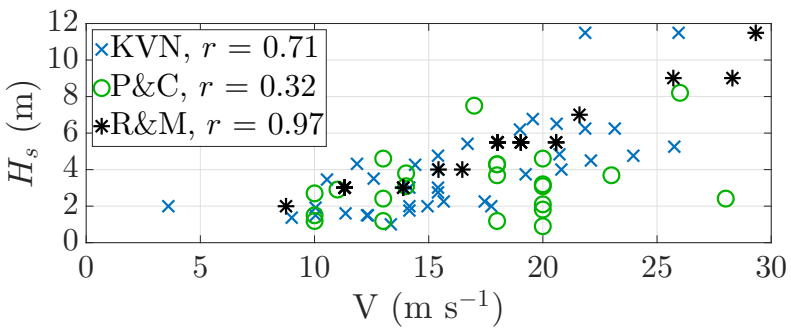

Figure 11. Observed wind speeds compared to observed wave heights for the 3 different data sets applied. Mean values between the start and end position of the KVN data set are plotted as crosses, P&C data are plotted as circles, and R&M data are plotted as asterisks. Wave heights in the R&M data are directly estimated from the wind speed, but in the other two data sets they are visually estimated independently of the observed wind speed. The KVN wave heights are also modified by applying a correction method of Gulev and Hasse (1998) described in more details by Samuelsen *et al.* (2017).

spray frequency and spray-cloud duration time are adjusted for the front position of a longer ship. The relative speed of the spray is also calculated from a mathematical expression of the gunwale of the KV Nordkapp ship, and for ships with very different gunwale shapes this approach is seemingly inaccurate. The freezing fraction applied in Over bulk2 is collected from the calculated fractions in the MINCOG applied on the KVN data set, and for smaller ships with lower free boards it is reasonable to assume that the freezing fraction should in general be lower than the value applied. However, in this model a higher icing rate is obtained if the freezing fraction increases when other parameters are held constant, since the overall heat from the spray is similarly reduced in this manner, i.e Q_d becomes less negative. This is clearly apparent when comparing the icing-rate BIAS of the Over bulk2 to the Over bulk model. The Over bulk model has a considerable lower freezing fraction than has the Over bulk2, and in addition it also considers a lower amount of heat fluxes, in total contributing to less freezing compared to the other model (not shown). Consequently, the underestimation in the Over bulk2 model for the P&C and R&M data may be due to other factors than the freezing fraction adjustment for a different ship, like the assumption of a droplet temperature equal to the sea-surface temperature, and the negligence of the turbulence generated by the local flow pattern in and around the ship, or in general the inaccuracies generated when applying a constant freezing fraction in marine-icing modelling.

For the combined data set of all 84 observations, the Over bulk2 and MINCOG are the only models with the MAE, RMSE, and r values that are better than the reference. The BIAS is also close to zero for these two models in this combined set of all observations. Particularly the Overland model which is the most common operationally-applied model (Table I), have notably the least accurate predictions of all models. One may argue that the original Overland model is adjusted for the "potential icing rate" instead of the event icing rate and therefore should yield higher rates. However, some of the icing rates with values around 40-80 cm h^{-1}, which are apparent in Figure 8, heat and spray fluxes with values of several thousands W m^{-2} (Figure 9) or g m^{-2} s^{-1} (Figure 10), are obviously unrealistic also if the observed icing-rate values are instantaneous.

5.2. Icing-rate sensitivity to important input parameters

Figures 12 and 13 illustrate how the predicted and the observed icing rates for all 84 icing events vary when six of the most important atmospheric, oceanographic, and ship parameters are plotted along the x-axis in the range of the values that are observed. Furthermore, the sensitivity of a selected variable to icing rate is plotted for predictions in which other parameters are set to standard constant conditions (Std). It is noticeable that for the observations in Figures 12 and 13 for all of the six variables, the icing rates are varying for a single constant value of a selected variable. This illustrates the importance of taking into account the complex interaction between the environmental parameters and the ship, and not just considering the variation in one or just a few of these parameters when forecasting icing. In Figure 12 a) it is apparent that the ModStall model predicts high values of icing rates for high waves and long wave periods specifically for other constant conditions. The reason for this sensitivity is that the spray flux is directly proportional to H_s and V_r^2 in the ModStall model. V_r^2 is a linear function of P_s^2. Since all of the spray acts towards cooling through the Q_d-term in this model (Figure 9), the icing rate will increase strongly since the heat flux and available spray are acting in the same direction. For the MINCOG model high waves and long periods lead to a high negative Q_d, and the icing rate is therefore reduced for high waves and long periods compared to the smaller waves and shorter periods. Time-averaging is also reducing the calculated spray flux in the MINCOG model compared to the spray flux from the ModStall model. Another interesting aspect of Figure 12 a) is the insensitivity in the Overland model and the Over bulk2 model to the wave parameters. Although one initially may argue that the Overland predictor (Eq.(6)) is dependent on wind speed which is also affecting the wave height, this wind-speed dependency is originally developed from the convective heat flux and the heat transported away from the freezing brine. The spray generation, on the other hand, is a part of the freezing fraction, which is assumed to be constant. Also from Figure 11 it is clear that the observed significant wave height is not only determined by the local wind speed; the low correlation in the P&C data is not unanticipated considering the positions of the ships in conjunction with the

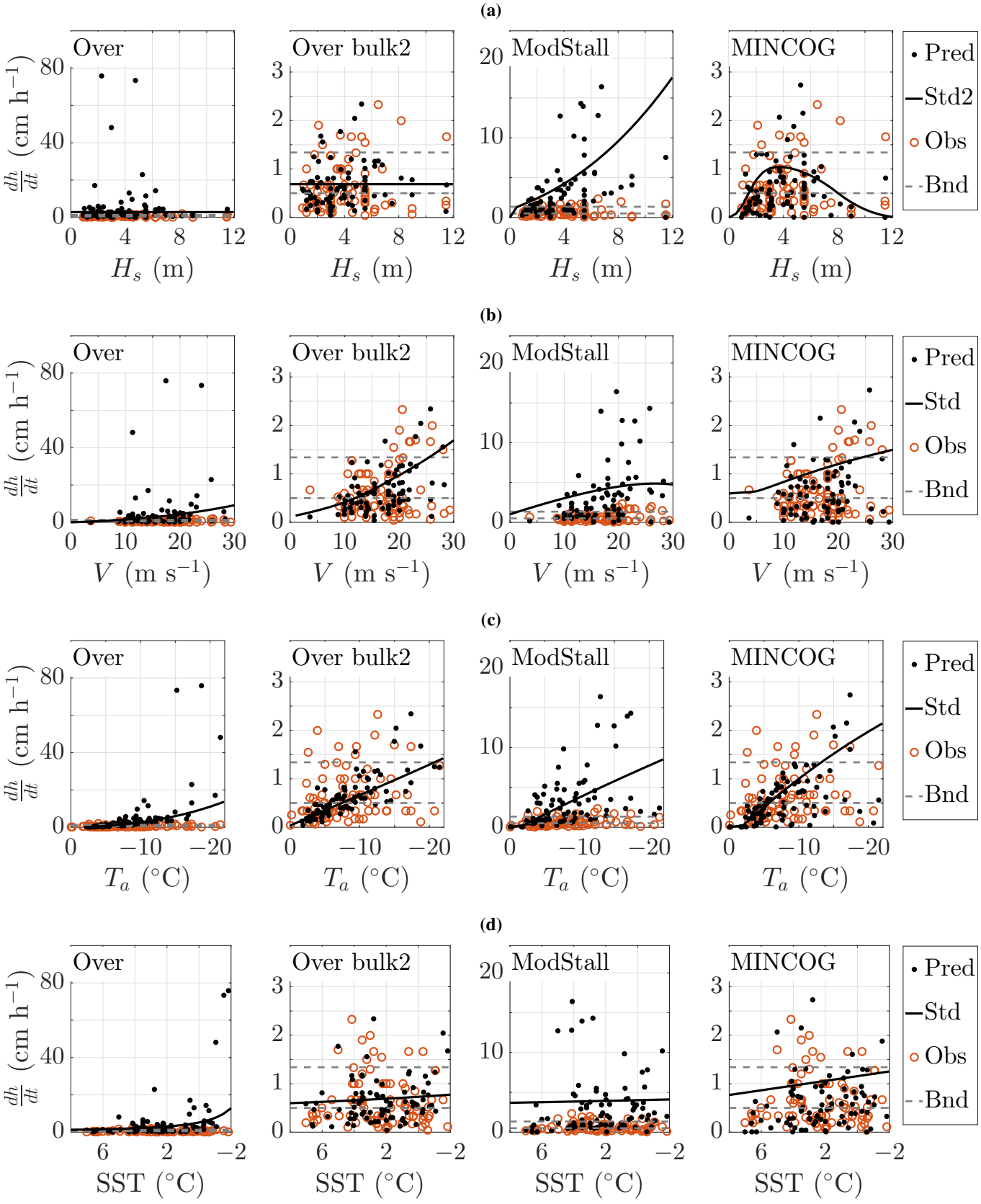

Figure 12. Icing-rate sensitivity to important atmospheric and oceanographic parameters for the main physical models. The following parameters are visualised: a) Significant wave height (H_s), b) Wind speed (V), c) Temperature (T_a), and d) Sea-surface temperature (SST). Predictions (Pred) are marked with dots, Observations (Obs) are marked with circles, and the icing-rate severity boundaries (Bnd) are marked with dashed lines. The icing-rate sensitivity of the different models to the selected variables is plotted for constant standard conditions (Std) and in a) the wave period is also a function of wave height (Std2). For the KVN data set only the mean values between the start and the end positions are applied.

wind direction with regard to the coastline (Figure 5). Fetch-limitation is obvious, and it is also surprising that these events are defined as open ocean events by both Pease and Comiskey (1985) and Overland *et al.* (1986). In the more advanced physical models like ModStall and MINCOG, the heat-balance equation is only a part of the icing-calculation process, since the spray generation, which

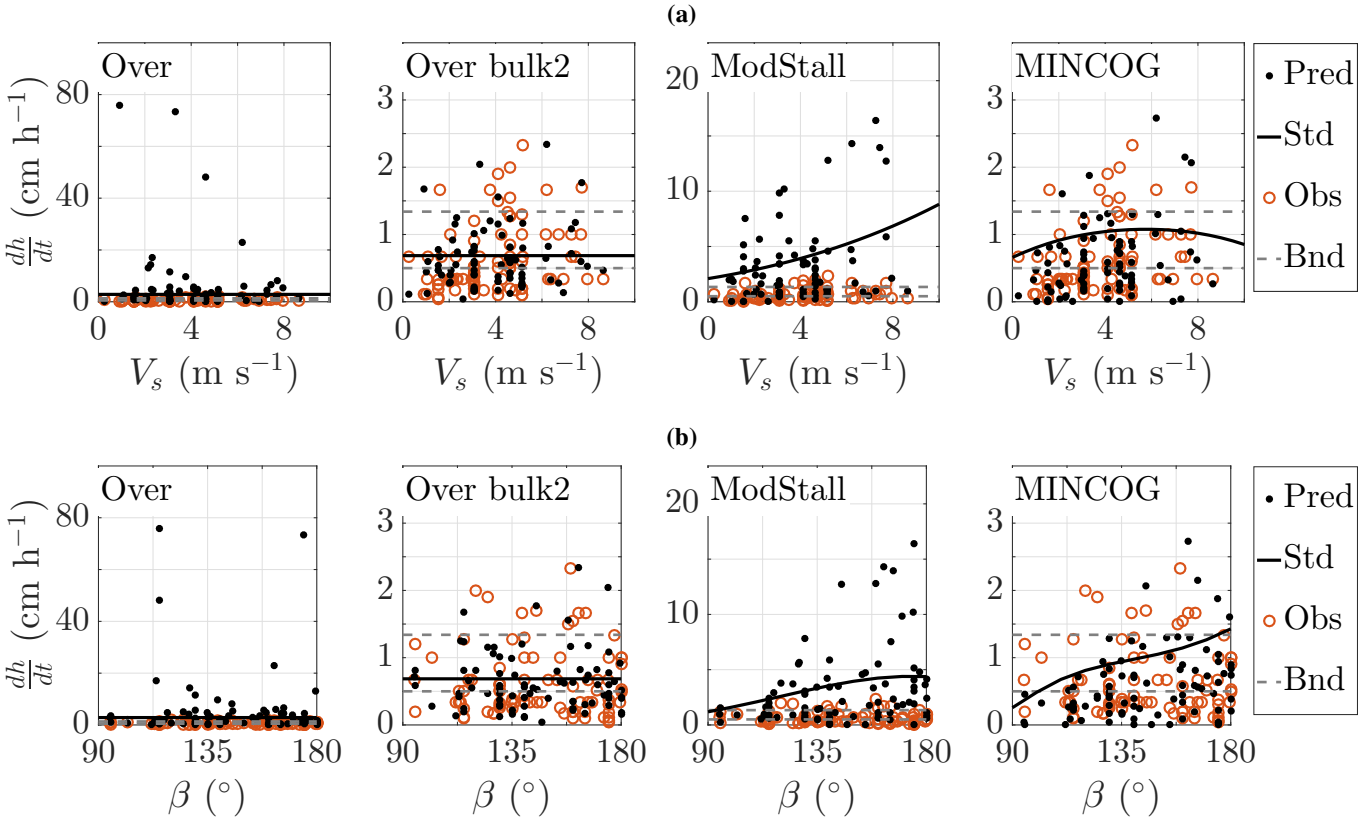

Figure 13. As Figure 12, but for the ship parameters a) Ship speed (V_s), and b) Angle between the ship and wind (β). In b) it is assumed that the direction between the wave and ship (α) is equal and is determined by β for the standard conditions.

is almost physically independent of the temperature, sets the upper limit for the icing flux. The assumption of applying a constant relationship between the icing flux and spray flux also imposes problems for the low-temperature events which is revealed from the observed temperatures and icing rates. From Figure 12 c) it is apparent that the most severe icing events do not occur for the lowest temperatures. A possible explanation for this seemingly observed incoherence, is that the lowest air temperature over sea areas occur near the ice edge or cold land areas during winter time. However, these areas do also have in general lower wave heights compared to areas further away from the coast or ice due to fetch limitation when the wind is blowing from the land or the solid packed sea ice. Fractional sea-ice cover in some of these areas will also reduce wave build-up. At a distance away from the land or sea ice there is strong mixing occurring between the relatively warm sea and the overlying cold air due to low static stability. This mixing will eventually raise the temperature. The claim that high waves do not occur for very low temperatures over sea areas is further underlined by approximately 145,000 ship observations in the Greenland Sea, the Barents Sea, and the Norwegian Sea collected in the period 1976 to 2007 (Figure 14 a)). In this figure there are for instance no occurrences of waves above 4 meter when the temperature is below -20 °C. A map and wind rose illustrating the position and the wind direction of these recorded 156 observations with temperatures below -20 °C further underline this argumentation (Figures 14 b) and c)). The dominating wind direction in these events are namely from around the north and the northeast in which the sea ice normally is located in periods of the year. For the observations south of Svalbard winds from around the north may originate from the cold land areas of this archipelago.

This opposite relationship between wave height and air temperature may also explain why the strongest icing events do not occur for the lowest sea-surface temperatures (Figure 12 d)). However, this does not imply that cold water impose less icing than warm water physically, but the sensitivity of icing is presumably higher to the spray-flux generation controlled by among other things the wave height. In other words low air temperatures and sea-surface temperatures even in moderate winds, do not imply the high icing rates suggested by the Overland model in Figure 12 since the spray generation in the areas these are observed in are too low. This is probably also a feasible explanation why the icing severity in the nomogram of Lundqvist and Udin (1977) based on experimental data in the

Baltic Sea decreases for decreasing air temperature for low values of temperature (Figure 3). Figure 15 in Lundqvist and Udin (1977) also highlights that most severe icing events do not occur for the lowest sea-surface temperatures. Figure 12 b) visualises that high wind speeds are important in order to get high icing rates. Actually all severe icing events are observed when the wind speeds are from around 20 m s^{-1}.

The model calculations of MINCOG also indicate that the icing-rate sensitivity to vessel speed alone is not that strong (Figure 13 a)). For the MINCOG model there is actually a decrease in icing rate for the highest vessel speeds in Figure 13 a). This issue is also supported by the observations. For the ModStall model the unrealistically high sensitivity of the icing rate to vessel speed is as a consequence the high sensitivity to the spray flux due to the positive and increasing values of Q_d for increasing R_w. The relative speed between the wind and ship, and the waves and ship, depend both on the vessel speed, and are included in the spray-flux term. Since the icing-rate predictions in the ModStall model are particularly sensitive to V_r, the predictions from this model are more sensitivity to vessel speed than are those of the other models. The last variable presented in conjunction with icing rate is the angle between the wind and a ship (β) which is plotted in Figure 13 b). The observations indicate that there is somewhat less icing when β is close to 90 °, than for higher angles. This observed sensitivity of the icing rate to the low β-angles is in general most apparent in the MINCOG model compared to the other models. For the ModStall model the overestimation related to the high values of the spray flux for high β-values is the most apparent signature. Icing-rate sensitivity to other variables like incoming sea-water salinity (S_w), incoming longwave radiation ($\downarrow$ LW), and incoming shortwave radiation ($\downarrow$ SW) is not presented, since the predicted and observed icing rates have low sensitivities to these variables, and specifically to the range of the values of these parameters applied in the current study. For instance there is only one observation with a S_w-value below 30 ppt.

5.3. Categorical icing-rate verification

In order to be able to verify the models ability to separate between events with ice accumulation and those without an accumulation, an additional category of no icing (N) is also applied to the analysis when verifying the icing rates divided into categories. Predictions encountering a value below 0.05 cm h^{-1} are considered as no icing. The reason for not using 0 cm h^{-1} as the boundary between no and light icing is the possibility of very low icing-rate predictions being included in the light-icing category when applying a threshold exactly equal to zero. In fact this application of a higher threshold than zero resulted in the circumstance that the lowermost icing-rate value in the R&M data is hereby classified as no icing instead of light icing since it has a value of 0.04 cm h^{-1}. Furthermore, all icing-rate values from the models and the observations are divided into 4×4 contingency tables with predictions versus observations. These four categories are no icing, light icing, moderate icing, and severe icing. The following multi-categorical verification scores are calculated from these contingency tables: Percent Correct (PC), Heidke Skill Score (HSS), Peirce Skill Score (PSS), and Gandin-Murphy Skill Score (GMSS) which is also named Gerrity score. The use of multi-categorical skill scores is a condense way of summarising the information from the contingency tables. Details about these scores may be found in Samuelsen *et al.* (2017) and in Wilks (2011) (Ch. 8). In short the PC, HSS, and PSS are scores taking into account the number of hits for each category relative to the total number of events with (HSS, PSS) or without (PC) adjustments for those hits occurring due to random chance. While PC, HSS, and PSS are characterised by rewarding hits on the diagonal of the contingency table, the Gandin-Murphy Skill Score (GMSS) considers all entries in the contingency table by applying a scoring weight to each element in the matrix based on sample climatology. Misses for the less common categories close to the diagonal are weighted higher than misses of the more common categories, or misses further away from the diagonal. As the HSS and PSS, GMSS is equitable and penalises hits occurring due to random chance alone. In order to test the icing models prediction ability relative to a random placement of numbers in a predicted category, a Monte-Carlo simulation is executed which distribute numbers randomly 10,000 times in the different categories, and the skill scores are calculated from these randomly distributed numbers. Two different random distributions are considered, one that distributes numbers equally

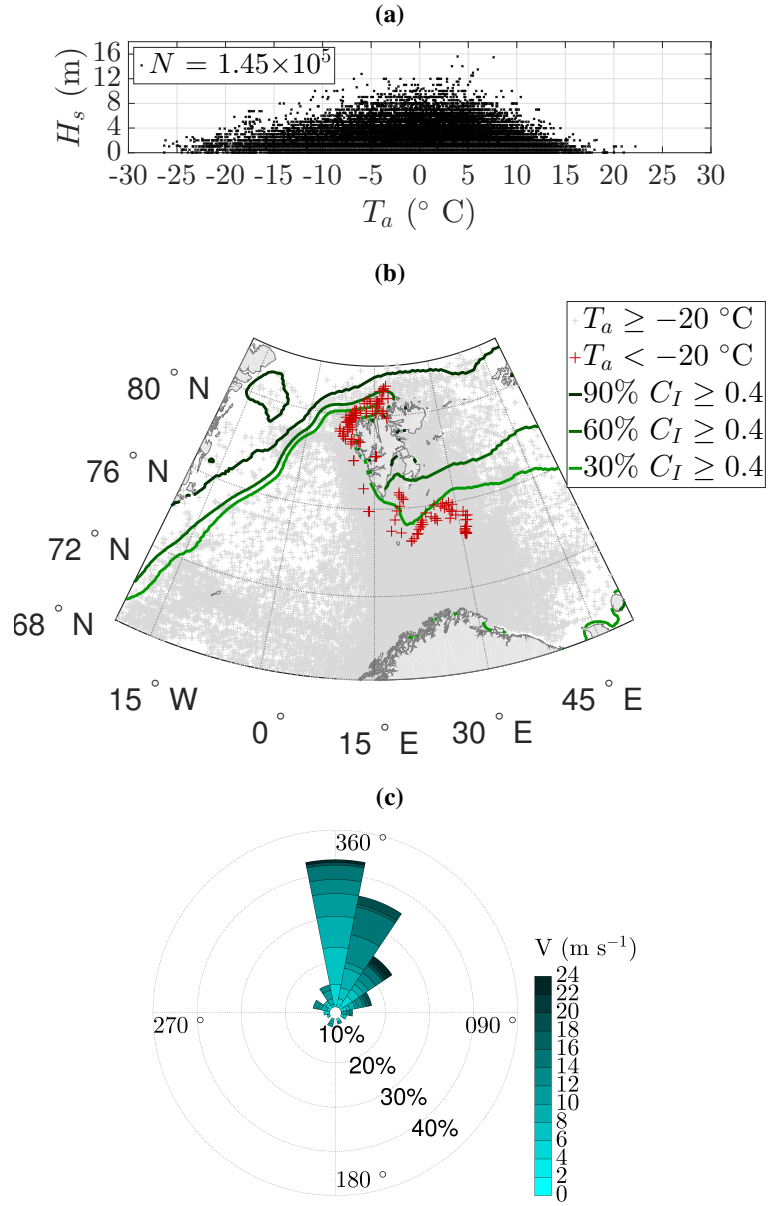

Figure 14. a) Observed wave heights and temperatures from ships operating north of 68 °N and between 20 °W and 50 °E in the Greenland Sea, the Barents Sea, and the Norwegian Sea collected in the period 1976 to 2007. For the visually-estimated wave heights a correction method of Gulev and Hasse (1998) is applied described in more details by Samuelsen *et al.* (2017).
b) The positions of the ship observations from a). Observations with temperatures below −20 °C are marked with a large cross. All other observations are marked with a smaller cross. Three selected relative frequencies of an ice concentration (C_I) more than 0.4 from all days in the years 1979 to 2007 provided by OSISAF (2015) are plotted with solid lines. Areas with $C_I < 0.4$ contain presumably enough water to generate waves, and this boundary is also applied for the selection of icing events in Samuelsen *et al.* (2017).
c) Wind rose for the observations in b) with $T_a < -20$ °C.

in each category, and one that distributes numbers according to the distribution of the observed icing rates in each category for the different data sets evaluated.

Table V describes the verification scores of both the six physical models and the three statistical methods evaluated. In general categorical verification is sensitive to the boundary values selected. For this reason a newly proposed definition of icing-rate categories adjusted towards the KV Nordkapp ship class is applied. Since the limits between light, moderate, and severe icing, when applying the 50th and 90th percentile of the CDF of the observed icing rates for the merged data set with several different ship types, are not far from the boundaries obtained when applying the KVN data alone (Figure 7), these new boundary values are also applied when verifying the merged data set in Table V. In order to compare the verification scores obtained by this new definition, the verification scores when applying the old definition used by Overland *et al.* (1986) and Samuelsen *et al.* (2017) are also evaluated. Moreover, verification scores when applying the original definition from Mertins, LU, and Overland (1990) are tested. The results of the scores from these latter evaluations are provided by the supplementary material of this article (Table S.III).

Table V. Verification scores for the 4×4 contingency tables with predicted versus observed icing rates distributed in the four categories no-icing, light icing, moderate icing, and severe icing according to the newly proposed definition of the thresholds between these categories (Section 4.2). The scores of the six physical models and the three nomograms are all listed below the reference values (Ref.) with the scores of the nomograms in the lowermost panel.

Modelname	KVN with no-icing (N = 67)				P&C and R&M (N = 47)				All data (N = 114)			
	PC	HSS	PSS	GMSS	PC	HSS	PSS	GMSS	PC	HSS	PSS	GMSS
New category definition												
Ref. (unif.)†	**0.34**	**0.11**	**0.13**	**0.19**	**0.36**	**0.12**	**0.16**	**0.33**	**0.32**	**0.09**	**0.09**	**0.14**
Ref. (obs.)††	0.42	0.13	0.13	0.14	0.53	0.18	0.18	0.14	0.37	0.10	0.10	0.10
Over	0.21	0.05	0.07	**0.28**	0.32	0.05	0.06	**0.45**	0.25	0.07	0.08	**0.22**
Over mod ff	0.22	0.05	0.07	**0.33**	0.30	−0.02	−0.02	**0.34**	0.25	0.03	0.04	**0.17**
Over bulk	0.31	−0.00	−0.00	0.06	0.34	−0.00	−0.00	0.23	**0.32**	−0.03	−0.03	0.03
Over bulk 2	0.30	0.07	0.08	**0.22**	**0.60**	**0.30**	**0.29**	**0.56**	**0.42**	**0.16**	**0.15**	**0.24**
ModStall	0.19	0.05	0.06	**0.26**	0.34	**0.16**	**0.21**	**0.56**	0.25	**0.10**	**0.12**	**0.29**
MINCOG	**0.40**	**0.17**	**0.18**	**0.30**	**0.51**	**0.21**	**0.21**	**0.44**	**0.45**	**0.20**	**0.20**	**0.26**
Mertins	**0.34**	**0.14**	**0.16**	**0.39**	**0.40**	**0.16**	**0.20**	0.14	**0.37**	**0.18**	**0.19**	**0.32**
LU	0.18	0.02	0.03	**0.26**	0.26	0.04	0.05	0.10	0.21	0.03	0.03	**0.20**
Sawada	0.19	0.04	0.04	**0.28**	**0.38**	**0.13**	**0.16**	0.18	0.27	0.09	**0.10**	**0.24**

† The 95 % upper limit of the ordered distribution organised from low to high of the different verification scores when applying Monte-Carlo simulations ($N = 10^4$) assuming that the predictions are uniformly distributed inside each category.
†† The 95 % upper limit of the ordered distribution organised from low to high of the different verification scores when applying Monte-Carlo simulations ($N = 10^4$) assuming that the predictions are distributed in the same manner as the observations.

From a diligent inspection of Table V it is apparent that the MINCOG model, the Over Bulk2 model, and the Mertins method are the only methods having better verification scores than all the reference scores generated from random values for both random distributions when all data points are considered. However, the Over bulk2 model, which had the overall best scores for continuous icing rates when all three data sets are consolidated (Table IV), has less accurate predictions when no icing events are included. When investigating the contingency table for the Over bulk2 model in detail for the 67 KVN events, it is revealed that only two of the 30 no-icing events for the KV Nordkapp ships are observed with an icing rate below 0.05 cm h^{-1} in this model (not shown). The MINCOG model on the other hand has the highest PC, HSS, and PSS values for the KVN data alone and when considering all data points. Regarding the scores of the nomograms which are not evaluated in the previous sections, LU is performing worse than most of the other models, while Sawada is comparable to the model performance of ModStall. Interestingly both Mertins and the LU methods have more accurate predictions when applying this new category definition for the predictions compared to the original definitions of these nomograms (Table V compared to Table S.III). Mertins has in fact a better GMSS value than the MINCOG model for both the KVN data and the consolidation of all data. This difference is due to somewhat more hits for the moderate and severe events in the contingency table of Mertins compared to MINCOG. However, Mertins has more overestimations of a single category than has the MINCOG model. Indeed it is noticeable that the GMSS value is generally very sensitive to hits for the severe events. This is the reason why the GMSS value being relatively high for the clearly overestimating models of Overland and ModStall, where most severe icing events are in fact predicted as severe. The scoring matrix is designed in such a way that the overpredictions of the three other categories than the severe category are not penalised as much as the credit gained by hitting the severe events. This is underlined even more when investigating the verification scores when applying the original Overland category definition (Table S.III) in which only the overestimating models of Overland, ModStall, and Mertins are performing better than the reference forecast for the GMSS.

5.4. *Data set uncertainty*

Indeed the uncertainty regarding the location on the ship in which the icing has been reported in the P&C and R&M data, might be an argument for not relying excessively on the verification results of the icing-rate calculations when applying input parameters from these two data sets. However, it is believed that the observations are collected in front of the ship on the superstructure, which are the common practice for icing observations according to WMO (1962) which state that ice accretion should be reported in the synop code when it is "encountered on the superstructure of the ship". In addition, it is not unreasonable to assume that the estimation from the MINCOG model of both the spray temperature and the spray speed is valid also for other ships with similar bow shapes as the KV Nordkapp. The distance from the bow in which freezing transpires in the front of the ship should neither be too far from that of KV Nordkapp. On the KV Nordkapp ships the distance from the gunwale to the location in which icing is measured is around 20 m from the bow and not far in the back of the 105 m long ships. This will probably make the calculations adjusted towards the KV Nordkapp ships more applicable for other shorter ships than otherwise.

An additional issue with the P&C and R&M data is the uncertainty regarding the applied meteorological and oceanographic parameters, specifically since more parameters are obtained from reanalysis data, and that reanalysis data with courser horizontal resolution is applied compared to the KVN data. There is also in particular a concern regarding the use of the wave height in the R&M data set derived from the wind speed. For this reason the verification results from the KVN data alone are probably the most reliable ones. On the contrary, the in general thorough selection and screening of the P&C and R&M data, and the supplementation of missing parameters by using reanalysis data, probably enhances accuracy and applicability of these two data sets compared to previous verification and modelling of icing rate applying these data (Brown and Roebber 1985; Overland *et al.* 1986; Roebber and Mitten 1987). Moreover, the verification results from these two data sets lead to generally the same conclusions as the results from the KVN data, i.e. that Overland and ModStall are overestimating icing rates and that Over bulk2 and the MINCOG model are the most accurate icing models verified in the current study.

6. Summary and conclusions

This study incorporates several unprecedented results regarding ship-icing prediction models and nomograms. Firstly, the study provides a comprehensive overview of the methods applied in operational weather forecasting. Secondly, the physical bases and assumptions of the icing models are critical evaluated by presenting and testing several modifications of the models including models in which most of the original assumptions have been changed. Hence, it is highlighted that the current version of the commonly-applied models of Overland (1990) and Modified Stallabrass (Henry 1995) are immensely overestimating the icing rates. Samuelsen *et al.* (2015) highlighted the challenges with these two models for a severe icing event related to a developing polar-low situation in the Barents Sea. The current study has fulfilled this evaluation by applying a more complete data set. This is probably the first time all of these methods are evaluated thoroughly against more comprehensive data sets including most atmosphere, ocean, and ship parameters involved in the icing process. The overestimation in the Overland (1990) model is particularly noticeable for low air and sea-surface temperatures. Nevertheless, it is discovered that instead of reducing the icing rates in the Overland model somewhat artificially by changing the polynomial fit or introducing a fetch factor, it is possible to formulate a simple physical Overland model providing more accurate predictions than the original one. This is obtained by applying a typical bulk-transfer coefficient for the heat exchange between the ocean and atmosphere, and including the evaporative and radiative heat flux in the calculation of icing rate. The evaporative heat flux is notably important in the calculation of the icing rates when applying input parameters from the P&C and R&M data sets in which the temperatures are higher than in the KVN data set (Figure 9). However, this new model, named Over bulk2, has also challenges regarding an in general overestimation of the icing rates in fetch-limited regions with very low air and sea-surface temperatures (Figure 12 c) and d)).

Regarding the modifications done in the model of Stallabrass (1980) in a model named Modified Stallabrass, it is apparent that the application of a longer droplet cooling time results in cooling from the sea spray instead of melting, which provides unrealistically high icing rates. The removal of the ship dependent variables, as is done in the original Modified Stallabrass model, is reducing this overestimation by lowering the calculated spray amounts, but the icing-rate predictions are still improbably high. In fact, the original Stallabrass (1980) model has better verification scores than the Modified Stallabrass model (Table IV vs. Table S.I, and Table V vs. Table S.II).

Simple nomograms with the definitions attached are not very accurate compared to the other methods. However, if applying a category definition which is more in agreement with the climatology of icing observations from the ships from which the nomograms are evaluated, these methods are more accurate (Table V vs. Table S.III). Indeed the nomogram of Mertins (1968) is providing reasonable results, specifically for the severe icing events and may be applied as a simplistic method when using the new category definition of this paper. Observations also indicate that a lower threshold for wind speed around 10.8 m s^{-1} (Beaufort 6) in icing events as applied in Mertins (1968), is not an unreasonable assumption based on the icing observations from the three data sets applied (Figure 12 b)). On the other hand this does not indicate that it is not possible for icing to occur at lower wind speeds. In addition, this nomogram does not apply wave height as a separate input parameter. As a matter of fact it is revealed that high waves and very low air temperatures rarely occur in conjunction over sea areas (Figure 14). For this reason models and nomograms that do not treat wind speed and wave height separately, will overestimate icing in such areas. Consequently it is preferable to apply methods taking this effect into account.

Finally, the overall best model from the results of this study is the MINCOG model. This is apparent when considering both the continuous and categorically verification scores, in addition to the predicted icing-rate sensitivity to important input parameters involved in the ship-icing process when compared to the sensitivity of the icing-rate observations to the same parameters. In general it is rather encouraging that this relatively simple, stationary model has more accurate predictions compared to the ModStall and Overland model, also when this model is tested against icing-rate observations from ship types that the model is not adjusted for. However, there are still uncertainties regarding the accuracy of the icing observations and data sets applied in the current study, and the use of the same spray-flux formulation in icing-rate calculations for different types of ships. This is probably the reason for the errors in general still being relatively high and the correlation relatively low for the MINCOG model, also when comparing errors and skill scores obtained by this model with naïv reference errors or skill scores generated from random numbers. A possible way to handle all this uncertainty in future marine-icing forecasts is to introduce ensembles by applying different expressions for the spray flux, spray temperature, heat-transfer coefficient or other uncertain factors. Including input parameters that stems from different ensembles as well, may then generate a large ensemble prediction system of icing. Whether such ensemble icing-rate forecasts will provide information that has high enough quality to be applicable in operational weather forecasting is then a question that needs to be addressed in further investigations. Such models must also be tested against hopefully more accurate icing-rate observations from a range of ship types from future marine-icing observation field programs.

Acknowledgements

This work is supported by Norwegian Research Council through the MAROFF program [grant number 226404]. The authors would like to thank all colleagues at UiT - The Arctic University of Norway, and collaboration partners at the Norwegian Meteorological Institute for providing crucial help, feedback, and discussions. The Norwegian deep water program is acknowledged for the use of the NORA10 hindcast data, the NOAA Earth System Research Laboratory - Physical Science Department for the use of reanalysis data for Alaska and the east coast of Canada from their Web site at `http://www.esrl.noaa.gov/psd/`, the European Centre of Medium range Weather Forecasting (ECMWF) for the application of the ERA-Interim data, in addition to the Department manager M. N. Jørgensen of the Norwegian Coast Guard (pers. comm., April 2014) for allowing publication of the observations from the KV Nordkapp ships. A

special gratitude goes to Dr Signe Aaboe at the Norwegian Meteorological Institute for helping out with extracting climatological ice-concentration data. The author would also like to thank the anonymous reviewers, Associate Professor Kåre Edvardsen, and Professor Rune Graversen for important and crucial feedback of the manuscript. Finally, the author declares that there is no conflict of interest regarding the work provided in the paper.

Appendix

Table A.I contains an explanation of the different symbols applied in the paper.

Table A.I. List of symbols

A	Albedo of freezing surface
BIAS	Mean error: $\frac{1}{n'}\sum_{i=1}^{n'}(P_i - O_i)$, n' number of events, P_i predictions, O_i observations
C_E	Bulk-transfer coefficient for moisture
C_H	Bulk-transfer coefficient for heat
C_I	Ice concentration (fraction)
c	Wave-phase speed (m s^{-1})
c_g	Wave-group speed (m s^{-1})
c_p	Specific heat capacity of air (1004 J kg^{-1} °C^{-1})
c_w	Specific heat capacity of sea water (4000 J kg^{-1} °C^{-1})
d_r	Droplet diameter (2.0×10^{-3} m)
D_W	Wave direction (°)
e_s	Saturation vapour pressure (hPa)
g	Gravitational acceleration (9.81 m s^{-2})
dh/dt	Icing rate (cm h^{-1})
h_a	Heat-transfer coefficient (W m^{-2} °C^{-1})
h_e	Evaporative heat-transfer coefficient (W m^{-2} °C^{-1})
H_s	Significant wave height (m)
k	Transfer coefficient (J m^{-3} °C^{-1})
k_a	Thermal conductivity of air (0.023 W m^{-1} °C^{-1})
L_{fs}	Latent heat of freezing of saline water (J kg^{-1})
L_v	Latent heat of vaporisation (2.5×10^6 J kg^{-1})
l_{wc}	Liquid water content in spray (kg m^{-3})
↓↑ LW	Incoming and outgoing longwave radiation (W m^{-2})
dM_i/dt	The accumulation rate of the total mass of ice (tonnes h^{-1})
MAE	Mean-absolute error: $\frac{1}{n'}\sum_{i=1}^{n'}\lvert P_i - O_i\rvert$
MASE	Mean-absolute-scaled error: $\mathrm{MAE}\times\left(\frac{1}{n'-1}\sum_{i=2}^{n'}\lvert O_i - O_{i-1}\rvert\right)^{-1}$
N	Number of elements
n	Freezing fraction
$\mathbf{n_1}$	Normal vector towards freezing plate
Nu_d	Droplet Nusselt number
P_s	Significant wave period (s)
p	Air pressure at mean sea level (hPa)
Q_c	Convective heat flux (W m^{-2})
Q_d	Heat flux from incoming water droplets (W m^{-2})
Q_e	Evaporative heat flux (W m^{-2})
Q_f	Heat flux released by freezing (W m^{-2})
Q_r	Radiative heat flux (W m^{-2})
r	Correlation coefficient
Re	Reynolds number
Re_d	Droplet Reynolds number
R_d	Gass constant of dry air (287 J kg^{-1} K^{-1})
R_i	Ice accretion flux (kg m^{-2} s^{-1})
R_H	Relative humidity of air (fraction)
R_S	Visually-estimated icing rate (code)
R_w	Spray flux (kg m^{-2} s^{-1})
RMSE	Root-mean-squared error: $\sqrt{\frac{1}{n'}\sum_{i=1}^{n'}(P_i - O_i)^2}$
RMSSE	Root-mean-squared-scaled error: $\mathrm{RMSE}\times\left(\frac{1}{n'-1}\sum_{i=2}^{n'}(O_i - O_{i-1})^2\right)^{-1/2}$
S_w	Salinity of sea water (ppt)
↓↑ SW	Incoming and reflected shortwave radiation (W m^{-2})
s	Distance from freezing plate to gunwale (m)
SST	Sea-surface temperature (°C)
T_a	Air temperature at ship level (°C)
T_d	Temperature of individual droplets (°C)
T_f	Freezing temperature of sea water (°C)
T_s	Freezing temperature of brine (°C)
T_{sp}	Temperature of the spray at the position where icing occurs (°C)
T_v	Virtual air temperature at ship level (°C)
t_{dur}	Time duration of spray cloud (s)
Δt	Droplet cooling time (s)
u^*	Friction velocity (m s^{-1})
V	Absolute wind speed (m s^{-1})
V_r	Relative speed between a ship and an oncoming wave (m s^{-1})
V_{gr}	Relative speed between a ship and wave groups (m s^{-1})
V_s	Ship speed (m s^{-1})
$\mathbf{W_r}$	Relative velocity between a ship and the wind (m s^{-1} and ° for direction)
W_{rn}	Normal component of $\mathbf{W_r}$ relative to a vertical plate (m s^{-1})
z	Height above sea level (m)
z_0	Roughness length (m)
$z_{0\mathrm{T}}$	Roughness length of temperature (m)
α	Angle between a ship and waves (°)
β	Angle between a ship and wind (°)
ϵ	Ratio of molecular weights of water and air (0.622)
κ	von-Karman constant (0.4)
$\mathcal{L}$	Likelihood function
ρ_a	Density of air (1.3 kg m^{-3})
ρ_i	Density of ice (890 kg m^{-3})
ρ_w	Density of sea water (1028 kg m^{-3})
σ	Stefan-Boltzmann constant (5.67×10^{-8} W m^{-2} K^{-4})
ψ_m	Stability correction term for momentum
ψ_h	Stability correction term for heat
ϕ_m	Non-dimensional wind shear
Φ	Parameter from Overland *et al.* (1986) equal to $c_w n^{-1} L_{\mathrm{fs}}^{-1}$ (°C^{-1})

Supporting information

Table S.I. As Table IV, but for the models Over mod, Over ff, Stall, and ModStall org not presented in Section 5.1 of the main article.

Table S.II. As table V, but for the models Over mod, Over ff, Stall, and ModStall org not presented in Section 5.3 of the main article.

Table S.III. As table V in Section 5.3, but the verifications scores for all methods are calculated with the original Overland category definition from Table III in the upper panel. In the lower panel the original category definitions are applied for the methods of Over1990, Mertins, LU, and LUsst.

References

ABS - American Bureau of Shipping. 2010. Guide for vessels operating in low temperature environments. Technical report, American Bureau of Shipping, Houston, TX 77060 USA. Updated February 2014. Accessed 06 March 2017. [Available online at: https://www.eagle.org/eagleExternalPortalWEB/ShowProperty/BEA%20Repository/Rules&Guides/Current/151_VesselsOperatinginLowTemperatureEnvironments/LTE_Guide].

Arya PS. 2001. *Introduction to Micrometeorology*, vol. 79. Academic Press, ISBN 9780120593545.

Bialek EL. 1966. Handbook of oceanographic tables. Technical report, U.S. Naval Oceanographic Office, Washington, D.C.

Borisenkov YP, Pchelko IG. 1975. Indicators for forecasting ship icing. Technical report, Cold Regions Research and Engineering Laboratory.

Borisenkov YP, Zablockiy G, Makshtas A, Migulin A, Panov V. 1975. On the approximation of the spray-cloud dimensions (In Russian). In: *Arkticheskii i Antarkticheskii Nauchno-Issledovatelskii Institut*, Gidrometeoizdat Leningrad, pp. 121–126.

Bourassa MA, Gille ST, Bitz C, Carlson D, Cerovecki I, Clayson CA, Cronin MF, Drennan WM, Fairall CW, Hoffman RN, *et al.* 2013. High-latitude ocean and sea ice surface fluxes: Challenges for climate research. *Bulletin of the American Meteorological Society* **94**(3): 403–423.

Brown RD, Agnew T. 1985. Characteristics of marine icing in Canadian waters. In: *Proceedings of the International Workshop on Offshore Winds and Icing*. pp. 78–94.

Brown RD, Roebber P. 1985. The ice accretion problem in Canadian waters related to offshore energy and transportation. Technical Report 85–13, Candian Climate Centre, Atmospheric Environment Service.

Buchinskiy VY. 1960. *Gololed i bor'ba s nim (Glaze and Its Control)*. Leningrad, Gidrometeoizdat.

Byrkjedal Ø, Harstveit K, Kravik R, Løvholm AL, Berge E. 2011. Analysis of the climate conditions for offshore wind power in Norwegian waters. In: *Proceedings of the European Wind Energy Association (EWEA)*. EWEA - The European Wind Energy Association: Brussels.

Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Hólm EV, Isaksen L, Kållberg P, Köhler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette JJ, Park BK, Peubey C, de Rosnay P, Tavolato C, Thépaut JN, Vitart F. 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society* **137**(656): 553–597, doi:10.1002/qj.828, URL http://dx.doi.org/10.1002/qj.828.

Desjardins S. 2013. The Fetch Factor: incorporating a fetch dependency into the Overland algorithm. [Accessed 06 March 2017. Available online at http://www.opc.ncep.noaa.gov/icing_rates/FetchFactor_Overland.pdf].

Dormand J, Prince P. 1980. A family of embedded runge-kutta formulae. *Journal of Computational and Applied Mathematics* **6**(1): 19–26, doi:http://dx.doi.org/10.1016/0771-050X(80)90013-3, URL http://www.sciencedirect.com/science/article/pii/0771050X80900133.

Eide LI. 1983. Environmental conditions in the Barents Sea and near Jan Mayen. Technical report, Norwegian Meteorological Institute.

Ekeberg OC. 2010. State-of-the-art on the marine icing models and observations. Techincal report (confidential) 2010-0745, DNV - Det Norske Veritas.

Fairall C, Bradley E, Hare J, Grachev A, Edson J. 2003. Bulk parameterization of air-sea fluxes: Updates and verification for the COARE algorithm. *Journal of Climate* **16**(4): 571–591, URL http://dx.doi.org/10.1175/1520-0442(2003)016<0571:BPOASF>2.0.CO;2.

Feit DM. 1987. Forecasting of superstructure icing for Alaskan waters. *Marine* **12**(2): 5–10. [Accessed 06 March 2017. Available online at: http://polar.ncep.noaa.gov/mmab/papers/tn12/OPC12.pdf].

Foken T, Nappo CJ. 2008. *Micrometeorology*. Springer Science & Business Media, ISBN 9783540746652.

Funk C. 2012. Freezing Spray and Ice Accretion on Vessels: A comprehensive Summary. In: *Proceedings of the National Conference on Undergraduate Research*. United States Coast Guard Academy, pp. 1506–1513.

Gemynthe PS. 2015. Estimate of risk and hazard on the maritime traffic in the waters of Greenland. Master's thesis, Technical University of Denmark. [Accessed 6 March 2017. Available online at http://www.kamikposten.dk/lokal/last/container/da/hvadermeningen/pdf/estimate_of_risk_and_hazard_on_the_maritime_traffic_in_the_waters_of_greenland.pdf.].

Gulev SK, Hasse L. 1998. North Atlantic wind waves and wind stress fields from voluntary observing ship data. *Journal of physical oceanography* **28**(6): 1107–1130, doi:10.1175/1520-0485(1998)028⟨1107:NAWWAW⟩2.0.CO;2, URL http://dx.doi.org/10.1175/1520-0485(1998)028<1107:NAWWAW>2.0.CO;2.

Hansen ES. 2012. Numerical modelling of marine icing on offshore structures and vessels. Master's thesis, NTNU - Norwegian University of Science and Technology. [Available online at http://hdl.handle.net/11250/246746.].

Hansen ES, Teigen SH. 2015. An efficient numerical model for marine icing. In: *Proceedings of the 23rd International Conference on Port and Ocean Engineering under Arctic Conditions*, 150. Norwegian University of Science and Technology, Trondheim, pp. 1–15. [Available online at http://www.poac.com/Papers/2015/pdf/poac15Final00150.pdf.].

Henry NL. 1995. Forecasting vessel icing due to freezing spray in Canadian east coast waters. Part I: Model physics. Technical report, Environment Canada, Newfoundland Weather Centre.

Hyndman RJ, Koehler AB. 2006. Another look at measures of forecast accuracy. *International journal of forecasting* **22**(4): 679–688, doi:10.1016/j.ijforecast.2006.03.001, URL http://www.sciencedirect.com/science/article/pii/S0169207006000239.

Iden KA, Reistad M, Aarnes OJ, Gangstø R, Noer G, Hughes NE. 2012. Knowledge about wind, waves, temperature, ice extension, visbility etc. - "Barents Sea SE" (in Norwegian). MET Report 11, Norwegian Meteorological Institute. [Accessed 06 March 2017. Available at https://met.no/Forskning/Publikasjoner/MET_report/2012/?module=Files;action=File.getFile;ID=4908].

Japan Meteorological Agency. 2017. Marine Warnings. Accessed 09 January 2017. URL http://www.jma.go.jp/en/seawarn/index.html.

Johansen K. 2013. Optimization of maritime operations in Arctic. Rating of improvement of weather routing/decision support systems for maritime operations in the Arctic region. Master's thesis, UiT - The Arctic University of Norway. [Available online at http://hdl.handle.net/10037/8180.].

Kachurin LG, Gashin LI, Smirnov IA. 1974. Icing rate of small displacement fishing boats under various hydrometeorological conditions. *Meteorologiya i Gidrologiya* **3**: 50–60. Moscow.

Kulyakhtin A, Kulyakhtin S, Loset S. 2016. The role of the ice heat conduction in the ice growth caused by periodic sea spray. *Cold Regions Science and Technology* **127**: 93 – 108, doi:http://dx.doi.org/10.1016/j.coldregions.2016.04.001, URL http://www.sciencedirect.com/science/article/pii/S0165232X16300532.

Kulyakhtin A, Tsarau A. 2014. A time-dependent model of marine icing with application of computational fluid dynamics. *Cold Regions Science and Technology* **104–105**(0): 33–44, doi:10.1016/j.coldregions.2014.05.001, URL http://www.sciencedirect.com/science/article/pii/S0165232X14000962.

Løset S, Shkinek KN, Gudmestad OT, Høyland KV. 2006. *Actions from ice on Arctic offshore and coastal structures*, ch. 6 Icing in the Ocean. LAN: St. Petersburg, ISBN 5811407033, pp. 191–206. Student's Books for Institutes of Higher Education. Special Literature.

Lozowski EP, Szilder K, Makkonen L. 2000. Computer simulation of marine ice accretion. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* **358**(1776): 2811–2845, doi:10.1098/rsta.2000.0687.

Lundqvist J, Udin I. 1977. Ice accretion on ships with special emphasis on Baltic conditions. Research Report 23, Winter Navigation Research Board, Swedish Administration of Shipping and Navigation, Finnish Board of Navigation., Norrköping, Sweden.

Makkonen L. 1987. Salinity and growth rate of ice formed by sea spray. *Cold Regions Science and Technology* **14**(2): 163–171, doi:10.1016/0165-232X(87)90032-2, URL http://www.sciencedirect.com/science/article/pii/0165232X87900322.

Makkonen L, Brown RD, Mitten PT. 1991. Comments on "Prediction of vessel icing for near-freezing sea temperatures". *Weather and Forecasting* **6**: 565–567.

Mertins HO. 1968. Icing on fishing vessels due to spray. *Marine Observer* **38**(221): 128–130.

MET Norway. 2015. Text forecast for high seas. Accessed 4 December 2015. Electronic, URL http://www.yr.no/hav_og_kyst/tekstvarsel/hav/.

Meteorological Office College. 1997. *Source book to the Forcasters' reference book*. The Met. Office, ISBN 0861803213.

Monin AS, Obukhov AM. 1954. Basic laws of turbulent mixing in the surface layer of the atmosphere. *Contrib. Geophys. Inst. Acad. Sci. USSR* **151**: 163–187.

Monterey G, Levitus S. 1997. Climatological cycle of mixed layer depth in the world ocean. *US Government Printing Office, NOAA NESDIS* **5**.

Moore GWK. 2013. A climatology of vessel icing for the subpolar North Atlantic Ocean. *International Journal of Climatology* **33**(11): 2495–2507, doi:10.1002/joc.3604, URL http://dx.doi.org/10.1002/joc.3604.

Nacional Oceanic and Atmospheric Administration. 2015. NCEP MMAB Global Superstructure Ice Accretion Guidance. Accessed 26 November 2015. [Available online at http://polar.ncep.noaa.gov/marine.meteorology/vessel.icing/].

Norwegian Maritime Authority. 2014. Database of ship incidents. Accessed 29 August 2016. Digital media. [Available online at https://www.sjofartsdir.no/ulykker-sikkerhet/ulykkesstatistikk/datauttrekk/.].

OSISAF. 2015. EUMETSAT Ocean and Sea Ice Satellite Application Facility. Global sea ice concentration reprocessing dataset 1978-2015 (v1.2, 2015). Accessed 16 October 2015. Electronic, URL http://osisaf.met.no.

Overland JE. 1990. Prediction of vessel icing for near-freezing sea temperatures. *Weather and Forecasting* **5**: 62–77, doi:10.1175/1520-0434(1990)005⟨0062:POVIFN⟩2.0.CO;2, URL http://dx.doi.org/10.1175/1520-0434(1990)005<0062:POVIFN>2.0.CO;2.

Overland JE, Pease CH, Preisendorfer RW, Comiskey AL. 1986. Prediction of vessel icing. *Journal of Climate and Applied Meteorology* **25**(12): 1793–1806, doi:10.1175/1520-0450(1986)025⟨1793:POVI⟩2.0.CO;2, URL http://dx.doi.org/10.1175/1520-0450(1986)025<1793:POVI>2.0.CO;2.

Panov VV. 1978. Icing of ships. In: *Polar Geography*, vol. 2, Taylor & Francis, pp. 166–186, doi:10.1080/10889377809388652, URL http://dx.doi.org/10.1080/10889377809388652.

Pease CH, Comiskey AL. 1985. Vessel icing data in Alaskan waters-1979-1984 dataset. NOAA Data Report ERL PMEL-14, Pacific Marine Environmental Laboratory.

Pedersen RN. 2015. QRA techniques on dynamic positioning systems during drilling operations in the Arctic: With emphasis on the dynamic positioning operator. Master's thesis, UiT - The Arctic University of Norway. [Available online at http://hdl.handle.net/10037/7916.].

Pierson WJ, Moskowitz L. 1964. A proposed spectral form for fully developed wind seas based on the similarity theory of sa kitaigorodskii. *Journal of geophysical research* **69**(24): 5181–5190.

Reijmer C, Meijgaard E, van den Broeke M. 2003. Roughness length for momentum and heat over Antarctica in a regional atmospheric climate model. In: *Proceedings of the Seventh Conference on Polar Meteorology and Oceanography and Joint Symposium on High-Latitude Climate Variations, 12–16 May 2003, Hyannis, Massachusetts*.

Reynolds RW, Smith TM, Liu C, Chelton DB, Casey KS, Schlax MG. 2007. Daily High-Resolution-Blended Analyses for Sea Surface Temperature. *Journal of Climate* **20**(22): 5473–5496, doi:10.1175/2007JCLI1824.1, URL http://dx.doi.org/10.1175/2007JCLI1824.1.

Roebber P, Mitten P. 1987. Modelling and measurement of icing in Canadian waters. Technical report, Canadian Climate Centre, Atmospheric Environment Service.

Ryerson CC. 2013. Icing management for coast guard assets. Technical report, DTIC Document. [Accessed 6 March 2017. Available online at http://hdl.handle.net/11681/5477.].

Samuelsen EM, Edvardsen K, Graversen RG. 2017. Modelled and observed sea-spray icing in Arctic-Norwegian waters. *Cold Regions Science and Technology* **134**: 54 – 81, doi:http://dx.doi.org/10.1016/j.coldregions.2016.11.002, URL http://www.sciencedirect.com/science/article/pii/S0165232X16303275.

Samuelsen EM, Løset S, Edvardsen K. 2015. Marine icing observed on KV Nordkapp during a cold air outbreak with a developing polar low in the Barents Sea. In: *Proceedings of the 23rd International Conference on Port and Ocean Engineering under Arctic Conditions*, 87. Norwegian University of Science and Technology, Trondheim, pp. 1–14. [Available online at http://www.poac.com/Papers/2015/pdf/poac15Final00087.pdf.].

Sawada T. 1962. Icing on ships and the forecasting method. (In Japanese). *Snow and Ice* (24): 12–14.

Sawada T. 1968. Ice accretion on ships in northern seas of Japan. *J. Meteorol, Soc. Japan* **46**: 350–354.

Shekhtman AN. 1971. The probability and intensity of the icing-up of ocean-going vessels. Technical report, DTIC Document.

Shellard HC. 1974. The meteorological aspects of ice accretion on ships. Technical Report 10, World Meteorological Organization. Marine Science Affairs Report.

Sollid MP. 2013. On a Decision-support System for Early Warning of the Risk of, and Growth-rate of, Icing on a Ship's Superstructure. Master's thesis, UiT - The Arctic University of Norway.

Stallabrass JR. 1979. Icing of fishing vessels: An analysis of reports from Canadian east coast waters. National Research Council Canada. Technical report, Laboratory Technical Report LTR-LT-98.

Stallabrass JR. 1980. Trawler icing - A compilation of work done at N.R.C. Mechanical Engineering Report MD-56, National Research Council Canada.

Stull RB. 1988. *An introduction to Boundary Layer Meteorology*, vol. 13. Springer Science & Business Media, ISBN 9027727694.

Tabata T, Iwata S, Ono N, Hope ER. 1967. Studies of ice accumulation on ships. Technical report, Directorate of Scientific Information Services, DRB Canada.

Tolman H. 2014. The WAVEWATCH III Development Group. User Manual and System Documentation of WAVEWATCH III version 4.18. Technical report, Nacional Oceanic and Atmospheric Administration (NOAA).

United States Coast Guard. 2008. Report of investigation into the sinking and loss of four crewmembers aboard the commercial fishing vessel Lady of Grace. Technical report, United States Coast Guard, Washington D.C. [Accessed 6 March 2017. Available online at http://www.uscg.mil/hq/cg5/cg545/docs/documents/LadyOfGrace.pdf.].

Wallace J, Hobbs P. 2006. *Atmospheric Science: An Introductory Survey*. International Geophysics, Elsevier Science, ISBN 9780080499536.

Wilks DS. 2011. *Statistical Methods in the Atmospheric Sciences*, *International Geophysics Series*, vol. 100. Academic Press, Third edn, ISBN 9780123850225.

Wise JL, Comiskey AL. 1980. Superstructure icing in Alaskan Waters. NOAA Special Report, Pacific Marine Evironmental Laboratory.

WMO. 1962. Commission for Synoptic Meteorology abridged final report of the third session. General summary. Rec. 26-28 CSM-III (Mar. 1962). Accessed 29 August 2016. WMO Report No.122. Rp.50, World Meteorological Organization (WMO). [Available online at https://www.wmo.int/pages/prog/amp/mmop/documents/publications-history/cbs/CSM03Rec26-28.pdf.].

WMO. 2015. Manual on Codes. International on codes. Manual 306. Volume I.1. Part A - Alphanumeric codes., World Meteorological Organization (WMO). [Available online at http://www.wmo.int/pages/prog/www/WMOCodes.html.].

Zakrzewski WP. 1986. Icing of ships, part 1: Splashing a ship with spray. Technical report, NOAA - National Oceanic and Atmospheric Administration.

Zakrzewski WP. 1987. Splashing a ship with collision-generated spray. *Cold Regions Science and Technology* **14**(1): 65–83, doi:10.1016/0165-232X(87)90045-0, URL http://www.sciencedirect.com/science/article/pii/0165232X87900450.

Zakrzewski WP, Lozowski EP. 1989. Soviet marine icing data. Technical report, Canadian Climate Centre, Atmospheric Environment Service.

Zakrzewski WP, Lozowski EP, Horjen I. 1989. The use of ship icing models for forecasting icing rates on sea-going ships. In: *Proceedings of the 10th International Conference on Port and Ocean Engineering under Arctic Conditions*. Luleå University of Technology, pp. 1454–1467. [Available online at http://www.poac.com/PapersOnline.html.].

Supporting information

associated with

Ship-icing prediction methods applied in operational weather forecasting

Eirik Mikal Samuelsen

Table S.I. As Table IV, but for the models Over mod, Over ff, Stall, and ModStall org not presented in Section 5.1 of the main article.

Name	KVN data (N = 37)				P&C and R&M data (N = 47)				All three data sets (N = 84)			
	BIAS	MAE	RMSE	r	BIAS	MAE	RMSE	r	BIAS	MAE	RMSE	r
	†	††	††	†††	†	††	††	†††	†	††	††	†††
Ref.		**0.44**	**0.62**	**0.27**		**0.48**	**0.64**	**0.24**		**0.48**	**0.66**	**0.18**
Over mod	3.74	3.78	6.73	−0.02	1.38	1.54	3.08	**0.36**	2.42	2.53	5.02	0.13
Over ff	3.29	3.41	6.94	−0.04	**1.49**	1.88	7.33	0.21	2.29	2.55	7.16	0.11
Stall	**0.05**	0.46	**0.59**	0.24	−0.18	**0.32**	**0.47**	**0.55**	**−0.07**	**0.38**	**0.53**	**0.39**
ModStall org	1.18	1.28	1.82	**0.55**	0.64	0.81	1.22	**0.40**	0.88	1.02	1.52	**0.47**

† Non-boldface is indicating that the mean error in the model is significantly (5% significance level) greater or lower than zero error (BIAS = 0). Boldface is indicating that there is not enough support from the data set to reject the null hypothesis that the BIAS in the model is greater or lower than zero (5% significance level).
†† The 95 % lower limit of the ordered MAE and RMSE when applying a Monte-Carlo simulation ($N = 10^4$) of a naïv reference error (Hyndman and Koehler 2006). Models with an MAE and an RMSE value below this reference value, have an MASE and an RMSSE significantly lower than 1 and these values are marked with boldface (see Section 5.1 for details).
††† The 95 % upper limit of no positive correlation ($r = 0$) assuming t-distribution. Models with an r-value marked with boldface are greater than this highlighted reference value.

Table S.II. As table V, but for the models Over mod, Over ff, Stall, and ModStall org not presented in Section 5.3 of the main article.

Modelname	KVN with no-icing (N = 67)				P&C and R&M (N = 47)				All data (N = 114)			
	PC	HSS	PSS	GMSS	PC	HSS	PSS	GMSS	PC	HSS	PSS	GMSS
New category definition (Section 4.2)												
Ref. (unif.)†	**0.34**	**0.11**	**0.13**	**0.19**	**0.36**	**0.12**	**0.16**	**0.33**	**0.32**	**0.09**	**0.09**	**0.14**
Ref. (obs.)††	0.42	0.13	0.13	0.14	0.53	0.18	0.18	0.14	0.37	0.10	0.10	0.10
Over mod	0.18	0.04	0.05	**0.23**	0.23	−0.02	−0.02	**0.40**	0.20	0.02	0.03	**0.16**
Over ff	0.25	0.08	0.10	**0.35**	0.34	0.00	0.00	**0.34**	0.29	0.08	0.08	**0.19**
Stall	0.27	0.01	0.01	**0.20**	**0.57**	**0.30**	**0.31**	**0.50**	**0.39**	**0.13**	**0.12**	**0.21**
ModStall org	0.22	0.02	0.02	**0.36**	**0.38**	**0.14**	**0.16**	**0.55**	0.29	0.08	0.09	**0.33**
LUsst	0.28	0.07	0.08	**0.19**	0.28	−0.03	−0.04	−0.09	0.28	0.03	0.03	0.07

† The 95 % upper limit of the ordered distribution organised from low to high of the different verification scores when applying Monte-Carlo simulations ($N = 10^4$) assuming that the predictions are uniformly distributed inside each category.
†† The 95 % upper limit of the ordered distribution organised from low to high of the different verification scores when applying Monte-Carlo simulations ($N = 10^4$) assuming that the predictions are distributed in the same manner as the observations.

Table S.III. As table V in Section 5.3, but the verifications scores for all methods are calculated with the original Overland category definition from Table III in the upper panel. In the lower panel the original category definitions are applied for the methods of Over1990, Mertins, LU, and LUsst.

Modelname	KVN with no-icing (N = 67)				P&C and R&M (N = 47)				All data (N = 114)			
	PC	HSS	PSS	GMSS	PC	HSS	PSS	GMSS	PC	HSS	PSS	GMSS
Overland category definition												
Ref. (unif.)[†]	**0.34**	**0.11**	**0.13**	**0.34**	**0.36**	**0.11**	**0.18**	**0.22**	**0.32**	**0.08**	**0.10**	**0.31**
Ref. (obs.)[††]	0.49	0.15	0.15	0.11	0.66	0.23	0.23	0.09	0.46	0.11	0.11	0.08
Over	0.21	0.05	0.07	**0.35**	**0.40**	**0.12**	**0.19**	**0.33**	0.29	**0.10**	**0.13**	**0.37**
Over mod	0.16	0.05	0.07	0.28	0.36	0.08	0.13	**0.31**	0.25	0.07	0.09	0.31
Over ff	0.22	0.01	0.01	**0.41**	**0.49**	**0.12**	0.16	**0.35**	**0.33**	0.07	0.08	**0.41**
Over mod ff	0.19	0.03	0.04	**0.38**	**0.49**	0.10	0.12	**0.33**	0.32	0.07	0.08	**0.35**
Over bulk	**0.45**	0.07	0.07	0.07	**0.43**	−0.02	−0.02	**0.24**	**0.44**	−0.00	−0.00	0.04
Over bulk2	0.27	−0.08	−0.09	0.05	**0.70**	**0.32**	**0.31**	**0.42**	**0.45**	0.05	0.05	0.08
Stall	0.31	0.01	0.01	0.15	**0.68**	**0.35**	**0.39**	**0.42**	**0.46**	**0.13**	**0.13**	0.14
ModStall org	0.25	0.03	0.04	**0.45**	**0.55**	**0.25**	**0.33**	**0.40**	**0.38**	**0.13**	**0.15**	**0.45**
ModStall	0.15	−0.00	−0.00	0.32	0.34	**0.13**	**0.22**	**0.31**	0.23	0.06	0.08	**0.35**
MINCOG	**0.46**	**0.21**	**0.23**	0.29	**0.60**	**0.22**	**0.25**	**0.40**	**0.52**	**0.23**	**0.24**	0.23
Mertins	0.25	0.04	0.06	**0.37**	**0.43**	**0.19**	**0.30**	0.02	**0.32**	**0.13**	**0.17**	**0.42**
LU	0.12	−0.01	−0.02	0.26	0.13	−0.04	−0.07	−0.15	0.12	−0.03	−0.04	0.24
LUsst	0.22	0.01	0.02	−0.01	0.32	0.05	0.08	−0.07	0.26	0.02	0.03	−0.01
Sawada	0.12	−0.02	−0.03	0.28	0.32	0.07	0.12	−0.05	0.20	0.04	0.05	0.29
Other category definitions												
Over 1990[$]	0.19	0.00	0.01	**0.35**	**0.45**	**0.11**	0.15	**0.37**	0.30	0.07	0.09	**0.38**
Mertins org[$$]	**0.39**	**0.15**	**0.19**	0.21	0.30	−0.02	−0.02	0.02	**0.35**	**0.10**	**0.12**	0.17
LU org[$$$]	0.21	0.01	0.01	0.23	**0.40**	**0.12**	0.14	0.13	0.29	0.05	0.06	0.18
LUsst org[$$$]	0.28	0.02	0.02	0.11	0.28	−0.12	−0.13	−0.07	0.28	−0.04	−0.04	0.03

[†] The 95 % upper limit of the ordered distribution organised from low to high of the different verification scores when applying Monte-Carlo simulations ($N = 10^4$) assuming that the predictions are uniformly distributed inside each category.
[††] The 95 % upper limit of the ordered distribution organised from low to high of the different verification scores when applying Monte-Carlo simulations ($N = 10^4$) assuming that the predictions are distributed in the same manner as the observations.
[$] Applying the Overland (1990) definition of the boundaries for the predictions as described in Section 2.2.1, and illustrated by the vertical dashed lines in Figure 1. The observations are however organised in the same manner as the Overland *et al.* (1986) category definition.
[$$] Applying the original category definition of Mertins from Table III.
[$$$] Applying the original category definition of LU from Table III.