Machine Learning Models with Uncertainty Quantification for Precipitation Type Prediction

NCAR
NATIONAL CENTER FOR ATMOSPHERIC RESEARCH

Dhamma Kimpara, Belen Saavedra
Charlie Becker, David John Gagne, Gabrielle Gantos, John S. Schreck

Problem Statement

Winter precipitation type greatly affects impacts of winter storms

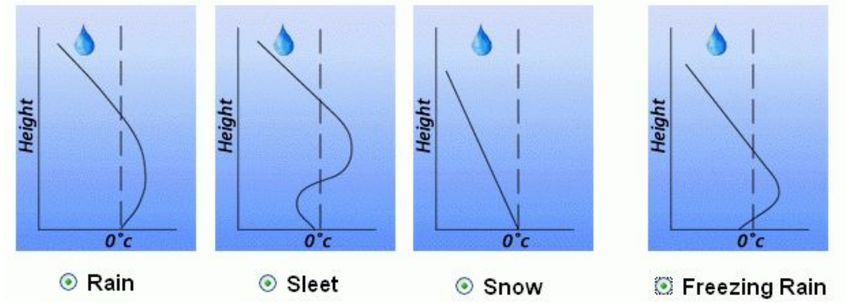


Figure 1: Prototypical sounding profiles for each winter precip. type

Data Sources

- Crowdsourced observations of precip. Type
- Rapid Refresh (RAP) Model
- Automated Surface Observing Stations

Modeling Goal: Capture Aleatoric and Epistemic Uncertainty

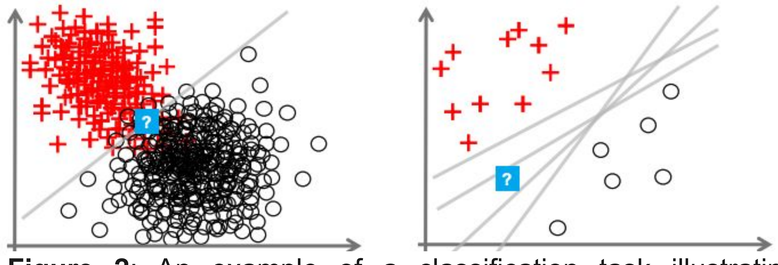
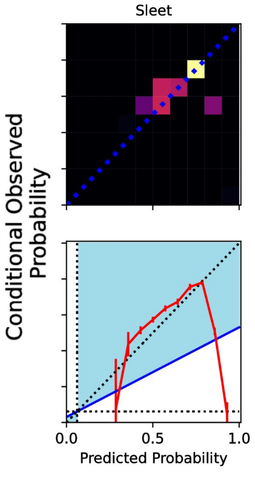


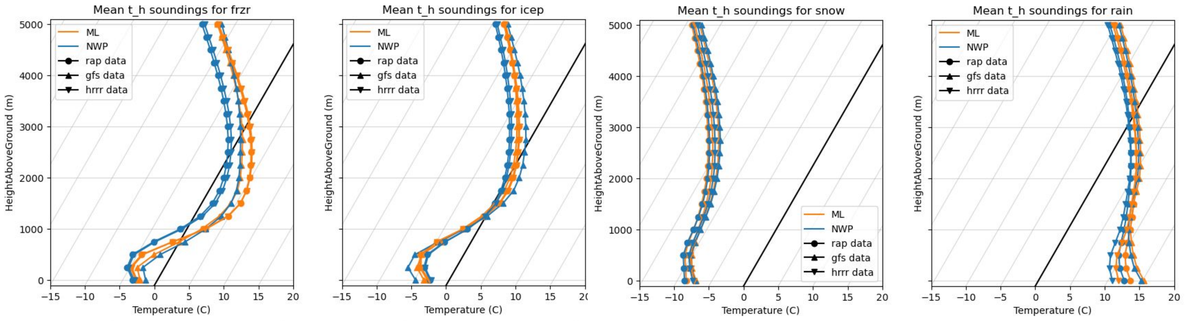
Figure 2: An example of a classification task illustrating aleatoric (inherent data uncertainty, left) vs epistemic uncertainty (lack of data, right).



Salient Issue: miscalibration for sleet

Figure 3: Calibration curves for the ML model on sleet predictions. The ML model does extremely poorly when predicting sleet with a prediction confidence > 0.7 (high confidence). All other calibration curves for the other precip. types are reasonable.

Evaluation of ML Models



	rap	gfs	hrrr
ML frac_abv_0	0.97	0.96	0.96
NWP frac_abv_0	0.79	0.99	0.78
ML num_obs	2.1E+06	7.2E+04	3.7E+07
NWP num_obs	3.5E+06	6.1E+04	6.2E+07

	rap	gfs	hrrr
ML frac_abv_0	0.71	0.66	0.72
NWP frac_abv_0	0.89	0.94	0.89
ML num_obs	3.0E+06	1.3E+05	5.6E+07
NWP num_obs	8.7E+05	7.0E+04	1.6E+07

	rap	gfs	hrrr
ML frac_abv_0	0.10	0.11	0.11
NWP frac_abv_0	0.08	0.10	0.10
ML num_obs	5.2E+07	1.7E+06	6.7E+08
NWP num_obs	5.2E+07	1.7E+06	6.7E+08

	rap	gfs	hrrr
ML frac_abv_0	1.00	1.00	1.00
NWP frac_abv_0	1.00	1.00	1.00
ML num_obs	3.3E+07	1.7E+06	4.2E+08
NWP num_obs	3.6E+07	1.8E+06	4.6E+08

Figure 4: Mean temperature sounding profiles of various models. Our ML model outperforms NWPs for freezing rain. However, for sleet (icep), ML only outperforms the RAP and HRRR. Snow and rain are highly accurate. (Statistics computed over 3 TB of data)

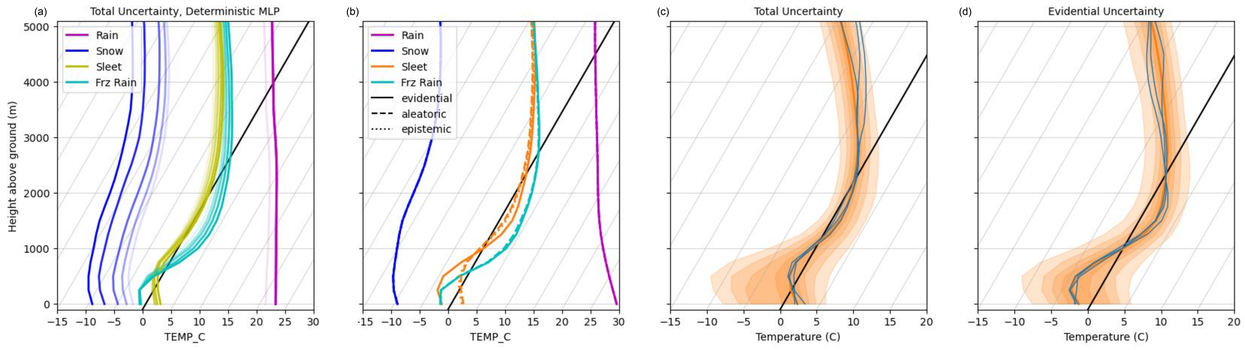


Figure 5: (a) Composite soundings are presented using the median of the 10th-90th percentile least uncertain predictions from the Monte-Carlo Dropout Method. (b) showcases the breakdown of the top 10% least uncertain composites (depicting the median) based on class and uncertainty from the evidential model. (c,d) illustrate the densities up to the 80% coverage region for top 10% least uncertain sleet predictions.

Main cause of miscalibration: Data quality

Figure 6: Density plots of composite soundings binned by crossings-of-0C characteristics. Row (a) are the soundings with sleet as the true label. Row (b) are the soundings where ML had a prediction confidence of 0.7 but was wrong. “One crossing” means a crossing from negative to positive temperature. All of the issues with calibration of sleet predictions can be explained by this issue of data quality.