

# Improved Diagnosis of Precipitation Type with LightGBM Machine Learning

HAOYU (RICHARD) ZHUANG<sup>a</sup>, FLAVIO LEHNER<sup>a,b,c</sup> AND ARTHUR T. DEGAETANO<sup>a,d</sup>

<sup>a</sup> *Department of Earth and Atmospheric Sciences, Cornell University, Ithaca, New York*

<sup>b</sup> *Climate and Global Dynamics Laboratory, National Center for Atmospheric Research, Boulder, Colorado*

<sup>c</sup> *Polar Bears International, Bozeman, Montana*

<sup>d</sup> *Northeast Regional Climate Center, Cornell University, Ithaca, New York*

(Manuscript received 11 July 2023, in final form 29 November 2023, accepted 21 January 2024)

**ABSTRACT:** Existing precipitation-type algorithms have difficulty discerning the occurrence of freezing rain and ice pellets. These inherent biases are not only problematic in operational forecasting but also complicate the development of model-based precipitation-type climatologies. To address these issues, this paper introduces a novel light gradient-boosting machine (LightGBM)-based machine learning precipitation-type algorithm that utilizes reanalysis and surface observations. By comparing it with the Bourgoiuin precipitation-type algorithm as a baseline, we demonstrate that our algorithm improves the critical success index (CSI) for all examined precipitation types. Moreover, when compared with the precipitation-type diagnosis in reanalysis, our algorithm exhibits increased F1 scores for snow, freezing rain, and ice pellets. Subsequently, we utilize the algorithm to compute a freezing-rain climatology over the eastern United States. The resulting climatology pattern aligns well with observations; however, a significant mean bias is observed. We interpret this bias to be influenced by both the algorithm itself and assumptions regarding precipitation processes, which include biases associated with freezing drizzle, precipitation occurrence, and regional synoptic weather patterns. To mitigate the overall bias, we propose increasing the precipitation cutoff from 0.04 to 0.25 mm h<sup>-1</sup>, as it better reflects the precision of precipitation observations. This adjustment yields a substantial reduction in the overall bias. Finally, given the strong performance of LightGBM in predicting mixed precipitation episodes, we anticipate that the algorithm can be effectively utilized in operational settings and for diagnosing precipitation types in climate model outputs.

**SIGNIFICANCE STATEMENT:** Freezing rain can have significant impacts on transportation and infrastructure, making accurate prediction of precipitation types crucial. In this study, we use a machine learning method known as LightGBM to predict precipitation types. We show that the new algorithm performs better than the existing methods for all precipitation types examined. Additionally, we compute a freezing-rain climatology over the eastern United States. Although the resulting climatology pattern corresponds well to observations, the algorithm overpredicts freezing-rain occurrence. We argue that this bias can be substantially reduced by increasing the precipitation cutoff from 0.04 to 0.25 mm h<sup>-1</sup>. Overall, this work highlights the potential of the LightGBM algorithm for both weather forecasting and diagnosing precipitation types in climate models.

**KEYWORDS:** Surface observations; Forecast verification/skill; Hindcasts; Classification; Decision trees; Machine learning

## 1. Introduction

Accurate prediction of winter precipitation type is crucial for both operational forecasting and understanding its change in a warming climate (Stewart et al. 2015). Freezing rain, for instance, is a hazardous weather phenomenon that can cause severe socioeconomic impacts (DeGaetano 2000). Although many precipitation-type algorithms can accurately diagnose rain and snow, for example, Bourgoiuin (2000), they do not perform as well with freezing rain and ice pellets (Reeves 2016). Compounded by the chaotic nature of the atmosphere, predicting precipitation type remains a challenge even at short lead times (Ralph et al. 2005).

Precipitation-type algorithms roughly fall into two categories. Explicit algorithms output precipitation types based on a cloud microphysics scheme, exemplified by the utilization of

the Thompson et al. (2008) microphysics scheme for precipitation-type forecasts in the High-Resolution Rapid Refresh model (Ikeda et al. 2013; Benjamin et al. 2016). In contrast, implicit algorithms predict precipitation type based on vertical atmospheric profiles (Ramer 1993; Bourgoiuin 2000; Reeves et al. 2016; Birk et al. 2021). There has been a notable shift toward the utilization of machine learning methods for implicit precipitation-type diagnosis. For example, Das et al. (2022) combined Global Precipitation Measurement data with six machine learning algorithms, such as XGBoost and 1D convolutional neural networks, to predict five precipitation types. Various studies have also leveraged the power of random forest to diagnose rain, snow, and sleet by using one or more of reanalysis products, observed thermodynamic profiles, and radar data (Pórolniczak et al. 2021; Shin et al. 2022; Lang et al. 2023).

Despite these advancements, it is important to note several limitations of the current methods. First, the implicit algorithms are developed on a small amount of sounding data (Bourgoiuin 2000; Reeves et al. 2016; Birk et al. 2021), and

Corresponding author: Haoyu (Richard) Zhuang, hz542@cornell.edu

only [Reeves et al. \(2016\)](#) attempted to account for model uncertainty, that is, the fact that models can only simulate the vertical atmospheric profiles with some inherent error. However, after introducing errors to the sounding profile, [Reeves et al. \(2016\)](#) reported a marked decrease in freezing-rain hit rate. This result points to an increasing need for a precipitation-type algorithm that performs well in models.

Second, while the adoption of machine learning methods for precipitation-type diagnosis enhances statistical robustness, none of the prior works aim to diagnose freezing rain ([Pótroniczak et al. 2021](#); [Shin et al. 2022](#); [Lang et al. 2023](#)). And yet, the research community has increasingly recognized the importance of understanding freezing-rain changes in a warming world ([IPCC 2022](#); [Cheng et al. 2007](#); [Lambert and Hansen 2011](#); [Bresson et al. 2017](#); [St-Pierre et al. 2019](#); [McCray et al. 2022](#)). Hence, an ideal precipitation-type algorithm should possess the capability not only to predict precipitation types in operational forecasts but also to diagnose freezing rain within the context of climate models.

Third, although researchers have successfully obtained a realistic freezing-rain climatology by combining implicit precipitation-type algorithms and dynamically downscaled regional climate models, for example, [McCray et al. \(2022\)](#), we argue that bulk statistics are not conclusive evidence of algorithm performance. For instance, although a precipitation-type algorithm may derive a realistic freezing-rain climatology from reanalysis products, there can be multiple compensating biases that produce a realistic climatology. Understanding the biases in models is therefore important for interpreting the derived climatology and its relationship with observations.

Our goal in this paper is to develop an accurate, machine learning-based precipitation-type algorithm that addresses these issues. [Section 2](#) describes the dataset used in this project, informs the reader about the light gradient-boosting machine (LightGBM) algorithm used to predict precipitation type, and discusses two applications of the algorithm. [Section 3](#) presents the results of these applications and discusses algorithm biases. [Section 4](#) interprets the LightGBM feature importance and summarizes the study.

## 2. Materials and methods

### a. Data

#### 1) PRECIPITATION-TYPE OBSERVATIONS

We use the precipitation-type observations from meteorological aerodrome reports (METARs) as the ground truth for our algorithm. Since winter precipitation occurs most frequently between November and March (NDJFM), our study focuses on these five months from 1979 to 2021. Although Automated Surface Observation Systems (ASOS) gradually replaced human observers in the late 1990s and early 2000s, some ASOS stations still routinely augment ASOS with human observations. Thus, to ensure the consistency of precipitation-type observations, we only include routinely augmented ASOS stations from all states east of the Rocky Mountains, with the exception of Florida, where winter precipitation rarely takes place.

To select the METAR stations that continuously provide human observations in the study domain, we sample five observation periods for every METAR station from 1979 to 2021. The five periods are 1–4 November 1979, 1–4 December 1989, 1–4 January 1999, 1–4 February 2009, and 1–4 March 2019. These five periods provide a decent coverage for the study period and allow us to robustly estimate the augmentation frequency of each station. Then, we calculate the fraction of nonaugmented observations (FNAO) in these five intervals by counting the number of records without the “AUTO” tag. Last, we classify the stations into four categories:

- 1) FNAO < 0.1 for all five observation periods (category-1 stations),
- 2) FNAO < 0.1 for four of the five observation periods (category-2 stations),
- 3) FNAO < 0.5 for all five observation periods (category-3 stations), and
- 4) FNAO < 0.5 for four of the five observation periods (category-4 stations).

The thresholds FNAO = 0.1 and FNAO = 0.5 are determined based on the assumption that stations with FNAO < 0.1 have all observations augmented, and that stations with FNAO < 0.5 have at least daytime observations augmented. All category-1 stations cannot be included, as stations in the northeastern United States are augmented much more routinely than those in the Great Plains. Instead, we would like to equally sample each geographical region to ensure that our algorithm can be generalized to the contiguous United States east of the Rocky Mountains. Thus, after classifying all stations into these four categories, we select the stations using a greedy algorithm with the constraint of

$$1.3\delta_{\text{lon}} + \delta_{\text{lat}} < 1, \quad (1)$$

where  $\delta_{\text{lon}}$ ,  $\delta_{\text{lat}}$  are the differences in longitude and latitude between each pair of stations, respectively. This constraint is used because 1.3° of longitude is equivalent in distance to 1° of latitude at 30°N. The greedy algorithm fills the entire domain with category-1 stations until no more stations can be added without violating the distance constraint. Then, we repeat the same procedure for stations from category 2 to category 4. This method yields a total of 169 stations for our study ([Fig. 1](#)).

#### 2) PRECIPITATION EVENT AGGREGATION

After obtaining the set of stations, we compile a training dataset with the following criteria. First, we require that the precipitation events are pure—that is, the precipitation events should contain only a single precipitation type, such as rain, snow, freezing rain, or ice pellets. Second, we assert that each precipitation event should last at least two hours. These two criteria are justified by the fact that short-duration and mixed precipitation episodes are not representative of each precipitation type. Third, we differentiate one precipitation event from another using a 6-h “null” period. For instance, the freezing-rain episode from 2000 to 2300 UTC on a particular day is considered to be a different event from the episode ending at 1300 UTC, because

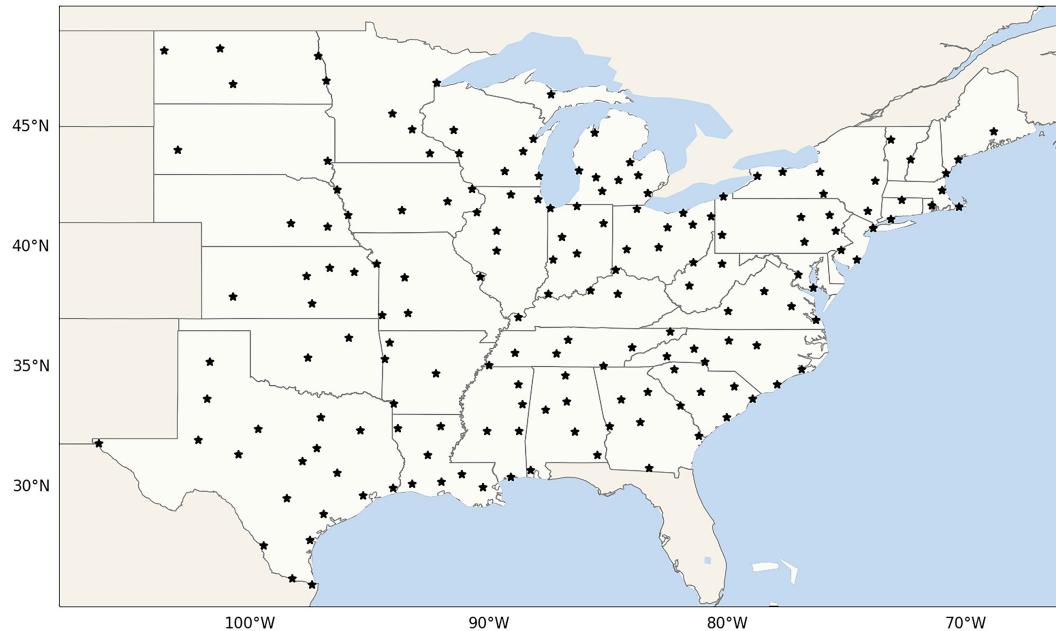


FIG. 1. The locations of the 169 ASOS stations employed in this study.

the two events are separated by a 7-h null period. With these requirements, we identify ~170 000 rain events, ~120 000 snow events, ~7000 freezing-rain events, and ~1500 ice-pellet events in NDJFM from 1979 to 2021.

To tackle the class imbalance problem, we down-sample the dominant precipitation classes (rain and snow) by randomly selecting 30 000 rain events, 30 000 snow events, 2500 freezing-rain events, and 1300 ice-pellet events for training and evaluating the algorithm. We use the hindcast experiment [see section 2e(2)] to determine the number of selected events for each precipitation type, thus ensuring a relatively unbiased training dataset.

### 3) ERA5 REANALYSIS

We use the hourly data from the ERA5 reanalysis (Hersbach et al. 2020) to select the predictors for our precipitation-type algorithm. The reanalysis product is run on 137 levels up to 1 Pa at a  $0.25^\circ$  horizontal resolution. To prepare the training data, we utilize the `.interp` function in Xarray to bilinearly interpolate the atmospheric fields from ERA5 to (i) the median time of each precipitation event and (ii) the location of each station, as each precipitation occurs at a unique location and time. For example, if a snow episode occurs from 1200 to 1430 UTC at Buffalo, New York, the atmospheric fields interpolated to  $42.9408^\circ\text{N}$ ,  $78.7358^\circ\text{W}$  at 1315 UTC will be used as the input data for the algorithm. It is worth mentioning that the outcomes are very similar when preparing the training data with or without temporal interpolation.

#### b. Precipitation-type algorithms

##### 1) THE BOURGOUIN PRECIPITATION-TYPE ALGORITHM

The Bourguin precipitation-type algorithm is a popular algorithm in operational forecast and climate models (Bourguin

2000). The algorithm utilizes the “area” method: it finds melting energy and refreezing energy by integrating the sounding profile against the  $0^\circ\text{C}$  line. It then classifies the precipitation types into rain (RA), snow (SN), freezing rain (FZRA), ice pellets (PL), a rain–snow mix (RASN), and a freezing rain–ice pellet mix (FZRAPL) based on melting energy, refreezing energy, and surface temperature. Although there have been efforts to improve the Bourguin algorithm, for example, through the use of wet-bulb temperature and a precipitation-generation layer (Birk et al. 2021), the original Bourguin algorithm remains one of the simplest and the most widely used precipitation-type algorithms (Bresson et al. 2017; Zarzycki 2018; Jeong et al. 2019; Matte et al. 2019; St-Pierre et al. 2019; McCray et al. 2022). Thus, the Bourguin algorithm is chosen as the first baseline in this paper, with the slight modification that it only outputs RA, SN, FZRA, and PL. The mixed classes RASN and FZRAPL are automatically merged into the four precipitation types based on the decision boundary described in Bourguin (2000) since our training data only contain pure precipitation episodes.

##### 2) ERA5 PRECIPITATION TYPES

The ERA5 global reanalysis product diagnoses a variety of precipitation types near  $0^\circ\text{C}$ , including rain, wet snow, dry snow, mixed rain and snow, freezing rain, and ice pellets. It makes a prediction based on the depth of the melting layer and consequently the fraction of liquid mass to total particulate mass (Owens and Hewson 2018). In this paper, we use the ERA5 precipitation type as the second baseline because the precipitation-type diagnosis from the European Centre for Medium-Range Weather Forecasts is widely used in operational settings. To achieve a one-to-one correspondence with our precipitation-type algorithm, we merge ERA5’s dry snow and wet snow into one class and discard all mixed-rain-and-snow

predictions. In other words, when ERA5 predicts RASN, the data point is deemed invalid such that it does not impact the performance of ERA5.

### 3) USING MACHINE LEARNING TO PREDICT PRECIPITATION TYPES

Machine learning techniques have been shown to excel in many atmospheric science problems, such as moist parameterization and numerical weather prediction (Herman and Schumacher 2018; Bi et al. 2023). Although machine learning sometimes receives criticism for the lack of interpretability, researchers have shown that tree-based algorithms and neural networks can offer valuable insights into the prediction process (O’Gorman and Dwyer 2018; Toms et al. 2020). For this reason, we test several algorithms that fall into these categories and compare their performance (see appendix A). We find that LightGBM has the highest area under the receiver-operating characteristic (AUC) score among all the algorithms and select it as the method of choice for this study. AUC measures the balance of false positive rate against hit rate.

LightGBM is a gradient-boosting decision tree (GBDT) framework (Ke et al. 2017). In general, GBDT iteratively combines weak learners into a strong learner by utilizing the gradient information from a predefined loss function (Chen and Guestrin 2016). While this characteristic gives it a considerable advantage over random forest that trains each individual tree separately, it incurs a higher computational cost and does not scale to large machine-learning problems. To address this issue, Ke et al. (2017) proposes the LightGBM framework that has numerous advantages over traditional GBDT. First, LightGBM discretizes continuous data into several bins in a histogram, which reduces the amount of time to traverse through the data points to find a perfect split. Second, LightGBM spends most of its training time on the “edge cases” in which the gradient of the loss function is large [referred to as gradient-based one-side sampling in Ke et al. (2017)]. This technique allows the classifier to focus on hard-to-separate cases, such as freezing rain and ice pellets, to achieve a higher accuracy in these categories. These two advances, among others, make LightGBM a powerful algorithm for numerous machine learning problems [see Ke et al. (2017) for a more detailed discussion].

To use LightGBM for precipitation-type predictions, we need to select a few predictors. After training and evaluating the algorithm using variables such as geopotential height and vertical velocity, we find that temperature and relative humidity best predict the precipitation types. Subsequently, we use 1000–500-hPa temperature (16 levels; 25-hPa interval up to 750 hPa and 50-hPa interval up to 500 hPa), surface temperature, 1000–500-hPa relative humidity (the same levels as temperature), and surface relative humidity to train LightGBM with a total of 34 predictors. The relative humidity is constrained to between 0% and 100% for best results with LightGBM.

Intuitively, LightGBM is made up of multiple decision trees that consist of if-else branches. As an illustrative example in Fig. 2, the first decision tree in LightGBM has two if-else branches: the first branch diagnoses snow using 950 hPa temperature, and the second branch differentiates between rain

and freezing rain with surface temperature. LightGBM then predicts the precipitation type by taking the majority vote from a suite of decision trees. We refer the readers to Chen and Guestrin (2016) for a more detailed discussion of the gradient tree boosting algorithm. The technical details of tuning LightGBM with 10-fold cross validation are explained in appendix B.

#### c. Performance metrics

We use five performance metrics in this paper: precision [success ratio (SR)], recall [probability of detection (POD)], F1 score, critical success index (CSI), and bias. We define each of them as follows:

$$P = \text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (2)$$

$$R = \text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (3)$$

$$F_1 = \frac{2}{R^{-1} + P^{-1}}, \quad (4)$$

$$\text{CSI} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} = (R^{-1} + P^{-1} - 1)^{-1}, \quad \text{and} \quad (5)$$

$$B = \text{bias} = R/P, \quad (6)$$

where TP, FP, and FN are true positives, false positives, and false negatives, respectively. We note the following: 1) Bias is defined as the ratio between recall and precision. A bias larger than 1 suggests that an algorithm overpredicts a precipitation type, and the opposite indicates underprediction. 2) The F1 score is a harmonic mean of precision and recall. The metric maximizes when bias  $\approx$  1, because  $R^{-1}$  and  $P^{-1}$  in the denominator act as a penalizing term when  $R$  or  $P$  gets too small. 3) CSI is a “sister” of F1 that shifts the denominator of F1 score by 1. An advantage of CSI is that it naturally occurs as contours on precision–recall plots, thus allowing us to use only one metric to interpret the result.

#### d. Defining hits in observations

While LightGBM is trained on precipitation events lasting more than two hours, the classifier should also achieve high performance with mixed and transient events. But since it could be challenging for reanalysis products to simulate the predominant precipitation type, we define hits using the notion of “any”—that is, if a precipitation type occurs within a given hour, we say that it is a hit regardless of the duration. However, as LightGBM only outputs one precipitation type, defining hits in this fashion could make our results vulnerable to inherent biases. For example, an algorithm biased toward predicting FZRA in FZRAPL would not achieve a high recall with PL, even though the prediction is partially correct. To tackle this problem, we adopt a looser definition of hits for mixed precipitation: if two precipitation types occur within the same hour, our algorithm would still be considered correct even if it only outputs one of the two types. Together, the two definitions enable us to examine, in unbiased fashion, whether

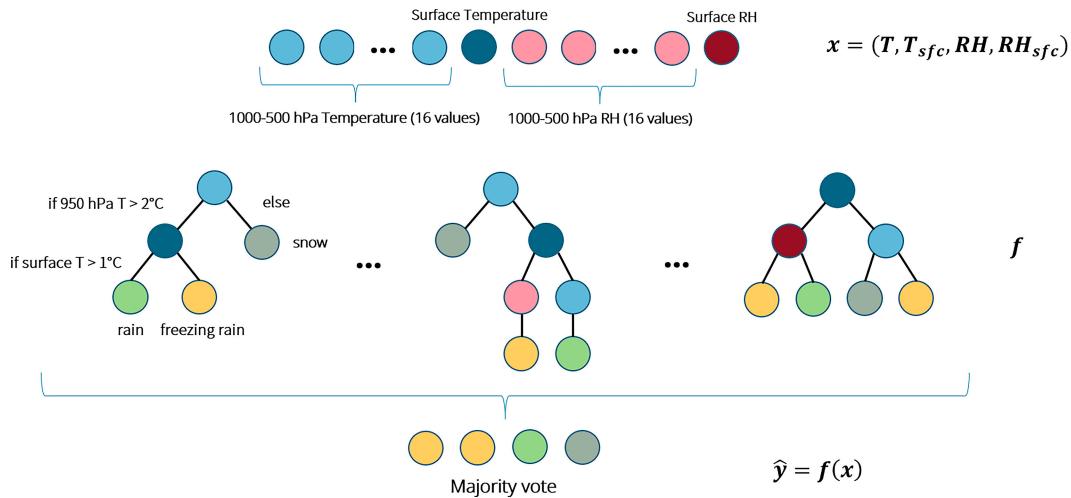


FIG. 2. A schematic diagram of the LightGBM classifier developed in this paper. The predictors are blue circles (1000–500-hPa temperature), dark-blue circles (surface temperature), pink circles (1000–500-hPa relative humidity), and red circles (surface relative humidity). The predictions are green circles (rain), yellow circles (freezing rain), and gray circles (snow). Ice pellets are omitted in this schematic diagram for simplicity.

our algorithm correctly predicts precipitation types in reanalysis products.

*e. Using bootstrapping to validate performance*

1) A FREEZING-RAIN CLIMATOLOGY

To further validate the performance of LightGBM, we use bootstrapping to generate a climatology of freezing rain and compare it with observations. As our study focuses on freezing-rain occurrence in NDJFM from 1979 to 2021, we divide the 215 months of reanalysis data into 43 buckets, with each bucket corresponding to a mix of five months from different years, for example, November 2012, December 1990, and January 2008. Then, after selecting a bucket, we exclude all precipitation events in the bucket from training data. We use the remaining subset of precipitation events to train LightGBM and apply it to the 5-month period. Last, we repeat the procedure for every bucket to produce a 43-yr climatology (1979–202; NDJFM). This approach ensures that LightGBM has not observed the correct labels a priori.

Additionally, we only count freezing-rain observations when ERA5 outputs an hourly precipitation rate of more than  $0.04 \text{ mm h}^{-1}$ , which is consistent with prior studies (McCray et al. 2022). The goal is to reduce drizzling bias in models to create a more realistic climatology. We also conduct a sensitivity test to see whether a higher precipitation threshold, that is more representative of the precision of precipitation measurements, leads to better performance.

2) PERFORMING A HINDCAST EXPERIMENT

Since we only validate the performance of the algorithm on precipitation events lasting more than 2 h, we also need a rough estimate of its accuracy in mixed precipitation episodes. We use bootstrapping to perform a hindcast experiment. Specifically, we randomly select 40 stations from the category-1

stations with a modified distance constraint—that is,  $1.3\delta_{\text{lon}} \pm \delta_{\text{lat}} < 2.5$ —and divide the 40 stations into 10 buckets such that each bucket has four stations. For each bucket, we remove the four stations from training data and use the remaining events to train LightGBM. By running the algorithm on hourly ERA5 data in NDJFM from 1979 to 2021, we utilize the two hit definitions in section 2d to assess its performance against the precipitation type provided by ERA5. Last, we require that both ERA5 and METARs record the four precipitation types that LightGBM aim to diagnose via the following criteria:

- 1) ERA5 resolves one of the four precipitation types (RASN is excluded),
- 2) ASOS records one of the four precipitation types, and
- 3) ERA5 records a minimum precipitation intensity of  $0.25 \text{ mm h}^{-1}$ .

These criteria allow us to accurately evaluate the errors resulting from precipitation typing because the error can be decomposed as

$$\varepsilon = \varepsilon_{\text{forecast}} + \varepsilon_{\text{precip}} + \varepsilon_{\text{ptype}}, \tag{7}$$

where  $\varepsilon$  is the total error,  $\varepsilon_{\text{forecast}}$  is the forecast lead time error,  $\varepsilon_{\text{precip}}$  is the precipitation error (i.e., when ERA5 falsely resolves precipitation or fails to resolve precipitation), and  $\varepsilon_{\text{ptype}}$  is the precipitation-type error. Since the forecast error is 0 in reanalysis products, the second constraint allows us to minimize  $\varepsilon_{\text{precip}}$  so that we could focus on the precipitation-type error.

3. Results

*a. Algorithm performance and the dependence on horizontal resolution*

LightGBM outperforms the Bourguin algorithm in nearly every category and demonstrates superiority in freezing-rain

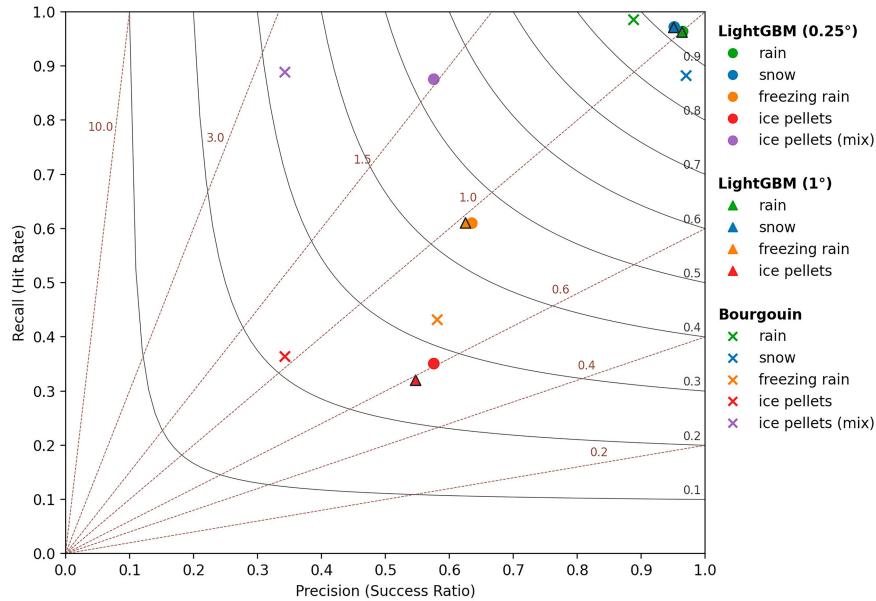


FIG. 3. The 10-fold cross validation performance of LightGBM ( $0.25^\circ$  and  $1^\circ$ ) and the Bourguoin precipitation-type algorithm ( $0.25^\circ$ ) with precipitation events described in section 2a. An explanation of cross validation methods can be found in appendix B. The  $1^\circ$  performance metrics are obtained by training and evaluating on the algorithm on  $1^\circ$  ERA5 data. The gray parabolas are contours of CSI, and the dashed brown lines are contours of bias (Roebber 2009).

prediction in particular (Fig. 3), as the CSI score (0.45) is higher than that of Bourguoin (0.33) (Table 1). Although the hit rate of LightGBM is slightly lower than that of Bourguoin for ice pellets, we note that maximizing success ratio over hit rate could lead to a higher overall CSI in mixed events because most ice-pellet precipitation episodes cooccur with rain or snow (Reeves 2016). Using observational estimates from hindcast experiments described in section 2b, we plot the expected precision and recall for mixed ice-pellet episodes (purple colors in Fig. 3). The inclusion of mixed events allows LightGBM to gain a significant advantage over the Bourguoin algorithm in ice-pellet prediction, with the CSI score increasing from 0.33 to 0.53 (Table 1). The precipitation typing from ERA5 is not shown because the median time of precipitation events can occur at any time, whereas ERA5 only outputs precipitation type on the hour.

To test the algorithm's dependence on horizontal resolution, we upscale the ERA5 data conservatively with the Python xESMF package to  $1^\circ$  resolution. We find that the

performance of LightGBM does not depend on the grid resolution, except for ice pellets, where precision and recall decrease by about 0.03 when coarsening the data (Fig. 3). It is noteworthy that this diminished accuracy may be explained by the loss of nuanced temperature and humidity fields essential for accurate ice-pellet predictions during the upscaling process.

Apart from the caveat, this result is generally congruent with Reeves (2016), who reported that the performance of common precipitation-type algorithms does not vary with the horizontal resolution ranging from 3 to 36 km. The invariance with grid spacing might stem from the fact that persistent precipitation events tend to be associated with slow-moving systems that span a large area. As a result, LightGBM would still be able to identify the quintessential characteristics of rain, snow, and freezing rain even if many small-scale features were smoothed out.

#### b. Hindcast experiment with ERA5

The results of the hindcast experiment are shown in Fig. 4, where we compare the performance of the LightGBM algorithm with the native precipitation typing in ERA5. We note that the recall rate from LightGBM is significantly higher than ERA5 for freezing rain and ice pellets for both the pure and mixed definitions, despite a marginally lower precision for freezing rain. In contrast, LightGBM obtains a much higher precision than ERA5 for ice pellets (Figs. 4b,d) and snow (not shown). The higher precision for snow leads to a higher mean F1 score (LightGBM

TABLE 1. A CSI comparison of the Bourguoin algorithm and the 10-fold cross validation score of LightGBM. For each precipitation type, boldface type indicates the best performance.

Precipitation type	Bourguoin	LightGBM
Rain	0.8767	<b>0.9309</b>
Snow	0.8560	<b>0.9263</b>
Freezing rain	0.3293	<b>0.4517</b>
Ice pellet	0.2144	<b>0.2787</b>
Ice pellet (mix)	0.3289	<b>0.5322</b>

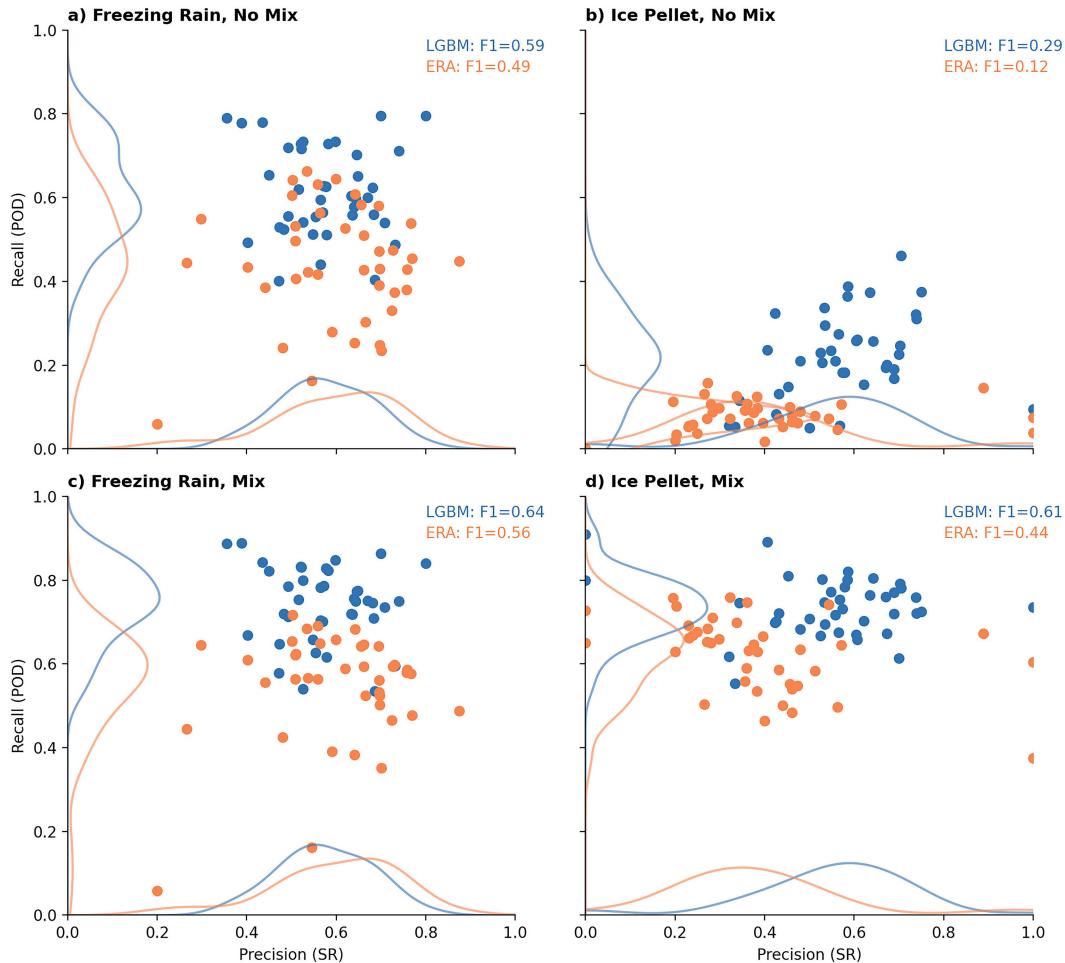


FIG. 4. Precision and recall of LightGBM and the ERA5 precipitation-type algorithm in the hindcast experiment for 40 stations. The four panels show the accuracy of the two algorithms in predicting (a) freezing rain; (b) ice pellets; (c) freezing rain, assuming that mixes are hits; and (d) ice pellets, assuming that mixes are hits. See section 2d for a detailed definition of hits. The curves are kernel density estimates of the two metrics for each precipitation-type algorithm. The texts on the upper-right corner in each plot indicates the average F1 score of all stations evaluated in the hindcast experiment.

0.923 vs ERA5 0.900), whereas the mean F1 score for rain is nearly identical for both algorithms (0.978 vs 0.973).

Furthermore, when precipitation errors are removed ( $\epsilon_{\text{precip}} = 0$ ), LightGBM shows limited bias in freezing rain for the no-mix definition (Fig. 4a). We note that the average precision and recall are similar, that is, bias  $\approx 1$ , which indicates that LightGBM does not overpredict freezing rain *on average*. Conversely, the underprediction in ice pellets arises from the trade-off between an unbiased prediction of mixed events and a high recall rate with ice pellets (Figs. 4b,d). In this paper, we advocate for prioritizing unbiased predictions of mixed events for their practical application in operational settings. However, if the aim is to produce an ice-pellet climatology, an alternative algorithm that achieves higher recall and a mean bias closer to 1 might be a more suitable choice.

### c. Observed and simulated climatology of freezing rain

We show the comparison between observations and the LightGBM-based climatology in Fig. 5. The observed climatology uses the “any” definition introduced in section 2d; we simply count whether freezing rain occurs at a given hour instead of analyzing whether it is the predominant precipitation type (Fig. 5a). The LightGBM climatology corresponds well to observations (Fig. 5b), with multiple local freezing-rain maxima over the Great Plains and the Appalachians. It also replicates the distinctive freezing-rain shadow to the west of the Appalachians (Bernstein 2000). Additionally, we notice that there is a salient freezing-rain maximum in the Black Hills in South Dakota and provide an explanation for its occurrence in section 4.

A major caveat of this result is that LightGBM significantly overpredicts freezing-rain occurrence for nearly the entire

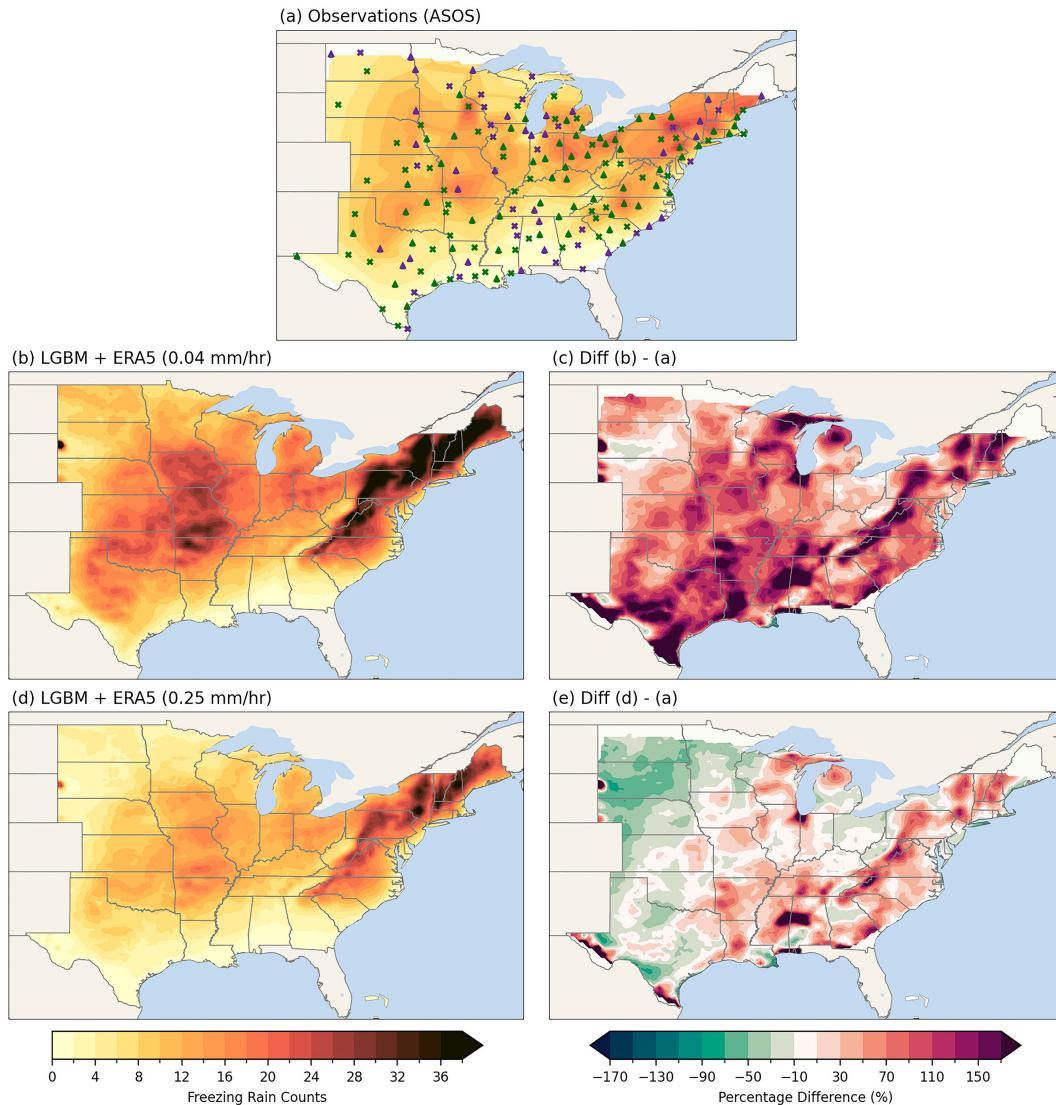


FIG. 5. (a) The observed annual mean freezing-rain frequency (NDJFM; 1979–2021). The green triangles and purple triangles respectively denote category-1 and category-2 stations, and the green crosses and purple crosses respectively denote category-3 and category-4 stations [see section 2a(1)]. (b) The annual mean freezing-rain frequency obtained by running LightGBM on ERA5 data (NDJFM; 1979–2021) with a precipitation cutoff of  $0.04 \text{ mm h}^{-1}$ . (c) The difference between the derived climatology and observation, namely, (b) – (a). (d) As in (b), but with a precipitation cutoff of  $0.25 \text{ mm h}^{-1}$ ; (e) as in (c), but with (d) – (a).

study domain (Figs. 5b,c). The overprediction bias can reach more than 170% in the southern Great Plains where annual mean freezing-rain hours are low. This bias can be eliminated by choosing a higher precipitation threshold of  $0.25 \text{ mm h}^{-1}$  (Figs. 5d,e). We argue that the  $0.25 \text{ mm h}^{-1}$  threshold is a better choice and can indeed help to reduce numerous biases (see discussion in section 3d).

#### d. Algorithm bias

##### 1) FREEZING-DRIZZLE BIAS AND PRECIPITATION BIAS

To understand the biases observed in the freezing-rain climatology, we revisit the forecast error decomposition in

Eq. (7) and note that the precipitation error  $\epsilon_{\text{precip}}$  is not 0 when producing the climatology. In other words, we do not enforce that freezing rain should only occur when METAR reports one of the four precipitation types. This treatment deviates from that in the hindcast experiment and contributes to freezing-rain biases in reanalysis products.

To delve into the precipitation error term  $\epsilon_{\text{precip}}$ , it is necessary to examine a new precipitation type: freezing drizzle. Unlike freezing rain, which is formed through the melting process, freezing drizzle is primarily produced by collision–coalescence (Cortinas et al. 2004). Because the hydrometeors are small when drizzling, the consensus is that precipitation intensity

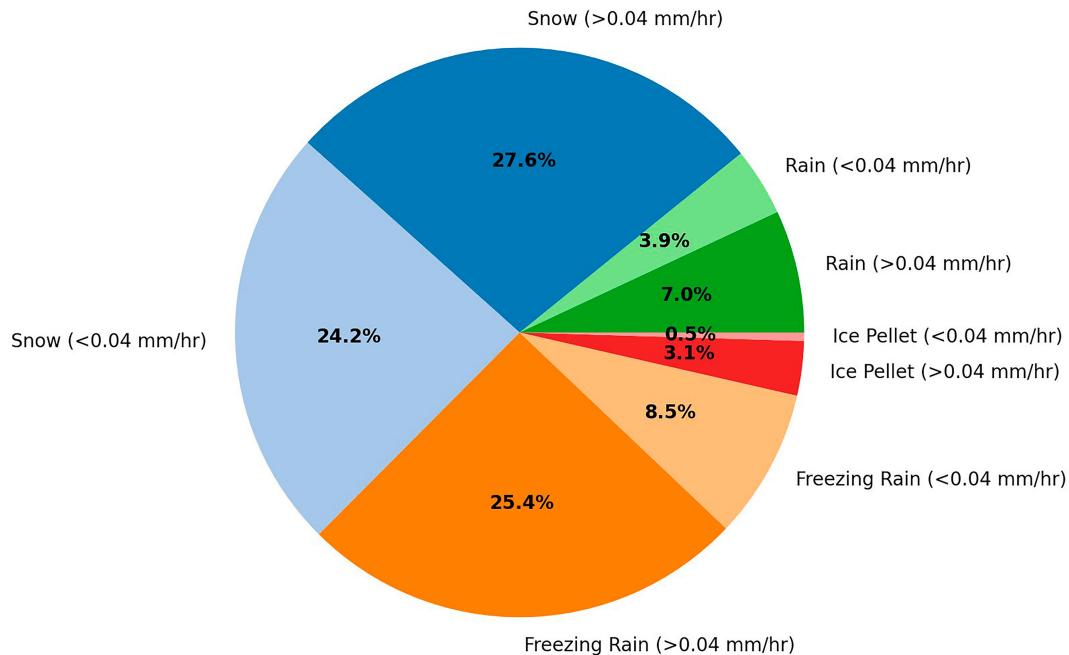


FIG. 6. The precipitation types resulting from running LightGBM on pure freezing-drizzle events lasting more than 2 h.

should be small, and thus freezing drizzle should have been filtered out of ERA5 using the  $0.04 \text{ mm h}^{-1}$  criterion. To test this assumption, we apply LightGBM to a list of pure freezing-drizzle events lasting more than 2 h and collect the predicted precipitation type (Fig. 6). The pie chart reveals that snow is the most similar precipitation type to freezing drizzle (51.8%), with freezing rain being the second (33.9%). But among the cases when freezing rain is predicted, about 75% have a precipitation rate exceeding  $0.04 \text{ mm h}^{-1}$ , which accounts for 25.4% of the freezing-drizzle events. Although this bound is likely pessimistic due to the inclusion of only long-duration events, the bias could contribute significantly to the simulated freezing-rain climatology, especially when freezing rain is the predominant type of freezing precipitation locally. Green Bay, Wisconsin, for instance, observes 57% of its total freezing precipitation as freezing drizzle but only 25% as freezing rain (Bernstein 2000). As a result, LightGBM may output freezing rain 50% more frequently than observation due to drizzling bias alone. We could further confirm this result by relaxing the second criterion in the hindcast experiment to calculate the situation in which

- 1) LightGBM predicts freezing rain and METAR reports one of the four precipitation types (*limited bias*),
- 2) LightGBM predicts freezing rain and METAR reports drizzle and freezing drizzle (*drizzling bias*), or
- 3) LightGBM predicts freezing rain and METAR reports no precipitation (*precipitation bias*).

Figure 7a demonstrates that freezing drizzle can contribute considerably to the local freezing-rain signal if a precipitation cutoff of  $0.04 \text{ mm h}^{-1}$  is used, particularly over the freezing-drizzle maximum in the northern Great Plains (Cortinas et al. 2004).

However, precipitation bias plays an even more prominent role in the simulated climatology, which stem from various sources including but not limited to *observation bias*, that is, METAR/ASOS fails to report precipitation when it should, the fact that reanalysis products cannot accurately capture the beginning of freezing-rain episodes due to evaporative cooling (McCray et al. 2020), or simply false alarms of precipitation events. Although a higher precipitation threshold can decrease these biases, it does not eliminate all of them (Fig. 7b). This result suggests that despite the high correlation between the observational and the LightGBM climatology (Fig. 5e), some of the predicted freezing-rain events still correspond to freezing-drizzle episodes or falsely resolved precipitation events. Regardless, the higher precipitation threshold of  $0.25 \text{ mm h}^{-1}$  facilitates an “apples to apples” comparison between the LightGBM climatology and observations to better evaluate model output.

## 2) REGIONAL BIAS

Aside from the precipitation and freezing-drizzle bias, we argue that the bias may also be attributed to the different synoptic weather patterns that lead to freezing rain in each region. For example, Bernstein (2000) discovered that freezing-rain events in the southeastern United States are associated with a deep, warm melting layer that yields a more distinctive freezing-rain signature. In contrast, when mixed precipitation events are developed farther inland, the inversion might become more subtle due to a shallower warm layer aloft.

This difference in turn leads to regional bias. In Fig. 8, we partition the stations into two halves based on latitude. Stations situated north of  $37^\circ\text{N}$  tend to cluster toward the lower

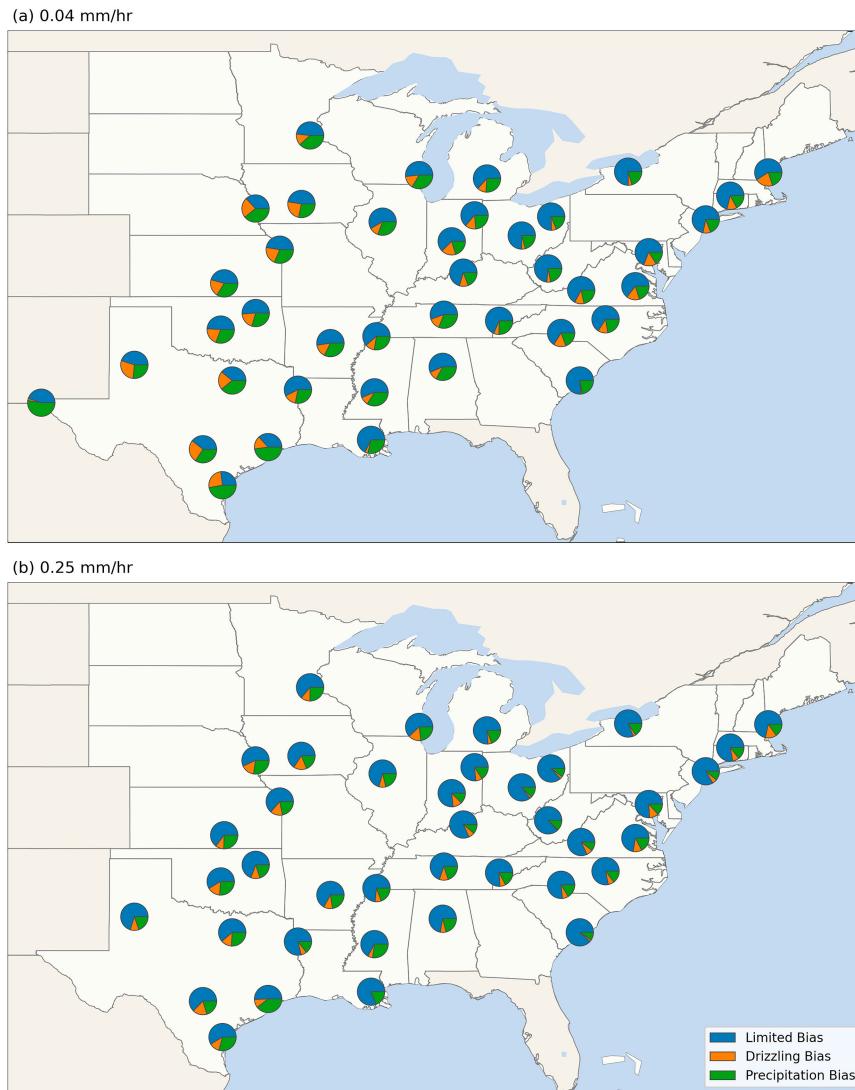


FIG. 7. The biases associated with each station in the hindcast experiment with a cutoff of (a) 0.04 and (b) 0.25 mm h<sup>-1</sup>.

portion of the precision–recall plot, indicating a tendency for underprediction bias. We note that this pattern is consistent with Fig. 5, with underprediction bias in the northern regions partially reconciled by the more prominent freezing-drizzle and precipitation bias in the Great Plains (Fig. 7). Notably, the observed pattern is mirrored in ERA5 precipitation diagnostics, where stations in the southern region consistently exhibit a higher recall-to-precision ratio than their northern counterparts.

In addressing the regional bias, we train a separate LightGBM classifier with latitude and longitude as the additional predictors. Interestingly, the result reveals a slightly increased precision and recall for freezing rain by approximately 1%, but a corresponding decrease in scores for other precipitation types—about 1% for ice pellets and 0.2% for rain and snow. Our interpretation is that latitude and longitude, while enhancing performance for freezing rain, may act as confounding variables for rain, snow, and ice pellets, ultimately diminishing their prediction accuracy. Therefore,

this experiment not only highlights the distinct spatial dependence of freezing rain but also underscores the potential challenges introduced by including latitude and longitude as additional parameters. Nonetheless, since LightGBM produces a curve of kernel density estimates with a smaller standard deviation than ERA5, machine learning methods should act as an interesting avenue to mitigate such regional biases.

### 3) ELEVATION BIAS

Algorithms that train on pressure-level data are susceptible to elevation bias since ERA5 interpolates atmospheric variables to below the surface layer (Trenberth et al. 1993). For instance, a freezing-rain event in western Nebraska might have a surface temperature and pressure of  $-1^{\circ}\text{C}$  and 900 hPa, respectively, but the 950- and 1000-hPa temperatures might well exceed  $0^{\circ}\text{C}$ . In contrast, a freezing-rain episode at Ohio

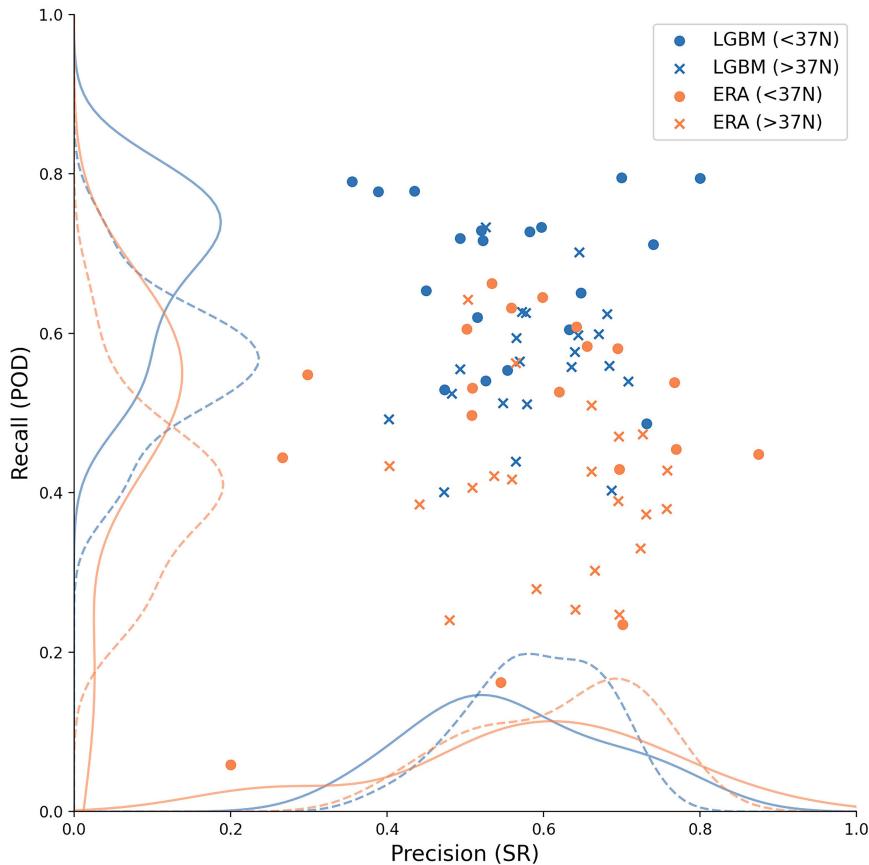


FIG. 8. As in Fig. 4a, but with stations north of 37°N and south of 37°N plotted separately. The solid lines indicate the kernel density estimates for stations south of 37°N, and the dashed lines indicate the kernel density estimates for stations north of 37°N.

Valley might have a 1000-hPa layer below 0°C because of a subfreezing surface layer. This difference introduces a positive temperature bias in the boundary layer, potentially contributing to a negative freezing-rain signal.

To test this hypothesis, we develop an additional classifier with temperature and relative humidity from the lowest 43 model levels as predictors and use it to derive a freezing-rain climatology (appendix D). No substantial difference is evident between the two classifiers (Fig. D1 of appendix D). This result indicates that LightGBM exhibits relative insensitivity to elevation bias and underscores that diverse biases in different regions stem from prevailing synoptic weather patterns. Further insights into why LightGBM demonstrates this robustness to elevation bias are explored in section 4.

#### 4) OBSERVATION BIAS AND INTERPOLATION BIAS

The substantial freezing-rain bias observed in the Great Plains, at the border of Mississippi and Alabama, and in the Appalachians may find an explanation in observation bias (Fig. 5). In Fig. 5a, we overlay the observation quality of each station onto the observed climatology. Notably, regions displaying a significant freezing-rain bias often coincide with category-3 and category-4 stations (green and purple crosses

in Fig. 5a), where observations are augmented only one-half of the time. It is important to highlight that this situation could introduce an underprediction bias, particularly in the pre-ASOS era when there were no automated means to infer the occurrence of freezing rain. Potential biases may still persist after the employment of ASOS observations, as the occurrence of freezing rain is inferred from the abnormal vibration frequency of the freezing-rain sensor. Conversely, regions with a high density of category-1 and category-2 stations, where human augmentation is frequent, exhibit lower freezing-rain bias. Consequently, the interpretation of bias in regions lacking category-1 and category-2 stations should be approached with caution.

## 4. Discussion and conclusions

### a. A physical interpretation of feature importance

LightGBM is highly explainable because it outputs *feature importance* (Fig. 9), which is obtained through the LightGBMClassifier in the Python LightGBM package with importance type set to “gain.” Feature importance measures the decrease in the loss function at each node of a tree that splits on a particular feature [see Eq. (7) in Chen and Guestrin (2016) for more details].

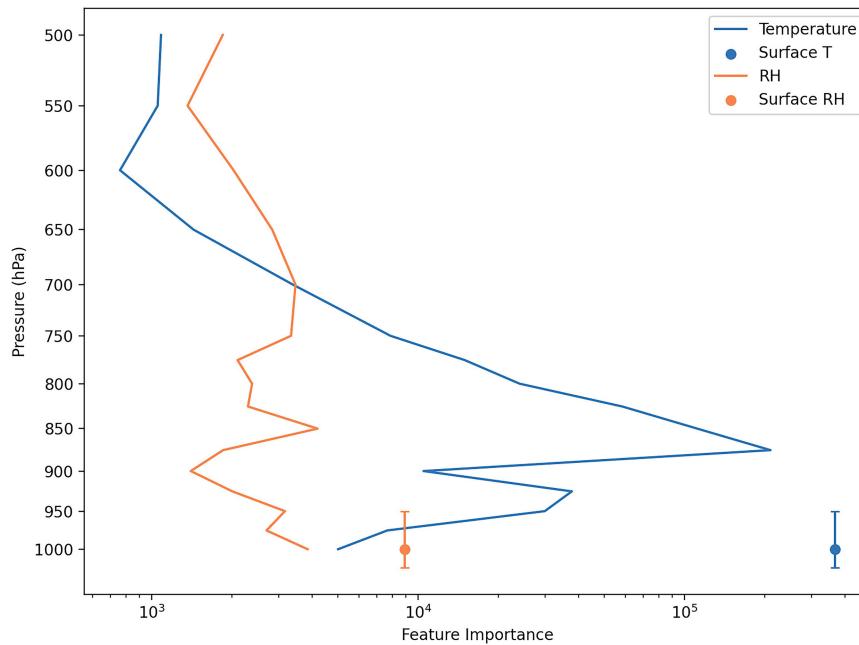


FIG. 9. Feature importance of the LightGBM classifier. The error bar on the surface temperature/surface relative humidity (950–1025 hPa) indicates a canonical range of values that the surface pressure can take when observing precipitation.

Two key implications emerge from Fig. 9. First, the two most important predictors are surface and 875-hPa temperature. This result aligns with the observation that temperature profiles tend to intersect the 0°C isotherm at 875 hPa for hard-to-separate precipitation cases. Intriguingly, surface temperature proves to be an order of magnitude more important than temperature at the lower levels, potentially contributing to the algorithm's robustness to elevation bias. The freezing-rain bias over the Black Hills can also be attributed to LightGBM potentially overemphasizing the importance of surface temperature. Although this bias can be eliminated by using model-level data as predictors (see Fig. C1 in appendix C), note that we lack a robust observational record in this location, and the primary focus of the study is not on diagnosing freezing-rain frequency in high terrain.

Second, relative humidity proves crucial for correctly diagnosing precipitation types. While temperature generally takes precedence over relative humidity, this dynamic reverses above 600 hPa. This shift underscores the significance of the precipitation generation layer in diagnosing precipitation types (Birk et al. 2021). Specifically, more freezing precipitation is observed if ice nucleation does not occur, and ice nucleation is directly dependent on temperature. Similarly, the algorithm demonstrates a strong dependence on surface relative humidity, echoing the understanding that dry, cool layers near the surface may play a significant role in ice-pellet formation (Bernstein 2000).

To test the sensitivity of each precipitation type to relative humidity, we train a LightGBM classifier with only temperature as predictors. This new algorithm shows a marked decrease in recall for freezing rain and ice pellets by 3% and 7%, respectively, while both recall and precision remain constant for rain and snow. Considering the collective influence of the precipitation-

generation layer and boundary layer moisture, we contend that explicitly incorporating relative humidity as predictors represents an enhancement over algorithms that only use temperature as predictors (Bourgouin 2000). This result highlights the importance of modeling relative humidity for effectively distinguishing between freezing rain and ice pellets.

#### b. Conclusions

In this study, we develop a LightGBM classifier to diagnose the precipitation type using ERA5 reanalysis. Our algorithm shows a considerable improvement over the Bourgouin algorithm and the ERA5 precipitation-type diagnosis, provided that precipitation episodes are correctly simulated. However, when a precipitation cutoff of 0.04 mm h<sup>-1</sup> is used, the algorithm substantially overpredicts freezing-rain frequency across the study domain due to either drizzling bias or precipitation bias. These biases can be reduced but not eliminated entirely by specifying the precipitation cutoff as 0.25 mm h<sup>-1</sup> to coincide with the precision of rainfall observations.

Therefore, it is crucial to realize the limitations of using precipitation-type algorithms to compute a freezing-rain climatology. Climate modelers and research scientists should investigate the inherent biases of each model before drawing quantitative conclusions about projected freezing-rain changes in a future climate. Although precipitation biases in models can remain unchanged with warming, making such assumptions requires us to carefully examine the synoptic conditions associated with these false positives. For example, freezing drizzle in the Great Plains can be caused by the passage of an Arctic cold front, as the intrusion of dry air in the midlayer inhibits the collision-coalescence process (Bernstein 2000; Fernández-González et al. 2014). While

the research community agrees that the number of cold spells will decrease with global warming, understanding the changes to this synoptic weather pattern and its relationship with freezing-drizzle bias will allow us to further reduce projection uncertainty.

Additionally, the method employed in this paper is applicable to mesoscale precipitation-type predictions at short forecasting lead times. For example, with High-Resolution Rapid Refresh (HRRR) hindcasts and surface observations, it might be possible to further increase the performance of LightGBM by selecting cloud and microphysics parameters that physically represent precipitation processes. Jensen et al. (2023) also demonstrated the potential of machine learning to diagnose supercooled large droplets, hence improving the predictability of freezing rain and freezing drizzle. In this regard, postprocessing model output with machine learning is an intriguing pathway to generate more accurate forecasts (Chantry et al. 2021).

LightGBM, however, has its limitations: it is a single-column method and neglects the large-scale conditions leading to these precipitation episodes. For this reason, we argue that the method could also be improved by incorporating additional predictors from the synoptic environment.

Nonetheless, we prove that LightGBM is a promising tool to predict precipitation types, interpret biases in the freezing-rain climatology, and understand physical processes that distinguish between freezing rain and ice pellets. A test case to use this algorithm to project future freezing-rain changes based on climate models will be examined in our next paper.

*Acknowledgments.* This work was supported by NOAA Contract 1332KP21FNEEN0023. Author Lehner acknowledges support from the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research (BER), Regional and Global Model Analysis (RGMA) component of the Earth and Environmental System Modeling Program under Award DE-SC0022070 and National Science Foundation (NSF) IA 1947282. The National Center for Atmospheric Research is sponsored by the National Science Foundation. Author Zhuang also acknowledges support from the Jane E. Brody

Undergraduate Research Award. The authors also thank the three anonymous reviewers for insightful comments that significantly improve the quality of this paper.

*Data availability statement.* ASOS data were obtained from the Iowa Environment Mesonet data server (<https://mesonet.agron.iastate.edu/request/download.phtml>). ERA5 data were obtained and are freely available from the Climate Data Store (<https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressure-levels?tab=overview>). Detailed instructions for utilizing the training data can be found at GitHub (<https://github.com/Z-Richard/JAMCzhuang24>).

## APPENDIX A

### Performance Comparison of Different Machine Learning Algorithms

In this paper, we test five different machine learning algorithms in Python: nonlinear support vector machines (SVM), 1D convolutional neural networks (1D-CNN), multilayer perceptron (MLP), gradient-boosting decision trees (XGBoost and LightGBM), and random forest (RF). We use PyTorch to train 1D-CNN; scikit-learn to train MLP, RF, and SVM; and XGBoost and LightGBM to train GBDT. The one-versus-one AUC score is reported in Table A1.

We note that, although it does not output AUC as a metric, the performance of nonlinear SVM falls short as compared with LightGBM for other metrics.

TABLE A1. A performance comparison of different machine learning algorithms. The boldface type indicates that LightGBM has the best performance of the five algorithms.

Algorithm	AUC
1D-CNN	0.7586
MLP	0.8431
<b>LightGBM</b>	<b>0.9378</b>
XGBoost	0.9370
RF	0.9345

## APPENDIX B

## APPENDIX C

**10-Fold Cross Validation**

When evaluating the performance of LightGBM, we divide the training data randomly into 10 stratified folds, with each fold containing 3000 rain events, 3000 snow events, 250 freezing-rain events, and 130 ice-pellet events. Then, we use nine folds to train the algorithm and one fold to validate the performance. After repeating the procedure 10 times, we take the average of the performance metrics in each fold and obtain the values shown in Fig. 3. This strategy is implemented using the *StratifiedKFold* class in scikit-learn.

**The LightGBM Climatology on 1° ERA5 Data**

A similar spatial pattern to Fig. 5d is obtained when we train and run the algorithm on 1° ERA5 data (Fig. C1). Although many fine-grain spatial variations are lost, the general pattern matches the 0.25° climatology well, which lends confidence in applying the algorithm to climate models at a similar grid spacing.

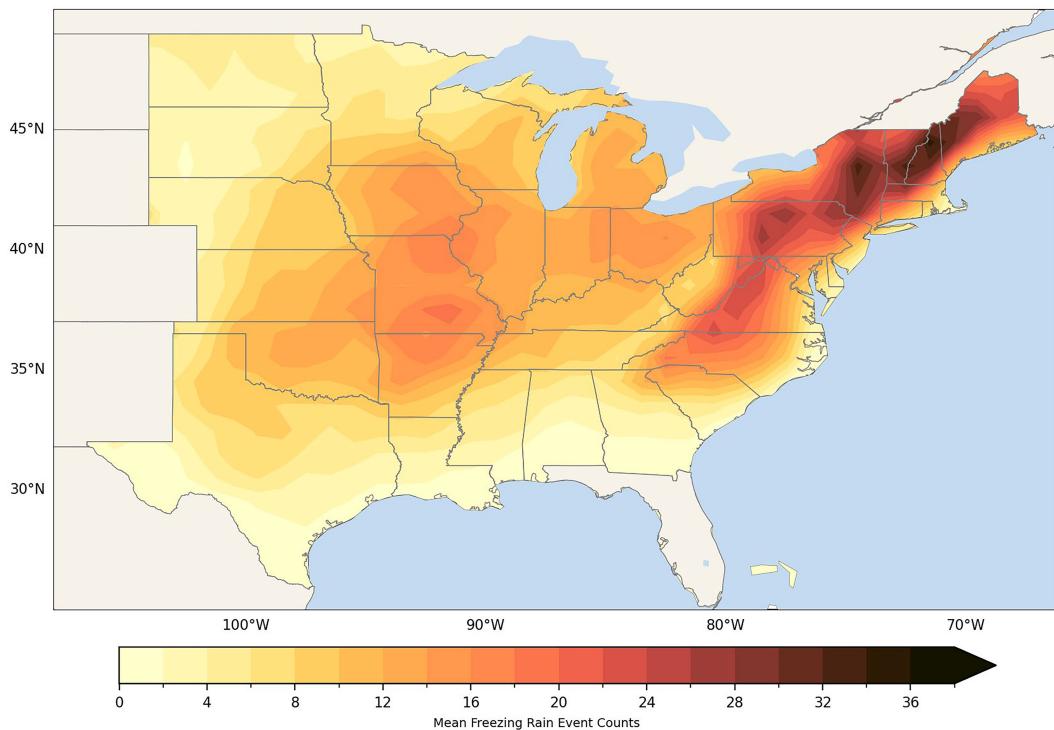


FIG. C1. As in Fig. 5d, but with LightGBM trained and run on 1° ERA5 data.

APPENDIX D

**Model-Level and Pressure-Level LightGBM Climatology**

Figure D1 shows the comparison between the derived climatology using model-level data and that with pressure-

level data. We argue that the two climatologies look nearly identical to each other, except for the absence of the topographical feature in the Black Hills for the model-level climatology.

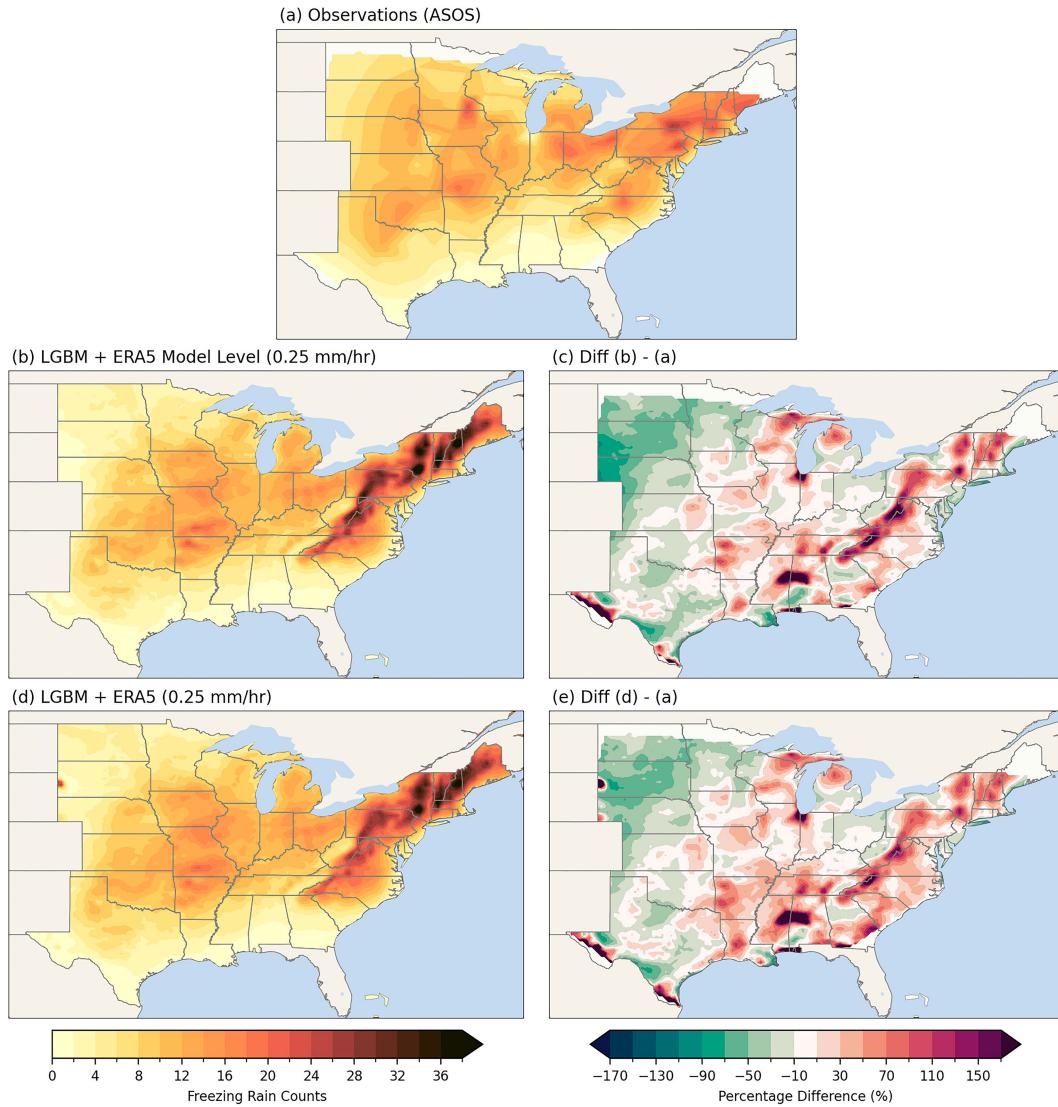


FIG. D1. (a) As in Fig. 5a, but without the superimposed stations; (b) The annual mean freezing-rain frequency obtained by running LightGBM on model-level ERA5 data (NDJFM; 1979–2021) with a precipitation cutoff of  $0.25 \text{ mm h}^{-1}$ ; (c) The difference between the derived climatology and observation, namely, (b) – (a). (d),(e) As in Figs. 5d and 5e, respectively.

## REFERENCES

- Benjamin, S. G., J. M. Brown, and T. G. Smirnova, 2016: Explicit precipitation-type diagnosis from a model using a mixed-phase bulk cloud–precipitation microphysics parameterization. *Wea. Forecasting*, **31**, 609–619, <https://doi.org/10.1175/WAF-D-15-0136.1>.
- Bernstein, B. C., 2000: Regional and local influences on freezing drizzle, freezing rain, and ice pellet events. *Wea. Forecasting*, **15**, 485–508, [https://doi.org/10.1175/1520-0434\(2000\)015<0485:RALIOF>2.0.CO;2](https://doi.org/10.1175/1520-0434(2000)015<0485:RALIOF>2.0.CO;2).
- Bi, K., L. Xie, H. Zhang, X. Chen, X. Gu, and Q. Tian, 2023: Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, **619**, 533–538, <https://doi.org/10.1038/s41586-023-06185-3>.
- Birk, K., E. Lenning, K. Donofrio, and M. T. Friedlein, 2021: A revised Bourgoiuin precipitation-type algorithm. *Wea. Forecasting*, **36**, 425–438, <https://doi.org/10.1175/WAF-D-20-0118.1>.
- Bourgoiuin, P., 2000: A method to determine precipitation types. *Wea. Forecasting*, **15**, 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).
- Bresson, E., R. Laprise, D. Paquin, J. M. Thériault, and R. de Elía, 2017: Evaluating the ability of CRCM5 to simulate mixed precipitation. *Atmos.–Ocean*, **55**, 79–93, <https://doi.org/10.1080/07055900.2017.1310084>.
- Chantry, M., C. Hannah, D. Peter, and P. Tim, 2021: Opportunities and challenges for machine learning in weather and climate modelling: Hard, medium and soft AI. *Philos. Trans. Roy. Soc.*, **A379**, 20200083, <https://doi.org/10.1098/rsta.2020.0083>.
- Chen, T., and C. Guestrin, 2016: XGBoost: A scalable tree boosting system. *KDD'16: Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, San Francisco, CA, Association for Computing Machinery, 785–794, <https://doi.org/10.1145/2939672.2939785>.
- Cheng, C. S., H. Auld, G. Li, J. Klaassen, and Q. Li, 2007: Possible impacts of climate change on freezing rain in south-central Canada using downscaled future climate scenarios. *Nat. Hazards Earth Syst. Sci.*, **7**, 71–87, <https://doi.org/10.5194/nhess-7-71-2007>.
- Cortinas, J. V., Jr., B. C. Bernstein, C. C. Robbins, and J. Walter Strapp, 2004: An analysis of freezing rain, freezing drizzle, and ice pellets across the United States and Canada: 1976–90. *Wea. Forecasting*, **19**, 377–390, [https://doi.org/10.1175/1520-0434\(2004\)019<0377:AAOFRF>2.0.CO;2](https://doi.org/10.1175/1520-0434(2004)019<0377:AAOFRF>2.0.CO;2).
- Das, S., and Coauthors, 2022: A comprehensive machine learning study to classify precipitation type over land from Global Precipitation Measurement Microwave Imager (GPM-GMI) measurements. *Remote Sens.*, **14**, 3631, <https://doi.org/10.3390/rs14153631>.
- DeGaetano, A. T., 2000: Climatic perspective and impacts of the 1998 northern New York and New England ice storm. *Bull. Amer. Meteor. Soc.*, **81**, 237–254, [https://doi.org/10.1175/1520-0477\(2000\)081<0237:CPAIOT>2.3.CO;2](https://doi.org/10.1175/1520-0477(2000)081<0237:CPAIOT>2.3.CO;2).
- Fernández-González, S., F. Valero, J. L. Sanchez, E. Gascón, L. López, E. García-Ortega, and A. Merino, 2014: Observation of a freezing drizzle episode: A case study. *Atmos. Res.*, **149**, 244–254, <https://doi.org/10.1016/j.atmosres.2014.06.014>.
- Herman, G. R., and R. S. Schumacher, 2018: Money doesn't grow on trees, but forecasts do: Forecasting extreme precipitation with random forests. *Mon. Wea. Rev.*, **146**, 1571–1600, <https://doi.org/10.1175/MWR-D-17-0250.1>.
- Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, <https://doi.org/10.1002/qj.3803>.
- Ikeda, K., M. Steiner, J. Pinto, and C. Alexander, 2013: Evaluation of cold-season precipitation forecasts generated by the hourly updating high-resolution rapid refresh model. *Wea. Forecasting*, **28**, 921–939, <https://doi.org/10.1175/WAF-D-12-00085.1>.
- IPCC, 2022: *Climate Change 2022: Impacts, Adaptation and Vulnerability*. H.-O. Pörtner et al. Eds., Cambridge University Press, 3056 pp., <https://doi.org/10.1017/9781009325844>.
- Jensen, A. A., C. Weeks, M. Xu, S. Landolt, A. Korolev, M. Wolde, and S. DiVito, 2023: The prediction of supercooled large drops by a microphysics and a machine learning model for the ICICLE field campaign. *Wea. Forecasting*, **38**, 1107–1124, <https://doi.org/10.1175/WAF-D-22-0105.1>.
- Jeong, D. I., A. J. Cannon, and X. Zhang, 2019: Projected changes to extreme freezing precipitation and design ice loads over North America based on a large ensemble of Canadian regional climate model simulations. *Nat. Hazards Earth Syst. Sci.*, **19**, 857–872, <https://doi.org/10.5194/nhess-19-857-2019>.
- Ke, G., Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T. Y. Liu, 2017: LightGBM: A highly efficient gradient boosting decision tree. *NIPS'17: Proc. 31st Int. Conf. on Neural Information Processing Systems*, Long Beach, CA, Curran Associates Inc., 3149–3157, <https://dl.acm.org/doi/10.5555/3294996.3295074>.
- Lambert, S. J., and B. K. Hansen, 2011: Simulated changes in the freezing rain climatology of North America under global warming using a coupled climate model. *Atmos.–Ocean*, **49**, 289–295, <https://doi.org/10.1080/07055900.2011.607492>.
- Lang, Z., Q. H. Wen, B. Yu, L. Sang, and Y. Wang, 2023: Forecast of winter precipitation type based on machine learning method. *Entropy*, **25**, 138, <https://doi.org/10.3390/e25010138>.
- Matte, D., J. M. Thériault, and R. Laprise, 2019: Mixed precipitation occurrences over southern Québec, Canada, under warmer climate conditions using a regional climate model. *Climate Dyn.*, **53**, 1125–1141, <https://doi.org/10.1007/s00382-018-4231-2>.
- McCray, C. D., J. R. Gyakum, and E. H. Atallah, 2020: Regional thermodynamic characteristics distinguishing long- and short-duration freezing rain events over North America. *Wea. Forecasting*, **35**, 657–671, <https://doi.org/10.1175/WAF-D-19-0179.1>.
- , J. M. Thériault, D. Paquin, and É. Bresson, 2022: Quantifying the impact of precipitation-type algorithm selection on the representation of freezing rain in an ensemble of regional climate model simulations. *J. Appl. Meteor. Climatol.*, **61**, 1107–1122, <https://doi.org/10.1175/JAMC-D-21-0202.1>.
- O'Gorman, P. A., and J. G. Dwyer, 2018: Using machine learning to parameterize moist convection: Potential for modeling of climate, climate change, and extreme events. *J. Adv. Model. Earth Syst.*, **10**, 2548–2563, <https://doi.org/10.1029/2018MS001351>.
- Owens, R., and T. Hewson, 2018: ECMWF forecast user guide. ECMWF, <https://doi.org/10.21957/m1cs7h>.
- Pórolniczak, M., L. Kolendowicz, B. Czernecki, M. Taszarek, and G. Tóth, 2021: Determination of surface precipitation type based on the data fusion approach. *Adv. Atmos. Sci.*, **38**, 387–399, <https://doi.org/10.1007/s00376-020-0165-9>.
- Ralph, F. M., and Coauthors, 2005: Improving short-term (0–48 h) cool-season quantitative precipitation forecasting: Recommendations from a USWRP workshop. *Bull. Amer. Meteor. Soc.*, **86**, 1619–1632, <https://doi.org/10.1175/BAMS-86-11-1619>.
- Ramer, J., 1993: An empirical technique for diagnosing precipitation type from model output. Preprints, *Fifth Int. Conf. on Aviation Weather Systems*, Vienna, VA, Amer. Meteor. Soc., 227–230.

- Reeves, H. D., 2016: The uncertainty of precipitation-type observations and its effect on the validation of forecast precipitation type. *Wea. Forecasting*, **31**, 1961–1971, <https://doi.org/10.1175/WAF-D-16-0068.1>.
- , A. V. Ryzhkov, and J. Krause, 2016: Discrimination between winter precipitation types based on spectral-bin microphysical modeling. *J. Appl. Meteor. Climatol.*, **55**, 1747–1761, <https://doi.org/10.1175/JAMC-D-16-0044.1>.
- Roebber, P. J., 2009: Visualizing multiple measures of forecast quality. *Wea. Forecasting*, **24**, 601–608, <https://doi.org/10.1175/2008WAF2222159.1>.
- Shin, K., K. Kim, J. J. Song, and G. Lee, 2022: Classification of precipitation types based on machine learning using dual-polarization radar measurements and thermodynamic fields. *Remote Sens.*, **14**, 3820, <https://doi.org/10.3390/rs14153820>.
- Stewart, R. E., J. M. Thériault, and W. Henson, 2015: On the characteristics of and processes producing winter precipitation types near 0°C. *Bull. Amer. Meteor. Soc.*, **96**, 623–639, <https://doi.org/10.1175/BAMS-D-14-00032.1>.
- St-Pierre, M., J. M. Thériault, and D. Paquin, 2019: Influence of the model horizontal resolution on atmospheric conditions leading to freezing rain in regional climate simulations. *Atmos.–Ocean*, **57**, 101–119, <https://doi.org/10.1080/07055900.2019.1583088>.
- Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev.*, **136**, 5095–5115, <https://doi.org/10.1175/2008MWR2387.1>.
- Toms, B. A., E. A. Barnes, and I. Ebert-Uphoff, 2020: Physically interpretable neural networks for the geosciences: Applications to Earth system variability. *J. Adv. Model. Earth Syst.*, **12**, e2019MS002002, <https://doi.org/10.1029/2019MS002002>.
- Trenberth, K. E., J. C. Berry, and L. E. Buja, 1993: Vertical interpolation and truncation of model-coordinate data. NCAR Tech. Note NCAR/TN-396+STR, 60 pp., <https://doi.org/10.5065/D6HX19NH>.
- Zarzycki, C. M., 2018: Projecting changes in societally impactful northeastern U.S. snowstorms. *Geophys. Res. Lett.*, **45**, 12 067–12 075, <https://doi.org/10.1029/2018GL079820>.