Blizzard Conditions in the Canadian Arctic: Observations and Automated Products for Forecasting

WILLIAM R. BURROWS^a AND CURTIS J. MOONEY^b

^a *Observation Based Research Section, Science and Technology Branch, Environment and Climate Change Canada, Edmonton, Alberta, Canada*

^b *National Lab-West, Meteorological Service of Canada, Edmonton, Alberta, Canada*

(Manuscript received 15 May 2020, in final form 16 February 2021)

ABSTRACT: Blizzard conditions occur regularly in the Canadian Arctic, with high impact on travel and life there. These extreme conditions are challenging to forecast for this vast domain because the observation network is sparse and remote sensing coverage is limited. To establish occurrence statistics we analyzed aviation routine weather reports (METARs) from Canadian Arctic stations between October and May 2014–18. Blizzard conditions occur most frequently in open tundra east and north of the boreal forest boundary, with the highest frequency found on the northwest side of Hudson Bay and over flat terrain in central Baffin Island. Except in sheltered locations, the reported cause of reduced visibility is blowing snow without precipitating snow in about one-half to two-thirds of METARs made by a human observer, even higher at some stations. We produce three products that forecast blizzard conditions from postprocessed NWP model output. The blizzard potential (BP), generated from expert's rules, is intended for warning well in advance of areas where blizzard conditions may develop. A second product (BH) stems from regression equations for the probability of visibility ≤ 1 km in blowing snow and/or concurrent snow derived by Baggaley and Hanesiak. A third product (RF), generated with the random forest ensemble classification algorithm, makes a consensus YES/NO forecast for blizzard conditions. We describe the products, provide verification, and show forecasts for a significant blizzard event. Receiver operator characteristic curves and critical success index scores show RF forecasts have greater accuracy than BP and BH forecasts at all lead times.

SIGNIFICANCE STATEMENT: Blizzard conditions, defined as visibility $\frac{1}{4}$ mile or less in blowing snow or concurrent snow and blowing snow with wind speed > 22 kt (11.3 m s^{-1}), are a regular occurrence in the Canadian Arctic. These extreme conditions can last for hours to days, significantly impacting travel and life. Forecasting is challenging due to sparse observations, lack of radar coverage and continuous satellite coverage, and polar night lasting up to six months. We derive recent occurrence statistics from METARs and describe three forecast products developed to address the need for timely automated forecast guidance. They have been well received by forecasters. One is driven by a machine-learning technique. As methods improve we expect the public to benefit in the future from increasingly better prediction.

KEYWORDS: Atmosphere; Arctic; Snow; Wind; Winter/cool season; Forecasting; Forecasting techniques; Operational forecasting; Automated systems; Classification; Decision trees; Expert systems; Machine learning; Visibility

1. Introduction

Blizzards are a regular occurrence in the Canadian Arctic from October to May. The combination of low temperature, strong wind, and low visibility in blowing snow makes this one of the most hazardous of winter weather events (Ricketts and Hudson 2001). Harsh conditions in blizzards can have a significant impact on local communities and on air and overland travel. Phillips (1990) describes a blizzard event at Iqaluit in

February 1979 that kept residents indoors for ten days. All forms of travel in the Arctic have been increasing rapidly in recent years as interest grows in the region (Arctic Council 2017). Blizzard conditions can significantly hamper, delay, or make such activity riskier, and may last anywhere from a few hours to several days, particularly north of the boreal forest boundary where open terrain makes blizzard conditions more likely to occur (Ricketts and Hudson 2001).

Several studies document environmental factors necessary for blowing snow which, when it becomes severe, results in blizzard conditions (e.g., Phillips 1990; Li and Pomeroy 1997; Pomeroy et al. 1997; Déry and Yau 1999; Ricketts and Hudson 2001; Baggaley and Hanesiak 2005; Savelyev et al. 2006; Hanesiak et al. 2010). The main environmental considerations are as follows: 1) below-freezing temperature, 2) source of snow (precipitation or snow on the ground), 3) wind speed near the surface strong enough to lift snow off the ground (Li and Pomeroy 1997) or suspend falling snow in the air longer, and 4) snowpack age (Baggaley and Hanesiak 2005). The Meteorological Service of Canada (MSC) issues a blizzard

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

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/WAF-D-20-0077.s1>.

Corresponding author: William R. Burrows, william.burrows@canada.ca

DOI: 10.1175/WAF-D-20-0077.1

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

warning 18 h in advance if the forecast guidance suggests a sustained period of temperature below 0°C, winds 40 km h⁻¹ (11.1 m s⁻¹) or stronger, and widespread visibilities 1/4 statute mile (SM) (approximately 400 m) or less for at least 6 h north of 60°N and 4 h south of 60°N. The visibility used to issue warnings is the aeronautical Meteorological Optical Range reported in aviation routine weather reports (METARs), measured 2.5 m above the ground according to the World Meteorological Organization standard. Particle size distribution and number density in blowing snow decrease strongly with height above the surface (Xiao and Taylor 2002). Though visibility may be 1/4 SM at 2.5 m above the ground it is likely to be significantly less at the level of surface vehicle traffic, say 1.5 m.

Predicting the onset and duration of Arctic blizzards is challenging. The area included in Canadian Arctic forecasts is immense. Surface observations are sparse and located mainly along coasts. Polar orbiting satellites do not provide continuous coverage in space and time. Visible satellite observations are not available during polar night. Where satellite observations are available it can be difficult to determine where blizzards are occurring because it is hard to distinguish blowing snow from snow-covered ground. There is no radar coverage aside from a research radar in Iqaluit (Joe et al. 2020). Verifiable forecasts over most of the Canadian Arctic are limited to a few small local areas with surface observations. Blowing snow processes are not simulated in Canadian numerical weather prediction (NWP) models run for use in operations, hence there is a need for automated guidance products postprocessed from NWP model output to be available for regular use by forecasters.

Tools for predicting blizzard conditions range from a few simple rules of thumb (Ricketts and Hudson 2001), to a comprehensive set of empirical diagnostic regression equations (Baggaley and Hanesiak 2005), to a sophisticated hydrological numerical blowing snow model (Déry and Yau 2001; Yang et al. 2010). The Canadian Center for Meteorological and Environmental Prediction (CCMEP) does not run these models to produce forecasts for use by operational forecasters. To address the need for products, the authors have engaged in developing research products in consultation with forecasters. As a result, we developed a suite of three different automated products that forecast blizzard conditions postprocessed from output generated by NWP models run by CCMEP. One product, the blizzard potential (henceforth BP), is a set of expert's rules that identify areas where blizzard conditions are likely occurring or may develop. Its intent is for "heads-up" guidance to suggest where forecasters should focus attention, and not for objective point forecasts. A second product (henceforth BH), is generated from regression equations derived by Baggaley and Hanesiak (2005) for the probability of visibility less than 1 km in blowing snow or concurrent blowing snow and precipitating snow (1 km is approximately 0.62 SM). A third product (henceforth RF) is derived by matching observations with a set of predictors calculated from NWP output, and applying the ensemble classification algorithm random forests (Breiman 2001) to predict the chances of blizzard conditions. The latter product assesses the lower-tropospheric environment predicted in NWP output to

decide if blizzard conditions are likely to occur, rather than predicting exact visibility and wind speed.

The aim of this paper is to present an analysis of the frequency of observed blizzard conditions in the Canadian Arctic in recent years, to describe the forecast products we have developed, show verification statistics, and describe product performance for a recent blizzard event. The location of specific geographical names referred to throughout the paper is shown in Fig. 1. A nonrefereed condensed version of this article intended for a general audience is Burrows and Mooney (2018). Examples of the BH and RF products discussed here are in Joe et al. (2020) (their Figs. 8 and 9). Details and methodology were not discussed in either of those articles.

2. Frequency of blizzard conditions in recent years

METARs that include visibility at Canadian Arctic observing stations for October–May 2015–17 were analyzed for occurrences of "near-blizzard" and "blizzard" conditions. We define *near-blizzard conditions* as visibility $\leq 1/2$ SM in snow and/or blowing snow with wind speed > 18 kt, and *blizzard conditions* as visibility $\leq 1/4$ SM in snow and/or blowing snow with wind speed > 22 kt (1 kt ≈ 0.51 m s⁻¹). There is no official definition for the term "near-blizzard," we use it here to describe conditions which are poor but do not quite meet the MSC blizzard visibility criterion. This was changed a number of years ago from $1/2$ SM to $1/4$ SM, so the descriptor "near-blizzard conditions" encompasses the old criterion. If blizzard or near-blizzard conditions were reported in a METAR within $\pm 1/2$ h of a synoptic hour they were deemed to occur at that hour. Figure 2 shows the normalized percent of hours that blizzard conditions were reported at each station, obtained by dividing the number of hours that blizzard conditions were reported by the number of hours that reports were available. Normalization was done to facilitate comparison between stations since many hours of observations are routinely missing at many observing stations in the Canadian Arctic, mainly for two reasons. First, METARs at most manned Arctic stations revert overnight to "AUTO" reports derived from unmanned instruments, and some airport stations revert to AUTO reports during daytime periods when no flights are expected. Second, an observer may stop observing and revert to AUTO reporting during severe conditions. The content of AUTO reports varies, and some do not report visibility.

Figure 2 reveals that blizzard conditions occur most frequently at sites east and north of the treeline. Blizzard conditions account for about 1%–3.5% of hourly observations along the northwest Hudson Bay coast and inland. Terrain there is relatively flat open tundra with many lakes. This region is well known to Canadian meteorologists as "blizzard alley." The frequency of blizzard conditions at most other stations north of the treeline is between about 0.5%–1.5% of hourly observations and is even lower in sheltered locations, particularly those in the rugged eastern and northern Arctic terrain. There, blizzard conditions may not occur at an observing site while they are simultaneously occurring in nearby open country. For example Dewar Lakes, located on high open tundra in the middle of Baffin Island, recorded a much higher frequency of

blizzard conditions (5.14%) than did other stations on the island. South of the treeline, where wind speeds near the ground are much lower due to sheltering by the boreal forest, the frequency of blizzard conditions is nearly zero.

Figure 3 shows the percent of hours blizzard conditions were classified in METARs as “clear-sky,” that is, blowing snow (no precipitating snow) was the only reason reported for reduced visibility. AUTO reports were not included

FIG. 3. Percent of all hours in METARs reported by a human observer (i.e., AUTO reports excluded) during October–May of 2014–18 with blizzard conditions that could be classified as “clear-sky,” that is, blowing snow only (no precipitating snow) was given as the reason for reduced visibility. Percentage ranges are color coded: $\geq 67\%$ in black, from $<67\%$ to $>33\%$ in red, and $\leq 33\%$ in blue.

(Fig. 4 shows the percent of all METARs reporting visibility at each station that were made by a human observer, that is, were not AUTO reports). Many AUTO machines reporting visibility in the Canadian Arctic have limited capability to determine when only blowing snow (BLSN) is occurring, and often report snow and blowing snow (SN BLSN) if blowing snow reaches the precipitation sensor even if no snow is falling. The percentages should be considered approximate since accuracy depends on the human observer’s ability to determine if only blowing snow is occurring. Overall, clear-sky blizzard conditions occur in about one-half to two-thirds or more of available manned METARs reporting blizzard conditions, except in sheltered areas. Higher percentages occur at sites in open tundra. The percent of clear-sky blizzard conditions reaches around 80% or more in the Kivalliq region on the northwest coast of Hudson Bay and inland to the northwest, at Alert on northern Ellesmere Island, at Tasiujak on the shore of Ungava Bay, and Kimmirut on southern Baffin Island. The importance of predicting wind accurately in the Arctic winter boundary layer is evident.

3. Blizzard and blowing snow forecast products

At MSC’s National Laboratory-West we run three products twice a day that produce hourly forecasts of blizzard conditions. Each is derived differently, each is driven with postprocessed output from CCMEP operational deterministic prediction system (DPS) models at three resolutions: 1) global (GDPS) at 15 km; 2) regional (RDPS) at 10 km; and the Canadian Arctic prediction system (CAPS) at 3 km. Model descriptions are in

Girard et al. (2014) and Milbrandt et al. (2016). The blizzard forecast products predict the environment where blizzard conditions are likely to occur over land and over ocean and large lake surfaces with more than 80% ice cover. Precipitation in blizzard conditions is snow, diagnosed from CCMEP model output using a hybrid diagnostic scheme based on the tephigram area method of Bourgouin (2000) and the top-down method (see COMET MetEd Program 2005). Liquid water content in CCMEP models is converted to snow with the Dubé (2003) algorithm. A brief description of the precipitation type diagnosis method is provided with computer code as supplemental material. Description of the three forecast products follows. In this discussion the term “nearest grid point” means nearest grid point having more than 30% land cover.

a. Blizzard potential (BP)

This product derives from rules developed by forecasters in MSC’s Prairie and Arctic Storm Prediction Center to provide a “head’s up” warning well in advance of where blizzard conditions may occur. Figure 5, showing the portion below 500 hPa of the sounding taken at 1200 UTC 29 March 2018 at Baker Lake, contains environmental aspects of blizzard conditions around which the rules are formulated. An AUTO METAR observation taken at the same time as the sounding did not include visibility but reported a 36-kt wind speed. The visibility is not known for certain because the observer ceased taking observations many hours earlier when conditions became extreme, but was likely near zero SM. Some features typical of blizzard conditions are seen in Fig. 5: 1) strong

FIG. 4. The percent of METARs reporting visibility that were manned, that is, made by a human observer and not identified as AUTO, during October–May 2014–18. Percentages are color coded: $\geq 50\%$ in black and $< 50\%$ in red.

temperature inversion and low temperature near the ground; 2) strong low-level wind, little direction change in a relatively deep layer; 3) a less stable, more moist mixed layer extending a small distance from the ground. This layer is caused by cooling and moistening of air in the near-surface layer by sublimation after snow is lifted off the ground by the strong wind (Déry and Yau 2001).

Rules for BP categories and expected weather with each category based on forecaster experience are in columns 1–3 of Table 1. Wind speeds in all categories are well above threshold values given in Li and Pomeroy (1997) for snow transport for all types of snowpack, although snowpack condition is not included in BP rules. Blizzard conditions are not expected with categories 1 and 2, and snow may or may not be precipitating, but drifting snow and blowing snow are possible which could cause problems with overland travel. Category 3 applies to the potential for blowing snow in strong winds when snow is not falling. Categories 4 and 5 apply to precipitating snow situations, with higher wind speed in category 5 than category 4. Categories 3 and 5 have the greatest likelihood of blizzard conditions due to the greater wind speed (≥ 30 kt). Close attention by forecasters is warranted when category 4 is predicted because blizzard conditions can easily develop due to changing visibility in precipitating snow or strengthening wind speed. Columns 4 and 5 in Table 1 show the percent of manned observations reporting drifting snow (DRSN) or blowing snow (BLSN) regardless of visibility in each BP category for 1 October 2015–31 May 2017, excluding June–September. Columns 6 and 7 show the same for 1 October 2018–31 May 2019, the verification period discussed in section 4. Data are for observations (AUTO reports excluded) pooled from seven stations located in open tundra (Resolute Bay, Baker Lake,

Cambridge Bay, Gjoa Haven, Arviat, Rankin Inlet, Coral Harbour). We selected these stations to minimize local terrain effects that would not be representative of weather prevalent in nearby open country. Observations (regular and special) were scanned in a one half-hour window on either side of each synoptic hour, and were stratified by the BP category at the nearest grid point predicted by the most recent RDPS forecast issued 6–17 h previously valid at the same time as each observation. We excluded forecasts issued less than 6 h previously to

FIG. 5. Vertical sounding taken at Baker Lake at 1200 UTC 29 Mar 2018. The horizontal axis unit is temperature ($^{\circ}\text{C}$), the vertical axis unit is pressure (hPa). The green line is dewpoint temperature, and the red line is air temperature. Wind speed and direction are plotted as barbs on the right, where a triangle represents 50 kt, a long barb represents 10 kt, and a short barb represents 5 kt.

TABLE 1. Definition of BP categories, expected weather associated with each category, and percent of manned-only METAR observations reporting drifting snow (% DRSN) or blowing snow (% BLSN), regardless of visibility, in the combined data from seven manned stations located in open tundra listed in section 3c, stratified by BP category. Columns 4–5 are for 1 Oct 2015–31 May 2017, excluding the months June–September; and columns 6–7 are for 1 Oct 2018–31 May 2019. Average wind means the average wind speed in the lowest six RDPS model levels shown in Table 2, which spans a layer of approximately 50 hPa (400 m) above the ground. The “N/A” in columns 2 and 3 for BP category 0 means does not apply.

BP category	Rule: temperature below 0°C, ≥ 1 cm snow on ground; AND	Expected weather	% DRSN 2015–17	% BLSN 2015–17	% DRSN 2018–19	% BLSN 2018–19
0	N/A	N/A	14.3	4.0	15.5	4.4
1	Avg wind speed 22–30 kt	Drifting snow; blowing snow possible if snowpack loose	29.5	49.9	34.3	44.4
2	Avg wind speed 22–30 kt; lapse rate of θ_w at one or more model levels $\leq 0.02^\circ\text{C hPa}^{-1}$	Drifting snow, or blowing snow if snowpack loose	23.2	58.8	29.4	46.2
3	Avg wind speed ≥ 30 kt; snow not precipitating	Blizzard conditions likely if snow is on the ground locally or is advected from upstream	13.4	79.5	6.7	85.3
4	Avg wind speed 18–30 kt; precipitating snow	Near-blizzard to possible blizzard conditions	33.6	36.9	35.4	40.7
5	Avg wind speed ≥ 30 kt; precipitating snow	Blizzard conditions	16.4	78.5	16.7	82.5

avoid potential problems due to NWP model spinup. The results shown in columns 4–7 of Table 1 are reasonable when compared with the expected weather for each BP category. Baggaley and Hanesiak (2005), Li and Pomeroy (1997), and others note that blowing snow is sensitive to wind speed, and the same can be seen in column 5 of Table 1. When the average wind speed is category 0 (less than 22 kt) the percent of observations reporting blowing snow is much less than for other BP categories. The effect of a less stable low-level lapse rate (BP category 2) increases the percent of blowing snow observations compared to drifting snow observations. The highest percent of blowing snow occurs for BP categories 3 and 5, in which the average wind in the lowest six RDPS model layers is ≥ 30 kt (see Table 2). The percent of blowing snow is less for BP category 4 than categories 3 and 5 since the average wind speed in category 4 is less than in categories 3 and 5. Based on blowing snow occurrence one could argue for changing the order of BP categories, but it is visibility that determines blizzard conditions.

Figure 6 is a box-and-whisker plot for the same observed visibilities used for Table 1 but including only those where DRSN, BLSN, or SN was reported, with observations stratified by their matching BP forecast category. Mean and median visibility are lowest in categories 3 and 5, again showing the importance of wind speed near the ground when forecasting the likelihood of blizzard conditions. BP forecasts in these categories represent the highest likelihood of blizzard conditions. The median visibility is considerably less than the mean visibility in all BP categories except 0 due to forecasts whose matching observation had little or no blowing snow (thus high visibility) falling into each category, hence the long whiskers on the high side.

b. Probability of reduced visibility in snow and blowing snow (BH)

The BH forecast product stems from an extension of work by Baggaley and Hanesiak (2005), in which they derived diagnostic

regression equations for the probability of visibility ≤ 1 km (0.62 SM) in blowing snow. Some interesting features of blowing snow processes came out of this work, namely the nonlinear relationship between the frequency of blowing snow, and temperature, wind speed [also discussed by Li and Pomeroy (1997)], and snowpack age. Baggaley and Hanesiak (2005) analyzed nearly 40 years of observations from 32 stations in the Canadian Arctic and Prairie provinces to obtain the probability of visibility ≤ 1 km in blowing snow, given 10-m wind speed, 2-m temperature when $< 3^\circ\text{C}$, and elapsed time since the last snowfall > 0.5 cm. The latter is a proxy for the snowpack ageing process. They divided the observations into groups with and without concurrent precipitating snow, subdividing these by six categories of increasing snowpack age. Wind speed thresholds for 5% and 95% probability of visibility ≤ 1 km in blowing snow alone were determined for each group. Regression curves for threshold wind speed in terms of temperature were fit to the data for each snowpack

TABLE 2. Eta and sigma values of CCMEP NWP model user grid vertical levels and approximate height above ground in an Arctic air mass.

Level No.	Eta	Sigma	Height above ground (m)
1	12 000	1.0000	0
2	11 950	0.9950	35
3	11 850	0.9850	115
4	11 733	0.9733	205
5	11 606	0.9606	300
6	11 467	0.9467	405
7	11 316	0.9316	515
8	11 151	0.9151	640
9	10 973	0.8973	780
10	10 780	0.8780	920
11	10 571	0.8571	1100

FIG. 6. Box-and-whisker plot for observed visibility when only drifting snow, blowing snow, or snow were reported in METARs at the seven open tundra sites listed in section 3a. Binned by the blizzard potential (BP) category of the most recent forecast issued by the RDPS model between 6 and 17 h previously at the nearest grid point that is valid at the same time as an observation. (top) The BP forecast analysis period discussed in section 3a and (bottom) the forecast verification period discussed in section 4. The upper and lower boundaries of each box are the 75th percentile and 25th percentile, respectively. Mean visibility is plotted as a red square, median visibility is plotted as a yellow horizontal line. The whiskers show approximately the complete range of the data: lower is 25th percentile $- 1.5 \times (75\text{th percentile} - 25\text{th percentile})$; upper is 75th percentile $+ 1.5 \times (75\text{th percentile} - 25\text{th percentile})$. The black circles are outliers.

age category. Threshold wind speed probabilities for a specific temperature and snowpack age are obtained by linear interpolation between the 5% and 95% curves. 100% probability is assigned to wind speeds higher than the 95% probability wind speed, 0% probability to wind speeds lower than the 5% probability wind speed.

Subsequent to this work D. Baggaley (2008, personal communication) used the Baggaley and Hanesiak (2005) data to derive the probability of visibility ≤ 1 km when blowing snow is occurring concurrently with falling snow. A relation for visibility in falling snow based on the observed precipitation rate was calculated. If the visibility due to falling snow is greater

than 1 km then a “visibility deficit” needed to reach 1-km visibility must be made up by blowing snow. Using this visibility deficit a probability for visibility ≤ 1 km in snow and blowing snow is obtained as follows. Regression equations for wind speeds giving 5% and 95% probability of 1-km visibility and 8-km visibility (the maximum allowed when blowing snow alone is reported) in snow and blowing snow were derived for the range of temperatures in the data. Given temperatures and wind speeds for the 8- and 1-km visibility, 5% and 95% probabilities are calculated. Interpolation is used to determine the 5% and 95% probability wind speeds for the visibility deficit. Finally, based on these windspeeds a linear interpolation of probability to the forecast wind speed determines the probability of 1-km visibility in concurrent snow and blowing snow.

We adapted the above algorithms to run with output from any NWP model. A forecast at each grid point is assigned as the maximum of two probabilities, one for blowing snow, one for concurrent snow and blowing snow. To show some characteristics of the BH forecasts in probability ranges, observations from section 3a were binned into 20% probability ranges of the most recent nearest gridpoint BH forecast issued by the RDPS model between 6 and 17 h previously valid at the observation time. Table 3 shows the percent of observations in each probability range that reported DRSN or BLSN regardless of visibility. The results show that as BH probability increases, the percent of observations reporting BLSN increases while the percent of observations reporting DRSN decreases. When the BH probability is greater than 40%, about 85% or more of the METARs reported BLSN. Figure 7 is a box and whisker plot of visibilities from manned observations with DRSN, BLSN or SN also reported, stratified by the same BH probability categories used in Table 3. It shows the mean and median visibility in the bins trend lower as forecast probability increases, with the lowest mean and median visibility in the 80%–100% probability bin. The distribution of visibility in the bins is skewed for BH probability greater than 20% since the mean visibilities are significantly higher than the median except for the 80%–100% probability bin, due to the presence of BH forecasts whose matching observation had little or no blowing snow (thus high visibility) falling into a bin.

c. Machine-learning product for forecasting the likelihood of blizzard conditions (RF)

The BP and BH forecast algorithms are derived with predictors sourced from observations, not from NWP output. When applied, their forecast accuracy is subject to errors inherent in the NWP model generating them. This type of forecast product is known as a “perfect prog” product. The RF product is derived by matching observations with predictors sourced from NWP model output, so a degree of model error will be accounted for in models built with this data. Its forecast product is known as a “model output statistics (MOS)” product. We use the RF models produced with 6–17-h RDPS forecast data to make forecasts at all lead times. This assumes that most of the NWP model bias error is accounted for in the 6–17-h forecast data, and the error does not change substantially with time. Burrows (1985) called this approach time-offset MOS (TOMOS).

TABLE 3. Percent of manned only METAR observations reporting drifting snow (% DRSN) or blowing snow (% BLSN), regardless of visibility, in the combined data from seven manned stations located in open tundra listed in section 3a, stratified by probability ranges of BH forecasts shown in column 1. Columns 2–3 are for 1 Oct 2015–31 May 2017, excluding the months June–September; and columns 4–5 are for 1 Oct 2018–31 May 2019.

BH probability	% DRSN 2015–17	% BLSN 2015–17	% DRSN 2018–19	% BLSN 2018–19
0%	16.2	6.2	17.3	6.7
>0%–20%	25.0	59.0	27.5	65.6
>20%–40%	16.5	74.7	15.6	80.4
>40%–60%	13.0	83.7	8.0	89.2
>60%–80%	6.6	93.0	1.8	95.6
>80%–100%	0.7	98.9	1.0	99.0

Rapidly increasing computer power makes it possible to apply sophisticated algorithms to construct prediction models from large datasets. For predictions covering the entire Arctic domain we constructed a learning database of observations for the October–May months of 2015–17 using METARs from seven manned stations located in open tundra listed in section 3a. Each METAR observation was matched with a set of predictors derived from RDPS model output, described below. Changes were made to the RDPS model during this period effective 1200 UTC 7 September 2016. Verification data from CCMEP showed there was negligible effect on forecasts near the surface in the Arctic (Environment and Climate Change Canada 2016), and we did not notice any effect on our forecasts. For the predictand we considered two categories of blizzard criteria in order to provide more leeway for users: 1) *near-blizzard conditions*: visibility $\leq \frac{1}{2}$ SM and wind speed > 18 kt; and 2) *blizzard conditions*: visibility $\leq \frac{1}{4}$ SM and wind speed > 22 kt. Both manned and AUTO METAR reports were included in the learning data since we were only concerned with the reported visibility, regardless if it was reduced due to BLSN or SN BLSN. Observations were classified as NO if these conditions were not reported in any observation in the ± 1 half-hour window on either side of a synoptic hour, and YES if they were reported. Each observation was matched with a set of 43 predictors calculated at the nearest RDPS grid point from the most recent RDPS forecast issued between 6 and 17 h previously valid at the same time as the observation. Forecasts issued 6 or fewer hours previous to observation time were excluded. Data for the seven stations were pooled. We also created datasets for each of 38 Arctic stations for building locally applicable models for point forecasts. For space consideration we limit discussion to models built with the combined data from the seven open tundra stations.

Predictors for our models are listed in Table 4. Some are direct RDPS model output fields, some are constructed from one or more output fields. These predictors are formulated to capture important aspects of the environment during blizzard conditions (some of which were discussed in section 3a). The regional blizzard affinity predictor (REGBLZ) is our ad hoc designation of the suitability of local terrain to produce blizzard conditions. It ranges from 1 for flat open tundra north of the treeline to 0 for mountainous rugged terrain within the boreal forest. The seven open tundra stations, on which the whole-domain RF models are based, are treeless (REGBLZ

set to 1). Also, sea ice temperature (I7) does not apply to inland stations, so these two predictors were not used. The remaining 41 candidate predictors were offered to the random forest algorithm to build prediction models. When building RF models

FIG. 7. Box-and-whisker plot for observed visibility when only drifting snow, blowing snow, or snow were reported in METARs at the seven open tundra sites listed in section 3a. Binned by the category of the most recent Baggeley–Hanesiak (BH) forecast issued by the RDPS model between 6 and 17 h previously at the nearest grid point that is valid at the same time as an observation. (top) The forecast analysis period discussed in section 3a and (bottom) the forecast verification period discussed in section 4. Category 0–5 ranges of BH probability are as follows: 0 (0%), 1 (>0%–20%), 2 (>20%–40%), 3 (>40%–60%), 4 (>60%–80%), and 5 (>80%–100%). The explanation of what is shown is as in Fig. 6.

TABLE 4. Acronym and explanation of predictors in the RF blizzard forecast models. Model vertical levels are shown in Table 2; θ_w is the wet-bulb potential temperature. BH represents the Baggeley–Hanesiak algorithm discussed in section 3b.

Predictor No.	Predictor acronym	Predictor description
1	REGBLZ	Local blizzard affinity assessed by boreal forest tree density (0–1)
2	HU2	Specific humidity, model level 2 (kg kg^{-1})
3	TT2	Temperature, model level 2 ($^{\circ}\text{C}$)
4	ES2	Dewpoint depression, model level 2 ($^{\circ}\text{C}$)
5	QC2	Mixing ratio, total condensate, model level 2 (kg kg^{-1})
6	UU2	Zonal wind component, model level 2 (kt)
7	VV2	Meridional wind component, model level 2 (kt)
8	UV2	Wind speed, model level 2
9	WW1	Vertical velocity, model level 1 (Pa s^{-1})
10	SI	Net infrared radiation flux at the ground (W m^{-2})
11	RT	Total precipitation rate (mm h^{-1})
12	I4	Water in the snowpack (kg m^{-2})
13	I7	Sea ice temperature (for nearshore stations only) ($^{\circ}\text{C}$)
14	FV	Upward surface latent heat flux (W m^{-2})
15	FS	Solar radiation flux absorbed at the surface (W m^{-2})
16	RB21	Bulk Richardson number using model levels 2 and 1 (nondimensional)
17	SD	Snow depth on the ground (cm)
18	CX	Sun angle above the horizon ($^{\circ}$)
19	OL	Monin–Obukhov length (m)
20	DPN	3-h pressure change (hPa)
21	NT	Total cloud cover (nondimensional)
22–26	SNW1–5	Snow accumulation (previous 1–2, 1–6, 1–12, 1–24, 1–48 h) (cm)
27	SNWMDL	BH snowpack age model (1–2, 3–6, 6–12, 12–24, 24–48, >48 h) (1–6)
28	BLSP	BH probability of visibility ≤ 1 km in a clear-sky blizzard (0–1)
29	S BSP	BH probability of visibility ≤ 1 km in concurrent snow and blowing snow (0–1)
30	PTYP	Precipitation type (0 = none, 1 = rain, 2 = snow, 3–9 = other types)
31	INVBOT	Model level of inversion bottom (level No.)
32	INVTOP	Model level of inversion top (level No.)
33	INVHT	Height of inversion bottom (m)
34	INVSTR	Inversion strength (temperature difference inversion top – bottom) ($^{\circ}\text{C}$)
35	INVDEP	Vertical distance from surface to inversion top (m)
36	MLTOP	Model level at the top of moist layer starting from the surface (level No.)
37	MLDEP	Vertical distance from surface to top of the moist layer (m)
38	HU3M2	Difference in specific humidity, model level 3 minus model level 2 (g kg^{-1})
39	RB32	Bulk Richardson number calculated using model levels 2 and 3 (nondimensional)
40	HL14	Dewpoint difference model level 1 minus model level 4 ($^{\circ}\text{C}$)
41	LLW18	Max wind speed in model levels 1–8 (kt)
42	DTWDP	No. of model levels below INVBOT that $d(\theta_w)/dp$ is $\leq 0.02^{\circ}\text{C hPa}^{-1}$ (nondimensional)
43	NESLVL	No. of model levels ES $< 2.0^{\circ}\text{C}$ in bottom 11 model levels

it is best not to predecide the importance of various predictors because there can be interdependencies between predictors and predictand that are not readily evident to a human. Highly correlated predictors do not pose a problem for decision-tree classification algorithms.

The blizzard condition predictand is categorical. Random forest is a recent algorithm for modeling categorical predictands developed by Breiman (2001). It is an extension of the classification and regression tree (CART) algorithm derived by Breiman et al. (1984). We will mention a few details about the random forest procedure here, but space does not allow us to provide more due to the algorithm's complexity. We recommend the interested reader consult the literature for details, for example Blouin et al. (2016), Rodriguez-Galiano et al. (2012), Cutler et al. (2007). Whereas CART builds a single decision tree from learning data, random forest constructs a group of uncorrelated decision trees by bootstrapping the

learning data and constructing a decision tree with each bootstrapped data sample. When a bootstrapped data sample is used to construct a decision tree, approximately two-thirds of its data are used to build the tree and the remaining one-third is left out for testing during the tree building process. As a decision tree builds, a subset of predictors is randomly chosen at each internal node, and the predictor giving the least error is used to split the node into two child nodes. When the ensemble of decision trees is applied for prediction, an overall prediction can be made from a threshold fraction (usually a majority) of the separate predictions made by each tree. For our binary YES/NO predictand, we henceforth use the term *YES vote ratio* to specify the fraction of decision trees in the random forest ensemble that voted YES.

To construct prediction models we used the “randomForest” package (Liaw 2018) available in the open source statistical

FIG. 8. GINI importance rank of predictors for (a) the second stage B34 prediction model and (b) the second stage B12 prediction model. Predictor definitions are in Table 4.

computing software environment R. We built two models, one for blizzard conditions (henceforth RF B34 model) and one for near-blizzard conditions (henceforth RF B12 model). In retrospect we could have built one three-category model to predict both predictands with one model. When applied for predicting blizzard and near-blizzard conditions this would give the same result for predictions of each predictand as individual predictions made by two separate models. Environmental conditions that cause $\frac{1}{4}$ SM visibility and $\frac{1}{2}$ SM visibility or in between ($\frac{3}{8}$ SM) are not much different, and visibility often fluctuates in the same event. Having two models gives more leeway to the forecaster, and the RF B12 model can be compared with BH forecasts, which predict the probability of visibility ≤ 1 km. Models were constructed in two stages. In the first stage all 41 predictors were offered to RF. When RF produces its decision trees it ranks the predictors by their overall GINI (gain in node impurity) measure of importance. GINI is a measure of the level of inequality in the learning data samples assigned to the two child nodes after their parent node is split (Breiman et al. 1984; Strobl et al. 2007). It is an inverse measure—the lower the GINI, the better the split. A model with the full complement of predictors in our case will demand considerable computer execution time when applied over the many grid points in a large domain. For real-time operational predictions we wanted to reduce execution time without seriously affecting prediction accuracy. In the second stage, predictors whose importance in the first stage is less than 0.2 times the maximum GINI importance of all predictors were eliminated, then new RF models were built with the reduced predictor set. The ad hoc 0.2 factor was found

through testing. This reduced the number of predictors to 11 for the RF B34 model and 9 for the RF B12 model.

Figure 8 shows the GINI importance ranking of predictors for the second stage RF B12 and RF B34 models. Lower GINI values for the RF B34 model than for the RF B12 model mean it has a lower likelihood of misclassification. The most important predictors are those associated with low-level boundary layer variables: wind speed (LLW18, UV2, UU2, VV2), temperature (TT2), depth of snow on the ground (SD), convective stability (RB32), moisture (HU2, HU3M2). Also important is BLSP (BH probability of visibility ≤ 1 km in blowing snow). This predictor depends partly on snowpack age using time elapsed since the last significant snowfall as a proxy. The degree of low-level convective stability (RB32) is more important for the RF B34 model than the RF B12 model, likely since lower visibility is required for blizzard conditions ($\frac{1}{4}$ SM) than for near-blizzard conditions ($\frac{1}{2}$ SM), and lower convective stability makes lifting snow from the ground easier. Temperature (TT2) is of greater importance for near-blizzard conditions (RF B12) than blizzard conditions (RF B34), likely because greater turbulence in higher wind speeds reduces dependency on temperature. The total precipitation rate predictor RT is less important than most other predictors since, as noted in section 2, the occurrence of clear sky blizzard conditions at stations in open Arctic tundra terrain, where wind speed is generally higher and fetch is greater, is significantly higher than at stations in complex terrain.

Table 5 shows the classification error for each model building run: the top two sections are the first and second stages of

TABLE 5. For combined learning data from the seven “open tundra” manned stations listed in section 3a. Classification error summaries (confusion matrices) for the RF B34 prediction models for blizzard conditions (visibility $\leq 1/4$ SM) using the initial set of 41 predictors and using the reduced set of 11 predictors. Classification error summaries for the RF B12 prediction models for near-blizzard conditions (visibility $\leq 1/2$ SM) using the initial set of 41 predictors and using the reduced set of 9 predictors. Verification of the reduced predictor set models when applied to data independent of the learning data (1 Oct 2018–31 May 2019) for the most recent 6–17-h RF B12 and RF B34 nearest gridpoint forecasts valid at the same time as an observation. Classifications are based on a majority vote (YES vote ratio ≥ 0.5).

Observed	Forecast		
	NO	YES	Classification error
B34 initial predictor set learning data			
NO	71 467	131	0.0018
YES	476	1012	0.3199
B34 reduced predictor set learning data			
NO	71 455	143	0.0020
YES	496	992	0.3333
B12 initial predictor set learning data			
NO	69 928	265	0.0038
YES	889	2004	0.3073
B12 reduced predictor set learning data			
NO	69 876	317	0.0045
YES	948	1945	0.3277
B34 reduced predictor set independent data			
NO	31 649	102	0.0032
YES	244	254	0.4900
B12 reduced predictor set independent data			
NO	31 120	154	0.0049
YES	499	476	0.5118

the RF B34 run; and the third and fourth sections are the first and second stages of the RF B12 run. Only a small increase in classification error results by reducing the number of predictors for both models. Based on a majority vote, classification error for NO forecasts is well below 1% for both models, classification error for YES forecasts is about 33%. These error rates indicate a reasonable chance of success for forecasts of blizzard conditions and near-blizzard conditions, given they are infrequent events.

4. Verification

For verification we chose the period 1 October 2018–31 May 2019, which is independent from the learning data period. This gave over 32 000 events. Classification error for second stage RF B12 and RF B34 models based on a majority vote (YES vote ratio ≥ 0.5) is shown in the bottom two sections of Table 5. Classification error for the most recent 6–17-h RF B12 and RF B34 nearest gridpoint forecasts valid at the same time as an observation is higher than for the learning

FIG. 9. For 1 Oct 2018–31 May 2019: (a) receiver operator characteristic (ROC) curves for the most recent 6–17-h RF B12, RF B34, and BH nearest gridpoint forecast valid at the same time as an observation for combined data from the seven open tundra stations listed in section 3a. YES vote ratios are treated as probability in this figure. Dots are plotted starting at 100% probability in the bottom-left corner and decreasing in 10% probability intervals to 0% at the top right. (b) Critical success index (CSI) scores for the same forecasts for threshold values of YES forecasts varying from 0 to 1.0 in increments of 0.1.

data. This may be due to natural variability between winter seasons.

Figure 9a shows receiver operator characteristic (ROC) curves for the most recent 6–17-h RF B12, RF B34, and BH nearest gridpoint forecast valid at the same time as an observation for combined data from the seven open tundra stations listed in section 3a. Definition of the ROC curve is found in Stanski et al. (1989). Here we treat the YES vote ratio as a probability. The RF and BH curves lie above the diagonal line (which indicates a no skill forecast) and so show more accuracy than a no skill forecast. The area under the ROC curve (AUC) is greater for both RF forecasts than the BH forecast, and so the RF forecasts have greater accuracy. When using a probability forecast one must choose a specific threshold value above which the answer is YES and below which the answer is NO. A ROC curve does not indicate the best probability threshold to use. To find an optimal threshold we calculated the critical success index (CSI), defined for a collection of categorical forecasts as hits/(hits + misses + false alarms) for each category. CSI rewards hits but penalizes misses and false alarms, giving a single score between 0 (no forecasts are hits) and 1 (all forecasts are hits). Figure 9b shows CSI scores for BH, and RF B12 and RF B34 YES vote ratio (referred to as probability) forecasts at probability threshold values from 0 to 1 for the forecasts used for Fig. 9a. The highest CSI scores indicate the optimal range to forecast YES is centered around 0.4 for RF forecasts and centered around 0.3 for BH forecasts.

FIG. 10. For the same data period and stations as in Fig. 9, ROC curves and accompanying CSI scores for RF B12, RF B34, and BH forecasts generated with RDPS hourly model data, with data combined for lead time segments of 13–24, 25–36, 37–48, 49–60, and 61–72 h, plotted as in Fig. 9.

Figure 10 shows ROC curves and CSI scores for various lead times for nearest gridpoint RDPS BH, RF B12, and RF B34 forecasts for the combined data from the seven open tundra stations. To create a sufficiently large database we combined hourly RDPS forecasts matched with observations in lead time segments of 13–24, 25–36, 37–48, 49–60, and 61–72 h. Results for 73–84 h are not shown for space considerations. AUC is greater for RF forecasts than for the BH forecast at all lead times, thus the RF forecasts have greater accuracy. AUC for the RF B34 forecast is greater than for the RF B12 and it has skill at all lead times, thus RF B34 forecasts are more accurate. This may be because there are fewer ways to achieve visibility as low as $\frac{1}{4}$ SM than $\frac{1}{2}$ SM in blowing snow. CSI scores for RF and BH forecasts reach the highest values to 24 h then slowly decrease after that. The optimal probability threshold for RF forecasts, as given by the highest CSI score, is centered around 0.4–36-h lead time, and is in the range 0.4–0.5 at longer lead times. For the BH forecast the optimal threshold is centered around 0.3 at all lead times. CSI scores for BP forecasts were 0.176, 0.170, 0.162, 0.160, 0.148, and 0.137, respectively, for 13–24-, 25–36-, 37–48-, 49–60-, 61–72-, and 73–84-h lead times. These are much lower scores than those for RF and BH forecasts, which is not surprising since the BP category forecasts are intended to provide a “heads up” warning well in advance of areas where blizzard conditions could develop, rather than to be specific point forecasts. The probability of detection for blizzard conditions by BP forecasts is 85% or better out to 60-h lead time and better than 75% beyond that, although the false

alarm rate is high. CSI scores for the BH and RF forecasts indicate they have a reasonably good chance for success, given that blizzard conditions occur infrequently.

5. Case study: 1200 UTC 29 March 2018

Forecasts by the three products valid at the same time, generated by different CCMEP NWP models for different lead times, are presented below for a significant blizzard event. Figure 11 shows a 0-h forecast by the RDPS model, valid at 1200 UTC 29 March 2018. A low pressure system with a strong circulation was centered on the northwest coast of Hudson Bay. Behind this low, strong northwesterly winds were impacting sites on the coast and inland.

Figure 12 shows a BP forecast generated by the GDPS model with lead time 108 h. Orange and red areas to the west and northwest of Hudson Bay give early warning that blizzard conditions could develop there in a few days, warranting close attention to this region by forecasters. Figure 13 shows a BH forecast produced from RDPS model output with an 84-h lead time. This forecast shows the probability continues to be very high that in 3.5 days widespread blizzard conditions will be occurring west and northwest of Hudson Bay. Figure 14 shows a forecast with 48-h lead time generated by the RF B34 model, produced with RDPS model output. The field shown is the *YES vote ratio*, that is, the fraction of RF decision trees deciding “yes, blizzard conditions will occur.” Blizzard conditions are still predicted where the YES vote ratio is ≥ 0.50

FIG. 11. The 0-h forecast by the RDPS model, valid at 1200 UTC 29 Mar 2018, of MSL pressure (hPa), 10-m wind, and 500–1000-hPa geopotential height difference (dam).

for a large region west and northwest of Hudson Bay. A small area of higher YES vote ratio lies north of Baker Lake. As time draws closer to an expected event, a high-resolution model becomes useful for providing a detailed forecast. Although our RF B34 model was built with predictors calculated from RDPS model output, in our experience it performs well when it is driven by CAPS model output. Figure 15 shows a 24-h forecast of the YES vote ratio by the RF B34 model generated from CAPS model output. There is more detail in the forecast compared to the forecast from the RDPS model in Fig. 14. The area with the highest YES vote ratio for blizzard conditions is still located west and northwest of Hudson Bay. Compared to Fig. 14 the area of higher values north of Baker Lake is larger,

FIG. 12. The 108-h blizzard potential (BP) forecast, valid at 1200 UTC 29 Mar 2018, overlaid by the average wind (kt) in the first six model levels shown in Table 2, which spans a layer approximately 400 m deep above the ground. The color code for BP categories defined in Table 1 is as follows: green, category 1; blue, category 2; orange, category 3; yellow, category 4; and red, category 5. Forecast produced from output of the CCMEP Global Deterministic Prediction System (GDPS) model. YBK marks the location of Baker Lake.

FIG. 13. The 84-h Baggaley–Hanesiak (BH) forecast, valid at 1200 UTC 29 Mar 2018, showing the maximum of 1) the probability of visibility ≤ 1 km in blowing snow, and 2) the probability of visibility ≤ 1 km in concurrent precipitating snow and blowing snow, using the full range of snowpack ages specified by Baggaley and Hanesiak (2005), overlaid by the average wind (kt) in the lowest six model levels shown in Table 2. Forecast produced from CCMEP Regional Deterministic Prediction System (RDPS) model output. YBK marks the location of Baker Lake.

but the area of values more than 0.80 is smaller than it was in Fig. 14. Some of the difference may be due to the shorter prediction time compared to the forecast in Fig. 14, but the higher resolution of the CAPS model does result in a more detailed prediction. Whether it is better is hard to say due to lack of observation stations in the affected area. Overall, even though the lead times and NWP models generating the forecasts are different, all the forecasts are consistent during the period leading up to 1200 UTC 29 March 2018, from the 108-h GDPS forecast to the 24-h CAPS forecast.

Verification of forecasts in Figs. 12–15 is difficult due to observing station scarcity. Figure 16 shows observations taken

FIG. 14. The 48-h forecast, valid at 1200 UTC 29 Mar 2018, of the RF B34 model's YES vote ratio for visibility $\leq 1/4$ statute mile in blizzard conditions, overlaid by the 10-m wind (kt). Forecast produced from output of the CCMEP Regional Deterministic Prediction System (RDPS) model. YBK marks the position of Baker Lake.

FIG. 15. The 24-h forecast valid, at 1200 UTC 29 Mar 2018, of the RF B34 model's YES vote ratio for visibility $\leq 1/4$ statute mile in blizzard conditions, overlaid by the 10-m wind (kt). The YES vote ratio is explained in the text. Forecast produced from output of the CCMEP high-resolution Arctic Prediction System model (CAPS). YBK marks the position of Baker Lake.

from METARs for the few stations located around northern Hudson Bay and to the north. Only stations where winds were >18 kt are shown since a lower wind speed does not trigger a forecast of near-blizzard conditions. No stations reporting lower wind speeds were in the area of interest northwest of Hudson Bay. The observations are in good agreement with the forecasts discussed above.

To compare forecasts with observations for a few days around this event we show meteograms for three observing stations (Baker Lake, Gjoa Haven, and Rankin Inlet). Figure 17a is a plot of nearest gridpoint BP, BH, and B34 forecasts, each with different projection times, in 12-h series for 0000 UTC 28 March 2018–0000 UTC 1 April 2018. Shown are 3-hourly 99–108-h BP forecasts by the GDPS model, hourly 73–84-h BH probability forecasts by the RDPS model, hourly 27–48 h RF B34 YES vote ratio forecasts by the RDPS model, and hourly RF B34 13–24-h YES vote ratio forecasts by the CAPS model. Forecast series are updated every 12 h after CCMEP NWP model runs generate new output. Figure 17b shows observed wind and gust speeds, and visibility each hour. Visibility is not reported in Baker Lake AUTO METARs, hence the large gap of missing visibilities during the event. Wind speed was also missing on 29 March for several hours in severe conditions. Figure 17b shows close negative correlation between wind speed and visibility. We infer that visibility was low at Baker Lake during the period of missing visibility. The GDPS BP forecasts in Fig. 17a show the most error, although they do tend to be category 3, 4, or 5 during most periods of low visibility. However, these are 99–108-h forecasts. BH forecasts were less accurate than RF forecasts but their lead time is longer. The overall accuracy of both RF forecasts is very good.

6. Summary and conclusions

The intention of this paper is twofold: 1) analyze METARs in recent years for occurrence frequency of blizzard conditions

FIG. 16. At 1200 UTC 29 Mar 2018, observed visibility (statute miles) and reason for restriction (SN = snow, BLSN = blowing snow, DRSN = drifting snow, IC = ice crystals, BR = fog, “-” = light, and “+” = heavy). Wind speed (kt) is shown followed by “//” for AUTO stations where only wind was reported. Station location is marked by “*.” Only the station location is shown for stations reporting wind speed less than 18 kt. The continuous green lines mark the boreal forest northern boundary.

in the Canadian Arctic; 2) describe products we provide to predict blizzard conditions, show verification and a case study. METARs reporting visibility from Canadian Arctic observing stations were analyzed for October through May in 2014–18 for occurrences of “near-blizzard” and “blizzard” conditions. Blizzard conditions occur mainly in open tundra east and north of the boreal forest boundary, and with nearly zero frequency to the south. The highest frequency of occurrence, 1%–3.5%, is on the northwest shore of Hudson Bay and northwest from there. Most other stations north of the treeline report blizzard conditions in approximately 0.5%–1.5% of available METARs, except much lower frequencies were reported by stations in sheltered locations, particularly in the eastern and northern Arctic. However, blizzard conditions may not occur at a sheltered site when they are simultaneously occurring in nearby open country, as evidenced by the 5% frequency at a station in high open tundra on central Baffin Island. At many stations when blizzard conditions were reported, in about half to two-thirds or more of human-observed METARs the stated cause of reduced visibility is blowing snow without precipitating snow, known as the “clear sky” blizzard. However on the northwestern shore of Hudson Bay and to the northwest, and at a few other stations, this situation occurs in approximately 80% or more of METARs reporting blizzard conditions. The importance of predicting wind accurately in the Arctic boundary layer for forecasting blizzard conditions is apparent.

Forecasting the onset and duration of blizzard conditions over the vast Canadian Arctic domain is challenging for several reasons. The three products discussed in this paper address the need for automated guidance covering the entire Arctic

FIG. 17. Series of forecasts updated every 12 h covering the period 0000 UTC 28 Mar 2018–0000 UTC 1 Apr 2018. Each forecast is valid at the time shown. BP 99–108-h forecasts driven by GDPS model output are plotted 3-hourly as black circles starting at 0300 UTC 28 Mar. Other forecasts plotted hourly starting at 0100 UTC 28 Mar are connected with colored lines: BH 73–84-h forecasts driven by RDPS model output (green), RF 37–48-h forecasts driven by RDPS model output (red), and RF 13–24-h forecasts driven by CAPS model output (blue). Location of stations is in Fig. 1. (b) Matching observations of visibility, wind speed, and gusts for the forecasts in (a). Missing observations could not be plotted. Non-missing visibilities greater than 5 SM are not plotted to conserve space.

domain. Each is derived by different methods, each is driven with postprocessed output from CCMEP operational global, regional, and high-resolution NWP models. The BP product is a collection of expert forecaster's rules covering six general categories of expected weather in differing severity from drifting snow to blowing snow, to near-blizzard conditions, to blizzard conditions. The intention of BP forecasts is to provide a "heads up" warning well in advance of areas where blizzard conditions might develop, rather than to be specific point forecasts. Another product, the BH probability forecast, stems from work by Baggaley and Hanesiak (2005). It provides more accuracy at points than does the BP product, and makes

probability forecasts. The BP and BH products are perfect prog forecast algorithms, hence the accuracy of their forecasts depends on the accuracy of NWP model output used to make their forecasts. Our third blizzard forecast product, RF, is a MOS product, so it does account for a degree of NWP model error. RF is a machine-learning product built with the random forest algorithm (Breiman 2001) that predicts the likelihood that blizzard conditions and near-blizzard conditions will occur. Classification errors for NO forecasts are much less than 1%, and for YES forecasts are about 33%, so the products have a reasonable chance for success given that blizzard conditions occur infrequently. Analysis of the most important

predictors for blizzard conditions in the RF models underscores the importance of predicting wind, lapse rate, and precipitation accurately in the Arctic winter boundary layer.

To assess model performance, compare forecasts, and determine optimal thresholds for BH and RF YES forecasts we analyzed the BP, BH, and RF forecasts produced from RDPS model output for an eight month period (one winter season). BH and RF forecasts were compared using receiver operator characteristic curves and critical success index scores. The results show RF forecasts have greater accuracy at all lead times, and the RF B34 forecasts are more accurate than the RF B12 forecasts at all lead times. The optimal probability threshold for making a YES forecast for RF forecasts, as given by the highest CSI score, is centered around 0.4 up to 36-h lead times and is in the range 0.4–0.5 at longer lead times. For BH forecasts the optimal threshold range is centered around 0.3 at all lead times. These CSI scores show the products have a reasonable chance for success given that blizzard conditions occur infrequently. CSI scores for BP forecasts are much lower than those for the RF and BH forecasts, but they are intended to be used for longer range area forecasts rather than point forecasts.

We show a sample of forecasts with different lead times by all three products driven by different CCMEP models during a major blizzard event. Blizzard conditions were consistently predicted to occur over the same large area on the northwest side of Hudson Bay. Meteograms displaying observations and four 12-hourly incremented time series, each for nearest grid-point forecasts taken at different lead times, are shown for three widely separated stations for a four day period centered around this event. The RF forecasts performed very well, and the BH and BP forecasts performed reasonably well considering their longer lead times. The RF forecasts for blizzard conditions have proven to be the most popular among operational forecasters, although all three forecast products are widely used. There have been many favorable comments from operational forecasters on all three.

We believe the automated products discussed here constitute a good start for filling the need of operational meteorologists for comprehensive guidance for forecasting blizzards for the vast Canadian Arctic domain. Besides producing forecasts utilizing all three methodologies we are experimenting with an ad hoc approach for showing areas where two and three products agree. We believe improvements can be made in blizzard prediction models. We continue to search for new ways to formulate suitable predictors and new methods to generate forecast models.

Acknowledgments. We are grateful to our ECCC colleagues Julian Brimelow, Gabriel Gascon, Ruping Mo, Zen Mariani, Tom Robinson, and the AMS reviewers for their comments and suggestions on an earlier version of this paper that helped to improve this manuscript. We thank Aaron Kennedy and Peter Taylor for offering many helpful comments and suggestions in their reviews.

Data availability statement. Aviation routine weather reports (METARs) are universally available. Output data from

our forecast models described here resides in our research account on the Government of Canada Science Network.

REFERENCES

- Arctic Council, 2017: Telecommunications infrastructure in the Arctic: A circumpolar assessment. Arctic Council, 92 pp., <https://oaarchive.arctic-council.org/handle/11374/1924>.
- Baggaley, D. G., and J. M. Hanesiak, 2005: An empirical blowing snow forecast technique for the Canadian arctic and Prairie provinces. *Wea. Forecasting*, **20**, 51–62, <https://doi.org/10.1175/WAF-833.1>.
- Blouin, K. D., M. D. Flannigan, X. Wang, and B. Kochtubajda, 2016: Ensemble lightning prediction models for the province of Alberta, Canada. *Int. J. Wildland Fire*, **25**, 421–432, <https://doi.org/10.1071/WF15111>.
- Bourguin, P., 2000: A method to determine precipitation types. *Wea. Forecasting*, **21**, 583–592, [https://doi.org/10.1175/1520-0434\(2000\)015<0583:AMTDPT>2.0.CO;2](https://doi.org/10.1175/1520-0434(2000)015<0583:AMTDPT>2.0.CO;2).
- Breiman, L., 2001: Random forests. *Mach. Learn.*, **45**, 5–32, <https://doi.org/10.1023/A:1010933404324>.
- , J. H. Friedman, R. A. Olshen, and C. J. Stone, 1984: *Classification and Regression Trees*. Wadsworth, 368 pp. <https://doi.org/10.1201/9781315139470>.
- Burrows, W. R., 1985: On the use of time-offset model output statistics for production of surface wind forecasts. *Mon. Wea. Rev.*, **113**, 2049–2054, [https://doi.org/10.1175/1520-0493\(1985\)113<2049:OTUOTO>2.0.CO;2](https://doi.org/10.1175/1520-0493(1985)113<2049:OTUOTO>2.0.CO;2).
- , and C. J. Mooney, 2018: Automated products for forecasting blizzard conditions in the Arctic. Polar Prediction Matters, HelmholtzBlogs, accessed 8 November 2018, <https://blogs.helmholtz.de/polarpredictionmatters/>.
- COMET MetEd Program, 2005: Topics in precipitation type forecasting, section 3: Top-down method. Accessed 14 May 2021, <https://www.meted.ucar.edu/norlat/snow/precipitype/>.
- Cutler, D. R., T. C. Edwards, K. H. Beard, A. Cutler, K. T. Hess, J. C. Gibson, and J. J. Lawler, 2007: Random forests for classification in ecology. *Ecology*, **88**, 2783–2792, <https://doi.org/10.1890/07-0539.1>.
- Déry, S. J., and M. K. Yau, 1999: A climatology of adverse winter weather events. *J. Geophys. Res.*, **104**, 16 657–16 672, <https://doi.org/10.1029/1999JD900158>.
- , and —, 2001: Simulation of an Arctic ground blizzard using a coupled blowing snow–atmosphere model. *J. Hydrometeor.*, **2**, 579–598, [https://doi.org/10.1175/1525-7541\(2001\)002<0579:SOAAGB>2.0.CO;2](https://doi.org/10.1175/1525-7541(2001)002<0579:SOAAGB>2.0.CO;2).
- Dubé, I., 2003: From mm to cm. . . study of snow/liquid water ratios in Quebec. Meteorological Service of Canada–Quebec Region, 127 pp.
- Environment and Climate Change Canada, 2016: Regional Deterministic Prediction System (RDPS), Update from version 4.2.0 to version 5.0.0: Description of changes and forecast performance up to 48h. Canadian Meteorological Centre Tech. Note, Section 4.2, 47 pp., https://collaboration.cmc.ec.gc.ca/cmc/cmoi/product_guide/docs/lib/technote_rdp5-500_20160907_e.pdf.
- Girard, C., and Coauthors, 2014: Staggered vertical discretization of the Canadian Environmental Multiscale (GEM) model using a coordinate of the log-hydrostatic-pressure type. *Mon. Wea. Rev.*, **142**, 1183–1196, <https://doi.org/10.1175/MWR-D-13-00255.1>.
- Hanesiak, J., and Coauthors, 2010: Storm Studies in the Arctic (STAR). *Bull. Amer. Meteor. Soc.*, **91**, 47–68, <https://doi.org/10.1175/2009BAMS2693.1>.

- Joe, P., and Coauthors, 2020: The Canadian Arctic weather science project: Introduction to the Iqaluit site. *Bull. Amer. Meteor. Soc.*, **101**, E109–E128, <https://doi.org/10.1175/BAMS-D-18-0291.1>.
- Li, L., and J. W. Pomeroy, 1997: Estimates of threshold wind speeds for snow transport using meteorological data. *J. Appl. Meteor.*, **36**, 205–213, [https://doi.org/10.1175/1520-0450\(1997\)036<0205:EOTWSF>2.0.CO;2](https://doi.org/10.1175/1520-0450(1997)036<0205:EOTWSF>2.0.CO;2).
- Liaw, A., 2018: Documentation for R package ‘randomForest,’ V 4.6-14. 29 pp., <https://cran.r-project.org/web/packages/randomForest/randomForest.pdf>.
- Milbrandt, J., S. Belair, M. Faucher, M. Vallee, M. Carrera, and A. Glazer, 2016: The pan-Canadian high resolution (2.5-km) deterministic prediction system. *Wea. Forecasting*, **31**, 1791–1816, <https://doi.org/10.1175/WAF-D-16-0035.1>.
- Phillips, D., 1990: *The Climates of Canada*. Environment Canada, 159 pp.
- Pomeroy, J. W., P. Marsh, and D. M. Gray, 1997: Application of a distributed blowing snow model to the Arctic. *Hydrol. Processes*, **11**, 1451–1464, [https://doi.org/10.1002/\(SICI\)1099-1085\(199709\)11:11<1451::AID-HYP449>3.0.CO;2-Q](https://doi.org/10.1002/(SICI)1099-1085(199709)11:11<1451::AID-HYP449>3.0.CO;2-Q).
- Ricketts, S., and E. Hudson, 2001: Climatology and forecasting hazardous weather in Canada’s Arctic. *Sixth Conf. on Polar Meteorology and Oceanography*, San Diego, CA, Amer. Meteor. Soc., 5B.15, <https://ams.confex.com/ams/Polar-AirSe/webprogram/Paper21123.html>.
- Rodriguez-Galiano, V. F., M. Chica-Olmo, F. Abarca-Hernandez, P. M. Atkinson, and C. Jeganathan, 2012: Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture. *Remote Sens. Environ.*, **121**, 93–107, <https://doi.org/10.1016/j.rse.2011.12.003>.
- Savelyev, S. A., M. Gordon, J. Hanesiak, T. Papakyriakou, and P. A. Taylor, 2006: Blowing snow studies in the Canadian Arctic shelf exchange study, 2003–04. *Hydrol. Process.*, **20**, 817–827, <https://doi.org/10.1002/hyp.6118>.
- Stanski, H. R., L. J. Wilson, and W. R. Burrows, 1989: Survey of common verification methods in meteorology. World Weather Watch Tech. Rep. 8, WMO Tech. Doc. 358, 114 pp., https://www.cawcr.gov.au/projects/verification/Stanski_et_al/Stanski_et_al.html.
- Strobl, C., A.-L. Boulesteix, A. Zeileis, and T. Hothorn, 2007: Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinfo.*, **8**, 25, <https://doi.org/10.1186/1471-2105-8-25>.
- Xiao, J., and P. A. Taylor, 2002: On equilibrium profiles of suspended particles. *Bound.-Layer Meteor.*, **105**, 471–482, <https://doi.org/10.1023/A:1020395323626>.
- Yang, J., M. K. Yau, X. Fang, and J. W. Pomeroy, 2010: A triple-moment blowing snow-atmospheric model and its application in computing the seasonal wintertime snow mass budget. *Hydrol. Earth Syst. Sci.*, **14**, 1063–1079, <https://doi.org/10.5194/hess-14-1063-2010>.