



Leveraging a radar-based disdrometer network to develop a probabilistic precipitation phase model in eastern Canada

Alexis Bédard-Therrien¹, François Anctil¹, Julie M. Thériault², Olivier Chalifour², Fanny Payette³, Alexandre Vidal³, and Daniel F. Nadeau¹

¹Département de génie civil et de génie des eaux, Université Laval, Québec, QC, Canada

²Département des sciences de la Terre et de l'atmosphère, Université du Québec à Montréal, Montréal, QC, Canada

³Hydro-Québec, Direction Planification de la conduite du système énergétique, Montréal, QC, Canada

Correspondence: Alexis Bédard-Therrien (alexis.bedard-therrien.1@ulaval.ca)

Received: 12 March 2024 – Discussion started: 29 April 2024

Revised: 13 September 2024 – Accepted: 21 December 2024 – Published: 28 February 2025

Abstract. This study presents a probabilistic model that partitions the precipitation phase based on hourly measurements from a network of radar-based disdrometers in eastern Canada. The network consists of 27 meteorological stations located in a boreal climate for the years 2020–2023. Precipitation phase observations showed a 2 m air temperature interval between 0–4 °C, where probabilities of occurrence of solid, liquid, or mixed precipitation significantly overlapped. Single-phase precipitation was found to occur more frequently than mixed-phase precipitation. Probabilistic phase-guided partitioning (PGP) models of increasing complexity using random forest algorithms were developed. The PGP models classified the precipitation phase and partitioned the precipitation accordingly into solid and liquid amounts. PGP_basic is based on 2 m air temperature and site elevation, while PGP_hydromet integrates relative humidity, surface pressure, and precipitation rate. PGP_full includes all previous data, along with atmospheric reanalysis data, the 1000–850 hPa layer thickness, and temperature lapse rate. The PGP models were compared to benchmark precipitation-phase-partitioning methods. These included a model with a single temperature threshold set at 1.5 °C, a linear-transition model with dual temperature thresholds of –0.38 and 5 °C, and a psychrometric balance model. Among the benchmark models, the single temperature threshold had the best classification performance (F1 score of 0.74) due to a low count of mixed-phase events. The other benchmark models tended to over-predict mixed-phase precipitation in order to decrease the partitioning error. All PGP models showed significant phase classification improvement by

reproducing the observed overlapping precipitation phases based on 2 m air temperature. PGP_hydromet and PGP_full displayed the best classification performance (F1 score of 0.84). In terms of partitioning error, PGP_full had the lowest RMSE (0.27 mm) and the least variability in performance. The RMSE of the single-temperature-threshold model was the highest (0.40 mm) and showed the greatest performance variability. An input variable importance analysis revealed that the additional data used in the more complex PGP models mainly improved mixed-phase precipitation prediction. The improvement of mixed-phase prediction remains a challenge. Relative humidity was deemed to be the least important input variable used due to consistent near-saturation water vapour conditions. Additionally, the reanalysis atmospheric data proved to be an important factor in increasing the robustness of the partitioning process. This study establishes a basis for integrating automated phase observations into a hydrometeorological observation network and for developing probabilistic precipitation phase models.

1 Introduction

Precipitation phase is a critical component in hydrological modelling. Simply put, the hydrological effect following either a snowfall or rainfall event is drastically different; snowfall accumulates in winter and melts later in the spring or during winter melt events, while rain can either infiltrate or become runoff, potentially increasing streamflow in the short term. The precipitation phase also affects snowpack charac-

teristics in different ways; a rain-on-snow event infiltrates the snowpack, increasing its liquid water content, and can create ice layers and change the snowpack internal characteristics (Singh et al., 1997; Wever et al., 2016), while snowfall increases the depth of the snowpack. During individual precipitation events, errors in snowpack water equivalent (SWE) and depth are mainly caused by errors in precipitation partitioning (Leroux et al., 2023). On a seasonal basis, the precipitation phase significantly affects the ablation of the snowpack, particularly due to its impact on the snow albedo (Essery et al., 2013; Günther et al., 2019). As such, the cumulative effects of misclassified precipitation have a significant impact on various seasonal values such as peak SWE, peak discharge date, and snow cover duration (Harder and Pomeroy, 2014).

As the climate warms, regions that typically experience winter snowfall are expected to face more rainfall in their winter precipitation, resulting in more rain-on-snow events (Jeong and Sushama, 2018; Ye et al., 2008; Musselman et al., 2018). Consequently, the proportion of runoff caused by rain-on-snow events during the winter is projected to increase in these regions (Jeong and Sushama, 2018; Musselman et al., 2018). Effective precipitation partitioning methods are more important than ever to anticipate potentially damaging events and to monitor water resources at the catchment scale. This need is also felt for field monitoring of precipitation quantities. Indeed, solid precipitation is much more sensitive to undercatch (underestimation due to the wind moving hydrometeors away from the gauge) than liquid precipitation (Rasmussen et al., 2012). Consequently, an inaccurate identification of the phase necessarily translates into an erroneous estimation of the precipitation quantity. Ehsani and Behrangi (2022) showed that undercatch for solid precipitation introduced a significant bias in gridded precipitation products at both the seasonal and annual scales at higher latitudes. This highlights the need to account for the precipitation phase at the synoptic scale, especially when using precipitation products to bias-correct satellite precipitation estimates (Behrangi et al., 2019; Ehsani and Behrangi, 2022).

The modelling of the precipitation phase in operational hydrological models is often based on a single near-surface air temperature threshold (Harpold et al., 2017). While simple to implement, this method cannot predict mixed precipitation events, which tend to occur when falling hydrometeors of different sizes coexist and melt at different rates (Thériault and Stewart, 2010). As an alternative to the single-threshold approach, one can use a linear relationship to account for mixed-phase precipitation events that occur between those two thresholds. Furthermore, these types of methods can be refined into curvilinear functions, which would theoretically yield a more accurate phase identification (Feiccabrino et al., 2013). In both cases, the classification error is reduced when compared to the single-threshold approach (Feiccabrino et al., 2013; Wen et al., 2013). The advantage of using temperature-threshold-based models comes mainly from

a data availability and computational requirement standpoint. Variables other than air temperature are known to influence the precipitation phase, such as relative humidity and atmospheric pressure (Behrangi et al., 2018; Dai, 2008; Jennings et al., 2018). Thus, precipitation partitioning models can be improved by using dew point temperature (e.g. Marks et al., 2013; Ye et al., 2013) or wet-bulb temperature (e.g. Ding et al., 2014; Wang et al., 2019; Behrangi et al., 2018) instead of relying solely on air temperature, increasing the spatial robustness of such models.

Phase-partitioning models tend to rely on near-surface hydrometeorological variables because this information is easily accessible. However, the hydrometeor's initial phase as it leaves the cloud, the shape and size distribution of the precipitation particles, and the properties of the atmosphere from the cloud to the ground all determine the precipitation phase (Feiccabrino et al., 2015). As they fall, the hydrometeors exchange latent and sensible heat with their surroundings, linking their phase to the temperature and vapour deficit as they fall through the atmosphere. Additionally, both heat fluxes are also affected by the ventilation of the hydrometeor, which depends on its fall speed and the surrounding wind velocities (Stewart, 1992).

Atmospheric temperature gradients can vary with time, and so the thickness of the melting atmospheric layer is also a key variable to consider as it affects the time the hydrometeor spends in conditions favourable to melting. Empirical models approximate this layer thickness by computing the height difference between two selected pressure levels (Feiccabrino et al., 2015). Precipitation rates can also increase the energy required to melt hydrometeors. Indeed, at high precipitation rates, there is a larger volume to melt, thus increasing the likelihood of solid precipitation at warmer temperatures (Froidurot et al., 2014; Thériault and Stewart, 2010). Therefore, by accounting for the characteristics of the atmospheric layer, microphysical models can determine the precipitation phase of falling hydrometeors (Thériault and Stewart, 2010). Other models are instead based on the statistical relationship between the hydrometeorological variables and the precipitation phase. Such models compute the probability of a precipitation phase occurring considering a set of environmental conditions.

The methodology for calculating the probability of phase occurrence varies across studies and includes, for example, a curvilinear function (Dai, 2008), logistic regression (Behrangi et al., 2018; Froidurot et al., 2014; Jennings et al., 2018), and machine learning algorithms (Shin et al., 2022). These methods output a precipitation type rather than a fraction of solid and liquid precipitation in the case of dual-threshold models. However, the use of these methods is limited when dealing with mixed-phase precipitation as they do not provide information on how the precipitation is partitioned. Fortunately, mixed-phase events are less common than single-phase events (Dai, 2008) and are thus often omitted from studies using probabilistic methods.

In addition to selecting the appropriate variables to include in a phase-partitioning model, the quality and availability of the validation dataset are critical aspects to consider. Indeed, the scarcity of validation data was cited by Harpold et al. (2017) as a major factor hindering the development of phase-partitioning models. Direct manual phase observations collected from trained observers have been used to validate precipitation partitioning models (e.g. Behrangi et al., 2018; Dai, 2008; Froidurot et al., 2014; Jennings et al., 2018). Jennings et al. (2023) have also shown the possibility of using crowdsourced precipitation phase data. While such datasets are extensive in time and space, they do not provide quantitative information on the snow and rain fractions in mixed-phase events, thus limiting the possible predicted precipitation phases to either solid or liquid.

High-frequency automatic measurements do not suffer from limitations caused by mixed-phase precipitation (Froidurot et al., 2014; Harpold et al., 2017) as the precipitation phase can be coupled with a concurrent precipitation amount. When both phase identification and precipitation gauge measurements are made at a high frequency, phase-separated precipitation can be compiled for hourly or shorter time steps, thus allowing for mixed-phase partitioning. One possible validation approach based on automatic data is to use precipitation measurements collocated with snow cover height measurements (Harder and Pomeroy, 2013; Marks et al., 2013). A more direct automatic approach is to use a disdrometer, which identifies the phase of the hydrometeor according to its size and falling speed. For instance, Wayand et al. (2016) utilized a disdrometer to associate precipitation phase with precipitation amounts, which helped evaluate multiple phase models. This combination of observations not only allows for the validation of a phase model but also addresses a major limitation of previous studies, namely the partitioning of precipitation in the case of mixed-phase events. Another important factor to consider is the time step of the validation data. While many conceptual hydrological models employ daily time steps to determine precipitation phase, sub-daily time steps greatly enhance the accuracy of the modelled phase (Feiccabrino, 2020; Harder and Pomeroy, 2013). Therefore, it is necessary to use sub-daily time steps, such as 15 min or hourly, as a significant portion of the phase model's performance depends on the time step of interest.

There are many ways to improve the representation of the precipitation phase for hydrological purposes. As pointed out in Harpold et al. (2017), the often too simple phase models need more hydrometeorological observations for a successful partitioning of the precipitation phase. Such observations include the relative humidity, as well as atmospheric information, like the temperature and humidity lapse rate. Additionally, an important research limitation comes from the lack of validation data. Direct observations, while commonly utilized, have limited applications for mixed-phase precipitation due to their qualitative nature. A preferable solution involves automated phase observations as they enable the coupling

with precipitation rate measurements. However, direct phase observation datasets indicate the existence of a temperature transition zone, where both snow and rain are possible. This highlights the limitations of simplistic phase models that fail to capture the complex nature of phase determination.

This study leverages a unique regional-scale radar disdrometer network coupled with precipitation observations to develop a probabilistic phase-partitioning model. The probabilistic model follows a phase-guided partitioning (PGP) in the form of a chain of random forest models. The precipitation phase is classified before partitioning to accurately replicate its intricate behaviour and to take advantage of the significant amount of validation data available through such a network. As such, the models classify the precipitation as either solid, liquid, or mixed phase. The predicted phase then dictates the partitioning into solid and liquid fractions. Additionally, multiple PGP models with lower data requirements are developed to evaluate the possibility of utilizing such models in practical operations. This study begins with a description of the precipitation dataset and hydrometeorological variables used, followed by the methodology used to develop the PGP models. Finally, the Results section presents an analysis of the dataset and evaluates the model's phase classification and partitioning performance, comparing it to benchmark models of differing levels of complexity.

2 Data

2.1 Surface hydrometeorological measurements

The disdrometer network used for this study was deployed on the north shore of the St. Lawrence River in the province of Quebec, Canada (Fig. 1). It is part of a larger hydrometeorological observation network operated by Hydro-Québec, a public utility responsible for the generation and distribution of electricity in Quebec. The network lies between latitudes of 47.23 and 52.13° N, and longitudes of 63.17 to 75.29° W, spanning an area of roughly 532 594 km². The 27 stations have been in operation for varying periods of time between the years 2019 and 2023, totalling 80 site-years. The site elevations range from 315 to 641 m above sea level (a.s.l.), with an average of 469 m a.s.l. The names, coordinates, and operational time frames for each station are presented in Appendix A.

The sites have a mean annual 2 m air temperature of 0.2 °C and a mean annual cumulative precipitation of 902 mm, calculated from 160 site-years of daily observations. Figure 2 illustrates the distribution of annual mean daily 2 m air temperature and precipitation, as well as the elevation of the sites. The annual precipitation decreases with latitude as the northern sites (> 51° N) experience an annual mean of 813 mm. The southernmost sites (< 49° N) receive more precipitation on average, with an annual mean of 1002 mm. However, the variability observed at these sites is much greater than ob-

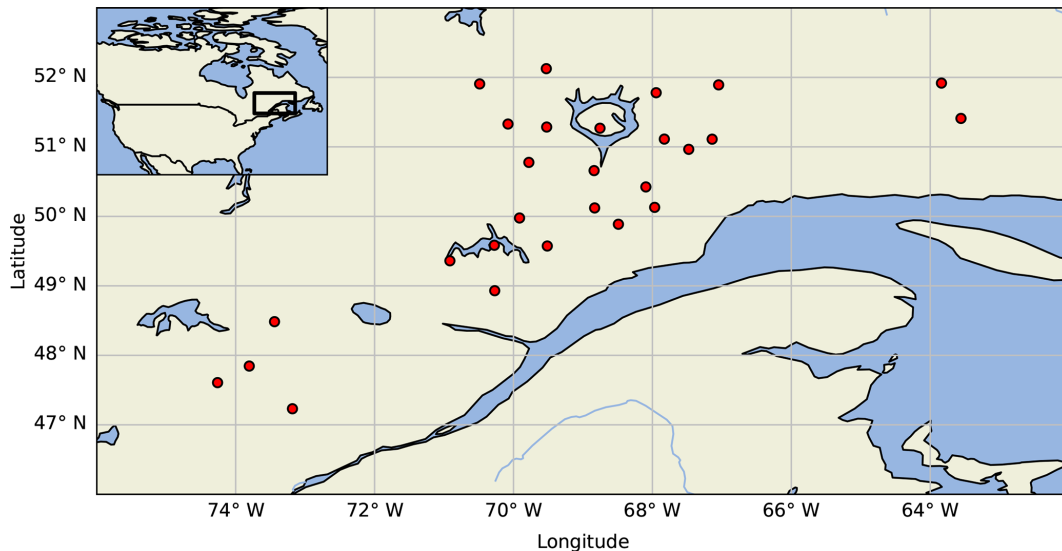


Figure 1. Location of the study sites in eastern Canada. The black square in the top-left inset corresponds to the map domain.

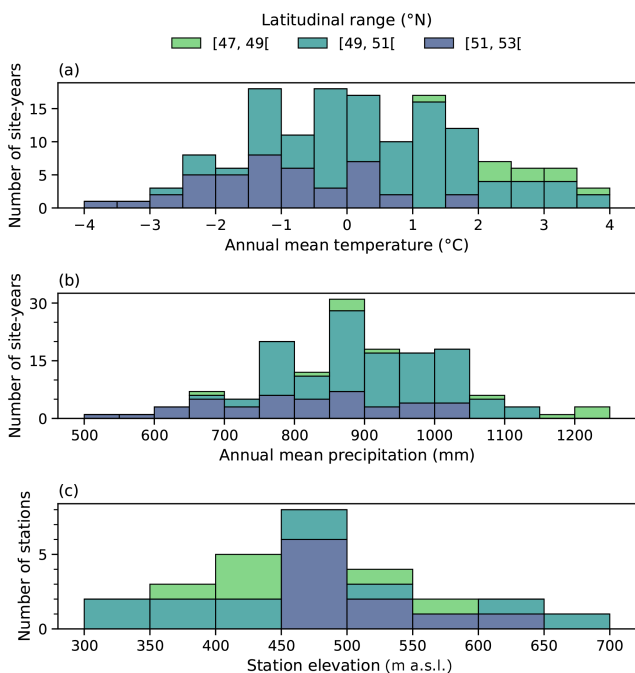


Figure 2. Distributions of the (a) annual mean temperature, (b) annual mean precipitation, and (c) elevation at the study sites, separated by latitudinal range.

served elsewhere. The sites' elevation follows a mostly normal distribution within a 400 m range of the mean, with elevations generally increasing northward. Following the Köppen climate classification, the sites are nearly evenly split between humid continental (Dfb) and humid subarctic (Dfc) climates. The study period spans from 1 October to 1 June of

the following year as the chances of snowfall are practically non-existent outside of these dates for the domain of interest.

Each site is equipped with a radar-based disdrometer (model WS100, Lufft), providing 15 min phase identification. The WS100 is a K-band (24 GHz) Doppler radar that classifies droplets into 11 size classes between 0.3 and 5.0 mm. The disdrometer assigns World Meteorological Organization (WMO, 2018; Table 4677) weather codes for no precipitation (code 0), rain (code 60), freezing rain (code 67), a mix of rain or drizzle and snow (code 69), and snowfall (code 70). The precipitation phase is identified according to the hydrometeor diameter–fall velocity relationships for water droplets outlined in Gunn and Kinzer (1949), as well as in Locatelli and Hobbs (1974) for solid-phase precipitation particles. Rain and snow fall velocity as a function of measured reflectivity from K-band Doppler radar was investigated in Atlas et al. (1973) and remains an area of active research (e.g. Garcia-Benadi et al., 2020; Kneifel et al., 2011; Löffler-Mang et al., 1999; Sarkar et al., 2015). For simplicity, the phase identifications derived from the diameter–fall velocity relationships are referred as observations in this study. In addition, each site also provides measurements of SWE using a passive gamma ray monitoring system (model CS725, Campbell Scientific) and of snow height using an ultrasonic sensor (model SR50A, Campbell Scientific).

The meteorological observations in this study come from weather stations operated by Hydro-Québec and SOPFEU, the province's wildfire prevention organization. The weather stations provide 15 min accumulated precipitation (model Pluvio² by OTT, equipped with a single-Alter shield), allowing the coupling of precipitation with concurrent disdrometer phase identification. The weather stations measure hourly air temperature (model CS109, Campbell Scientific) and relative humidity (model HMP155a, Campbell Scien-

tific), with sensors mounted 2 m above the ground. Additionally, wind speed and direction are monitored at a 15 min interval with a ground propeller anemometer (model 05103, Young) mounted 2 and 10 m above the ground.

Most of the study sites and weather stations are located in close proximity to each other. Specifically, 67 % of the study sites are within 3 km of or are collocated with the nearest weather station. The remaining stations are between a median distance of 7 km and a maximum distance of 12 km. The only exception is the Auxlous station and the nearest weather station, which are separated by 28 km. To account for the elevation differences between the study sites and weather stations, the air temperature measurements were adjusted using international standard atmosphere methods. The Discussion section will address the uncertainty related to the distance between the study site and the weather station. The temporal resolution and detailed specifications about the study sites' instruments are provided in Table 1.

2.2 Reanalysis products

Hourly atmospheric data from the ECMWF Reanalysis v5 (ERA5) (Hersbach et al., 2023) are added to this study's dataset to account for the energy transfer to falling hydrometeors in the atmospheric levels closest to the surface. Furthermore, this will help in assessing the potential performance gain of incorporating gridded data despite the spatial scale discrepancy with local observational data. The added data include temperature profiles for pressure levels of 1000 and 850 hPa. The corresponding geopotential height of these levels is also added to the dataset. The values from the nearest $0.25^\circ \times 0.25^\circ$ grid cell, roughly $28 \text{ km} \times 18 \text{ km}$ at the study sites' latitudes, are assigned to every study site. Additionally, the hourly surface atmospheric pressure from ERA5-Land (Muñoz Sabater, 2019) is added to the dataset as it was not measured at the weather stations used in this study. The atmospheric pressure from the nearest $9 \text{ km} \times 9 \text{ km}$ grid space is assigned to every study site. From these data, the thickness Δz between the 1000 and 850 hPa layers (m) is calculated with

$$\Delta z = \frac{z_{850} - z_{1000}}{g}, \quad (1)$$

where z_{850} and z_{1000} correspond to the geopotential heights ($\text{m}^2 \text{ s}^{-2}$) at the top and bottom of the layer, and g is the gravitational acceleration (9.81 m s^{-2}). The layer thickness between the two pressure levels is correlated with the mean temperature of the layer and is also a commonly used variable in operational meteorological models (Feiccabrino et al., 2015). The pressure levels were selected based on their successful use in the classification of the precipitation phase at the surface in prior studies (e.g. Bourgoïn, 2000; Shin et al., 2022). The temperature lapse rate Γ ($^\circ\text{C km}^{-1}$) is also calculated:

$$\Gamma = -\frac{\Delta T}{\Delta z} \times 1000, \quad (2)$$

where ΔT corresponds to the temperature difference between the 850 and 100 hPa layers ($^\circ\text{C}$).

3 Methodology

3.1 Precipitation data processing

The observed 15 min precipitation amounts were compiled at an hourly time step. Each 15 min precipitation data segment was coupled with a disdrometer phase identification. Both valid, non-zero values were required for the data segment to be included in the analysis. A first filter was applied, where hourly precipitation rates $< 0.2 \text{ mm h}^{-1}$ were considered to be erroneous trace amounts, following standard WMO methodology (WMO, 2018). A second filter was also applied where precipitation rates $> 110 \text{ mm h}^{-1}$ were considered to be erroneous (Smith et al., 2022). A neutral aggregating filter (Ross et al., 2020) was then applied to eliminate noise and diurnal oscillations in the precipitation data. Additionally, hourly precipitation exceeding 30 mm h^{-1} was visually inspected. Any data not consistent with nearby stations were considered to be invalid.

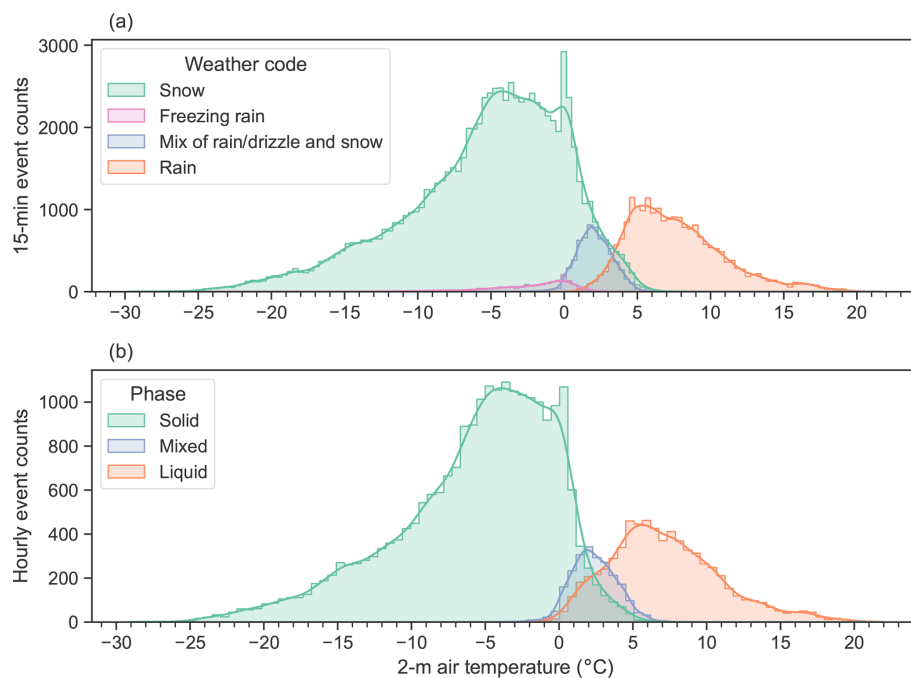
The disdrometers used in this study can identify freezing rain and a mix of rain and snow in addition to snow and rain. However, as most hydrological models only interpret the effect of snow and rain, this study focuses on the prediction of solid and liquid precipitation. Therefore, the disdrometer identifications of freezing rain and of a mix of rain or drizzle and snow were aggregated with, respectively, snow and rain events. For example, if hourly precipitation has a fraction of rain and a mix of rain or drizzle and snow, it would be considered to be completely liquid after the aggregation. The selected phase aggregation aims to group the phases that are most similar in terms of hydrological influence and average occurring temperature. These assumptions are supported by an analysis of the effect of each precipitation phase on the snowpack properties (height and snow water equivalent), as detailed in Appendix B.

When solid precipitation was identified, the universal transfer functions of Kochendorfer et al. (2017) were applied to adjust for wind-induced gauge undercatch. To do so, local hourly wind speed and temperature measurements at gauge height were used, which were shown to provide appropriate corrections for sites in boreal climates (Pierre et al., 2019). The solid and liquid 15 min precipitation were then compiled at hourly time steps and were partitioned into liquid- and solid-precipitation fractions, totalling 31 905 data points. The resulting phase partitioning was used to classify the phase of each precipitation event as solid, liquid, or mixed, with respective dataset proportions of 71 %, 22 %, and 7 %.

Figure 3a shows the phase occurrence of the coupled 15 min precipitation data along an interpolated 15 min 2 m air temperature. The phase occurrence in Fig. 3b shows that it is mostly the mixed and liquid precipitation that are affected

Table 1. Study site instrument details.

Model, manufacturer	Temporal resolution	Observation	Specifications
WS100, Lufft	15 min	Precipitation phase	See Sect. 3.1 and Appendix B
Pluvio ² , OTT	15 min	Precipitation rate	Resolution: 0.1 mm Accuracy: ± 0.05 mm
CS109, Campbell Scientific	60 min	Near-surface air temperature	Accuracy: ± 0.2 °C (from 0 to 70 °C), increasing to ± 0.5 °C at -50 °C
HMP155a, Campbell Scientific	60 min	Near-surface relative humidity	Accuracy: $\pm (1.0 + 0.008 \times \text{reading})$ % RH (from -20 to 40 °C)
05103, R.M. Young	15 min	Wind speed 2 m above ground	Wind speed threshold: 1.0 m s^{-1} Accuracy: $\pm 0.3 \text{ m s}^{-1}$
SR50A, Campbell Scientific	60 min	Snow height	Resolution: 0.25 mm Accuracy: ± 1 cm or 0.4 % of distance to target
CS725, Campbell Scientific	6 h	Snowpack SWE	Resolution: 1 mm Accuracy: ± 15 mm (from 0 to 300 mm)

**Figure 3.** Event counts of (a) the 15 min precipitation phase identified by the disdrometers and (b) the aggregated hourly precipitation phases according to the 2 m air temperature.

by the aggregation and that very few freezing rain events aggregated with snow events result in the snow and solid-phase distributions being very similar. The aggregation of the mix of rain or drizzle and snow with rain results in an increase in liquid precipitation in the 0–5 °C range. Mixed-phase precipitation occurs in the same air temperature range as that of the mix of rain or drizzle and snow, suggesting that this phase is often present in mixed-phase precipitation, thus validat-

ing the aggregation. A cursory analysis of the mixed-phase precipitation events revealed that events with a phase transition between snow and a mix of rain or drizzle and snow account for roughly 75 % of the mixed-phase events. Transitions from rain to snow are infrequent and represent roughly 15 % of the mixed-phase precipitation events, while the remainder includes other phase combinations.

3.2 Model performance evaluation

The models presented in this study are evaluated for their ability to correctly predict the precipitation phase using a variety of performance metrics. First, the metrics used to quantify the predictive ability of the models are the precision (PRE) and the recall (REC), as well as the F1 score (Rokach et al., 2023). The combination of precision and recall is commonly used to evaluate model classification performance as the metrics indicate different information. The precision indicates the proportion of correct predictions for a given phase, while the recall indicates the probability of detection for a given phase. By definition, model precision and recall are inversely proportional. The assessment of both metrics informs whether a model overpredicts or underpredicts a given class. For instance, low precision and high recall indicate a class overprediction, while high precision and low recall indicate a class underprediction. Therefore, a model that achieves good performance in both metrics is desirable.

$$\text{PRE} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (3)$$

$$\text{REC} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4)$$

In the above, TP, FN, and FP are the true-positive, false-negative, and false-positive counts, respectively, for a given phase. The F1 score, being the harmonic mean of the precision and recall, is a useful metric to quantify the general performance of the model.

$$\text{F1} = 2 \left(\frac{\text{PRE} \times \text{REC}}{\text{PRE} + \text{REC}} \right) \quad (5)$$

These metrics are computed for each precipitation phase separately. A general score is also computed by weighing each phase's score according to its proportion in the dataset. As such, the weighted F1 score is used as a general classification performance metric as it combines both precision and recall and harshly penalizes a poor score in either of them while also considering the dataset imbalance.

Second, the model partitioning performance is evaluated based on the predicted solid- and liquid-precipitation amounts. The metrics used are the coefficient of determination R^2 and the RMSE. Due to the slightly asymmetric phase distribution and overlap between the phases shown in Fig. 3, different R^2 values are calculated for the solid- and liquid-precipitation. Thereby, the metrics are calculated on the phase-separated precipitation rather than on the precipitation fraction as the precipitation phase could be solid, liquid, or mixed for a given temperature. Because of the partitioning between solid or liquid precipitation, the RMSE is equal to the root mean squared of the misclassified precipitation. Therefore, the RMSE is the same for both solid and liquid precipitation, and a single score is presented. Finally, the partitioning performance metrics are performed on different subsets of the dataset using a K -fold method. The K -fold

validation method is commonly used to assess the variability of model performance with machine learning methods. By using different subsets of the dataset to train and validate the model K times, a more general performance can be assessed. Because of the fewer liquid- and mixed-precipitation events compared to solid-precipitation events, the K fold is also stratified to maintain phase proportions between training and validation sets from fold to fold. As such, in the case of the partitioning validation, the variability of the precipitation amounts from fold to fold must be considered. The performance metrics are repeated until the variance of the partitioning performance metrics stabilizes. In this case, the validation was performed with 5-fold validation and was repeated six times for a total of 30 validation folds.

3.3 Phase-guided probabilistic precipitation phase model

Machine learning algorithms are powerful tools for building classification and regression models. Random forests (Breiman, 2001) are commonly utilized in the environmental sciences due to their simple implementation and lower susceptibility to overfitting compared to other models. The model is based on decision trees, where variables are randomly chosen at each node to create a prediction. Therefore, the decision trees, each unique due to randomness, provide predictions that are ultimately aggregated to generate a final well-informed prediction.

Given the overlapping phases of the dataset, a random forest (RF) classifier is used to predict the precipitation phase with a probabilistic approach. The procedure to develop the RF model is illustrated in Fig. 4. To address the phase type imbalance, the data were adjusted by undersampling the solid phase and increasing the weight of both liquid- and mixed-precipitation phases in the dataset. To achieve this, only data points with air temperatures between -4 and 8 °C were kept in the analysis. The phase proportions resulting from the undersampling are 60 % solid, 26 % liquid, and 14 % mixed. The data were then split using an 80/20 ratio between the training and validation sets, respectively, resulting in 13 339 data points for training and 3335 for validation. To account for the prevalence of solid-precipitation samples, the training and validation sets were stratified to maintain the aforementioned phase proportions between the two subsets (60 % solid, 26 % liquid, 14 % mixed). Hyperparameters were optimized on the training set to increase the model performance and to reduce the chance of overfitting by using a stratified 5-fold cross-validation and maximizing for a weighted F1 score. The RF classifier was then retrained on the entire training dataset and was ready for use on the validation set.

While the precipitation partitioning is straightforward when the predicted phase is either solid or liquid, it is less so for a predicted mixed phase, where a solid- and liquid-precipitation fraction must be assigned. Thus, in the case of a predicted mixed phase, an RF regression model is developed

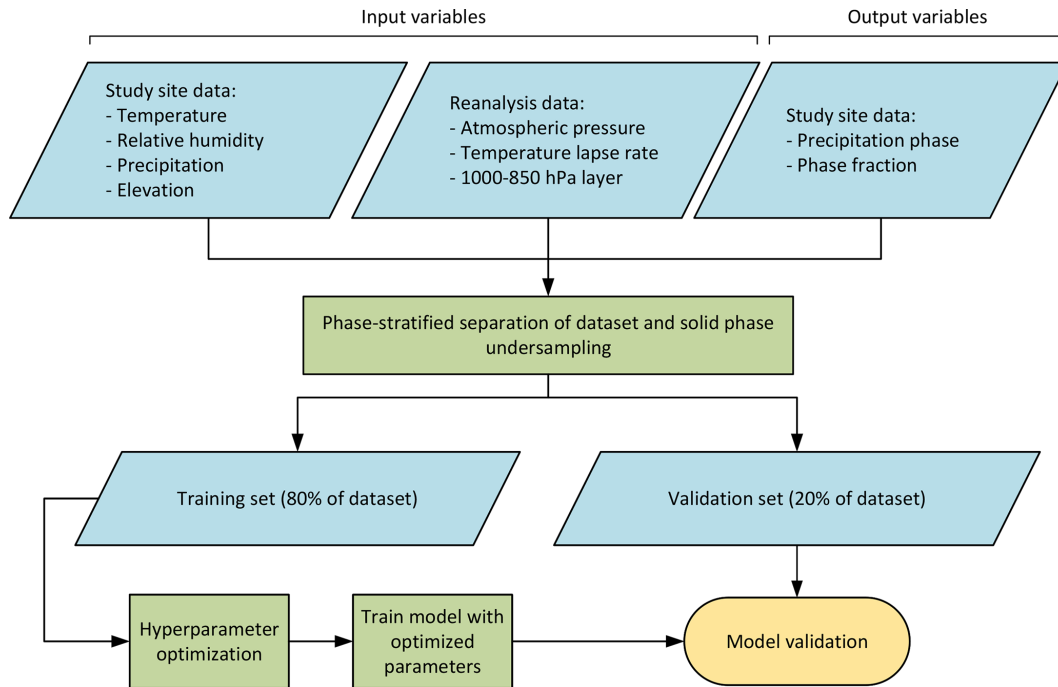


Figure 4. Development and validation methodology for the phase-guided-partitioning model.

following the same steps described above. The loss function used to optimize the regression model parameters is the mean-squared error (MSE) to increase the penalty on larger errors. The phase-guided-partitioning model predicts a precipitation phase, as well as a solid- and liquid-precipitation partitioning according to the predicted phase, with the complete process illustrated in Fig. 5.

Multiple PGP models using a combination of atmospheric variables were developed. The subsets of input variables of the PGP models accommodate different levels of data availability, ranging from the strictest minimum data requirements (e.g. in an operational context) to atmospheric variables, with each subset fully incorporating the previous subsets (see Table 2). This approach will help to quantify the impact of some atmospheric variables that are not measured at surface weather stations. The simplest model, PGP_basic, includes only 2 m air temperature and site elevation. Next, PGP_hydromet includes all related near-surface hydrometeorological data, such as relative humidity, atmospheric pressure, and precipitation rate. Finally, the PGP_full model, as discussed in the previous sections, incorporates atmospheric data from reanalysis, specifically the thickness of the 1000–850 hPa layer and the temperature lapse rate.

3.4 Benchmark phase-partitioning models

The benchmark models used for this study are common methods of increasing complexity found in hydrological models. First, a single 2 m air temperature threshold (ST) model is used as a baseline comparison. This model sepa-

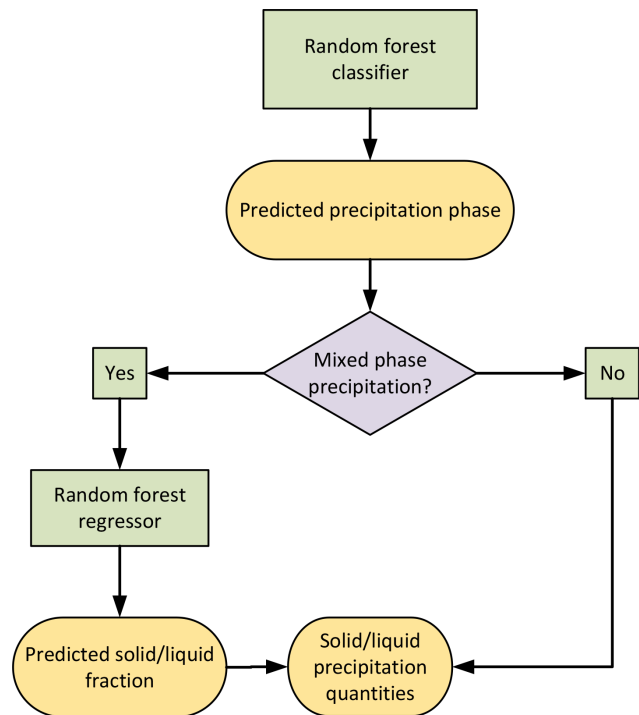


Figure 5. Phase-guided-partitioning model structure.

Table 2. Listing of the input variables used in the tested PGP models.

PGP model	Input variables
PGP_basic	2 m air temperature, elevation
PGP_hydromet	2 m air temperature, elevation, relative humidity, atmospheric pressure, precipitation rate
PGP_full	2 m air temperature, elevation, relative humidity, atmospheric pressure, precipitation rate, 1000–850 hPa layer thickness, temperature lapse rate

rates precipitation into solid and liquid phases based on a calibrated air temperature threshold. While this type of model is widely used, it is generally associated with larger partitioning errors on a seasonal basis (Harpold et al., 2017). Second, a linear-transition (LT) model is used. It allows for mixed-phase precipitation while still being of low complexity. LT partitions the solid and liquid precipitation according to a linear relationship between a snow and rain temperature threshold. Finally, the psychrometric-energy-balance (PB) model is used, which is a phase-partitioning method based on the mass energy and energy balance of a sublimating ice sphere that integrates the relative humidity to estimate the hydrometeor temperature (Harder and Pomeroy, 2013). The estimated hydrometeor temperature is then used as an input in a two-parameter curvilinear relationship. All three benchmark models are calibrated individually with the least squares method, minimizing the error with the observed solid-phase fractions. The models and their calibrated values are given in Appendix C. The precipitation phase can then be inferred from the predicted fractions. In other words, a mixed phase will be predicted in the instance of non-zero solid and liquid fractions, otherwise the predicted phase is either solid or liquid. Previous studies that employed probabilistic models based on direct phase observations (e.g. Behrangi et al., 2018; Jennings et al., 2018) were not included as benchmark models. Mixed-phase precipitation is typically excluded from such studies as there is no effective method to accurately partition the precipitation due to the categorical nature of direct phase observations. The above considerations make such models difficult to compare with the PGP models presented in this study.

3.5 Input variable importance analysis

A common way to interpret input variable importance for a machine learning model is to use permutation importance, which helps in decreasing the black-box aspect of machine learning algorithms (McGovern et al., 2019). The performance of the model is computed according to a chosen scoring scheme. Each variable of the model is then shuffled individually. The goal of this step is to break the relationship between a variable and the desired prediction. After each shuffle, a performance score is calculated to show the decrease

in model performance. This process is then repeated several times to account for data variability. Thus, the relative importance of each input variable to the model can be quantified with the resulting performance decrease. Permutation importance analysis provides only the importance of an input variable to the model and not the inherent information provided by that variable. However, when shuffling a variable that is highly correlated to another, the model can still find the shuffled variable's information when performing permutation importance analysis. In practice, this is an important consideration as it means the importance of either or both input variables can be lower. This analysis offers insight into the crucial variables for the PGP models and how they can be further improved.

4 Results

4.1 Dataset analysis

The distribution of the hydrometeorological variables categorized by precipitation phase is displayed in Fig. 6. The temperature distributions show a significant overlap between all three phases from 1.5 to 3.6 °C, similarly to that reported in Jennings et al. (2023). Mixed-precipitation probability peaks at approximately 2.4 °C. The distribution of relative humidity reveals that precipitation is associated with near-saturation water vapour conditions, with a median value of 97 %, regardless of the precipitation phase. The mean precipitation rate is generally low at 0.9 mm h⁻¹. The median precipitation rate for mixed-phase events is generally the highest at 0.8 mm h⁻¹, followed by that of liquid-phase events at 0.7 mm h⁻¹ and that of solid-phase events at 0.6 mm h⁻¹. Atmospheric pressure distributions are similar for both liquid- and mixed-phase precipitation events. The mean air pressure during the solid-precipitation events is comparable to that of the other phases, but there are more events between 90 and 92 kPa. The distribution for the thickness of the 1000–850 hPa layer closely mirrors that of air temperature, given their general correlation. The temperature lapse rate averages 4.9 °C km⁻¹, and distributions, especially of solid precipitation, show a bias toward the standard atmospheric lapse rate of 6.5 °C km⁻¹.

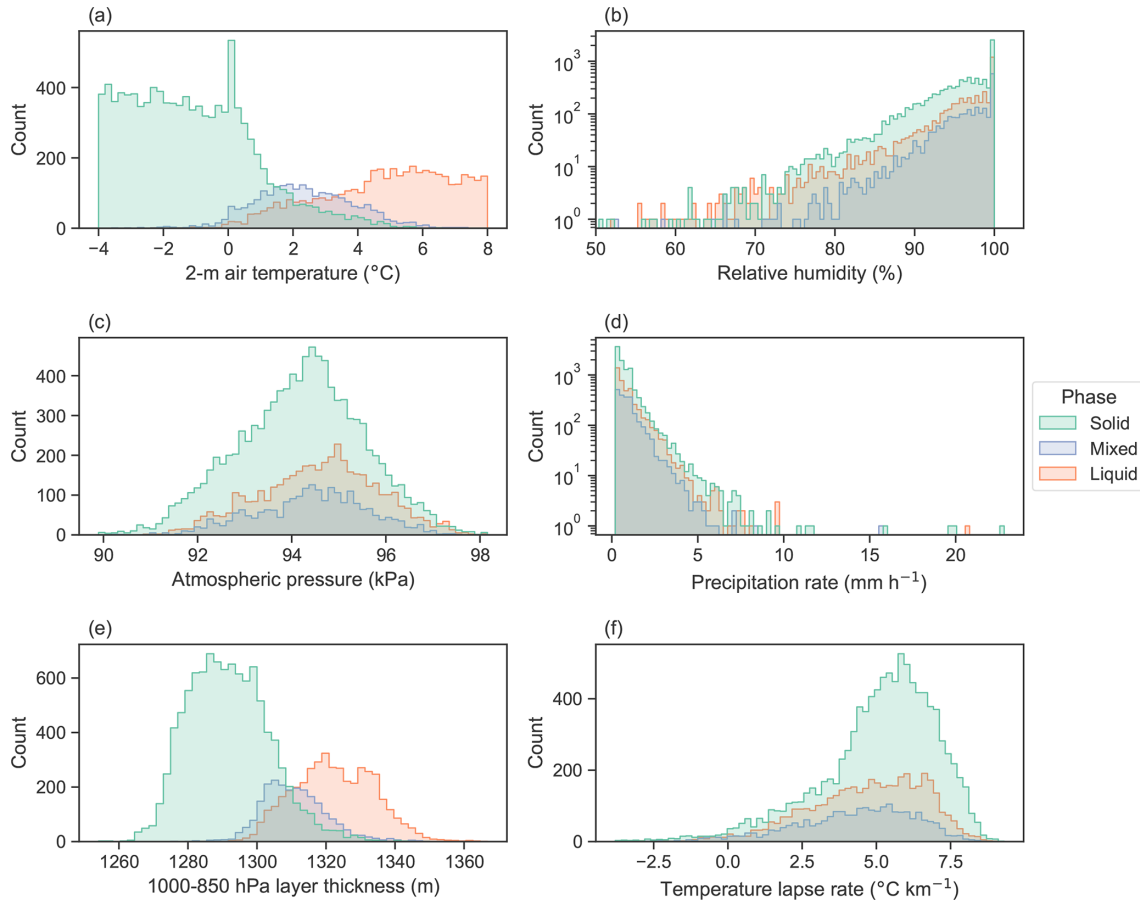


Figure 6. Input distributions separated by phase of (a) 2 m air temperature, (b) relative humidity, (c) atmospheric pressure, (d) precipitation rate, (e) 1000–850 hPa layer thickness, and (f) air temperature lapse rate.

The overlap of the phase distributions for each input variable, most notably the air temperature, indicates that a probabilistic approach is appropriate for predicting the precipitation phase. Indeed, between approximately 0 and 4 °C, solid and liquid precipitation may occur separately or coexist. According to findings in previous studies, precipitation over land is more likely to occur in a single phase than in mixed-phase precipitation (Dai, 2008; Froidurot et al., 2014), as is the case in this study, where only 13 % of the precipitation data points are mixed phase. There is, however, a narrow 2 m air temperature range, between 2 and 2.5 °C, where mixed-phase probability exceeds the probability of single-phase precipitation. An appropriate phase-partitioning model must thus accurately predict the phase in the temperature interval where solid-, liquid-, and mixed-precipitation occurrences overlap while also providing accurate partitioning when needed.

4.2 Phase classification

Figure 7 shows the phase density distribution of the benchmark models and the PGP models in comparison to the ob-

Table 3. Weighted classification scores for single-threshold (ST), linear-transition (LT), psychrometric-balance (PB), and phase-guided-partitioning PGP models. PGP model details are summarized in Table 2.

Model	F1	Precision	Recall
ST	0.74	0.71	0.79
LT	0.71	0.88	0.66
PB	0.31	0.88	0.30
PGP_basic	0.82	0.86	0.80
PGP_hydromet	0.84	0.85	0.83
PGP_full	0.84	0.84	0.84

servations. The corresponding classification scores of the models, which were weighted to reflect the precipitation phase proportions in the dataset, are presented in Table 3. The phase density distributions show the limitations of benchmark phase-partitioning models, namely that the mixed phase is absent or overrepresented compared to the observations. However, ST performs well in all three metrics due to the low likelihood of mixed-phase occurrence. When eval-

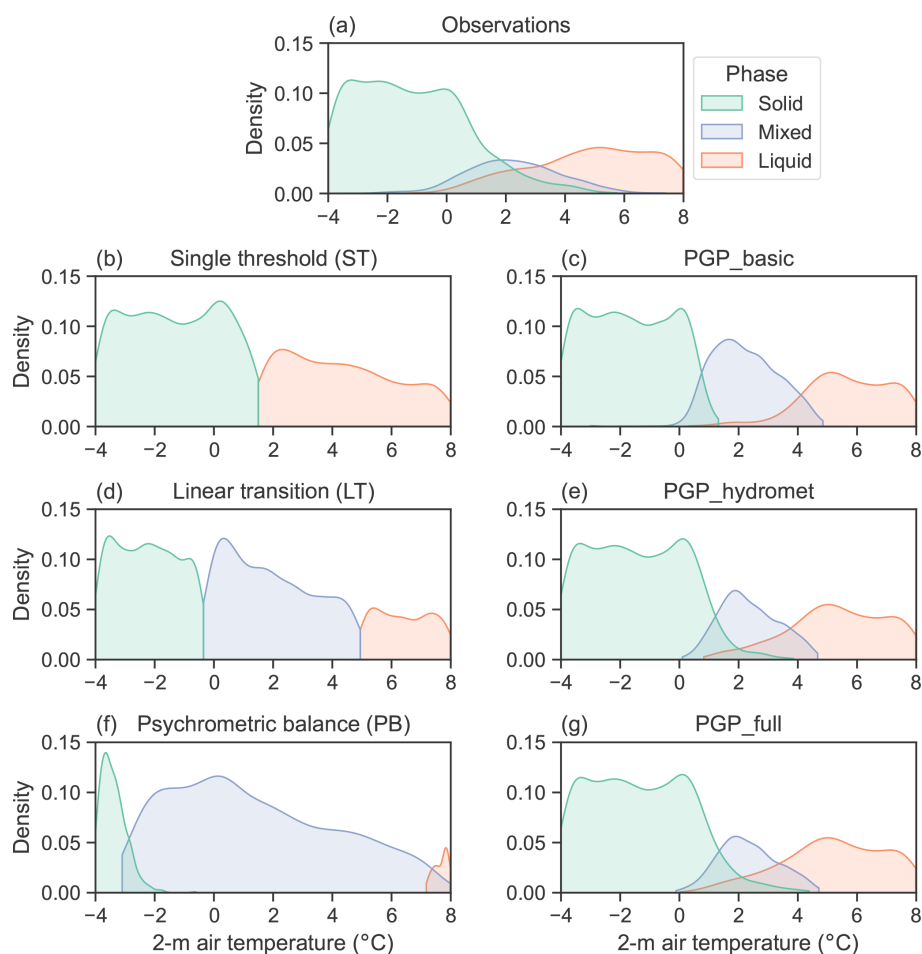


Figure 7. Hourly phase distributions according to 2 m temperature of the (a) observations, (b) single threshold, (c) PGP_basic, (d) linear transition, (e) PGP_hydromet, (f) psychrometric balance, and (g) PGP_full. PGP model details are summarized in Table 2.

uating the overall classification performance using the F1 score, LT follows ST because of a disparity between precision and recall that affects its F1 score. The lower recall score for LT can be attributed to its overprediction of the less frequent mixed phase, which, in turn, negatively affects the recall of other phases. This enhances the model's weighted precision by decreasing the number of false positives in non-mixed-phase prediction. The same reasoning can be more extensively applied to PB's weighted scores. The mixed phase's overlap with other phases significantly decreases the model's overall recall. Due to the relationships used to create the benchmark models, the overlap between all three phases is not accurately represented. By including relative humidity, PB can model phase overlap, but this does not improve the modelled phase distribution density with respect to the observations.

The weighted F1 score for the PGP models shows that they have a more robust general performance as they have high weighted precision and recall scores while having a small disparity between both scores. The PGP models reproduce

the observed phase overlap well but overpredict the mixed phase slightly, affecting both the solid- and liquid-phase predictions. PGP_basic exhibits the greatest mixed-phase overprediction, while the difference between PGP_hydromet and PGP_full is marginal. This result suggests possible improvements to PGP models, particularly for mixed-phase precipitation.

The phase-separated classification metrics provide further insight into the performance of the models, as shown in Fig. 8. The F1 score provides an overall performance for each phase prediction. PGP_full has the best F1 scores for both the solid and liquid phases, while PGP_hydromet has a higher F1 for the mixed phase. PGP_basic is generally the third best-performing model in terms of F1 score, except for the solid phase, where ST outperforms it. While it is not able to predict the mixed phase, ST has the highest scores for the liquid- and solid-precipitation phases out of the benchmark models. This is probably because mixed-phase precipitation events constitute only roughly 13 % of the samples; this low proportion does not significantly decrease the model's per-

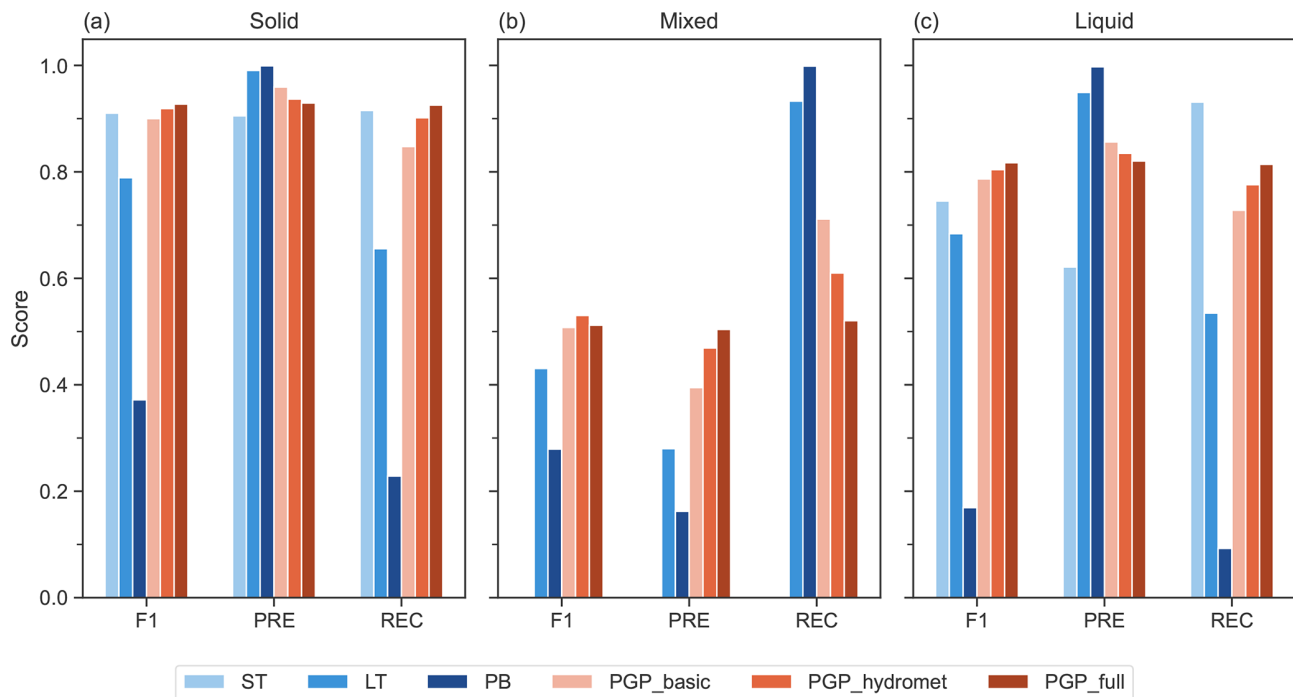


Figure 8. Model phase classification metrics separated by (a) the solid, (b) the mixed, and (c) the liquid phase. PGP model details are summarized in Table 2.

formance. LT performs slightly worse than ST for both solid- and liquid-phase F1 scores but has the highest mixed-phase F1 score out of the benchmark models.

PB's poor F1 scores are explained by the overlaps between the phases shown in Fig. 7. The model allows the predicted mixed phase to overlap with the predicted solid and liquid phases, which is the opposite behaviour compared to the observed phase density, where mixed-phase precipitation mostly exists in the solid- and liquid-phase overlap. Given the modelled phase density and the resulting classification scores, the phase prediction abilities of both LT and PB suffer from overprediction of the mixed phase. This is evidenced by the significant disparity between the precision and recall scores of LT and PB for the mixed phase. A high recall score signifies that the model minimizes the number of false negatives, which negatively affects the model's precision. Thus, overpredicting the mixed phase greatly reduces the models' precision for the mixed phase while greatly increasing their recall for the mixed phase. Conversely, the conservative prediction of the liquid and solid phases increases the precision of the model but decreases the recall for both phases.

Although they are not always the best models in terms of either precision or recall, the PGP models have the best general performance, making them more reliable for phase prediction. Thus, PGP models significantly reduce phase identification error by showing high precision and recall, with small disparities for solid- and liquid-phase prediction. However, PGP's main distinguishing feature is its general ability

to predict mixed-phase precipitation. Furthermore, the disparity between the precision and recall scores for the mixed phase is much smaller than for the other models studied, indicating that the overprediction of the mixed phase is much less severe for PGP.

4.3 Precipitation partitioning

Figure 9 displays how the regression metrics vary across validation folds. The precipitation rate variability has a significant impact on ST's performance, making its ability to partition precipitation highly variable from winter to winter. In contrast, LT and PB exhibit better performance than ST due to their ability to partition the solid and liquid phases, with much less variability in performance. The variability of R^2 for liquid precipitation is lower for LT and PB than for solid precipitation because fewer of these events occur. For all regression metrics, LT and PB have similar performances. This is most likely due to the very humid environment, which decreases the difference between the 2 m air temperature and the hydrometeor temperature computed for PB.

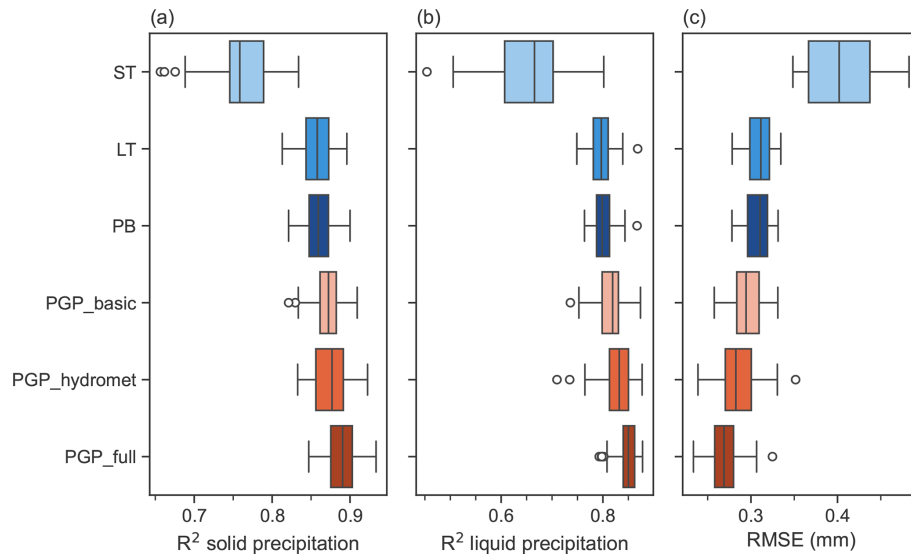


Figure 9. Model regression performance in (a) R^2 for solid precipitation, (b) R^2 for liquid precipitation, and (c) RMSE. PGP model details are summarized in Table 2.

Table 4. Average regression scores for single-threshold (ST), linear-transition (LT), psychrometric-balance (PB), and phase-guided-partitioning (PGP) models. PGP model details are summarized in Table 2.

Model	R^2 solid	R^2 liquid	RMSE (mm)
ST	0.76	0.65	0.40
LT	0.86	0.80	0.31
PB	0.86	0.80	0.31
PGP_basic	0.87	0.81	0.30
PGP_hydromet	0.88	0.83	0.29
PGP_full	0.89	0.85	0.27

The average regression metrics in Table 4 show the partitioning performance of the various models. All models have a high R^2 for solid precipitation, likely due to the abundance of solid precipitation. However, model performance decreases for liquid precipitation R^2 , with ST being significantly lower than for other models. This trend is also observed for the RMSE. While ST is the worst-performing model, LT, PB, and PGP_basic perform similarly in all regression metrics. The inclusion of hydrometeorological data in PGP_hydromet leads to a slight increase in performance. Lastly, the inclusion of atmospheric data in PGP_full improves performance compared to the other models.

Generally, the performance of PGP_basic is similar to that of LT and PB, with slight differences. PGP_basic is more variable in its performance for solid-precipitation R^2 and RMSE. This variability can be attributed to the misclassification of precipitation events due to its limited input variables. The R^2 scores for PGP_hydromet are less variable than for PGP_basic, while its RMSE is the most variable out

of the PGP models. PGP_full exhibits the lowest variability for both R^2 and is the only PGP model with RMSE variability similar to the benchmark models LT and PB.

The broader RMSE score range of the PGP models highlights the impact of misidentified phases. Misidentification can be more costly than for a benchmark model that systematically separates precipitation into solid and liquid phases for temperatures where mixed-phase events are possible. Furthermore, models such as LT and PB achieve partial accuracy in phase partitioning by forcing mixed-phase precipitation, but if a PGP model misclassifies the phase, the entire precipitation event may be incorrectly partitioned. However, PGP models do show that phase identification prior to phase partitioning can reduce the overall error of a model for both solid and liquid precipitation. This suggests that improved phase identification, specifically with mixed-phase prediction, could greatly enhance the accuracy of precipitation partitioning from PGP models.

4.4 Input variable importance

Figure 10 shows the correlation matrix of PGP_full input variables. While the correlation of most input variable combinations is low, the 2 m air temperature and 1000–850 hPa layer thickness are highly correlated. The layer thickness is affected by environmental temperatures as the air density is inversely proportional to its temperature. Therefore, as temperatures increase, the distance between two pressure levels also increases. There is a moderate negative correlation between elevation and air pressure, probably because of the small range of study site elevations. The temperature lapse rate has a small correlation with almost all features.

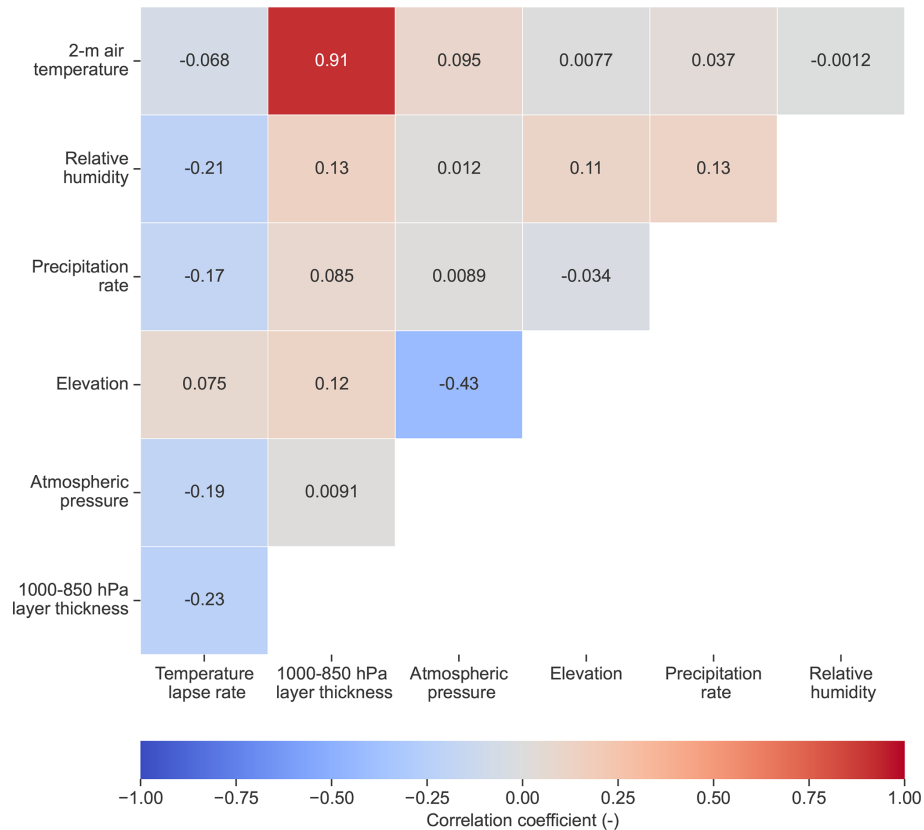


Figure 10. Correlation matrix of PGP_full input variable pairs.

The scoring scheme for permutation importance must be carefully selected according to the model and use case. In this instance, the PGP models tend to overpredict the mixed phase, which also negatively impacts their ability to predict the other phases. In turn, this also affects the models' partitioning errors, which indicates that their overall performance is reliant on accurate phase classification. For these reasons, the chosen scoring scheme for the permutation importance is the weighted F1 score, used to consider the classification of the imbalanced phase dataset.

Figure 11 shows the permutation importance of PGP_full input variables and the resulting decrease in the weighted F1 score in relation to the validation set. The 2 m air temperature is the most important variable, with its permutation decreasing the score by more than 0.2. The second most important variable is the 1000–850 hPa layer thickness, a result shared by Shin et al. (2022). Because of its high correlation with the 2 m air temperature, it is difficult to interpret the real importance of this variable. In terms of classification performance, the addition of this variable seems to provide small improvements, as shown by the differences in classification metrics between PGP_hydromet and PGP_full.

For the remaining variables, the importance decreases sharply. However, while the individual importance of the variables is low, they improve the phase classification when

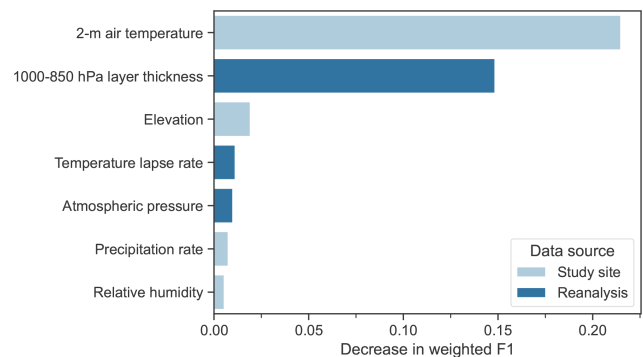


Figure 11. Permutation importance of PGP_full input variable, showing the decrease in the model's weighted F1 score.

combined. This suggests that the additional variables used in PGP_full are likely to improve mostly mixed-phase prediction, which is supported by the model's performance in Sect. 4.2. The elevation is used to approximate the atmospheric pressure of a site and can improve phase partitioning (e.g. Ding et al., 2014; Behrangi et al., 2018). Furthermore, atmospheric pressure is often cited as an important variable for phase partitioning (e.g. Behrangi et al., 2018; Dai, 2008; Jennings et al., 2018). The thinner air in low-pressure envi-

ronments allows snow to reach the ground faster. The temperature lapse rate provides key information regarding the amount of energy the hydrometeors absorb before reaching the ground.

The precipitation rate has a minor impact on the performance of the model. Nevertheless, it may hold significance for the prediction of the mixed phase. The precipitation rate is linked to the precipitation phase as it increases the energy required to completely melt falling precipitation (Froidurot et al., 2014; Thériault et al., 2010). However, its effect is minimal, most likely due to the small proportion of mixed-phase events. Finally, the model ranks relative humidity as the least important feature. This outcome is unexpected because relative humidity was shown to have a significant effect on phase partitioning (e.g. Behrangi et al., 2018; Jennings et al., 2018). One explanation could be the high percentage of data points near water vapour saturation, resulting in the variable being less regionally significant than more heterogeneous regions such as mountain ranges. Besides, this could account for the PB model's underwhelming classification accuracy since it utilizes relative humidity to determine the precipitation phase and due to the fact that it was developed for the drier climate near the Canadian Rockies.

5 Discussion

5.1 Model performance and input variable importance

The classification and regression metrics of the PGP models show that phase classification prior to phase partitioning reduces the partitioning error of solid and liquid precipitation while also providing a more reliable phase prediction than benchmark models. The use of radar-based disdrometer measurements enabled the partitioning step of the model by providing precipitation fractions for mixed-phase events, a flaw mentioned in other studies (Froidurot et al., 2014; Jennings et al., 2018). Out of the benchmark models, ST displayed the best classification performance despite not allowing for mixed-phase precipitation. The tendency of overpredicting the mixed-phase precipitation of both LT and PB reduced their overall classification performance. This general behaviour was also observed in Leroux et al. (2023), where simpler methods outperformed methods based on a precipitation phase fraction. However, ST showed significantly worse partitioning performance compared to LT and PB. The limitations of precipitation-fraction-based models are highlighted by the fact that LT and PB were the worst-performing phase classification benchmark models despite being the best partitioning benchmark models. These models are calibrated to minimize partitioning error, but, in doing so, they are biased toward predicting mixed-phase precipitation. The mixed-phase prediction of the benchmark models could be artificially constrained to reduce overprediction and to improve classification performance. Such constraints

would, however, increase the benchmark models' partitioning error, given that they were calibrated according to solid-precipitation fractions. Therefore, there is a trade-off between classification and partitioning error for precipitation-fraction-based models such as LT and PB.

PGP_basic, while showing an improvement in phase classification, did not significantly outperform the partitioning of the benchmark models LT and PB. PGP_hydromet showed improved phase classification, notably for the mixed phase, and partitioning. PGP_full showed a further increase in overall performance while also reducing the partitioning error variability. However, all PGP models tended to overpredict the mixed phase, as shown in Fig. 7. Reducing the overprediction of the mixed phase is a persistent challenge in improving precipitation phase modelling, as noted in previous studies (Casellas et al., 2021; Leroux et al., 2023).

The permutation importance analysis showed that most input variables used, apart from the 2 m air temperature, are of low importance. However, the classification performance improvements in PGP_hydromet and PGP_full show the cumulative importance of the additional variables used, most notably for mixed-phase prediction. Despite many studies demonstrating its impact on precipitation phase, relative humidity was found to be the least important factor. This is likely due to the regional homogeneity, with most observations occurring near liquid water vapour saturation. The site elevation was considered to be important for phase classification, even though it is a constant variable. This suggests that an atmospheric pressure estimated by the elevation could provide enough relevant information to improve phase classification. Still, out of the hydrometeorological variables, the atmospheric pressure had the most impact on phase classification performance. This is in line with other studies that found that it has a significant impact on the precipitation phase (e.g. Behrangi et al., 2018; Dai, 2008; Jennings et al., 2018), although this is generally to a lesser extent than relative humidity when considering the regional variability.

The precipitation rate's low importance is most likely because it affects mixed-phase prediction and thus has little impact on the overall performance. According to Thériault et al. (2010), higher precipitation rates raise the likelihood of larger hydrometeors, which require more energy to melt. Consequently, there is an increased likelihood of mixed-phase precipitation occurring in the form of partially melted hydrometeors. As noted by Feiccabrino et al. (2015), higher precipitation rates can lead to snowfall happening at warmer temperatures due to the presence of unstable air below the isothermal layer.

The permutation importance of the 1000–850 hPa layer thickness is second to the 2 m air temperature. However, because of the high correlation of the pair of variables combined with the moderate classification improvement in PGP_full, the real importance of the 1000–850 hPa layer is most likely to be low. While the importance of the temperature lapse rate is low, the partitioning results demonstrated

that incorporating gridded atmospheric variables alongside local observations led to a reduction in the variance of the regression performance. This finding is noteworthy because few studies, as pointed out by Harpold et al. (2017), have explored the impact of incorporating atmospheric reanalysis data into phase modelling. Froidurot et al. (2014) indicated that models using atmospheric data did not greatly improve the phase prediction, as is the case in this study's classification performance. Furthermore, Dai (2008) emphasizes the terrain-dependent nature of lapse rates. Thus, even though the study sites in the region are relatively similar in terms of terrain, the importance of lapse rates in the modelling process was still significant, contrarily to the fairly homogenous relative humidity measurements.

5.2 Coupled precipitation data uncertainty

There are uncertainties regarding the results due to the dataset and assumptions employed. Hydrological models commonly limit the precipitation phases to solid or liquid. Nonetheless, this dataset includes a considerable quantity of mixed snow and rain or drizzle events, and it is uncertain how hydrological models should handle this precipitation phase. The phase aggregation step considered the behaviour of the snowpack following the different phases detected by the disdrometers. Following a mix of rain or drizzle and snow, the SWE and snow height tended to decrease, a snowpack response similar to that following a rain event. It can be inferred that this type of precipitation is likely to be dominated by rainfall given the warm temperature at which it occurs and the ensuing effects on the snowpack. However, it is probable that this interpretation is specific to the disdrometers used in this study unless evidence to the contrary emerges. However, phase identification errors have the potential to introduce uncertainties into the results, notably in the case of mixed-phase precipitation. To measure this uncertainty, it is recommended that studies be conducted using collocated WS-100 disdrometers and other well-documented options such as laser disdrometers to assess the differences between ground-truth-providing instruments (Harpold et al., 2017).

Another source of uncertainty arises from the coupling of precipitation amounts and phase observations. Fehlmann et al. (2020) demonstrated that laser disdrometers have low missed-event and false-alarm rates for sub-daily integration times compared to precipitometers, but no such study was conducted with the radar-based disdrometers of this study. Additionally, the study by Fehlmann et al. (2020) was carried out in a site sheltered from the wind, implying that the wind-induced gauge undercatch could not be studied. In turn, the wind could influence the missed-event and false-alarm rates of this study's instruments. In this study, data segments where either the precipitation gauge or disdrometer did not detect any precipitation were discarded. Figure 12 displays the hit rate of both the instruments at the initial 15 min intervals and shows that the instruments' hit rates are generally

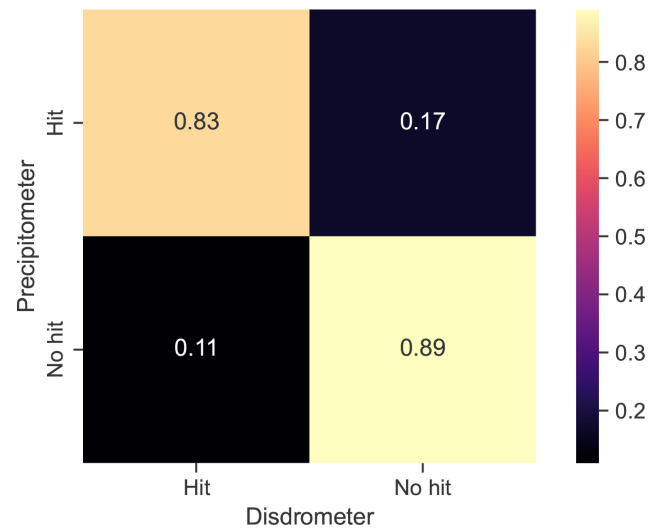


Figure 12. Confusion matrix of the precipitometer and disdrometer 15 min precipitation hit rate, normalized over the precipitometer observations. The upper-left metric is the probability of detection, the upper-right metric is the miss rate, and the lower-left metric is the false-alarm rate.

in agreement. The instrument hits were normalized over the precipitation gauge observations to compute relevant agreement metrics. The precipitation gauge is considered to be ground-truth as it would be used in conjunction with a precipitation phase model, and phase observations are rare in an operational context.

Precipitation data segments of 0.1 mm coincide with 38 % of the disdrometer misses. Assuming that a significant portion of this precipitation data would be labelled as trace amounts when resampled at the hourly interval (< 0.2 mm), the probability of missed events is likely to be lower in reality. Multiple factors could account for disagreements between the instruments, including the effect of wind, which likely varies from instrument to instrument, and the fact that certain stations are not collocated or nearby. However, studies on the environmental effects on the disdrometer performance are lacking, and this would require more detailed investigation, as outlined in Harpold et al. (2017).

Figure 13 shows the variation of the instrument agreement according to the distance between stations. The station pairs are divided into four distance categories: less than 3 km, 4 to 7 km, 7 to 12 km, and more than 12 km. Generally, the stations separated by less than 3 km show better agreement, with a few outliers. However, the instrument agreement does not seem to decrease with distance as the 4 to 7 km category exhibits the poorest agreement. Notably, the Auxlous station, separated by 28 km from the nearest weather station, has a probability of detection of 0.79, a false-alarm rate of 0.14, and a miss rate of 0.21. These instrument agreement metrics are only slightly worse than the metrics of the 15 min dataset in Fig. 12. This suggests that instrument agreement is

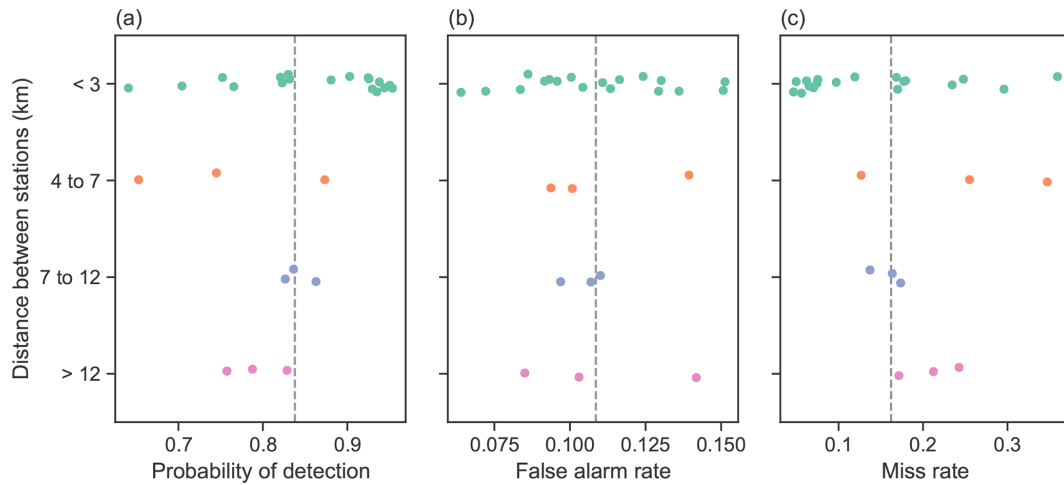


Figure 13. Comparison of the (a) probability of detection, (b) false-alarm rate, and (c) miss rate according to the distance between the study site and paired weather stations. The dashed grey line corresponds to the metric computed based on the full dataset in Fig. 12.

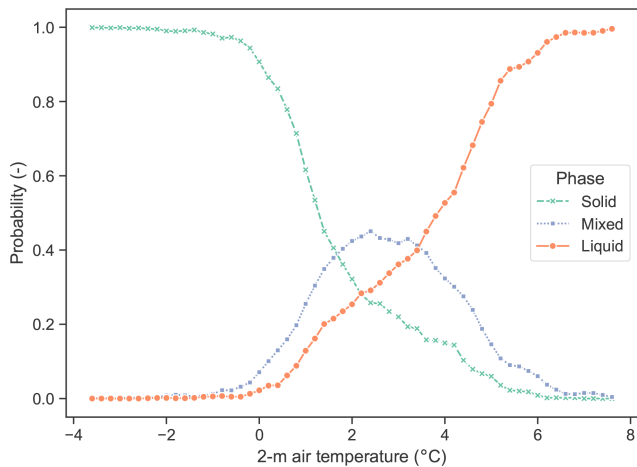


Figure 14. Observed probability of occurrence for the solid, liquid, and mixed phases.

linked to site-specific conditions rather than to distance between stations. However, by discarding data points where the instruments do not agree, we ensure that precipitation events are consistent across study sites and weather stations. In addition, the coupling of instruments from nearby stations brings the spatial scale of the observational data closer to the scale of the reanalysis data.

5.3 Data validation across studies

The phase observations from this study can be compared to other studies that use different validation data, such as direct observations (e.g. Behrangi et al., 2018; Dai, 2008; Jennings et al., 2018). However, it can be difficult to compare the phase occurrence according to the 2 m air temperature as the datasets in such studies often exclude mixed-phase precipita-

tion. Consequently, the mixed phase is usually not analyzed in detail. One method to simply compare phase-partitioning models is the critical-threshold air temperature value CT_a , which is defined as the critical temperature threshold where both the solid and liquid phases have a 50 % chance of occurrence. In the case of this study, we define a different critical threshold for the solid phase (CT_S) and the liquid phase (CT_L), as well as a temperature where the probability for the mixed phase is highest (P_m). Figure 14 shows the probability of occurrence of the phases at the study sites separated into 0.2 °C bins. The resulting thresholds are CT_S of 1.3 °C , CT_L of 3.8 °C , and P_m of 2.4 °C for a mixed-phase probability of 0.44. It is also noteworthy that P_m is roughly where the probability of solid and liquid precipitation is equal. Because of this study's aggregation step, CT_S should be similar to CT_a values from other studies as the aggregation mostly affected the probability of mixed and liquid precipitation, and CT_L will be much warmer than CT_a .

In Behrangi et al. (2018), the average hourly CT_a of 1.58 °C aligns roughly with the study's CT_S and the calibrated 2 m air temperature threshold of the benchmark ST. One of the main conclusions from the study was that the wet-bulb temperature model is more robust than the dry-bulb temperature model as the CT_a can vary significantly from site to site. Humid conditions lead to a cooler CT_a , while drier conditions have the opposite effect. As such, the CT_a for humid conditions would be approximately equal to the mean value minus the standard deviation, resulting in 1.18 °C , and is thus closer to this study's CT_S of 1.3 °C . Additionally, the upper limit of CT_a of 2.16 °C in Behrangi et al. (2018) closely matches P_m . This finding lends credibility to the disdrometer phase identification and the phase aggregation step as it indicates the temperature range in which both solid and liquid phases are possible.

In Dai (2008), the overland 3-hourly CT_a of 1.2°C is comparable to this study's CT_S despite the different time steps. The chance for the mixed phase in this study is much higher and more likely at warmer temperatures than in Dai (2008), where a peak 14.3 % chance of mixed rain and snow at 1.4°C overland is reported. However, Ding et al. (2014) have shown that the probability of mixed-phase precipitation at the daily time step greatly increases under humid conditions, particularly near saturation. Such an analysis would, however, be required at the hourly time step to confirm this behaviour. The reasoning for the increase in mixed-phase precipitation probability is that the increase in relative humidity decreases evaporative cooling and favours a transition from snow to rain. In contrast, the temperature difference between the hydrometeors and the air decreases as humidity rises, which decreases sensible heat transfer and hinders the transition from snow to rain. The relatively homogenous conditions of the study sites could explain the differences in mixed-phase precipitation probability, while the analysis in Dai (2008) lumped together a large number of stations.

The findings in Jennings et al. (2018) show a much lower 3-hourly CT_a of 0.7°C for precipitation in 90 %–100 % relative humidity and 0.9°C for precipitation occurring in 90–105 kPa humidity and pressure conditions, under which the majority of this study's precipitation occurs. The greater difference between these CT_a and CT_S values could be due to several reasons. First, the 3-hourly CT_a should theoretically be lower than the hourly CT_a . As the time step increases, the occurrence of mixed-phase precipitation increases due to the higher likelihood of a phase transition. Second, the different types of validation data could explain why CT_a is generally lower than CT_S . Phase identification errors, particularly near the solid–liquid-phase transition, could also differ between direct observations and radar disdrometers.

Overall, the radar-based disdrometer measurements are similar to the findings of previous studies, although, generally, with slightly warmer conditions of occurrence for solid precipitation. However, more research is needed to properly quantify the uncertainties associated with this type of disdrometer. In addition, models based on automated phase observations may differ from those based on direct observations, especially as the time step can vary from study to study. This also highlights the importance of the verification step performed after aggregating mixed snow and rain or drizzle with rainfall as their effect was deemed to be closer to that of rainfall.

6 Conclusion

The study used phase measurements from radar-based disdrometers to train probabilistic models to classify and partition precipitation data for a network of study sites in eastern Canada. The study sites were located in predominantly boreal climates and at similar elevations, ranging from 315 to

641 m above sea level. The mean annual 2 m air temperature was around 0.2°C , and the cumulative annual precipitation was significant at 902 mm. The humidity conditions for the data points used in the study were generally close to water vapour saturation. The utilization of automated measurements enabled partitioning of precipitation for mixed-phase events, which were previously limited to direct phase observations. The studied PGP models showed an improvement in phase partitioning with prior phase classification compared to benchmark models of varying complexity. PGP provides more accurate phase classification, which can benefit hydrological modelling at both local and watershed scales. It successfully reproduced the phase overlap between 1.5 and 3.5°C , as seen in Fig. 7, where mixed-phase probability was the highest.

The classification performances show a substantial enhancement in phase classification as opposed to benchmark models, which were designed to minimize errors in phase partitioning. Additionally, the PGP models reduced partitioning error, especially PGP_hydromet and PGP_full. However, due to prior classification, partitioning performance is highly dependent on classification performance. As a result, the less complex PGP_basic had increased error variability. According to the input variable importance analysis, atmospheric pressure was the second most important near-surface variable for phase classification. The reanalysis atmospheric data reduced the partitioning error variability of PGP_full in comparison to the other PGP models. As for relative humidity, it was deemed to be the least important hydrometeorological feature for phase classification due to the regional homogeneity of the study sites. Overall, these findings demonstrate that automated phase observations enhance PGP method development and significantly improve precipitation phase classification, even with limited hydrometeorological information. The incorporation of reanalysis atmospheric data further enhances the accuracy of local observations, pointing toward potential operational applications for such methods.

The methodology of this study could be applied to other environments, including drier conditions or a broader spectrum of environments. Further research should include a comprehensive comparison of the radar-based disdrometers used in this study with other phase validation techniques to assess potential limitations. Research is also needed to improve the prediction of the mixed phase. Other variables such as wind speed could be considered as high wind speed can have a cooling effect on precipitation. Additionally, the impact of using a model that combines both phase classification and partitioning on snowpack accumulation and basin mass and energy dynamics should be investigated.

Appendix A: Study site details

Table A1 shows the details for the study sites used from the Hydro-Québec observational network.

Table A1. List of study site coordinates, elevations, and time frames for when the station was in operation.

Station name	Latitude (° N)	Longitude (° W)	Elevation (m a.s.l.)	Operational time frame
Argent	50.776574	69.778857	641	2020–2023
Auxlouis	51.905890	70.487430	537	2020–2023
Baubert	51.409460	63.556960	541	2020–2022
Betsia_M	49.976570	69.913268	403	2019–2023
Cabituqg	49.573333	69.513611	491	2019–2023
Conrad	47.607249	74.261131	433	2020–2023
Diamand	47.231302	73.183080	373	2020–2023
Garemang	51.110640	67.139860	778	2020–2023
Hartj_G	51.779310	67.945980	460	2020–2023
Lacroi_G	51.328710	70.079390	621	2019–2023
Laflam_G	48.930225	70.270496	519	2019–2023
Lbardo_G	51.111896	67.828433	486	2019–2023
Levasseu	51.268480	68.754960	466	2020–2023
Louis	49.886620	68.489740	315	2020–2023
Louise_G	50.658526	68.839767	397	2019–2023
Moucha_M	52.125510	69.527780	565	2019–2022
Noirs	50.121160	68.831730	385	2020–2023
Parleur	51.285470	69.522370	485	2020–2021
Perdrix	50.129460	67.967450	315	2022–2023
Pipmua_G	49.360520	70.915810	566	2020–2023
Porto	49.584230	70.275910	413	2020–2023
Roussy_G	50.423470	68.094360	456	2020–2023
Rtoulnus	50.964750	67.476760	688	2020–2022
Sauterel	51.917820	63.838040	459	2020–2023
Stmarg_G	51.891980	67.046360	461	2020–2023
Wabistan	48.484572	73.441157	565	2020–2023

Appendix B: Disdrometer phase identification validation

Snow water equivalent (SWE) and snow height observations were compiled from the entire network on a winter-by-winter basis. If more than 30 % of a winter's snowpack observations were missing at a station, that winter is not included in this analysis. The resulting data subset consists of 11 winter sites, with a total of 53 520 hourly data points. The hourly data were then separated into precipitation events. The following filters were applied to the events:

- duration ≥ 3 h
- mean 2 m air temperature between -5 and 5°C
- total precipitation ≥ 0.5 mm
- mean SWE ≥ 15 mm.

This filtering step aimed to exclude short events and events that occurred either in warmer conditions, where phases other than rain are uncommon, or in the absence of snow cover detectable by the instrumentation in place. As such, 235 precipitation events were retained. In addition to the data points encompassing each event, the following hours were added until the next update of the SWE observations (at most 6 h). The events were then classified according to their main precipitation phase, i.e. the phase associated with at least half the total precipitation of the event.

The mean 2 m air temperature, SWE variation (ΔSWE), and snow height variation (ΔSH) are compiled from precipitation events according to the main precipitation phase of the event (Table B1). The effects of rain and the mix of rain or drizzle and snow events on the snowpack are similar: an SWE increase accompanied by an SH decrease. In addition, the average temperature of mixed snow and rain or drizzle events is significantly above the freezing point, where rainfall is more likely to occur than snowfall. In the case of freezing rain, the average temperature during the events is more similar to snowfall. Although freezing rain does not generally increase the SH, it contributes to the solid component of the snowpack as it freezes on contact. Thus, the phase aggregation of this study was based on the hydrological impact and temperature range of freezing rain and the mix of rain or drizzle and snow.

Table B1. Precipitation event characteristics separated by phase.

Main phase	Event count	Mean temperature ($^{\circ}\text{C}$)	ΔSWE (mm)	ΔSH (cm)
Snow	192	-2.0	3.0	3.3
Rain	12	3.6	3.5	-1.8
Mix of rain or drizzle and snow	19	1.5	5.9	-1.5
Freezing rain	12	-1.7	3.2	0.2

Appendix C: Benchmark model description

The single-threshold model (ST) used to compute the solid-precipitation fraction $f_{\text{snow}} (-)$ functions as follows:

$$f_{\text{snow}} = \begin{cases} 1 & T_a \leq T_K \\ 0 & T_a > T_K \end{cases}, \quad (C1)$$

where T_a is the temperature ($^{\circ}\text{C}$), and T_K is the calibrated temperature threshold ($^{\circ}\text{C}$). The linear-transition model (LT) uses two calibrated thresholds to calculate f_{snow} :

$$f_{\text{snow}} = \begin{cases} 1 & T_a \leq T_{\text{snow}} \\ \frac{T_{\text{rain}} - T_a}{T_{\text{rain}} - T_{\text{snow}}} & T_{\text{snow}} < T_a < T_{\text{rain}} \\ 0 & T_a \geq T_{\text{rain}} \end{cases}, \quad (C2)$$

where T_{rain} and T_{snow} are the calibrated rain and snow thresholds ($^{\circ}\text{C}$). Finally, the psychrometric-energy-balance model (PB) (Harder and Pomeroy, 2013) calculates f_{snow} as follows:

$$f_{\text{snow}} = \frac{1}{1 + b + cT_i}, \quad (C3)$$

where b and c are calibrated values, and T_i is the temperature of an unventilated hydrometeor ($^{\circ}\text{C}$). Based on the mass balance of a sublimating ice sphere, T_i is calculated iteratively with the following function:

$$T_i = T_a + L_t \frac{D}{\lambda_t} (\rho_{T_a} - \rho_{T_i}), \quad (C4)$$

where L_t is the latent heat of sublimation or vaporization (Jk^{-1}); D is the diffusivity of water vapour ($\text{m}^2 \text{s}^{-1}$); λ_t is the thermal conductivity of air ($\text{J m}^{-1} \text{s}^{-1} \text{K}^{-1}$); and ρ_{T_a} and ρ_{T_i} are the water vapour density of the surrounding air and on the hydrometeor’s surface, respectively (kg m^{-3}). The procedure to compute the variables is as detailed in Harder and Pomeroy (2013). D is computed following Thorpe and Mason (1966):

$$D = 2.06 \times 10^{-5} \left(\frac{T_a}{273.15} \right)^{1.75}. \quad (C5)$$

The vapour pressure e (kPa) is computed as in Dingman (2015):

$$e = \frac{\text{RH}}{100} \times 0.611 \exp \left(\frac{17.37T}{237.3 + T} \right), \quad (C6)$$

where RH is the relative humidity (%), and T is the air temperature ($^{\circ}\text{C}$). ρ is computed following the ideal gas law:

$$\rho = \frac{m_w e}{RT}, \quad (C7)$$

where m_w is the molecular weight of water ($0.01801528 \text{ kg mol}^{-1}$), and R is the universal gas constant ($8.31441 \text{ J mol}^{-1} \text{ K}^{-1}$). The air thermal conductivity λ_t is computed as in List (1951):

$$\lambda_t = 0.000063T_a + 0.00673. \quad (C8)$$

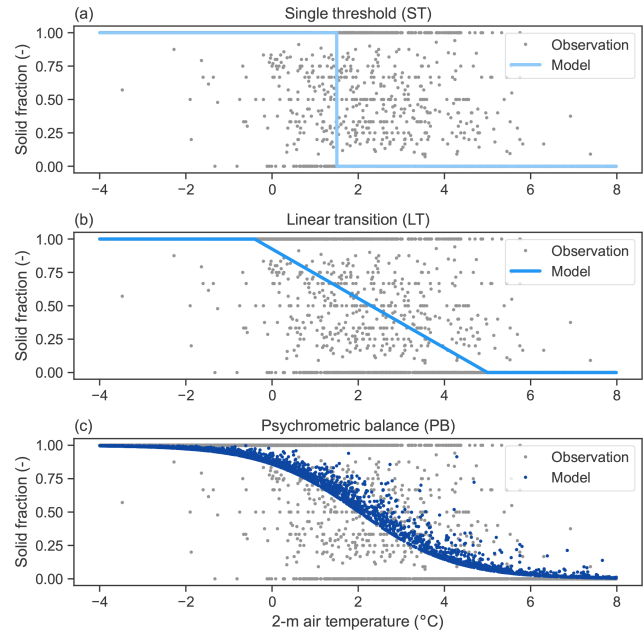


Figure C1. Observed solid-precipitation fraction according to the 2 m air temperature and the modelled solid-precipitation fraction of the (a) static threshold, (b) linear transition, and (c) psychrometric balance.

Table C1. Benchmark model calibrated parameters.

Model	Calibrated parameters
Single threshold (ST)	$T_K = 1.50$
Linear transition (LT)	$T_{\text{snow}} = -0.38$ $T_{\text{rain}} = 5.00$
Psychrometric balance (PB)	$b = 6.34$ $c = 0.39$

Finally, the latent heat of sublimation ($T_a < 0$) and vaporization ($T_a \geq 0$) is computed as follows (Yau and Rogers, 1996):

$$L_t = \begin{cases} 1000(2834.1 - 0.29T_a - 0.004T_a^2) & T_a < 0 \\ 1000(2501 - 2.31T_a) & T_a \geq 0 \end{cases}, \quad (C9)$$

Table C1 shows the calibrated parameters for the models presented in this section. The calibration was made on the same training set used for the PGP models. Figure C1 shows the simulated solid fraction for the benchmark models, as well as the observed solid fraction.

Data availability. The data used to train and validate the models in this study are available at <https://doi.org/10.5281/zenodo.10790810> (Bédard-Therrien et al., 2024). Supplementary data or models used in the analysis are available from the corresponding author upon reasonable request. All data are subject to Hydro-Québec’s Cre-

ative Commons Attribution – Non-Commercial 4.0 International licence (<https://www.hydroquebec.com/documents-data/open-data/licence.html>, last access: 12 March 2024).

Author contributions. ABT, DFN, and FA designed the research. FP and AV provided the study site data, as well as technical guidance. ABT, with the help of OC, performed the data clean-up and corrections. ABT analyzed the data and devised the precipitation-partitioning models with inputs from DFN, FA, and JMT. ABT wrote the paper with inputs from DFN, FA, and JMT, as well as comments from OC, FP, and AV.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher’s note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

Acknowledgements. We would like to express our gratitude to Hydro-Québec for making this project possible by sharing the data used for the analysis. We thank James Feiccabrino and the two anonymous reviewers for their helpful comments and insights. We also thank Mathias Bédard-Therrien for their guidance in the development of machine learning models.

Financial support. This project was financially supported by the Quebec Ministry of Public Safety (grant no. FLUTEIS-CPS18-19-26) and by the Natural Sciences and Engineering Research Council of Canada through the Alliance program (grant no. ALLRP 549108–19, “Climate projection for hydrological applications in cold regions region, EVAP-2”), with Hydro-Québec and Ouranos as industrial partners.

Review statement. This paper was edited by Shraddhanand Shukla and reviewed by James Feiccabrino and two anonymous referees.

References

Atlas, D., Srivastava, R. C., and Sekhon, R. S.: Doppler radar characteristics of precipitation at vertical incidence, *Rev. Geophys.*, 11, 1, <https://doi.org/10.1029/RG011i001p00001>, 1973.

Bédard-Therrien, A., Anctil, F., Theriault, J. M., Chalifour, O., Payette, F., Vidal, A., and Nadeau, D.: Data for “Leveraging a Disdrometer Network to Develop a Probabilistic Precipitation Phase Model in Eastern Canada”, Zenodo [data set], <https://doi.org/10.5281/zenodo.10790810>, 2024.

Behrangi, A., Yin, X., Rajagopal, S., Stampoulis, D., and Ye, H.: On distinguishing snowfall from rainfall using near-surface atmospheric information: Comparative analysis, uncertainties and hydrologic importance, *Q. J. Roy. Meteor. Soc.*, 144, 89–102, <https://doi.org/10.1002/qj.3240>, 2018.

Behrangi, A., Singh, A., Song, Y., and Panahi, M.: Assessing Gauge Undercatch Correction in Arctic Basins in Light of GRACE Observations, *Geophys. Res. Lett.*, 46, 11358–11366, <https://doi.org/10.1029/2019GL084221>, 2019.

Bourgouin, P.: A method to determine precipitation types, *Weather Forecast.*, 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), 2000.

Breiman, L.: Random forests, *Mach. Learn.*, 45, 5–32, <https://doi.org/10.1023/A:1010933404324>, 2001.

Casellas, E., Bech, J., Veciana, R., Pineda, N., Rigo, T., Miró, J. R., and Sairouni, A.: Surface precipitation phase discrimination in complex terrain, *J. Hydrol.*, 592, 125780, <https://doi.org/10.1016/j.jhydrol.2020.125780>, 2021.

Dai, A.: Temperature and pressure dependence of the rain-snow phase transition over land and ocean, *Geophys. Res. Lett.*, 35, <https://doi.org/10.1029/2008GL033295>, 2008.

Ding, B., Yang, K., Qin, J., Wang, L., Chen, Y., and He, X.: The dependence of precipitation types on surface elevation and meteorological conditions and its parameterization, *J. Hydrol.*, 513, 154–163, <https://doi.org/10.1016/j.jhydrol.2014.03.038>, 2014.

Dingman, S. L. (Ed.): *Physical hydrology*, 3rd edn., Waveland press, ISBN 10 1-4786-1118-9 ISBN 13 978-1-4786-1118-9, 2015.

Ehsani, M. R. and Behrangi, A.: A comparison of correction factors for the systematic gauge-measurement errors to improve the global land precipitation estimate, *J. Hydrol.*, 610, 127884, <https://doi.org/10.1016/j.jhydrol.2022.127884>, 2022.

Essery, R., Morin, S., Lejeune, Y., and B. Ménard, C.: A comparison of 1701 snow models using observations from an alpine site, *Adv. Water Resour.*, 55, 131–148, <https://doi.org/10.1016/j.advwatres.2012.07.013>, 2013.

Fehlmann, M., Rohrer, M., von Lerber, A., and Stoffel, M.: Automated precipitation monitoring with the Thies disdrometer: biases and ways for improvement, *Atmos. Meas. Tech.*, 13, 4683–4698, <https://doi.org/10.5194/amt-13-4683-2020>, 2020.

Feiccabrino, J. M.: Precipitation phase uncertainty in cold region conceptual models resulting from meteorological forcing time-step intervals, *Hydrol. Res.*, 51, 180–187, <https://doi.org/10.2166/nh.2020.080>, 2020.

Feiccabrino, J., Gustafsson, D., and Lundberg, A.: Surface-based precipitation phase determination methods in hydrological models, *Hydrol. Res.*, 44, 44–57, <https://doi.org/10.2166/nh.2012.158>, 2013.

Feiccabrino, J., Graff, W., Lundberg, A., Sandström, N., and Gustafsson, D.: Meteorological knowledge useful for the improvement of snow rain separation in surface based models, *Hydrology*, 2, 266–288, <https://doi.org/10.3390/hydrology2040266>, 2015.

Froidurot, S., Hingray, B., Zin, I., and Gautheron, A.: Sensitivity of Precipitation Phase over the Swiss Alps to Different Meteorological Variables, *J. Hydrometeorol.*, 15, 685–696, <https://doi.org/10.1175/JHM-D-13-073.1>, 2014.

Garcia-Benadi, A., Bech, J., Gonzalez, S., Udina, M., Codina, B., and Georgis, J.-F.: Precipitation type classification of micro rain radar data using an improved doppler

- spectral processing methodology, *Remote Sens.*, 12, 4113, <https://doi.org/10.3390/rs12244113>, 2020.
- Gunn, R. and Kinzer, G. D.: The terminal velocity of fall for water droplets in stagnant air, *J. Meteorol.*, 6, 243–248, [https://doi.org/10.1175/1520-0469\(1949\)006<0243:TTVOFF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1949)006<0243:TTVOFF>2.0.CO;2), 1949.
- Günther, D., Marke, T., Essery, R., and Strasser, U.: Uncertainties in Snowpack Simulations—Assessing the Impact of Model Structure, Parameter Choice, and Forcing Data Error on Point-Scale Energy Balance Snow Model Performance, *Water Resour. Res.*, 55, 2779–2800, <https://doi.org/10.1029/2018WR023403>, 2019.
- Harder, P. and Pomeroy, J.: Estimating precipitation phase using a psychrometric energy balance method, *Hydrol. Process.*, 27, 1901–1914, <https://doi.org/10.1002/hyp.9799>, 2013.
- Harder, P. and Pomeroy, J. W.: Hydrological model uncertainty due to precipitation – phase partitioning methods, *Hydrol. Process.*, 28, 4311–4327, <https://doi.org/10.1002/hyp.10214>, 2014.
- Harpold, A. A., Kaplan, M. L., Klos, P. Z., Link, T., McNamara, J. P., Rajagopal, S., Schumer, R., and Steele, C. M.: Rain or snow: hydrologic processes, observations, prediction, and research needs, *Hydrol. Earth Syst. Sci.*, 21, 1–22, <https://doi.org/10.5194/hess-21-1-2017>, 2017.
- Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., and Thépaut, J.-N.: ERA5 hourly data on pressure levels from 1940 to present, Copernicus Climate Change Service (C3S) Climate Data Store (CDS) [data set], <https://doi.org/10.24381/cds.bd0915c6>, 2023.
- Jennings, K. S., Winchell, T. S., Livneh, B., and Molotch, N. P.: Spatial variation of the rain-snow temperature threshold across the Northern Hemisphere, *Nat. Commun.*, 9, 1148, <https://doi.org/10.1038/s41467-018-03629-7>, 2018.
- Jennings, K. S., Arienzo, M. M., Collins, M., Hatchett, B. J., Nolin, A. W., and Aggett, G.: Crowdsourced Data Highlight Precipitation Phase Partitioning Variability in Rain-Snow Transition Zone, *Earth Space Sci.*, 10, e2022EA002714, <https://doi.org/10.1029/2022EA002714>, 2023.
- Jeong, D. I. and Sushama, L.: Rain-on-snow events over North America based on two Canadian regional climate models, *Clim. Dynam.*, 50, 303–316, <https://doi.org/10.1007/s00382-017-3609-x>, 2018.
- Kneifel, S., Maahn, M., Peters, G., and Simmer, C.: Observation of snowfall with a low-power FM-CW K-band radar (Micro Rain Radar), *Meteorol. Atmos. Phys.*, 113, 75–87, <https://doi.org/10.1007/s00703-011-0142-z>, 2011.
- Kochendorfer, J., Nitu, R., Wolff, M., Mekis, E., Rasmussen, R., Baker, B., Earle, M. E., Reverdin, A., Wong, K., Smith, C. D., Yang, D., Roulet, Y.-A., Buisan, S., Laine, T., Lee, G., Aceituno, J. L. C., Alastrué, J., Isaksen, K., Meyers, T., Brækkan, R., Landolt, S., Jachcik, A., and Poikonen, A.: Analysis of single-Altitude-shielded and unshielded measurements of mixed and solid precipitation from WMO-SPICE, *Hydrol. Earth Syst. Sci.*, 21, 3525–3542, <https://doi.org/10.5194/hess-21-3525-2017>, 2017.
- Leroux, N. R., Vionnet, V., and Thériault, J. M.: Performance of precipitation phase partitioning methods and their impact on snowpack evolution in a humid continental climate, *Hydrol. Process.*, 37, e15028, <https://doi.org/10.1002/hyp.15028>, 2023.
- List, R. J. (Ed.): Smithsonian meteorological tables, Smithsonian miscellaneous collections, Smithsonian Institution Press, 114, 1–527, 1951.
- Locatelli, J. D. and Hobbs, P. V.: Fall speeds and masses of solid precipitation particles, *J. Geophys. Res.*, 79, 2185–2197, <https://doi.org/10.1029/JC079i015p02185>, 1974.
- Löffler-Mang, M., Kunz, M., and Schmid, W.: On the Performance of a Low-Cost K-Band Doppler Radar for Quantitative Rain Measurements, *J. Atmos. Ocean. Tech.*, 16, 379–387, [https://doi.org/10.1175/1520-0426\(1999\)016<0379:OTPOAL>2.0.CO;2](https://doi.org/10.1175/1520-0426(1999)016<0379:OTPOAL>2.0.CO;2), 1999.
- Marks, D., Winstral, A., Reba, M., Pomeroy, J., and Kumar, M.: An evaluation of methods for determining during-storm precipitation phase and the rain/snow transition elevation at the surface in a mountain basin, *Adv. Water Resour.*, 55, 98–110, <https://doi.org/10.1016/j.advwatres.2012.11.012>, 2013.
- McGovern, A., Lagerquist, R., John Gagne, D., Jergensen, G. E., Elmore, K. L., Homeyer, C. R., and Smith, T.: Making the Black Box More Transparent: Understanding the Physical Implications of Machine Learning, *B. Am. Meteorol. Soc.*, 100, 2175–2199, <https://doi.org/10.1175/BAMS-D-18-0195.1>, 2019.
- Muñoz Sabater, J.: ERA5-Land hourly data from 1950 to present, Copernicus Climate Change Service (C3S) Climate Data Store (CDS) [data set], <https://doi.org/10.24381/cds.e2161bac>, 2019.
- Musselman, K. N., Lehner, F., Ikeda, K., Clark, M. P., Prein, A. F., Liu, C., Barlage, M., and Rasmussen, R.: Projected increases and shifts in rain-on-snow flood risk over western North America, *Nat. Clim. Change*, 8, 808–812, <https://doi.org/10.1038/s41558-018-0236-4>, 2018.
- Pierre, A., Jutras, S., Smith, C., Kochendorfer, J., Fortin, V., and Anctil, F.: Evaluation of Catch Efficiency Transfer Functions for Unshielded and Single-Altitude-Shielded Solid Precipitation Measurements, *J. Atmos. Ocean. Tech.*, 36, 865–881, <https://doi.org/10.1175/JTECH-D-18-0112.1>, 2019.
- Rasmussen, R., Baker, B., Kochendorfer, J., Meyers, T., Landolt, S., Fischer, A. P., Black, J., Thériault, J. M., Kucera, P., Gochis, D., Smith, C., Nitu, R., Hall, M., Ikeda, K., and Gutmann, E.: How Well Are We Measuring Snow: The NOAA/FAA/NCAR Winter Precipitation Test Bed, *B. Am. Meteorol. Soc.*, 93, 811–829, <https://doi.org/10.1175/BAMS-D-11-00052.1>, 2012.
- Rokach, L., Maimon, O., and Shmueli, E.: Machine Learning for Data Science Handbook, 3, Springer Cham, <https://doi.org/10.1007/978-3-031-24628-9>, 2023.
- Ross, A., Smith, C. D., and Barr, A.: An improved post-processing technique for automatic precipitation gauge time series, *Atmos. Meas. Tech.*, 13, 2979–2994, <https://doi.org/10.5194/amt-13-2979-2020>, 2020.
- Sarkar, T., Das, S., and Maitra, A.: Assessment of different rain-drop size measuring techniques: Inter-comparison of Doppler radar, impact and optical disdrometer, *Atmos. Res.*, 160, 15–27, <https://doi.org/10.1016/j.atmosres.2015.03.001>, 2015.
- Shin, K., Kim, K., Song, J. J., and Lee, G.: 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>, 2022.
- Singh, P., Spitzbart, G., Hübl, H., and Weinmeister, H. W.: Hydrological response of snowpack under rain-on-snow events: a

- field study, *J. Hydrol.*, 202, 1–20, [https://doi.org/10.1016/S0022-1694\(97\)00004-8](https://doi.org/10.1016/S0022-1694(97)00004-8), 1997.
- Smith, C. D., Mekis, E., Hartwell, M., and Ross, A.: The hourly wind-bias-adjusted precipitation data set from the Environment and Climate Change Canada automated surface observation network (2001–2019), *Earth Syst. Sci. Data*, 14, 5253–5265, <https://doi.org/10.5194/essd-14-5253-2022>, 2022.
- Stewart, R. E.: Precipitation types in the transition region of winter storms, *B. Am. Meteorol. Soc.*, 73, 287–296, [https://doi.org/10.1175/1520-0477\(1992\)073<0287:PTITTR>2.0.CO;2](https://doi.org/10.1175/1520-0477(1992)073<0287:PTITTR>2.0.CO;2), 1992.
- Thériault, J. M. and Stewart, R. E.: A Parameterization of the Microphysical Processes Forming Many Types of Winter Precipitation, *J. Atmos. Sci.*, 67, 1492–1508, <https://doi.org/10.1175/2009JAS3224.1>, 2010.
- Thériault, J. M., Stewart, R. E., and Henson, W.: On the Dependence of Winter Precipitation Types on Temperature, Precipitation Rate, and Associated Features, *J. Appl. Meteorol. Climatol.*, 49, 1429–1442, <https://doi.org/10.1175/2010JAMC2321.1>, 2010.
- Thorpe, A. and Mason, B.: The evaporation of ice spheres and ice crystals, *Brit. J. Appl. Phys.*, 17, 541, <https://doi.org/10.1088/0508-3443/17/4/316>, 1966.
- Wang, Y. H., Broxton, P., Fang, Y., Behrangi, A., Barlage, M., Zeng, X., and Niu, G. Y.: A Wet-Bulb Temperature-Based Rain-Snow Partitioning Scheme Improves Snowpack Prediction Over the Drier Western United States, *Geophys. Res. Lett.*, 46, 13825–13835, <https://doi.org/10.1029/2019GL085722>, 2019.
- Wayand, N. E., Stemberis, J., Zagrodnik, J. P., Mass, C. F., and Lundquist, J. D.: Improving simulations of precipitation phase and snowpack at a site subject to cold air intrusions: Snoqualmie Pass, WA, *J. Geophys. Res.-Atmos.*, 121, 9929–9942, <https://doi.org/10.1002/2016JD025387>, 2016.
- Wen, L., Nagabhatla, N., Lü, S., and Wang, S.-Y.: Impact of rain snow threshold temperature on snow depth simulation in land surface and regional atmospheric models, *Adv. Atmos. Sci.*, 30, 1449–1460, <https://doi.org/10.1007/s00376-012-2192-7>, 2013.
- Wever, N., Würzer, S., Fierz, C., and Lehning, M.: Simulating ice layer formation under the presence of preferential flow in layered snowpacks, *The Cryosphere*, 10, 2731–2744, <https://doi.org/10.5194/tc-10-2731-2016>, 2016.
- WMO: 6.1.2 Units and scales, in: *Measurement of Meteorological Variables*, 2023 edition ed., Guide to instruments and methods of observation, Volume 1, WMO, Geneva, 574, 2018.
- Yau, M. K. and Rogers, R. R.: *A short course in cloud physics*, 3rd edn., Elsevier, ISBN 9780080570945, 1996.
- Ye, H., Yang, D., and Robinson, D.: Winter rain on snow and its association with air temperature in northern Eurasia, *Hydrol. Process.*, 22, 2728–2736, <https://doi.org/10.1002/hyp.7094>, 2008.
- Ye, H., Cohen, J., and Rawlins, M.: Discrimination of Solid from Liquid Precipitation over Northern Eurasia Using Surface Atmospheric Conditions, *J. Hydrometeorol.*, 14, 1345–1355, <https://doi.org/10.1175/JHM-D-12-0164.1>, 2013.