

# A Statistical Forecast Model for Extratropical Cyclones Including Intensity and Precipitation Type

REBEKAH CAVANAGH<sup>a,b</sup> AND ERIC C. J. OLIVER<sup>a</sup>

<sup>a</sup> *Dalhousie University, Halifax, Nova Scotia, Canada*

<sup>b</sup> *Meteorological Service of Canada, Environment and Climate Change Canada, Toronto, Ontario, Canada*

(Manuscript received 22 February 2023, in final form 3 July 2023, accepted 7 August 2023)

**ABSTRACT:** Winter extratropical cyclones (ETCs) are dominant features of winter weather on the east coast of North America. These storms are characterized by high winds and heavy precipitation (rain, snow, and ice). ETCs are well predicted by numerical weather prediction models (NWP) at short- to midrange forecast lead times, but prediction on seasonal time scales is lacking. We develop a set of multiple linear regression models, using stepwise regression and cross validation, to predict the number of storms expected to affect a specific location throughout the winter storm season. Each model in the set predicts a specific storm type (e.g., snow, rain, or bomb storms). This set of models is applied in a probabilistic forecast framework that uses the probability density function of the prediction in combination with climatological mean storm activity. The resulting forecast makes statements about the likelihood of below-average, average, or above-average activity for all storms and for each of the type-specific subsets of storms. Though this forecast framework could in theory be applied anywhere, we demonstrate its skill in forecasting the characteristics of the winter storm season experienced in Halifax, Nova Scotia, Canada.

**SIGNIFICANCE STATEMENT:** Winter storms are a disruptive but inevitable part of life on the eastern coast of North America all the way from the Carolinas to Labrador. Knowing each fall what to expect for the upcoming winter storm season is not only a matter of public interest, but also of great public safety and financial importance. Here we develop a model that uses the state of the atmosphere over the month of September to forecast the upcoming winter storm characteristics for a specified region of interest. Our model uses a multiple linear regression approach to make skilled forecasts including probability statements about the level and type of storm activity. Forecasts can be used to inform planning for the winter ahead.

**KEYWORDS:** Extratropical cyclones; Winter/cool season; Regression analysis; Probability forecasts/models/distribution; Seasonal forecasting; Statistical forecasting

## 1. Introduction

Extratropical cyclones (ETCs) are salient features of midlatitude weather and extreme cases threaten life and property in eastern North America and around the world every winter. Hurricane-force winds can down power lines, mixed precipitation can cause treacherous road conditions, storm surges can flood coastlines, and countless homes can be snowed in as these powerful synoptic-scale low pressure systems pass through a region. Fortunately, not every storm passing through has these extreme consequences. As with hurricanes, their severity and frequency varies greatly from storm to storm and from winter to winter. ETC tracks and impacts are typically well predicted by numerical weather prediction (NWP) models and operational forecasters on time scales of a few days. Prediction on this time scale is crucial for emergency preparedness and human safety. However, a seasonal forecast can provide the opportunity for

more extensive planning, preparation, and damage mitigation in the months leading up to the storm season.

Seasonal forecasting of ETCs trails behind that of the well-established field of Atlantic hurricane seasonal forecasting. The development of seasonal hurricane models, led by researchers at Colorado State University (Klotzbach and Gray 2009), began with purely statistical modeling (Gray 1984) and more recently moved into the realm of statistical–dynamical modeling (Klotzbach et al. 2020). In general, current seasonal hurricane forecasts, such as the outlook produced by NOAA, are skillful (Klotzbach et al. 2019) and used by the general public. Current seasonal ETC forecasts are mainly derived from global circulation models (GCMs; Perkins and Hakim 2020), empirical orthogonal functions (EOFs; Feng et al. 2019), or teleconnections with large-scale climate modes (DeGaetano et al. 2002). The most influential teleconnections for east coast winter storms are the NAO and ENSO. While they show some effects on seasonal activity, the combined variability explained by these predictors is still small (DeGaetano et al. 2002). The region with greatest predictability in seasonal ETC forecasting is located in the North Pacific, where many studies have identified enhanced predictability (Feng et al. 2019; Yang et al. 2015; Befort et al. 2019). Seasonal activity of windstorms near Europe have also been predicted with some skill (Befort et al. 2019); however, most models fail to skillfully predict storms in the western North

---

Cavanagh's current affiliation: Meteorological Service of Canada, Environment and Climate Change Canada, Toronto, Ontario, Canada.

---

*Corresponding author:* Rebekah Cavanagh, rebekah.cavanagh@dal.ca

DOI: 10.1175/MWR-D-23-0041.1

© 2023 American Meteorological Society. This published article is licensed under the terms of the default AMS reuse license. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy ([www.ametsoc.org/PUBSReuseLicenses](http://www.ametsoc.org/PUBSReuseLicenses)).

Atlantic on a seasonal scale (Befort et al. 2019; Yang et al. 2015; Feng et al. 2019). One existing model for extratropical cyclone prediction which specifically focusses along the U.S. coast was developed by DeGaetano et al. (2002). It produces skillful forecasts of overall seasonal storm activity, but does not add any detail around the expected storm types. The majority of models are characterized by generality: the spatial extent of the prediction area is large, there is little spatial nuance added in the forecast, and little to no information about precipitation typing or storm intensity is included. The prediction efforts presented here are focused on two aspects lacking in previous studies. We add the capacity to forecast storms based on impacts and do so in areas where GCMs and other prediction models have limited skill: the east coast of North America.

This paper demonstrates a methodology to produce a storm-type- and location-specific forecast of ETCs for an upcoming winter season. The winter ETC field is spatially variable and nuanced, making skilled large-scale prediction models difficult to create. Therefore, we control for the spatial variability of the storm field by focusing on a specific region. By considering a region small enough such that the storm field varies uniformly in space within the region, we can focus specifically on analyzing the temporal variability in that area, limiting the complexity of the problem and potentially allowing for a higher level of predictability. We used the region around Halifax, Nova Scotia, Canada, to demonstrate a methodology which, in principle, may be applied to develop a forecast at any location. The model is developed with the application of operational usage in mind. Therefore, we elect to use predictors from the immediately preceding fall season so they can be observed and processed with adequate time to produce a forecast before the following winter season. With a goal of supplementing public forecasting, it is desirable to have a model that can give outputs that a typical citizen would value. For this reason, the focus is placed on forecasting frequency of storms and storm types, rather than trying to produce specifics such as track locations. The main ETC effects that the average citizen is concerned with are precipitation type, whether or not there will be high winds, and the overall intensity of the storms (e.g., a rapidly deepening bomb storm as in Sanders and Gyakum 1980). We then develop a multiple linear regression prediction model for each subset of storm type. Model outputs are combined with their uncertainties to develop a probabilistic seasonal forecast. Forecast results are communicated as the likelihood of high, average, or low storm-track activity through the season for storms of each type.

The structure of the paper is as follows. The data sources for the project are given in section 2. Next, the development of the storm-track dataset is described in section 3, which ends by presenting the predictand Halifax storm-track time series. Section 4 explains the development of the model including the selection of predictors and model fitting. Then, a practical application of the statistical model is proposed in section 5, including the results of the probabilistic winter storm forecast for Halifax. Finally, the discussion and conclusions are presented in section 6.

## 2. Data

### a. Atmospheric reanalysis

The storm data used in this study are derived from the output of an atmosphere reanalysis. Reanalysis data were obtained from the European Centre for Medium-Range Weather Forecasts Reanalysis v5 (ERA5) hourly single levels data from 1979 to the present (Hersbach et al. 2018b). The variable used is mean sea level pressure, at a temporal resolution of 3 h, starting at 0000 UTC. The grid spacing of the model is  $0.25^\circ$ . Data were obtained between  $25^\circ$  and  $70^\circ\text{N}$  and between  $110^\circ\text{W}$  and  $0^\circ$ . In this study, the extended winter storm season is defined as 1 November of the first year to 31 March of the following year. Data were obtained from 1 November 1979 to 31 March 2019 (40 winter storm seasons). The convention used when naming a season is to refer to it by the year in which it began. For example, the 1979 season spans the period of time from 1 November 1979 to 31 March 1980.

When considering potential predictors, we use 2-m air temperature (T2M), 1000–500 hPa thickness (T500), mean sea level pressure (MSLP), 500-hPa geopotential height, wind at 250 hPa ( $u$  component: U250;  $v$  component: V250; magnitude: WND250), and total precipitable water vapor (TPWV) from the ERA5 hourly data on pressure levels from 1979 to 2019 (Hersbach et al. 2018a). Each of these variables is considered over a spatial extent that is bounded by the longitudes of  $103^\circ$  and  $22^\circ\text{W}$  and latitudes of  $25^\circ$  and  $63^\circ\text{N}$ . Within this region, the data are obtained on a  $1^\circ \times 1^\circ$  grid. Though these potential predictor fields are available at  $1/4^\circ$  spatial resolution, we use the coarser resolution to reduce computational cost.

### b. Weather station

Hourly observations were obtained from the Environment and Climate Change Canada (ECCC) weather station at the Halifax Stanfield International Airport over the same time period as the ERA5 data. We used the wind speed measurement, which is measured by an anemometer at 10 m above the ground. We also used the current conditions recorded by the onsite observer to determine precipitation type. These observations describe any weather phenomena occurring such as reductions to visibility and precipitation.

## 3. Storm-track methodology

We begin by developing a dataset of storm tracks that will be used to build the predictand time series. The principles of the storm-tracking algorithm used here originate in the ocean eddy tracking algorithm developed by Chelton et al. (2011) as applied by Oliver et al. (2015), which has been adapted here to track atmospheric storms.

### a. Storm-track dataset

#### 1) DETECTION

Storms are detected in a mean sea level pressure (MSLP) field that was preprocessed with a Gaussian filter with a  $1^\circ$  radius to remove high wavenumber variability. At a single point in time the algorithm loops through a series of critical MSLP

levels from highest to lowest. The array of critical MSLP levels begins at 1048 hPa and goes down to 920 hPa at decrements of 4 hPa. At each level, the MSLP field is separated into pixels with values above the critical level and pixels with values below the critical level.

Storms are then identified as contiguous regions of pixels below the critical level that meet the following criteria:

- (i) There is at least 1 local minimum (i.e., a viable storm center).
- (ii) There are at least 9 pixels and not more than 6000 pixels within a region.
- (iii) The relative amplitude of the storm, or difference between pressure at the edge of the storm and in the center of the storm, is at least 1 hPa.
- (iv) The distance between any pair of pixels is no more than a maximum value based on the area of the storm. If the area of a storm is assumed to take the shape of an ellipse with eccentricity 0.9, the maximum distance of any two pixels would be equal to the length of the major axis. If two pixels in the region are further than this calculated distance, we determine the region does not represent a storm.

After a storm is identified, the location of its center is recorded along with the date and time. The pixels comprising the storm are then removed from consideration at all lower critical MSLP levels to avoid double detection.

## 2) TRACKING

Once possible storm centers have been identified at each time step, they are stitched together across time to form tracks. For each storm center at time  $t$ , the algorithm looks for a storm center at time  $t + 1$  within a radius of 240 km. This works out to a maximum propagation speed of  $80 \text{ km h}^{-1}$ . If there is one storm center found at time  $t + 1$  within a 240 km radius of the time  $t$  location, the time  $t + 1$  and time  $t$  storm centers are stitched together as part of a track. If multiple centers are detected, the closest one to the time  $t$  location is chosen. If no centers are found, the storm is considered to be terminated. The details of the storm are then saved and no further time steps can be added to its track. These details include location, central pressure, deepening rate, speed, and storm area every 3 h. Once the storm-track dataset have been created for the whole study period, track locations are reinterpolated to hourly resolution.

## 3) POSTPROCESSING

After the tracking is complete, the dataset is refined by removing storms that are not representative of the systems in which we are interested. Specifically, storms are removed if they have (i) duration less than 24 h, (ii) genesis north of  $60^\circ\text{N}$ , or (iii) location above 1000 m above sea level. The reasoning for each criterion is given below.

- (i) Duration less than 24 h.  
This is a common practice used to remove very short term, noisy features that have been picked up by the tracking algorithm (Hoskins and Hodges 2002; Massey

2012; Neu et al. 2013; Pinto et al. 2016; Raible et al. 2008).

- (ii) Genesis north of  $60^\circ\text{N}$ .  
This was implemented for the purposes of another study that shared the storm-track dataset. The other study focused on climatology extending as far north as Iceland. This part of the methodology prevents the Icelandic low from increasing the storm-track density in that region, but retains the contribution of extratropical cyclones that pass through that area. It has no impact on the set of storms used in this particular study.
- (iii) Location more than 1000 m above sea level.  
Orographic effects on air circulations in mountainous regions create mesoscale low pressure centers that differ from the larger ETCs which are the focus of this study. The filtering removes all centers that are detected at a location where the surface elevation is greater than 1000 m above sea level before the tracking begins. Some of these rotational features may develop into ETCs of interest. However, if they do, the algorithm picks them up once they are below 1000 m elevation so they are not lost.

### b. Storm time series

We must define a predictand time series in order to develop a prediction model. The model takes a regionally specific approach by focusing on storms that affect Halifax, Nova Scotia. Storms are selected from the full storm-track dataset described above for inclusion in the predictand time series based on a single criterion: that their storm center location is within 750 km of Halifax for at least one time step (Fig. 1a). Given this set of storm tracks, an annual time series of total Halifax winter storms is calculated (Fig. 2a). Each unique storm track can contribute to the total counts in the time series exactly one time, regardless of the number of times it enters or exits the 750 km radius.

The time series is additionally separated into seven subsets based on precipitation, winds, and pressure tendency at the Halifax International Airport. It is assumed that the weather experienced at Halifax is due to the storm when the storm is within 1000 km of the weather station given the typical length scale of ETCs is on the order of 1000 km. When assessing hours of precipitation due to the storm or hours of high winds due to the storm, the precipitation and/or wind are assessed at Halifax airport only for the time steps when the storm is within 1000 km of the station. The combination of 750 and 1000 km radii allowed us to minimize the number of storms included that had little effect on the conditions at Halifax while retaining sufficient precipitation information from storms that did have an impact. The seven subsets of storms that pass within 750 km of Halifax are defined as follows:

- (i) Storms with at least 3 h of precipitation (total precip)  
If a storm is within 1000 km of Halifax for at least 3 h, and precipitation is recorded at the Halifax airport for at least three of the hours that it is within that radius, the storm is determined to be a precipitating storm (Fig. 1b). Remarks

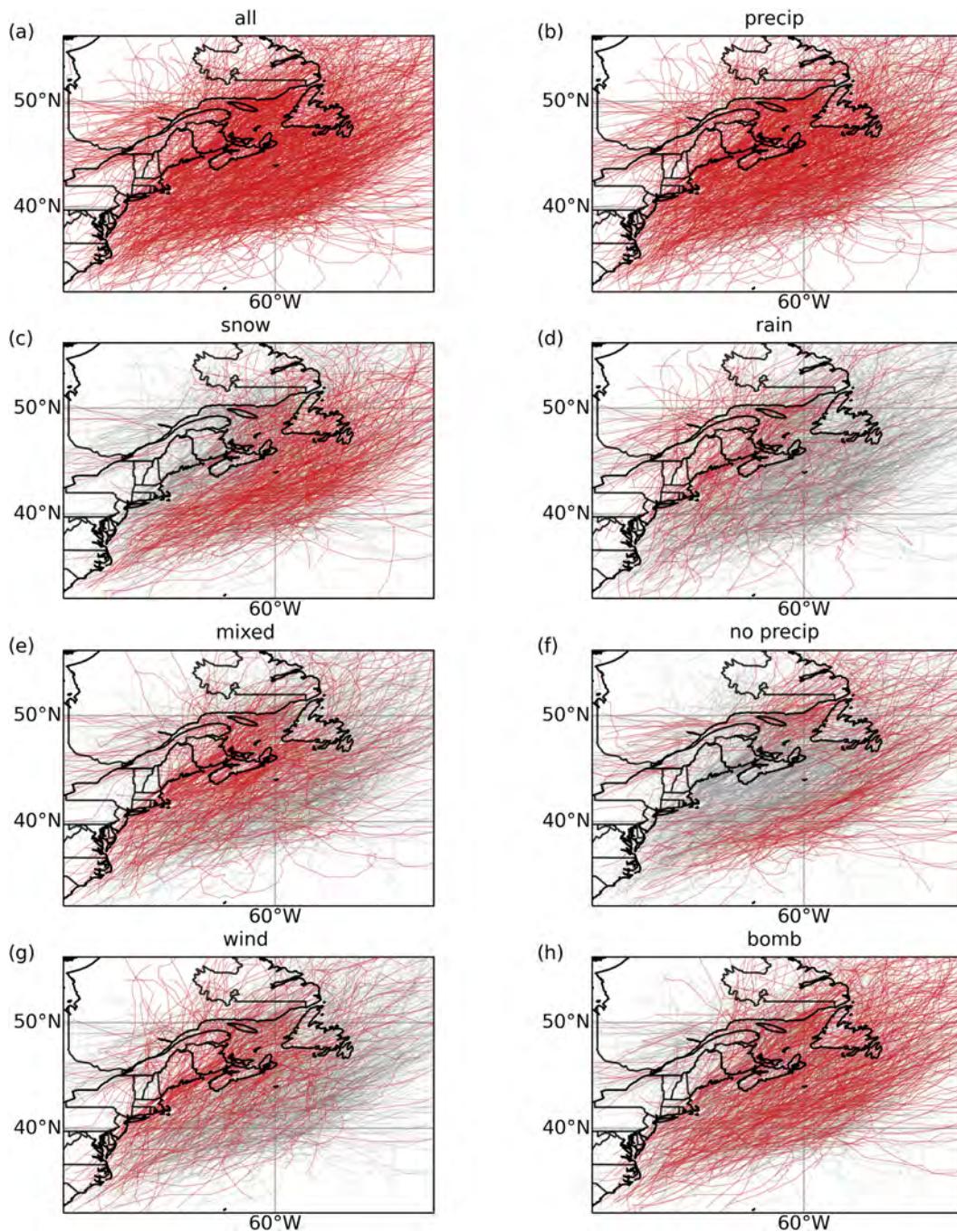


FIG. 1. Tracks of all storms passing within 750 km of Halifax (gray) overlaid with storms of each subseries (red): (a) total, (b) total precip, (c) snow, (d) rain, (e) mixed, (f) no precip, (g) high wind, and (h) bomb.

that qualify as precipitation include the words “rain,” “drizzle,” “freezing,” “snow,” or “ice pellets.”

(ii) Snow storms (snow)

All storms with at least 3 h of precipitation are further classified as either a snow, rain, or mixed storm. To be a snow storm, the observation remarks must include either “snow” or “ice pellets” for at least 90% of time steps during which precipitation is recorded,

with the restriction that precipitation can only be recorded when the storm is within 1000 km of Halifax (Fig. 1c).

(iii) Rain storms (rain)

To be classified as a rain storm, the observation remarks must include either “rain” or “drizzle” for at least 90% of time steps at which precipitation is recorded when the storm is within 1000 km of Halifax (Fig. 1d).

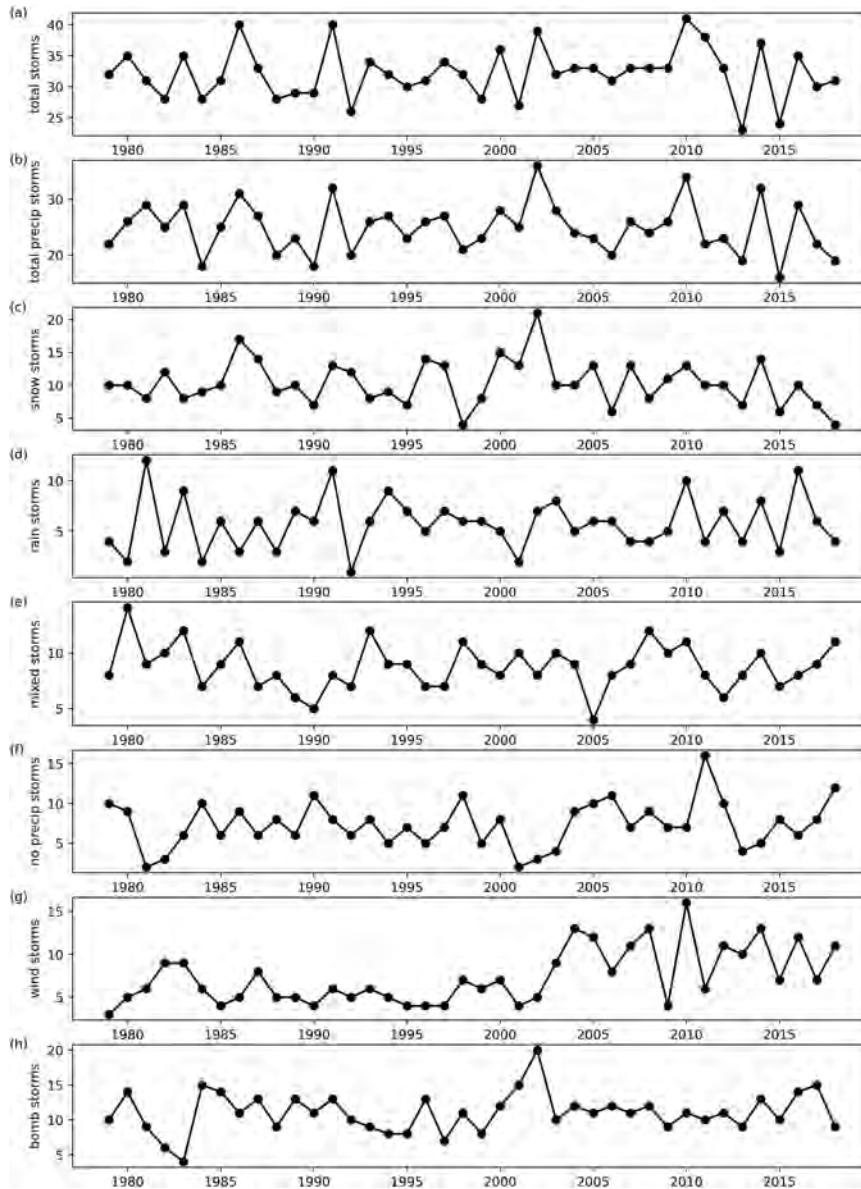


FIG. 2. Time series of (a) all winter storms affecting Halifax and the seven subseries of storms: (b) total precip, (c) snow, (d) rain, (e) mixed, (f) no precip, (g) high wind, and (h) bomb.

(iv) Mixed precipitation storms (mixed)

A storm can be grouped into the mixed precipitation category one of two ways. If a precipitating storm does not fit into the snow or rain storm categories, it is marked as a mixed precipitation storm because it is not predominantly one type or the other. Alternatively, a storm will be placed in this category if it records a “freezing” remark in at least 90% of the time steps at which precipitation is recorded within 1000 km of Halifax (Fig. 1e).

(v) Storms with less than 3 h of precipitation (no precip)

Any system with less than 3 h of precipitation recorded in the weather remarks at the Halifax airport when it is within 1000 km of the station is categorized as a storm with no precipitation (Fig. 1f). This category separates out

storms that are geographically close enough to Halifax that they could have an impact on the region, but either are not large enough or developed enough to have a significant precipitation impact. These storms may have other impacts such as storm surge or high winds, although in our dataset there are no storms that fall in both the high wind and no precipitation categories. It also catches storms that are not within 1000 km of Halifax for at least three time steps, since such a storm does not have enough time steps to meet the precipitation requirements.

(vi) High wind storms (wind)

If an hourly sustained wind speed greater than  $44 \text{ km h}^{-1}$  is recorded at Halifax while the storm is within 1000 km, it

TABLE 1. Statistics of Halifax storm time series.

Type	Mean (count)	Variance (count <sup>2</sup> )	Linear trend (count yr <sup>-1</sup> )	<i>p</i> value (trend)	Percentage of total storms
Total	32.2	16.91	0.01	0.8364	100
Total precip	24.85	20.23	-0.03	0.6155	77
Snow	10.32	11.62	-0.04	0.4404	32
Rain	5.75	6.69	0.02	0.6365	18
Mixed	8.78	4.07	-0.01	0.6713	27
No precip	7.35	8.23	0.04	0.2754	23
Wind	7.38	10.43	0.16	0.0002	23
Bomb	11.05	8.15	0.04	0.3573	34

is classified as a high wind storm (Fig. 1g). The threshold of  $44 \text{ km h}^{-1}$  is the value of the 98th percentile of all wind measurements at Halifax throughout the study period. This category is not a subset of the precipitation storms, therefore these storms may or may not also be found in a precipitation category or in bomb storms.

(vii) Rapidly deepening storms (bomb)

A storm is classified as rapidly deepening if the central pressure drops at a rate of at least  $12 \text{ hPa} (\sin \phi / \sin 45^\circ)$  in 12 h, where  $\phi$  is the latitude of the storm center (Fig. 1h). This is consistent with the definition of a bomb storm set by Sanders and Gyakum (1980). This category has no requirements pertaining to precipitation or wind speeds, although bomb storms are known to be associated with high impact weather.

Annual time series of storm counts are calculated for each of these categories or storm types (Fig. 2). Rain,

snow, and mixed storms sum to the precipitation storms. The precipitation storms and the no precipitation storms sum to the total storms. Each storm within the high wind storms and bomb storms categories will also be found in either the precipitation or no precipitation storms categories and possibly in one of the subsets of precipitation types. The basic statistics of these time series are summarized in Table 1. None of the time series have a statistically significant linear trend except for the high wind storm time series, which increases at a rate of 0.16 storms per season ( $p = 0.0002$ ).

#### 4. Model methodology and development

##### a. Overview

A multiple linear regression (MLR) model is a simple, but powerful method of modeling future behavior within a system that combines the effects of multiple variables to

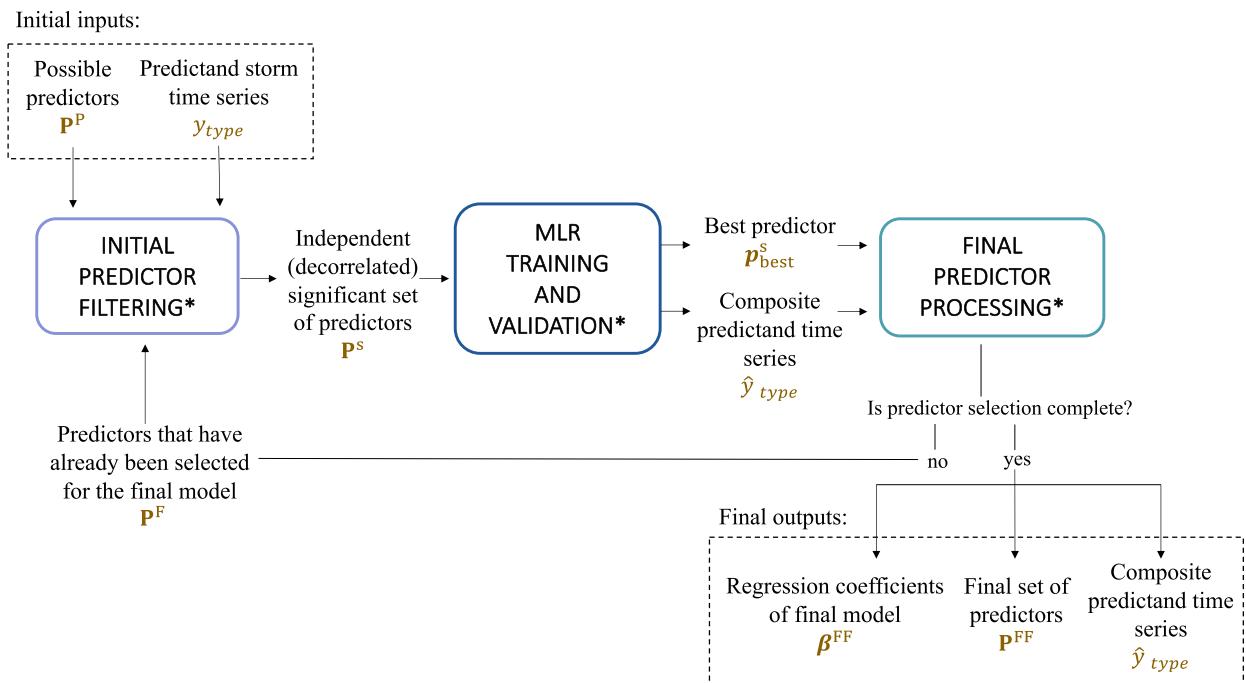


FIG. 3. Predictor selection process overview. Processes indicated with an asterisk are further explained subsequently in Fig. 4 (initial predictor filtering), Fig. 5 (MLR training and validation), and Fig. 6 (final predictor processing).

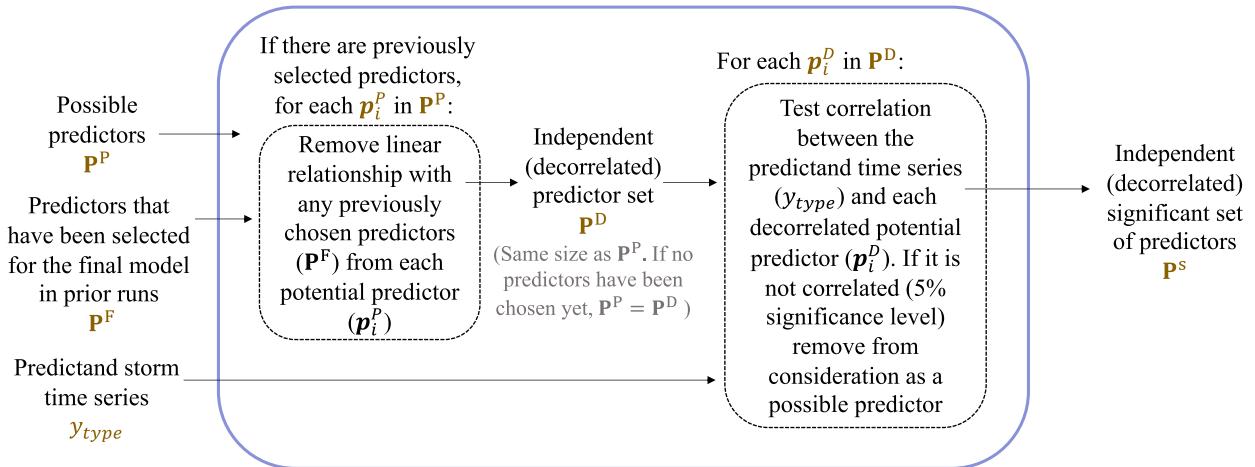


FIG. 4. Initial predictor filtering schematic.

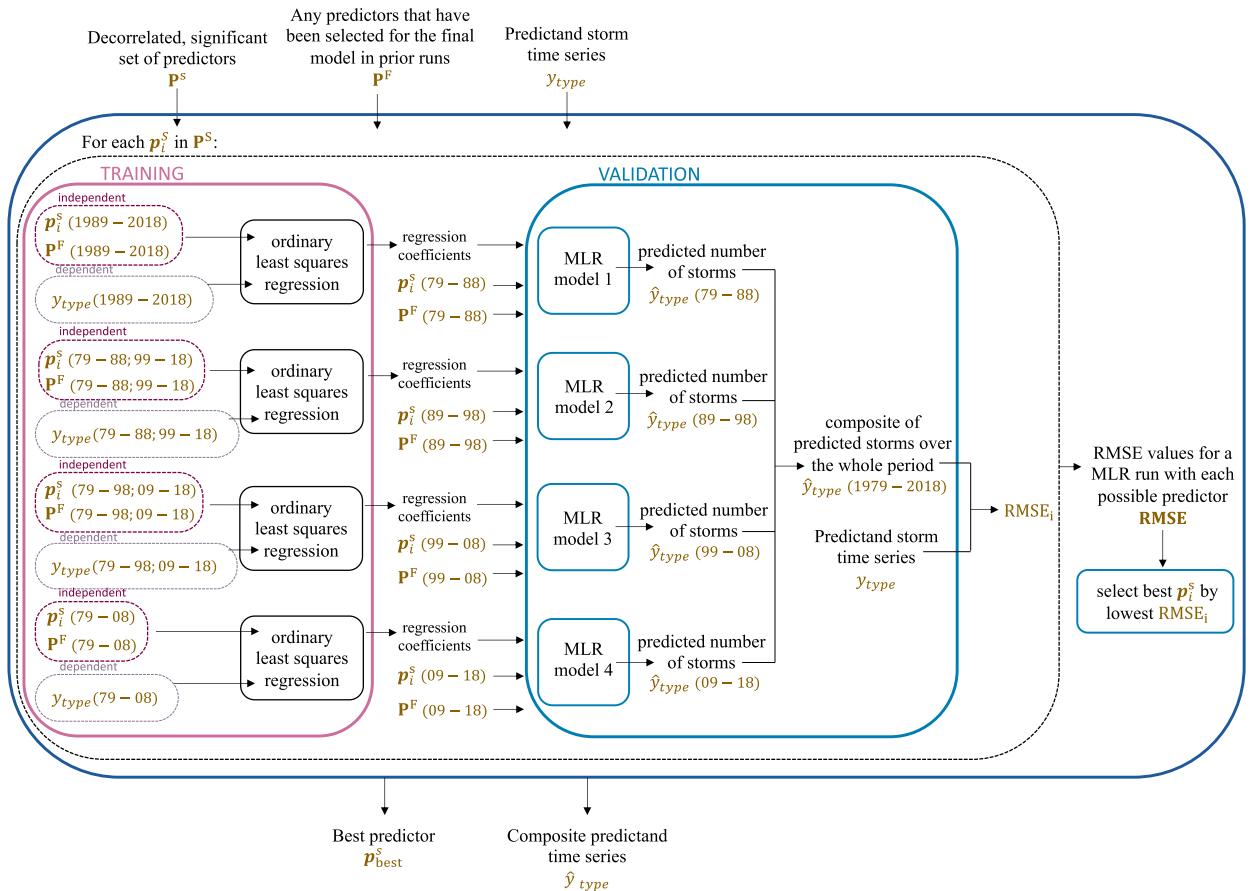


FIG. 5. Multiple linear regression model training and validation process diagram.

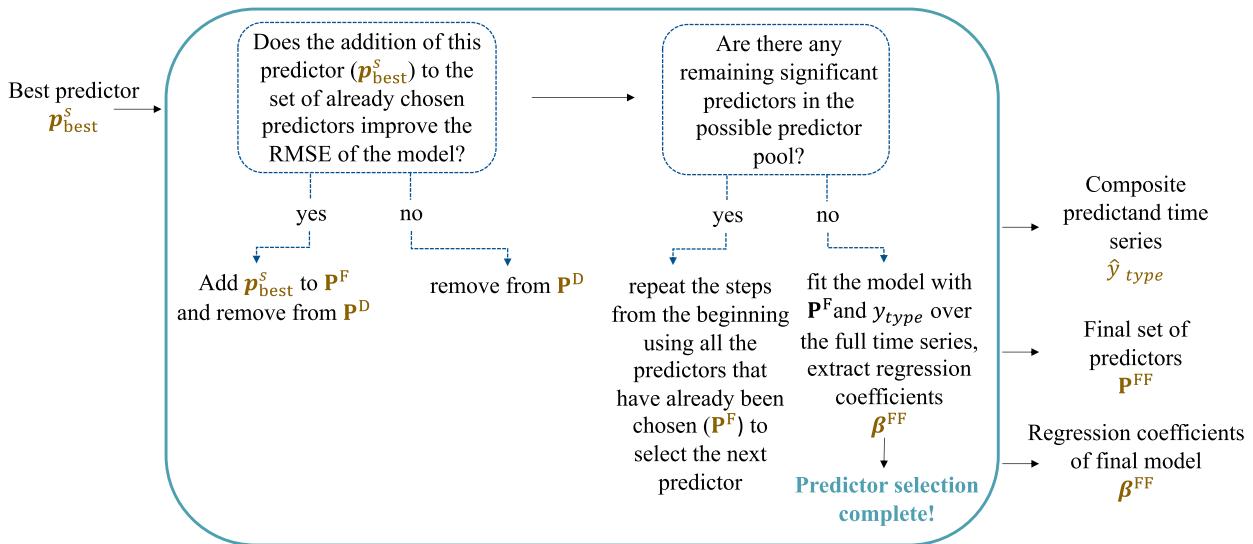


FIG. 6. Final predictor processing flowchart.

predict an outcome. For this task, we use a constant-lag MLR model. Given a predictand ( $y$ ) and a set of predictors ( $\mathbf{P}$ ) time series, the model takes the following form:

$$y_{i+\tau} = \beta_o + \boldsymbol{\beta} \mathbf{P}_i^T + \varepsilon \quad (1)$$

where  $i$  indicates current time step,  $\tau$  is some lag so that  $(i + \tau)$  is a future time,  $\boldsymbol{\beta}$  is a  $1 \times m$  vector of regression coefficients,  $\mathbf{P}_i^T$  is a  $m \times 1$  vector of predictors at time step  $i$ ,  $\beta_o$  is the intercept, and  $\varepsilon$  is the residual or the portion of  $y_{i+\tau}$  not explained by a linear relationship with  $\mathbf{P}_i$ .

With practical usage of the prospective model in mind, a short time lag is chosen. It is likely that dynamical predictors from the previous season are too far removed in time to have predictive power over the next season. We decide to use predictors that are averaged over the month of September immediately preceding the winter storm season (November–March). All the data required to forecast storm-track activity in the upcoming season would, in practice, be available before the first day of the season and the time lag is small enough that one can reasonably expect the dynamics of the system to be relevant for the storm season.

The set of predictand time series was determined in section 3. The next step is to define a set of predictors. We expect known drivers of ETCs to be useful in predicting the storms of interest. The variables investigated as possible predictors are 2 m air temperature, 500 hPa thickness, mean sea level pressure, 500 hPa geopotential height, wind at 250 hPa ( $u$  component,  $v$  component, and magnitude), and total precipitable water vapor. These variables are chosen for their relationships with baroclinic zones, moisture, upper-level divergence, and vorticity advection, and are drivers known to affect the development and propagation of ETCs. In addition, the spatial gradient of each of these variables was calculated using second-order central differencing and added to the pool to make a total of 16 potential variables.

### b. Stepwise regression and cross validation

The potential variable fields each have 3198 grid points within them. With 16 variables at over 3000 locations, there are over 50 000 possible predictors  $\mathbf{p}_i^P$  in the predictor pool,  $\mathbf{P}^P = \{\mathbf{p}_i^P | i = 1, 2, \dots, N\}$ , where  $i$  is the predictor index and  $N$  the number of possible predictors in the pool. Each possible predictor  $\mathbf{p}_i^P = [p_{i,1}^P, p_{i,2}^P, \dots, p_{i,40}^P]$  is a 40-yr time series (1979–2018) of the mean September value of the variable,  $p_{i,k}^P$  at a specific location, where  $k$  indicates the index of the year. The time series in this set are not independent. In fact, in some cases they are very highly correlated especially for possible predictors at proximate locations. Removal of this correlation will be addressed in the selection process.

With the number of possible predictors being greater than the number of observations, the MLR problem is underdetermined. It is not possible to develop an MLR using the entire predictor pool. To solve this problem, cross validation is used to select an appropriate set of predictors for the model. The process of selecting the best predictors for each storm type model from this broad pool involves three main steps: initial predictor filtering, MLR training and validation, and final predictor processing (Fig. 3). These steps are repeated until the addition of any remaining possible predictors fails to improve the model.

The initial pool of predictors used when building the model is common to all the subseries of storms (Fig. 4). However, the pool undergoes an initial filtering process as the first step of predictor selection for each subseries-specific model. We remove the correlations between the possible predictors and all predictors that have already been selected for use in the model. This prevents the inclusion of redundant information in the model and makes the final coefficients clearer to interpret. To do so, the linear relationship between the set of previously chosen predictors ( $\mathbf{P}^F$ ) and each possible predictor ( $\mathbf{p}_i^P$ ) is obtained through ordinary least squares regression. This relationship is then subtracted from the predictor time

TABLE 2. Location (lat, lon) and variable type of predictors selected for each type of subseries seasonal storm count model.

Predictand ( $y_{type}$ )	Predictor 1 ( $P_1^{FF}$ )	Predictor 2 ( $P_2^{FF}$ )	Predictor 3 ( $P_3^{FF}$ )	Predictor 4 ( $P_4^{FF}$ )
Total	Gradient TPWV 57°N, 309°E	Gradient TPWV 49°N, 327°E		
Total precip	Gradient TPWV 57°N, 310°E	Gradient T2M 29°N, 258°E		
Snow count	Gradient T2M 48°N, 332°E	Gradient V250 42°N, 331°E	Gradient TPWV 53°N, 311°E	Gradient TPWV 26°N, 277°E
Rain count	Gradient TPWV 37°N, 331°E	Gradient TPWV 63°N, 284°E	Gradient TPWV 48°N, 321°E	
Mixed count	Gradient T500 63°N, 313°E	Gradient TPWV 29°N, 293°E	Gradient T2M 40°N, 310°E	Gradient U250 29°N, 277°E
No precip count	Gradient TPWV 42°N, 279°E	Gradient T2M 29°N, 264°E	Gradient V250 40°N, 319°E	
Wind count	T2M 28°N, 291°E	Gradient U250 27°N, 281°E	T2M 49°N, 310°E	
Bomb count	Gradient TPWV 30°N, 292°E	Gradient V250 55°N, 298°E	WND250 32°N, 309°E	T500 61°N, 322°E

series ( $p_i^D$ ) to leave the independent portion of the time series ( $p_i^D$ ) for possible use as a predictor. This step is only necessary when selecting the second or greater predictor for the model. Each predictor time series in the decorrelated pool ( $P^D$ ) is then correlated with the predictand time series of that subseries model ( $y_{type}$ ). Any possible predictor that does not have a statistically significant correlation with the predictand at a 95% confidence level is removed from the pool. The output from the initial predictor filtering step is a reduced set of possible predictors ( $P^S$ ) that are independent of any previously selected predictors and significantly correlated with the predictand time series ( $y_{type}$ ).

The multiple linear regression model is then built by selecting predictors one by one from the filtered pool of possible predictors (Fig. 5). The process of selecting the best predictors from our large predictor pool is based on their correlation

with predictands. In the first round of predictor selection, the model does not yet have any predictors chosen ( $P^F = 0$ ). Thus, the process starts by testing each possible predictor ( $p_i^S$ ) in a single variable linear regression. In the second round of predictor selection, a two variable linear regression model is used with the previously chosen predictor ( $P^F = p_i^f$ ) and the possible predictor ( $p_i^S$ ) as the independent variables. In subsequent selection rounds, the number of independent variables in the multiple linear regression continues to grow with  $P^F$ . The performance of the possible predictor is evaluated based on the RMSE of the predictions made from the test MLR, which is the MLR that is generated by including the possible predictor in question together with any previously selected predictors. The test model is trained over a 30-yr period (Fig. 5, training), and then used to forecast storm activity in the remaining 10 years of our 40-yr study period (Fig. 5,

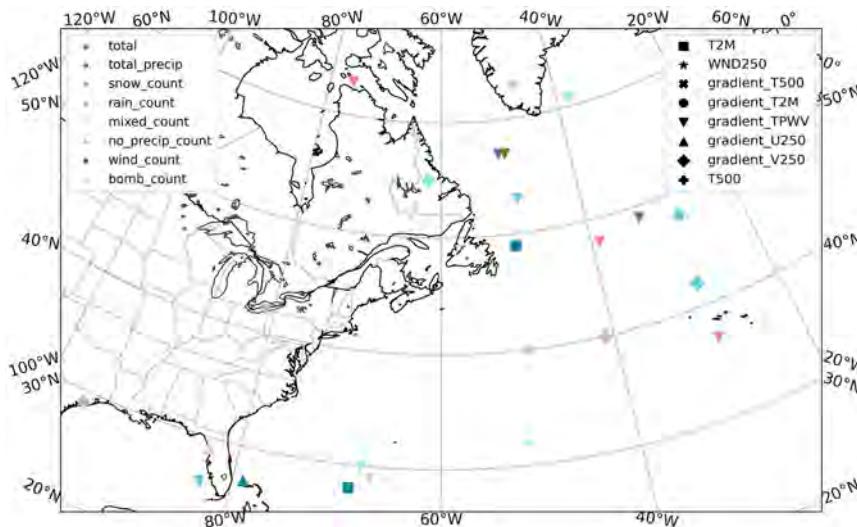


FIG. 7. Locations of predictors chosen for each seasonal storm count model. Predictand storm type is shown by color, and predictors are indicated with marker shape.

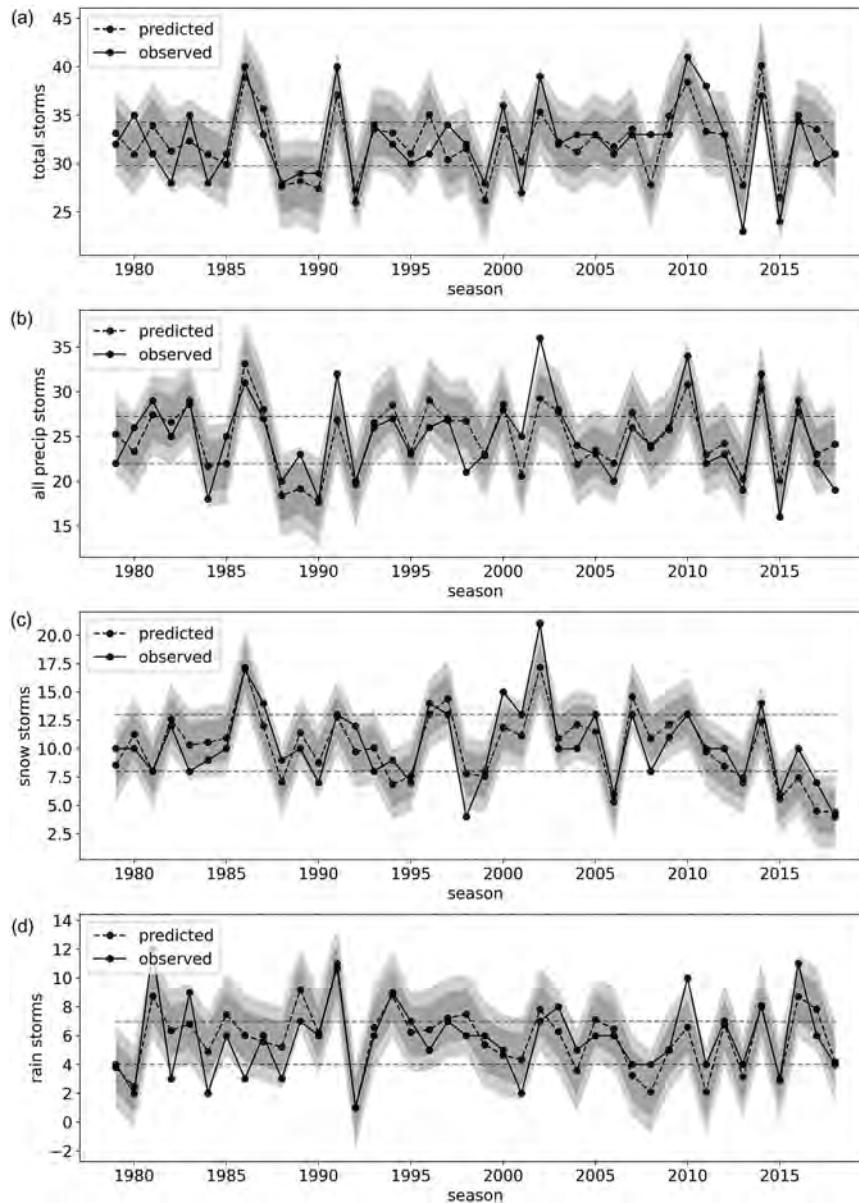


FIG. 8. Model output time series validation for (a) all storms, (b) precipitation storms, (c) snowstorms, and (d) rain storms. Forecasted storm count time series (dashed black line) shown with a 66% prediction interval indicated by the dark gray shaded area and a 90% prediction interval indicated by the entire light and dark gray shaded areas. Observed storm counts (solid) shown for comparison. For each type, dashed gray lines indicate the first and third quartiles of the observed storm tracks per season.

validation). This process is repeated four times with unique fitting and prediction periods each time, which allows for the predicted storm counts for each year in the 40-yr period ( $\hat{y}_{\text{type}}$ ) to be obtained without the forecast of any individual season being informed by the observations from that season ( $y_{\text{type}}$ ). It also prevents choosing a predictor that by chance does very well over one time period, but not over another and is therefore not a robust predictor. The 40-yr composite of forecasted storm counts over the validation time periods

( $\hat{y}_{\text{type}}$ ), and the observed storm counts over the study period ( $y_{\text{type}}$ ) are used to calculate the root mean squared error of the forecasts for each possible predictor ( $\text{RMSE}_i$ ). The possible predictor ( $\mathbf{p}_i^s$ ) that combines with the previously chosen predictors ( $\mathbf{P}^f$ ) to form the MLR with the lowest composite RMSE is chosen as the “best predictor” ( $\mathbf{p}_{\text{best}}^s$ ).

The last step is final predictor processing (Fig. 6). In this step, the algorithm either selects another predictor or finalizes the model. If the addition of the best predictor improves the

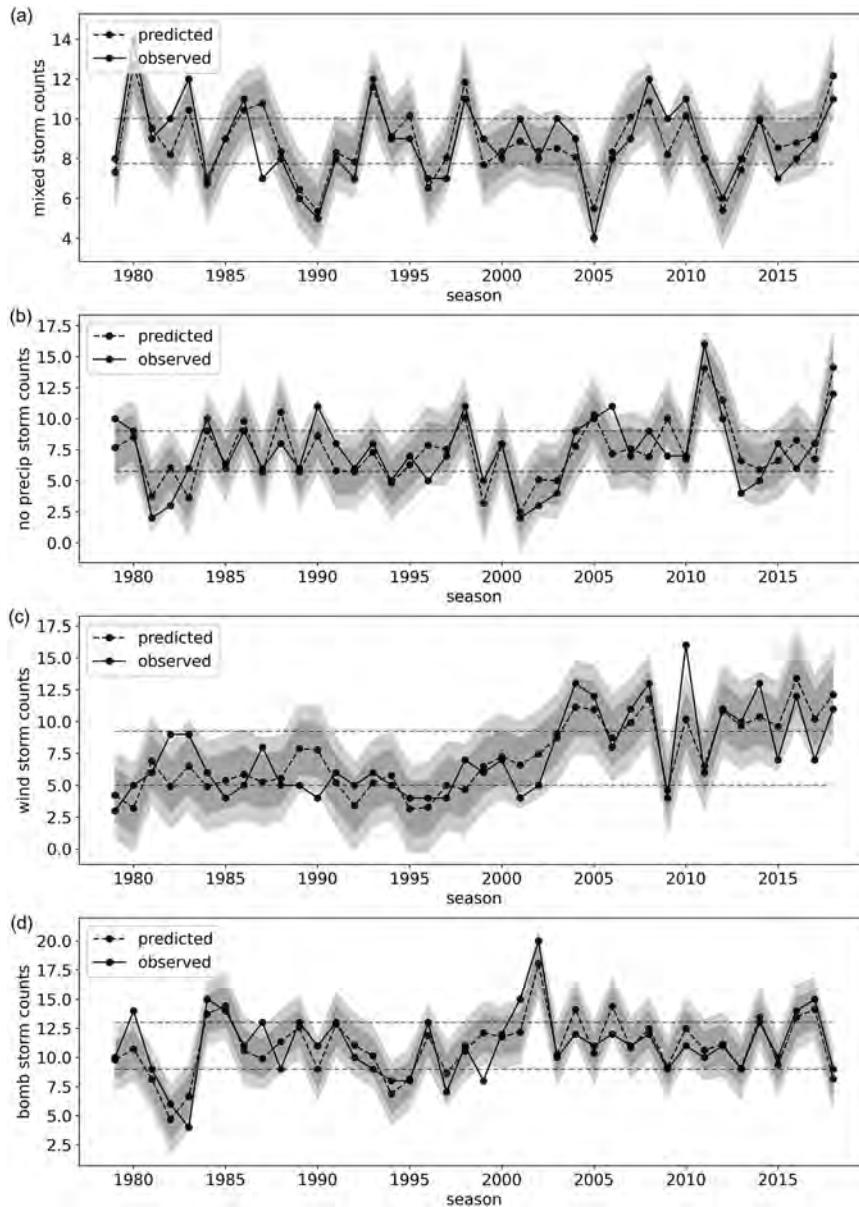


FIG. 9. As in Fig. 8, but for (a) mixed precipitation storms, (b) storms without precipitation, (c) wind storms, and (d) bomb storms.

RMSE of the model,  $\mathbf{p}_{\text{best}}^S$  is added to  $\mathbf{P}^F$ . Regardless of its usefulness, the best predictor is removed from future consideration in the possible predictor pool,  $\mathbf{P}^P$ . If any possible predictors remain in the possible predictor pool, the next step is to return to the beginning of the selection process to choose another predictor for the model. This process continues, adding new predictors to the model until eventually there are no possible predictors left. At this point, all original possible predictors that have not been added to the model cannot decrease the RMSE of the model if they are added as another predictor. The final model parameters can then be determined. The final set of predictors ( $\mathbf{P}^{FF}$ ) is the set of predictors

that have been selected until this point in the process. The final regression coefficients ( $\beta^{FF}$ ) result from using ordinary least squares to fit a model with the full time series of  $\mathbf{P}^F$  as the independent variables and  $y_{\text{type}}$  as the dependent variable.

The set of predictors ( $\mathbf{P}^{FF}$ ) chosen for each subseries model are given in Table 2. The models range in size from having two to four predictors. The most commonly selected predictor is gradient of total precipitable water vapor and the second most common predictor field is the gradient of 2 m air temperature. The frequency of these predictors is not surprising. The pattern of total precipitable water vapor in the atmosphere is a

TABLE 3. Validation results for all eight predictand models.

Predictand	RMSE	NRMSE	Corr(obs, pred)
Total	2.53	0.76	0.79
Total precip	2.65	0.72	0.81
Snow	1.76	0.58	0.86
Rain	1.61	0.74	0.78
Mixed	1.09	0.62	0.84
No precip	1.72	0.68	0.80
Wind	1.98	0.72	0.79
Bomb	1.53	0.61	0.84

reflection of upper-level flow and integrates many atmospheric fields. This makes it a likely choice in the prediction model since it combines the effects of many predictive features, such as temperature, moisture, vorticity, and midlevel winds into a single predictor. Recalling that the primary energy source of ETCs is baroclinicity explains the prevalence of 2 m temperature gradient among the final predictors. Baroclinic zones exist where there are strong spatial differences in temperature which is represented in the 2 m temperature gradient field. The locations of the predictors are quite variable, but many are located downstream from Halifax in the climatological storm track (Fig. 7).

### c. Model fitting and validation

To assess the performance of the final models (Table 2), we evaluate the composite predicted time series ( $\hat{y}_{\text{type}}$ ). For each prediction, two prediction intervals are also calculated (66% and 90% intervals). A prediction interval (PI) gives the range within which a predicted value is expected to fall based on the uncertainty of the model's ability to predict a specific value rather than simply predict the mean. It incorporates the sample uncertainty typically expressed by a confidence interval (CI) associated with the prediction of the mean in addition to the uncertainty associated with the new prediction. Due to this difference, a PI is always wider than a CI. In this model, PI is calculated for a given significance level ( $\alpha$ ) and time step ( $i$ ) according to the following equation:

$$\text{PI}_{\hat{y}} = \hat{y} \pm t_{\text{df}}^{\alpha} \hat{\sigma}_{\hat{y}_i}, \quad (2)$$

where  $\hat{y}$  is the predicted value of storms in the season;  $t_{\text{df}}^{\alpha}$  is the critical  $t$  value for the specified significance level and the degrees of freedom (df) of the model; and  $\hat{\sigma}_{\hat{y}_i}$  is the

standard error of the  $i$ th predicted value (Helwig 2017). The standard error is defined mathematically as follows:

$$\hat{\sigma}_{\hat{y}} = \sqrt{\text{MSE}\{1 + (\mathbf{P}_i^{FF})^T [(\mathbf{P}^{FF})^T \mathbf{P}^{FF}]^{-1} \mathbf{P}_i^{FF}\}} \quad (3)$$

where MSE is the mean squared error and  $\mathbf{P}_i^{FF}$  is the set of predictor values at the  $i$ th time step.

To validate the model, the 40-yr forecasted storm activity time series and prediction intervals are compared with the observed storm activity over those 40 seasons (Figs. 8 and 9). We use RMSE, normalized root mean squared error (NRMSE), and cross correlation are used to quantify the fit of each subseries prediction model (Table 3). All predicted model time series are correlated with their corresponding observation time series at a value of at least 0.78. Our best model as determined by cross correlation is the snow storm model at a correlation of 0.86. The lowest RMSE is recorded by the mixed storm model. However, to compare RMSE across models, the value should be normalized. The NRMSE is calculated by dividing the RMSE by the standard deviation of each subseries ( $\sigma_{\text{type}}$ ). When this is taken into account, one can see that while results are comparable across models, the snow storms model is best with the NRMSE equal to 0.58. Overall, the models forecast seasonal storm activity with some skill showing the chosen predictors have predictive value.

For future use of these models, we produce eight final equations for forecasting winter seasonal storm activity in the Halifax area. The regression coefficients are obtained from fitting the predictors and predictand time series over the whole 40-yr study period (Table 4). For six of the eight MLRs, the intercept ( $\beta_0^{FF}$ ) is similar to the mean storm activity value for that storm type. The value of the intercepts for the wind and bomb storms, however, are much lower to account for the positive trend in these time series.

## 5. Forecast application and results

### a. Probabilistic forecast framework

A forecast made directly from the output of the MLRs has considerable uncertainty as shown by the prediction intervals on the model outputs. While this precludes the use of the models for deterministic forecasting, the model outputs can still be utilized for a probabilistic forecast. The framework for such usage is outlined in this subsection. Rather than forecast

TABLE 4. Model parameters for each type of subseries forecast model.

Predictand	Intercept ( $\beta_0^{FF}$ )	Coef 1 ( $\beta_1^{FF}$ )	Coef 2 ( $\beta_2^{FF}$ )	Coef 3 ( $\beta_3^{FF}$ )	Coef 4 ( $\beta_4^{FF}$ )
Total	+42.0	-10.9	-3.48	—	—
Total precip	+29.1	-13.8	-9.02	—	—
Snow	+20.9	-20.4	-4.13	+3.24	+1.42
Rain	+8.86	-3.77	+4.04	-2.00	—
Mixed	+11.3	-0.0350	-1.36	+3.37	+0.966
No precip	+6.86	+2.60	-4.34	-2.57	—
Wind	-1825.00	+4.96	+2.43	+1.23	—
Bomb	-176.00	+2.35	-2.95	-0.547	+0.00357

TABLE 5. Classifications of storm activity. The upper threshold ( $Q_{type}^3$ ) separating above average and below average is the 75th percentile of the observed storm time series or the third quartile and the lower threshold ( $Q_{type}^1$ ), which separates average and below average is the 25th percentile of the or the first quartile.

Storm-track activity category	Range
Above average	$\hat{y} > Q_{type}^3$
Average	$Q_{type}^1 < \hat{y} < Q_{type}^3$
Below average	$\hat{y} < Q_{type}^1$

an exact number of storms, we create categories of storm-track activity that give context to the forecast outputs. These categories are “above-average,” “average,” and “below-average” activity (Table 5). The threshold values that separate these categories are defined based on the quartiles of the observed storm time series for each storm type. The upper threshold is the third quartile ( $Q_{type}^3$ ) and the lower threshold is the first quartile ( $Q_{type}^1$ ; illustrated with dashed lines in Figs. 8 and 9). The above-average storm season range lies above the upper threshold and the below-average category lies below the lower threshold. The range considered average lies between the first and third quartiles, i.e., the interquartile range.

The probabilistic forecast determines the likelihood that the number of storms in a given season falls within each of these three ranges. To do so, the prediction interval (PI) is utilized. The PI follows a Student’s  $t$ -distribution probability density function, which is centered on the forecasted value ( $\hat{y}_{type}$ ) and dependent on the degrees of freedom of the model. To determine the likelihood that the number of storms in a storm season will fall within each of the three activity categories, we assess the percent of the forecast distribution that falls above, between, and below the thresholds. This gives the probability as a percentage for each

possible activity level (Fig. 10). The numerical probabilities are also categorized. The language of our proposed probabilistic forecast follows the IPCC (2014) probability language (Table 6).

The resulting language of the forecast would be, for example, it is *very unlikely* that Halifax has a *below-average* number of *rain storms* this season. The italicized text indicates the parts of the forecast that would change based on the model output each season and the predictand being examined. Mathematically, the example above is stating that 90% of the probability density function of the forecast lies above the third quartile of rain storms recorded over the 40-yr study period.

b. Halifax winter storm forecast

This probabilistic framework is applied to forecast the likelihood of above-average, average, and below-average storm activity in the 2019/20 and 2020/21 storm seasons, which were not used in any step of the model development process (Tables 7 and 9). The forecasts are also compared with the observed storm activity in those seasons (Tables 8 and 10). The probabilistic forecast shows some skill with an accurate activity category being predicted half the time. For three of the forecasts categorized as misses (2019 total precip, 2020 precip, and 2020 snow), the worded forecast said the probability was “about as likely as not” for two activity categories. The observed activity did fall within one of these two categories; however, it was not numerically the most probable of the two. Even though the observed activity category was not the category with the highest numerical probability, the probabilistic framework communicated an accurate forecast.

6. Discussion

We have proposed a probabilistic forecast model for winter extratropical cyclones that affect a given location of interest

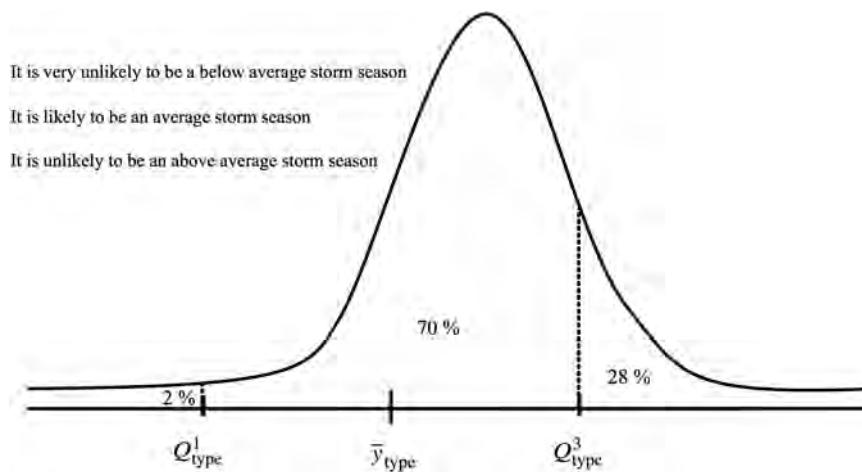


FIG. 10. Probabilistic forecast schematic for an example storm season. The probability density function of the forecast is illustrated with a normal distribution curve and compared with the storm activity categories. The percentage of the area under the curve that lies within each category is used to make the written forecast statements to the top left of the curve.

TABLE 6. Relationship between probability language and statistical likelihood.

Probability terminology	Numerical probability
Very likely	≥90%
Likely	67%–89%
About as likely as not	34%–66%
Unlikely	11%–33%
Very unlikely	≤10%

separated into categories based on precipitation type, wind speed, and deepening rate. Due to the high degree of spatial and temporal variability of winter ETCs, the predictand field was simplified to be regionally focused on Halifax, NS, allowing the temporal variability in the field to be quantified and predicted without complications of spatial variability. We developed eight MLR forecast models, one to forecast all storms and one each for the seven subsets of unique storm types. Using the September mean values of known ETC drivers, the most skilled predictors were methodically selected for each model. Most often, the best predictors were found to be the gradient of total precipitable water vapor and 2 m air temperature and its gradient field. The best model as determined by  $R^2$  is the snow storm model ( $R^2 = 0.73$ ). The seasonal forecast is built upon this set of eight moderately skilled, storm-type-specific MLR models. Through the whole development process, the intent was to create a forecast that communicates information of value to the general public. This was ultimately achieved with probabilistic descriptions that compare projected activity and the uncertainty in the projections to usual winter storm season activity in clear, digestible statements. While this project focused on Halifax, the framework presented could, in principle, be applied at any location.

The ETC research field is currently lacking in regionally nuanced seasonal forecasting of ETCs and in forecasts that specify storm characteristics, such as precipitation type or severity. Existing seasonal forecast models based on climate models, EOF-based predictions, and teleconnection-based predictions

have limited spatial resolution and limited information about storm characteristics. This makes it difficult to deliver results that are meaningful to the public when applying them to a specific location or region. The regionally specific, multitype probabilistic forecast model presented here addresses the need for a focused and digestible ETC forecast by providing skillful and detailed forecasts of ETC activity differentiated according to storm impacts. Existing models make broad sweeping statements such as, “The storm track will experience a northward shift this winter season.” Our model adds spatial and impact-related detail to make more specific and applicable forecasts such as, “It is likely to be an above-average storm season this winter. It is highly likely to be above average in terms of snow storms, likely to be below average for rain storms, and highly likely to be average in numbers of wind storms and bomb cyclones.” Focusing the spatial extent of the predictand and subtyping storms according to impacts has allowed for increased detail and enhanced usefulness for the general public.

Accurate and timely seasonal forecasting of ETCs can increase disaster preparedness and mitigate human and economic loss. Our probabilistic forecast gives a practical projection of likely winter storm season characteristics one month ahead (October) of the upcoming season (November–March) to allow for ample preparation to minimize losses. The possible predictors were chosen based on atmospheric variables that are known to drive storm-track activity through concurrent temporal relationships. However, the relationships are applied in this MLR at a specified lag. This was a practical choice made primarily on the basis of usability. The skill exhibited by the predictors demonstrates that a nonconcurrent relationship does exist; however, the physical basis for the mechanisms or processes underpinning those relationships are not developed. A further investigation into the predictor–predictand relationships at the lag used in the forecast model would give better insights into the mechanisms at play and build a stronger theoretical basis for the model. Such an analysis might reveal that predictors at a slightly different time lag have a stronger physical connection or that spatial averaging or other filtering of

TABLE 7. Probabilistic forecast of 2019 winter storm activity for eight subseries. The most probable activity category for each type is italicized. Boldface indicates when the observed storm activity falls within the forecasted highest probability category.

	Probability of above average	Probability of average	Probability of below average
Total	Unlikely 0.2847	About as likely as not <b>0.5930</b>	Unlikely 0.1223
Total precip	About as likely as not <i>0.6038</i>	About as likely as not 0.3789	Very unlikely 0.0173
Snow	Unlikely 0.1437	Likely <b>0.8006</b>	Very unlikely 0.0557
Rain	Unlikely 0.1114	About as likely as not <b>0.6081</b>	Unlikely 0.2804
Mixed	Very unlikely 0.0018	Very unlikely 0.0820	Very likely <b>0.9162</b>
No precip	Likely <b>0.8827</b>	Unlikely 0.1147	Very unlikely 0.0026
Wind	Very likely <i>0.9960</i>	Very unlikely 0.0040	Very unlikely 0.0000
Bomb	Likely <i>0.8008</i>	Unlikely 0.1981	Very unlikely 0.0010

TABLE 8. Validation of 2019 winter season predictions. Comparison of predicted ( $y_{\text{hat}}$ ) and observed ( $y_{\text{obs}}$ ) storm counts for each type with accuracy of the prediction given in the bottom row. The observed storm counts must fall within the most probable predicted activity category for the forecast to be considered a hit.

	Total					No		
	Total	precip	Snow	Rain	Mixed	precip	Wind	Bomb
$y_{\text{hat}}$	32.78	27.97	11.01	4.96	5.9	11.23	16.09	14.37
$y_{\text{obs}}$	32.0	22.0	10.0	5.0	7.0	10.0	8.0	7.0
	Hit	Miss	Hit	Hit	Hit	Hit	Miss	Miss

the predictor field gives more robust predictor–predictand relationships. Here we chose a purely statistical approach which leverages lagged relationships between storm activity and large-scale predictor fields. Another promising avenue for future work is combining dynamical and statistical modeling. Incorporating, for example, a dynamical seasonal forecast could allow for a model that instead leverages concurrent (zero lag) predictor–predictand relationships, which may be stronger than lagged relationships and therefore lead to a more skilled model. Though this technique was considered, we opted for a purely statistical model 1) to preserve simplicity and maintain usability, and 2) because a preliminary lagged cross-correlation analysis shows no significantly different predictor–predictand correlations at seasonal lags.

Another limitation can be found in the lack of intermodel relationships. Since by definition there are mathematical relationships between the categories of storms (e.g., precip + no precip = total storms), the models could be constrained to preserve these relationships. This might mean letting the total storms forecast simply be the sum of the precipitation and no precipitation forecasts, rather than forecasting total storms by its own individual model. Because the actual use of the model outputs is in a probabilistic forecast, the outputs do not need to be perfectly congruent. However, if an exact quantitative forecast of the storm types was derived from the MLRs, the number of storms in an aggregate category (e.g., precip storms) should be equal to the sum of its constituents (e.g.,

snow, rain, and mixed storms). If this is the desired result, a logistic regression may be a more appropriate technique.

Finally, the base of the forecasting framework is a set of MLRs that have room for improvement. Efforts toward further refinement of the predictor pool, more sophisticated predictor selection techniques, or even an alternative base model to the MLR could be expected to improve the deterministic forecast component of the system. Since our storm time series are counts (i.e., positive integers) an alternative model that one might consider is Poisson regression. We find that the use of Poisson regression does not in general provide improved predictions over linear regression (not shown) and also that the likelihood of our model producing nonsense predictions (i.e., negative storm counts) is small. Further, the calculation of a prediction interval for Poisson regression, which is critical to the outcomes of this study, presents a major challenge as there is no unique way to do so and those that exist are cumbersome to implement (Kim et al. 2022). Though we acknowledge the potential of a superior alternative, we have checked the assumptions of linear regression and confirmed they hold for our data making linear regression a suitable choice (see the appendix). For both ease of use and practicality we found that linear regression worked well for us in this study.

Model improvement could also come in the predictor selection. Here we used stepwise regression with cross-validation in order to select our predictors. Stepwise regression, in particular classic forward selection, is known to have some undesirable properties including, e.g., small standard errors of the parameter estimates making CIs on the parameters too narrow and  $R^2$  values that are biased high (Harrell 2001). Forward selection is known to be especially problematic with high levels of collinearity in predictors. We took care to avoid this pitfall by removing correlations between predictors already chosen for the model and any potential predictors being considered for addition. Further, the use of cross-validation was incorporated to protect against overfitting, which is another known downfall of the forward selection process. Our resulting model predictions have good correlations (~0.8 on

TABLE 9. As in Table 7, but for the 2020 winter season.

	Probability of above average	Probability of average	Probability of below average
Total	Unlikely 0.1437	About as likely as not 0.6007	Unlikely 0.2555
Total precip	Very unlikely 0.0136	About as likely as not 0.3416	About as likely as not 0.6447
Snow	About as likely as not 0.5567	About as likely as not 0.4395	Very unlikely 0.0037
Rain	Very unlikely 0.0769	About as likely as not <b>0.5437</b>	About as likely as not 0.3794
Mixed	Very unlikely 0.0001	Very unlikely 0.0178	Very likely 0.9820
No precip	Unlikely 0.1486	About as likely as not 0.6363	Unlikely 0.2151
Wind	Likely <b>0.7919</b>	Unlikely 0.2040	Very unlikely 0.0041
Bomb	Very unlikely 0.0160	About as likely as not <b>0.5679</b>	About as likely as not 0.4161

TABLE 10. As in Table 8, but for the 2020 season.

	Total				No			
	Total	precip	Snow	Rain	Mixed	precip	Wind	Bomb
$y_{\text{hat}}$	31.46	20.98	13.27	4.53	5.11	7.15	11.02	9.35
$y_{\text{obs}}$	41.0	27.0	10.0	6.0	11.0	14.0	15.0	10.0
	Miss	Miss	Miss	Hit	Miss	Miss	Hit	Hit

average) but these are not suspiciously high such that they would indicate overfitting. The number of predictors selected for the models is at most four which, considering the size of the potential predictor pool ( $\mathcal{O} \sim 10\,000$ ), suggests the method is appropriately selective. Other methods could be used to alleviate these issues in different ways, such as Least Absolute Shrinkage and Selection Operator (LASSO) regression which minimizes the residual sum of squares with one key constraint placed on the model. The constraint is a penalty term which dictates the sum of the coefficients must be less than a constant. This effectively eliminates predictors by forcing some coefficients to zero (Tibshirani 1996). While the LASSO technique is a powerful tool, we decided that exploring its use was outside the scope of this study. Future work focused on improving the predictor selection process or the base model could significantly improve the forecast system as a whole. Nonetheless, the larger framework presented here is well designed.

We established the structure of a probabilistic forecast that delivers estimates of storm type and frequency for the upcoming winter season with 1 month of lead time. It uses publicly available ERA5 data, has a practical lead time, can be readily applied at any location, and gives contextual, probabilistic forecast statements in a manner that is digestible for the general public. By leveraging the probability distribution of the model output and the typical characteristics of a winter storm season in delivering our forecast, we demonstrated a valuable method of incorporating both context and mathematical uncertainty in an accessible way. Publishing such a forecast before an upcoming winter season has great potential to mitigate losses, improve financial planning, and enable better overall preparedness for the impending trials of the winter ahead.

*Acknowledgments.* The authors would like to acknowledge ECMWF for providing ERA5 data. The authors acknowledge funding from the National Sciences and Engineering Research Council of Canada Discovery Grant RGPIN-2018-05255, which made this research possible. The authors would like to recognize the invaluable contributions of the late Dr. Keith Thompson who generously provided his feedback and wisdom throughout the course of the project.

*Data availability statement.* The work here relies on ERA5 data that may be downloaded from <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview>

and <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressure-levels?tab=form> and ECCC station data that can be found at [https://climate.weather.gc.ca/historical\\_data/search\\_historic\\_data\\_e.html](https://climate.weather.gc.ca/historical_data/search_historic_data_e.html). Alternatively the ECCC data can be downloaded following the instructions found here: <https://drive.google.com/drive/folders/1WJCDU34c60ffOnG4rv5EPZ4IhhW9vZH> using station IDs 6358 and 50620, which correspond to climate IDs 8202250 and 8202251. The processing of this data and development of the model described here can be found at <https://github.com/bekahcav/WinterETCSeasonalModel>. This study has not generated any new data.

## APPENDIX

### Assumptions of Linear Regression

We have checked assumptions of linear regression as follows:

- (i) **Linearity:** there exists a linear relationship between the independent variable  $x$  and the dependent variable  $y$ . We have plotted each predictor against the predictand. We visually assessed that linear relationships are present (not shown).
- (ii) **Independence of errors:** there is no correlation between consecutive residuals. Autocorrelation functions for residuals in all models fits show no statistically significant correlations at nonzero lags ( $\alpha = 0.05$ ).
- (iii) **Homoskedasticity:** the residuals have constant variance and do not depend on the value of the predictor. This was assessed using the White test (White 1980). The null hypothesis (that the errors are homoskedastic) could not be rejected for all subseries ( $p = 0.05$ ; Table A1).
- (iv) **Expectation of errors is zero.** We used a  $t$  test to assess this and the null hypothesis (that the expected value of the sample is equal to zero) was not rejected for all subseries ( $p = 0.05$ ; Table A2).

TABLE A1. Results of White test for homoskedasticity for each storm series.

Storm type	$p$ value
Total	0.5149
Total precip	0.9855
Snow	0.0619
Rain	0.381
Mixed	0.9318
No precip	0.5443
Wind	0.7406
Bomb	0.9888

TABLE A2. Results of test for zero mean of errors.

Storm type	Mean of residuals	<i>p</i> value
Total	−0.0211	0.9586
Total precip	0.0497	0.9072
Snow	−0.0996	0.7248
Rain	0.0851	0.743
Mixed	0.0045	0.9797
No precip	0.0252	0.9278
Wind	−0.1099	0.7308
Bomb	−0.0672	0.7855

## REFERENCES

- Befort, D. J., and Coauthors, 2019: Seasonal forecast skill for extratropical cyclones and windstorms. *Quart. J. Roy. Meteor. Soc.*, **145**, 92–104, <https://doi.org/10.1002/qj.3406>.
- Chelton, D. B., M. G. Schlax, and R. M. Samelson, 2011: Global observations of nonlinear mesoscale eddies. *Prog. Oceanogr.*, **91**, 167–216, <https://doi.org/10.1016/j.pocean.2011.01.002>.
- DeGaetano, A. T., M. E. Hirsch, and S. J. Colucci, 2002: Statistical prediction of seasonal East Coast winter storm frequency. *J. Climate*, **15**, 1101–1117, [https://doi.org/10.1175/1520-0442\(2002\)015<1101:SPOSEC>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<1101:SPOSEC>2.0.CO;2).
- Feng, X., B. Huang, and D. M. Straus, 2019: Seasonal prediction skill and predictability of the Northern Hemisphere storm track variability in Project Minerva. *Climate Dyn.*, **52**, 6427–6440, <https://doi.org/10.1007/s00382-018-4520-9>.
- Gray, W. M., 1984: Atlantic seasonal hurricane frequency. Part II: Forecasting its variability. *Mon. Wea. Rev.*, **112**, 1669–1683, [https://doi.org/10.1175/1520-0493\(1984\)112<1669:ASHFPI>2.0.CO;2](https://doi.org/10.1175/1520-0493(1984)112<1669:ASHFPI>2.0.CO;2).
- Harrell, F. E., 2001: *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. Springer, 568 pp.
- Helwig, N. E., 2017: Multiple linear regression. University of Minnesota Tech. Note, 88 pp., <http://users.stat.umn.edu/~helwig/notes/mlr-Notes.pdf>.
- Hersbach, H., and Coauthors, 2018a: ERA5 hourly data on pressure levels from 1979 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), accessed 1 April 2020, <https://doi.org/10.24381/cds.bd0915c6>.
- , and Coauthors, 2018b: ERA5 hourly data on single levels from 1979 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), accessed 1 April 2020, <https://doi.org/10.24381/cds.adbb2d47>.
- Hoskins, B. J., and K. I. Hodges, 2002: New perspectives on the Northern Hemisphere winter storm tracks. *J. Atmos. Sci.*, **59**, 1041–1061, [https://doi.org/10.1175/1520-0469\(2002\)059<1041:NPOTNH>2.0.CO;2](https://doi.org/10.1175/1520-0469(2002)059<1041:NPOTNH>2.0.CO;2).
- IPCC, 2014: *Climate Change 2014: Synthesis Report*. Cambridge University Press, 151 pp.
- Kim, T., B. Lieberman, G. Luta, and E. A. Peña, 2022: Prediction intervals for Poisson-based regression models. *Wiley Interdiscip. Rev.: Comput. Stat.*, **14**, e1568, <https://doi.org/10.1002/wics.1568>.
- Klotzbach, P. J., and W. M. Gray, 2009: Twenty-five years of Atlantic Basin seasonal hurricane forecasts (1984–2008). *Geophys. Res. Lett.*, **36**, L09711, <https://doi.org/10.1029/2009GL037580>.
- , and Coauthors, 2019: Seasonal tropical cyclone forecasting. *Trop. Cyclone Res. Rev.*, **8**, 134–149, <https://doi.org/10.1016/j.tcr.2019.10.003>.
- , L.-P. Caron, and M. Bell, 2020: A statistical/dynamical model for North Atlantic seasonal hurricane prediction. *Geophys. Res. Lett.*, **47**, e2020GL089357, <https://doi.org/10.1029/2020GL089357>.
- Massey, N., 2012: Feature tracking on the hierarchical equal area triangular mesh. *Comput. Geosci.*, **44**, 42–51, <https://doi.org/10.1016/J.CAGEO.2012.03.012>.
- Neu, U., and Coauthors, 2013: IMILAST: A community effort to intercompare extratropical cyclone detection and tracking algorithms. *Bull. Amer. Meteor. Soc.*, **94**, 529–547, <https://doi.org/10.1175/BAMS-D-11-00154.1>.
- Oliver, E. C. J., T. J. O’Kane, and N. J. Holbrook, 2015: Projected changes to Tasman Sea eddies in a future climate. *J. Geophys. Res. Oceans*, **120**, 7150–7165, <https://doi.org/10.1002/2015JC010993>.
- Perkins, W. A., and G. Hakim, 2020: Linear inverse modeling for coupled atmosphere-ocean ensemble climate prediction. *J. Adv. Model. Earth Syst.*, **12**, e2019MS001778, <https://doi.org/10.1029/2019MS001778>.
- Pinto, J. G., S. Ulbrich, T. Economou, D. B. Stephenson, M. K. Karremann, and L. C. Shaffrey, 2016: Robustness of serial clustering of extratropical cyclones to the choice of tracking method. *Tellus*, **68A**, 32204, <https://doi.org/10.3402/tellusa.v68.32204>.
- Raible, C. C., P. M. Della-Marta, C. Schwierz, H. Wernli, and R. Blender, 2008: Northern Hemisphere extratropical cyclones: A comparison of detection and tracking methods and different reanalyses. *Mon. Wea. Rev.*, **136**, 880–897, <https://doi.org/10.1175/2007MWR2143.1>.
- Sanders, F., and J. R. Gyakum, 1980: Synoptic-dynamic climatology of the “bomb.” *Mon. Wea. Rev.*, **108**, 1589–1606, [https://doi.org/10.1175/1520-0493\(1980\)108<1589:SDCOT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1980)108<1589:SDCOT>2.0.CO;2).
- Tibshirani, R., 1996: Regression shrinkage and selection via the lasso. *J. Roy. Stat. Soc.*, **58B**, 267–288, <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>.
- White, H., 1980: A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, **48**, 817–838, <https://doi.org/10.2307/1912934>.
- Yang, X., and Coauthors, 2015: Seasonal predictability of extratropical storm tracks in GFDL’s high-resolution climate prediction model. *J. Climate*, **28**, 3592–3611, <https://doi.org/10.1175/JCLI-D-14-00517.1>.