673 Correcting Errors in Mesoscale Convective System Quantitative Precipitation Forecasts of Intensity, Timing, and Location Using Machine Learning Algorithms

Tyreek Frazier, Aniruddha Pathak, William A. Gallus Jr., Somak Dutta, Kristie Franz

IOWA STATE UNIVERSITY

Introduction

- Convection-allowing models have been shown to improve simulations of heavy rain with warm-season convection
  - They improve the skill in quantitative precipitation forecasts (QPF), but still struggle with precipitation intensity, timing, and location which negatively impacts flood forecasts
  - The goal of our research is to create a machine learning (ML) based post-processor for ensemble QPF that will predict likely errors to improve uncertainty estimates

Methods

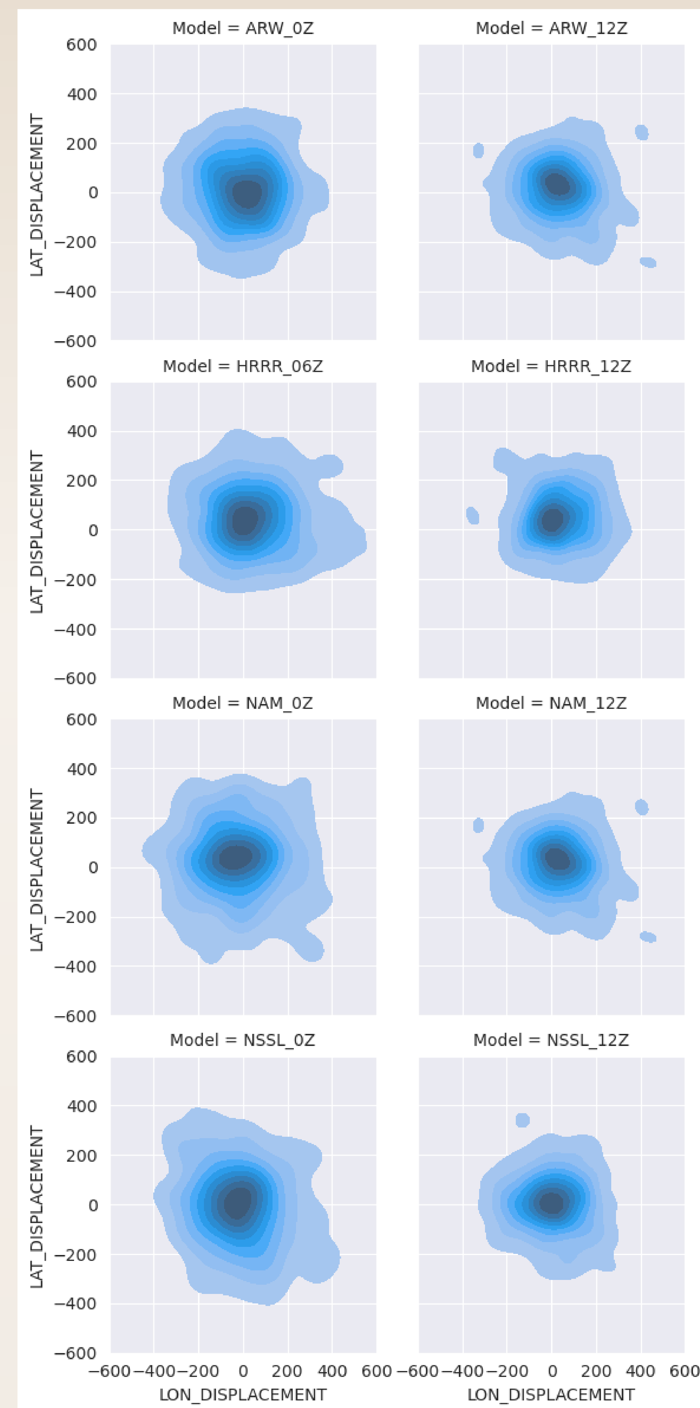
- Method for Object-based Diagnostic Evaluation (MODE) was used to examine mesoscale convective system (MCS) QPF errors for 384 cases within the High-Resolution Ensemble Forecast members (HREF) during the months of May-October of 2018-2023 and several metrics provided by MODE are being used as predictors for the training of ML algorithms
- To include additional predictors for ML training, 22 atmospheric parameters were retrieved from the Storm Prediction Center mesoanalysis; Parameters were collected from a 5x5 grid centered on the centroids of the forecast objects
- ML algorithms currently being trained & very early results prior to feature selection & parameter tuning:
  - **Linear Regression**
    - N/S Displacement Prediction; *R^2: 0.377*
    - E/W Displacement Prediction: *R^2: 0.319*
    - RMSE: *[149.34734438509773, 125.27514623919052] (N/S, E/W) [in km]*
  - **Ridge Regression**
    - RMSE: *[136.8165712838063, 117.09144858634627] (N/S, E/W) [in km]*
  - **XGBoost**
    - RMSE: [139.55530801637607, 125.66827209547353] *(N/S, E/W) [in km]*
  - **Random Forest**
    - RMSE: [123.05427623331967, 117.43077514510055] *(N/S, E/W) [in km]*

Acknowledgements

This work was funded by NOAA grant NA23OAR4590377-T1-01.Thanks are given to Kyle Hugeback, Samuel Ritter, and Joshua Schwarz for their assistance with this research.

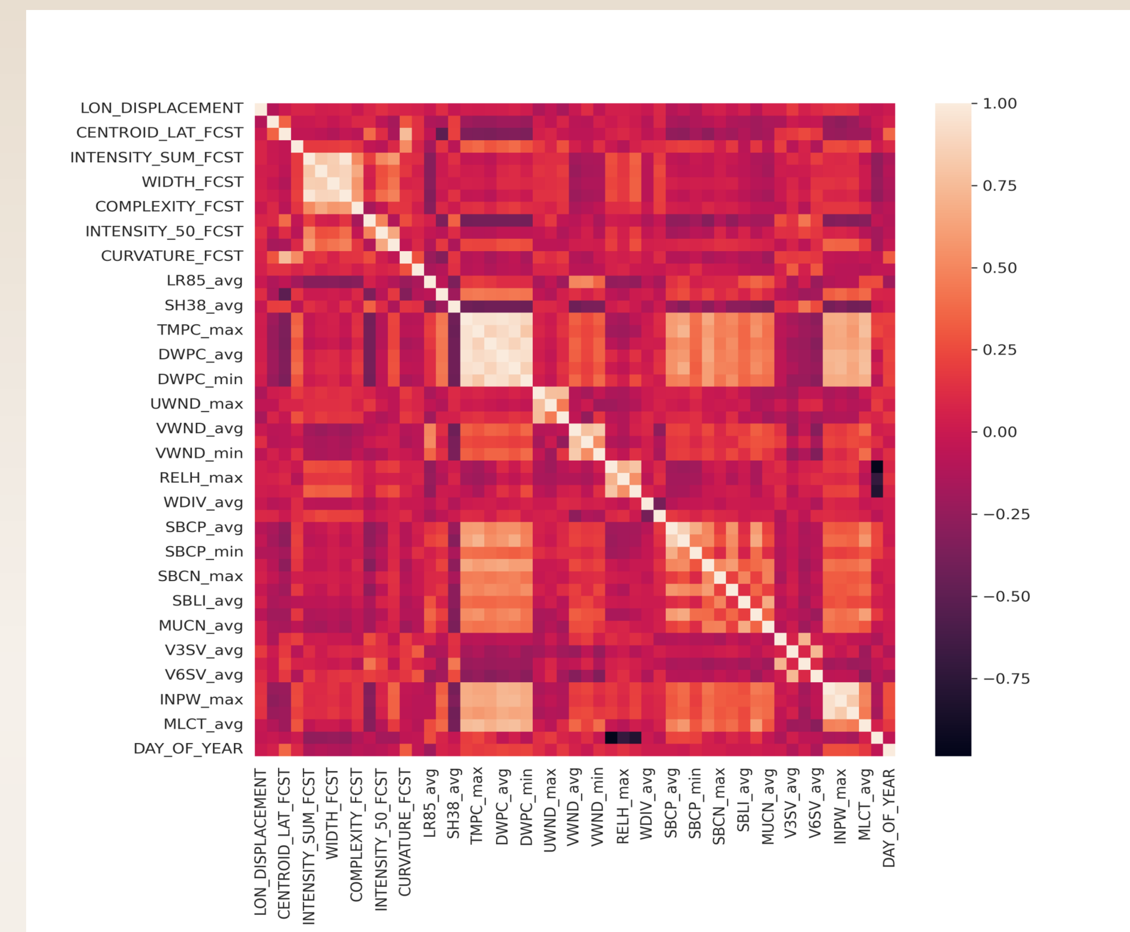
Data Analysis and MODE Output

Kernel Density Estimations of Displacements for Each Model



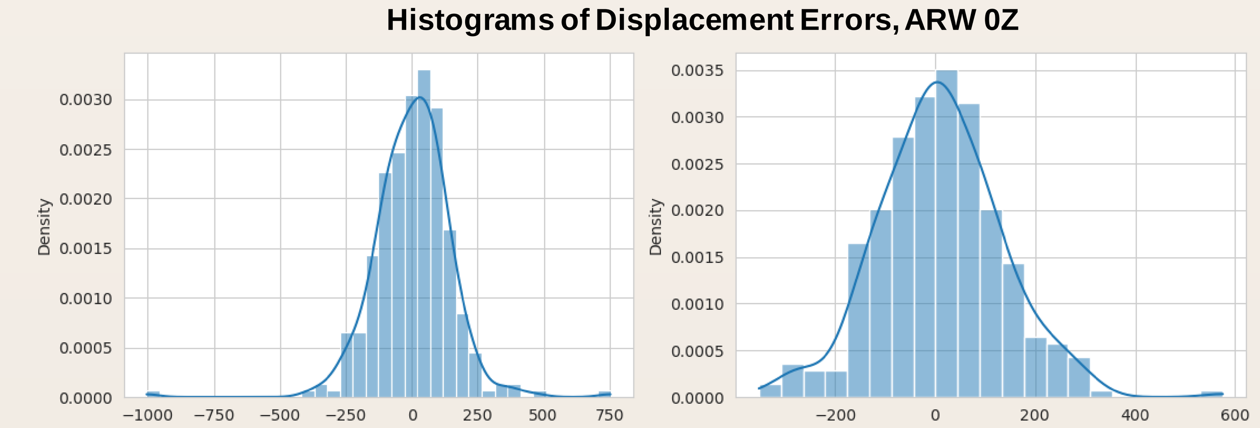
Darker shades represent a higher concentration of displacement errors (in km) with that value. Earlier model runs generally show a broader distribution of longitudinal displacement errors, with the HRRR 06Z being the only notable exception.

Correlation Heat Map Between Variables

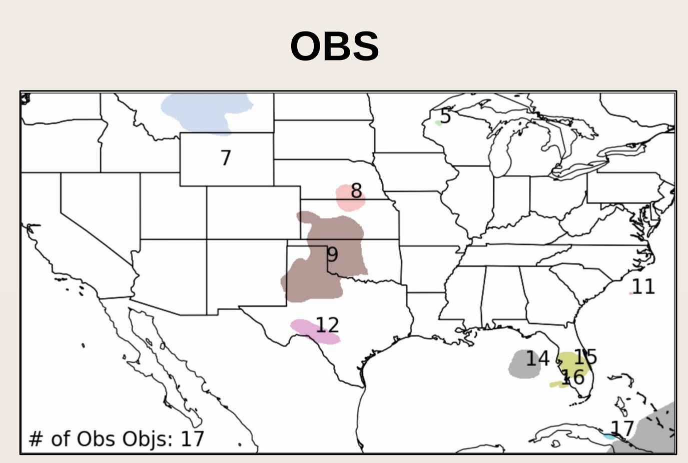


Plot illustrating correlation between the various predictors; Lighter colors represent positive correlation and vice-versa.

Histograms of Displacement Errors, ARW 0Z

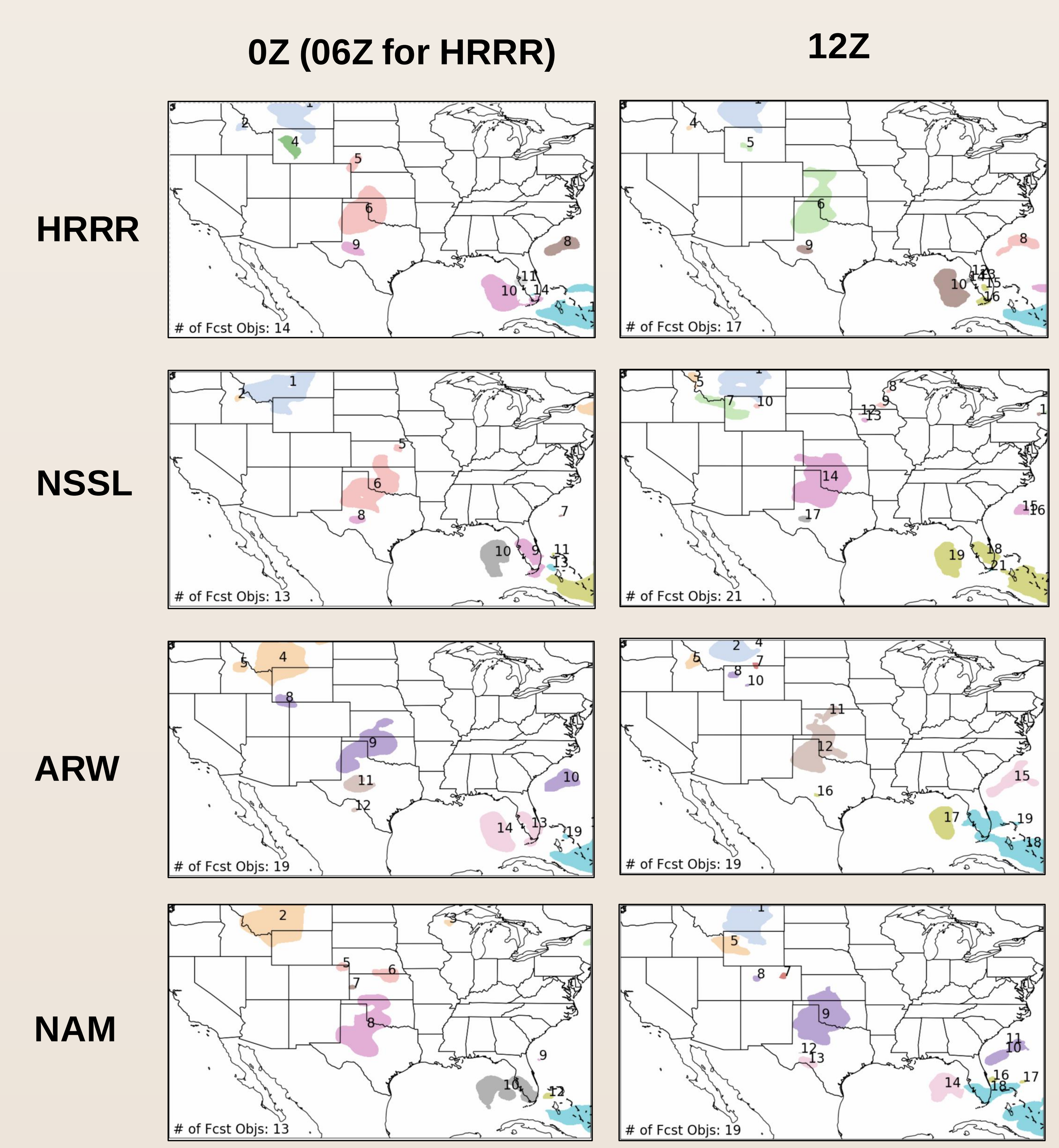


Histograms showing the distributions of displacement errors (in km) in the east/west and north/south directions. In this member, there is a slight tendency towards eastward displacement errors.

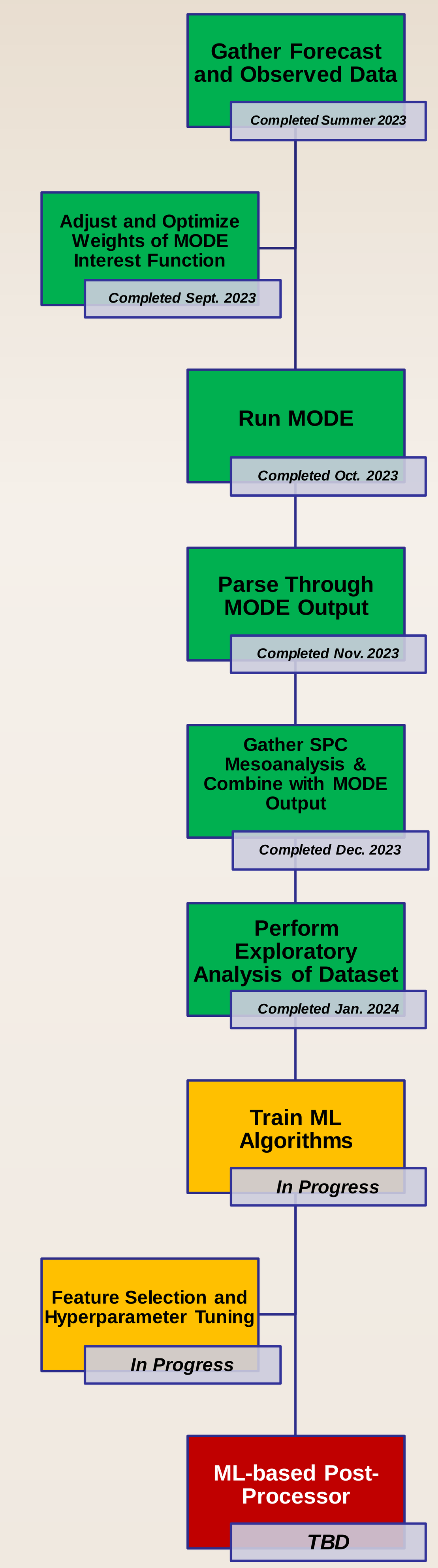


Event of interest is object 12. MCS traveled from the Texas Panhandle to S. Central Texas.

MODE Output Example Case: June 3, 2023



Progress Overview



Future Work

- **Functional ML-based post-processing tool**
  - Prototype to be tested at FFaIR experiment
  - Verification via application to the Flooded Locations and Simulated Hydrographs (FLASH) system
- **Develop guidance for Weather Prediction Center to apply tool to forecasts**