MRMS-based Hail Sizing and Classification Using Different, Large Databases

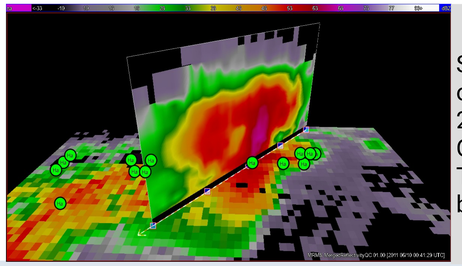
Jose E. Aguilar, Skylar S. Williams and Kiel L. Ortega
OU/CIMMS & NOAA/OAR/NSSL

Background

Current studies conducted at NSSL on hail prediction algorithms look at the performance different data sets have when used to train machine learning algorithms. This study focuses on the performance of five regression algorithms trained, validated, and tested *Storm Data* hail reports. Ideally, the performance of the trained algorithms would surpass those of the currently used hail-sizing algorithm MESH.

Data

- *Storm Data* hail reports that span from 2005 to 2011 from all across the US.
- The data set is composed of 70 float-point attributes of different atmospheric measurements -- such as Vertically Integrated Liquid, height of wet bulb, MUCAPE, low & medium level shear, among others.
- About 66,039 Instances are available, which makes the data set small in size.
- There is a considerable amount of NaN and invalid values in the data set caused by errors in the measurements (irreparable).
- Radar data comes from the MYRORSS database.



Sample Cross-Section of a storm from June 9, 2009 at Northwest Oklahoma.
The storm structure can be easily appreciated.

Methods

Data Preparation:

- During data exploration, only 40 out of the 70 attributes were acceptable signals for the model. The leftover 30 attributes had a very high amount of invalid and NaN values.
- Only 65,203 out of 66,039 instances had optimal integrity for the machine learning step.

Machine Learning Step:

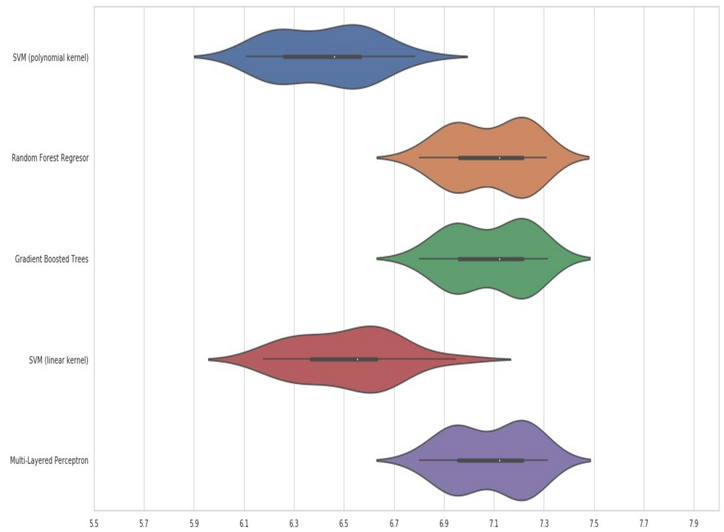
- five different models from the python Scikit-Learn (v. 0.21.2) library were trained and cross-validated with 20 iterations using negative Mean Absolute Error (MAE) as performance measure:

  1)Random Forest Regressor
  2)Gradient Boosting Tree Regressor
  3)Support Vector Machine with Polynomial kernel
  4)Support Vector Machine with Linear kernel
  5)Multi-Layer Perceptron Regressor

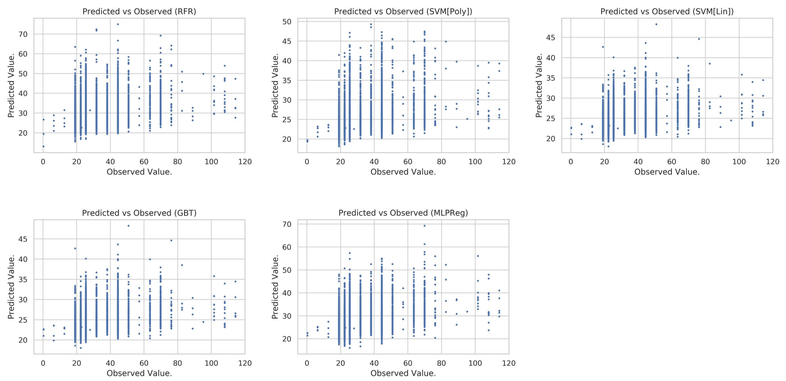
- The selected model was Random Forest Regressor due to its lower bias towards smaller hail.

Results

- All of the models trained had a bias towards smaller-sized hail. This is seen as a drop in performance when generalizing towards hail larger than 50-70mm.
- The prediction pattern is similar to a higher quality data base.
- About %77 of the data set measured hail falls in the range of **19.05mm to 25.4mm.** This is a probable reason for the models to be biased towards smaller hail sizes.

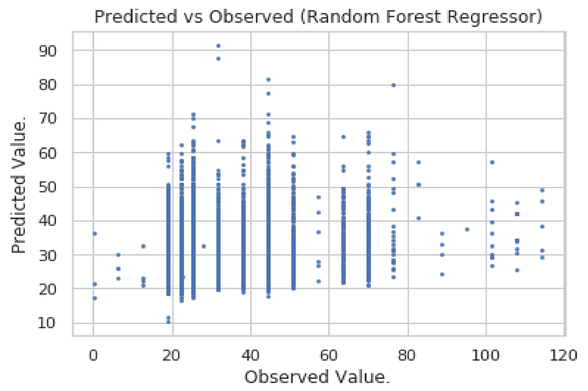


The Support Vector Machine with Polynomial and Linear kernels seemed to do the best out of the 5 models tested, but the graph Predicted vs Observed tells a different story when predicting hail sizes above 35mm.



It is possible to see how the models predict with a bias towards hail being smaller than 50-70mm. This is a problem because it should be larger hail a priority to predict with higher precision.

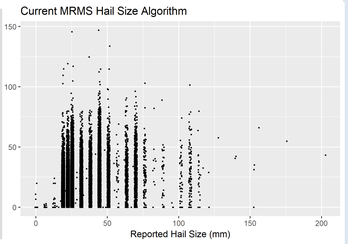
The Observed vs Predicted graphs shows how the Random Forest Regressor model does a slightly better job at predicting hail sizes larger than 65mm.



Conclusions & Future Work

- While *Storm Data* is a small and error-heavy data set, the results of the algorithms were similar to those of previous studies that used higher-resolution data sets such as SHAVE (Severe Hazards Analysis & Verification Experiment).
- The models had a bias towards smaller-sized hail, possibly due to lack of representative data for severe hail.
- Further investigation in more machine learning algorithms, hyper-parameter tuning, and data cleaning could potentially fix the current bias.

Storm data hail reports compared to MYRORSS MESH.
This pattern of regression is seen in the results from the models.



This poster was prepared by Jose E. Aguilar with funding provided by NOAA/Office of Oceanic and Atmospheric Research under NOAA-University of Oklahoma Cooperative Agreement #NA16OAR4320115, U.S. Department of Commerce. The statements, findings, conclusions, and recommendations are those of the authors and do not necessarily reflect the view of NOAA or the U.S. Department of Commerce.

Contact:
jose.aguilar@noaa.gov