UAlbany CSTAR Winter Precipitation Type Probabilistic Forecasts
Product Description Document (PDD)
December 2022

Part 1: Background and Goals

A. **Product Description-** The UAlbany CSTAR Winter Precipitation Type Probabilistic Forecasts are based on a machine learning model (Random Forest) that processes both observational data and numerical weather model output from across New York State. These forecasts provide additional insight about precipitation type by processing data using a new methodology as well as by combining multiple data sources to allow forecasters to examine the information together instead of separately. The probabilities are displayed alongside radar or model reflectivity to indicate where areas of precipitation are occurring. Probabilities represent **if** precipitation is falling, what is the chance it is a specific precipitation type. Probabilities shown in the products can either be for a specific precipitation type (Figure A) or as the dominant precipitation type at a specific location (Figure B). Examples of each type of product are shown below:

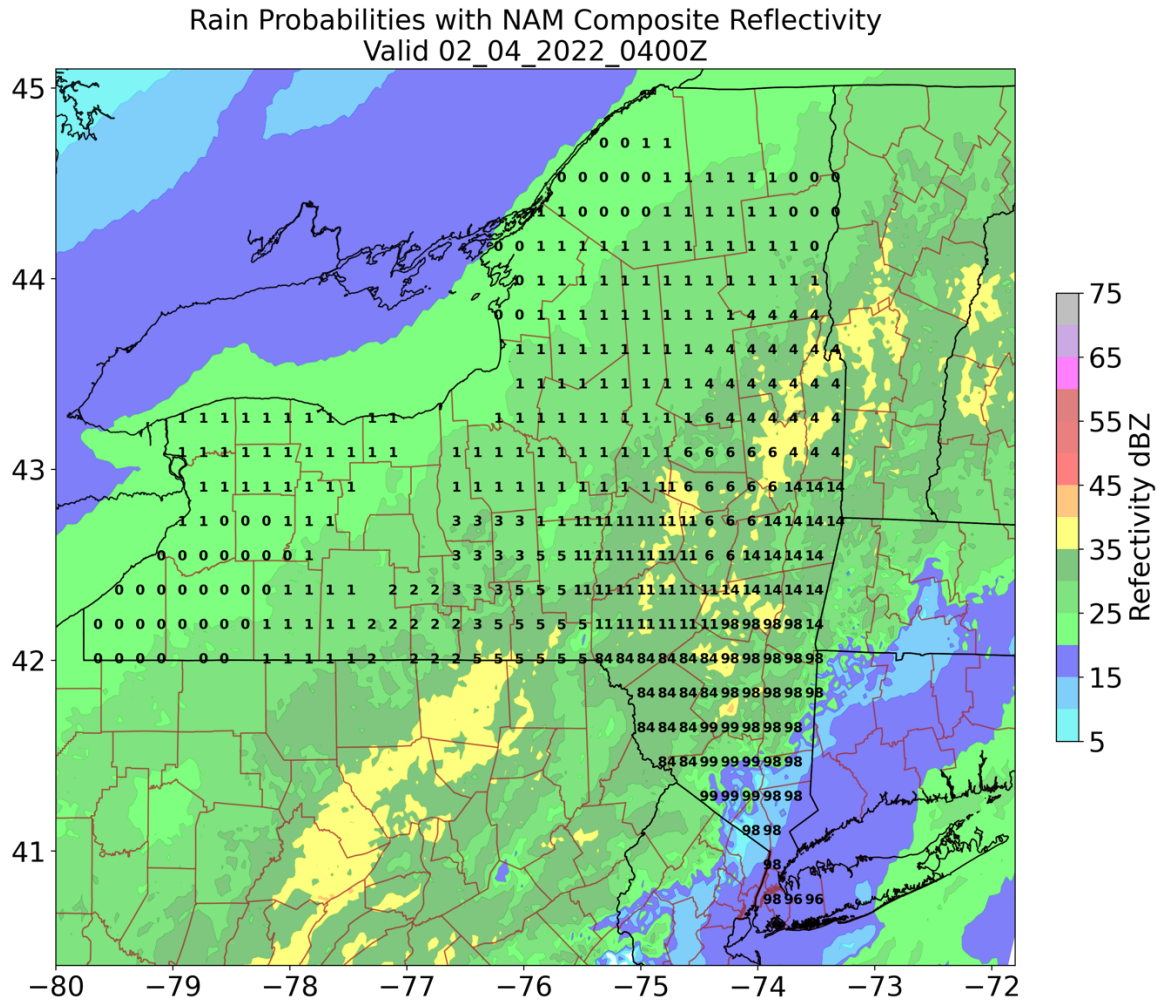

Figure A: Map of only Rain probabilities from NAMNEST dataset

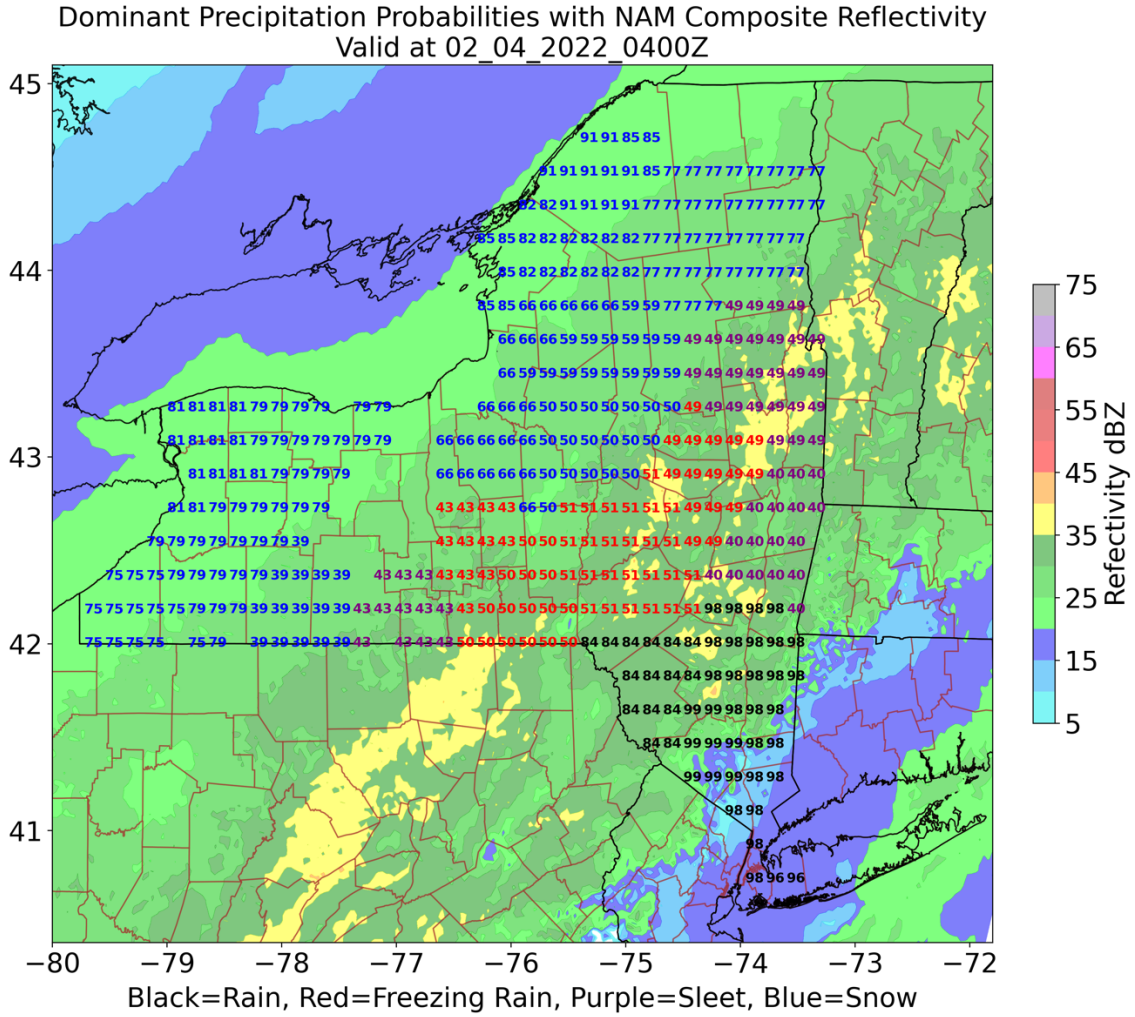

Figure B: Map of dominant precipitation type probabilities from the NAMNEST dataset. Probabilities are color coded by precipitation type.

B. **Purpose-** The goal of the UAlbany CSTAR Winter Precipitation Type Probabilistic Forecasts is to provide additional information about the likelihood of winter precipitation types occurring through synthesizing multiple data sources. This work can allow forecasters to see and understand where there could be more or less certainty for locations of different precipitation types. This can allow them to spend more time focusing in on the regions where the forecast is more challenging. In addition, it can help forecasters view multiple sources of information processed together which can form different perspectives on what precipitation type can be occurring.

C. **Audience-** These products are intended for WFOs and other NWS forecasters that have areas of interest in New York. It can be helpful as well for neighboring WFOs to get a sense of what could happen before it reaches their CWAs.

D. **Presentation Format-** The experimental products are available across all of New York as web-based images located here: https://www.atmos.albany.edu/student/filipiak/op/index.html. There are 4 different base datasets used to analyze and make forecasts from, and they update on a variety of timescales. Products are made between November 1st and March 31st. Below is a list and table for information about when new forecasts become available.

- New York State Mesonet (NYSM) and Upper-Air: Available Hourly between 0-4 UTC and 12-16 UTC
- NYSM and HRRR: Available Hourly

Model Initialization	0Z	3Z	6Z	9Z	12Z	15Z	18Z	21Z
RF NAMNEST Available	4Z	N/A	10Z	N/A	16Z	N/A	22Z	N/A
RF HRRR Available	3Z	6Z	9Z	12Z	15Z	18Z	21Z	0Z+1DAY

E. **Contact and Additional Information-** For additional information, please visit the following links:

- VLAB Quick Reference:
- Brian Filipiak's Thesis:
- Operational Website: https://www.atmos.albany.edu/student/filipiak/op/

Any further questions or comments can be directed to:

Brian Filipiak
bfilipiak@albany.edu

Kristen Corbosiero
kcorbosiero@albany.edu

Ross Lazear
rlazear@albany.edu

Part 2: Technical Description

A. Data Sources- For these products, 4 different data sources were used to create the different datasets that were run through the random forest algorithm to create forecasts. NYSM observations were used as surface data in the nowcasting datasets. Upper-air radiosonde profiles were matched with NYSM observations to create one nowcasting dataset. NAMNEST forecast data is used to make forecasts with 12 hours of lead time. HRRR forecast data is used to make forecasts with 12 hours of lead time and is used in combination with NYSM data for nowcasting products.

B. **Methodology-** Community Collaborative Rain, Hail and Snow (CoCoRaHS) Network Reports were used to identify mixed precipitation (freezing rain and sleet) events in New York between January 2017 and September 2020. In addition to the mixed precipitation events, rain and snow events during the cold season (October-April) over the same time period were also identified using CoCoRaHS reports. The cases identified with one of the 4 precipitation types (freezing rain, sleet, rain and snow) were examined and verified using New York State Mesonet, radar, and satellite observations. This process resulted in about 2600 verified cases forming the training dataset to make all future predictions from.

   Data to help train the machine learning algorithm was incorporated into the CoCoRaHS reports through locating the nearest Mesonet site, radiosonde site, HRRR vertical profile site, and NAMNEST vertical profile site and using the data they provide at the nearest time to the event. Machine learning was applied to the verified CoCoRaHS reports to generate probabilistic values for each precipitation type in nowcasts or forecasts. The random forest algorithm is an ensemble decision tree machine learning method where numerous decision trees are used to make a prediction based on previous learned data. In our algorithm, the random forest classifier method is used to generate predictions for each class or type of precipitation. This means that the predictions made by the random forest represents what type of precipitation is currently or forecasted to be occurring as opposed to how much of that precipitation will fall. The probabilities for each location are based on each tree in the forest voting for a precipitation type and coming to a consensus. The random forest is run 50 times, so the final displayed probability is the average across the 50 runs.

C. **Product Guidance-** Forecasts are made on a 20-km grid across New York where each of the 4 data sources are matched appropriately at each location. The final matched datasets given to the random forest are filtered by the 95th percentile 2-meter or surface temperature because at this point, the datasets are extreme cases and are not suitable to be used for accurate forecasts. For the NYSM and Upper Air data, NAMNEST data, and HRRR data, this threshold is about 45°F, and for the NYSM and HRRR data, it is about 42°F.

   The output of the random forest probabilities is plotted on the maps of New York and surrounding states with county outlines and radar or model reflectivity overlay. A blank map or blank areas on a map indicate that there is missing data or that those points have been filtered out because of temperature.