P455
January 11, 2016
2:30 – 4:00 pm
American Meteorological Society

Improving MRMS Cloud-to-Ground Lightning Probabilities

Tiffany Meyer[1,2], Kristin Calhoun[1,2] and David John Gagne[3]

[1]Cooperative Institute of Mesoscale Meteorological Studies/University of Oklahoma
[2]National Oceanic and Atmospheric Administration/National Severe Storms Laboratory
[3]Center for Analysis and Prediction of Storms/University of Oklahoma

Current MRMS Cloud-to-Ground Probability in next 30 min in the National Weather Service

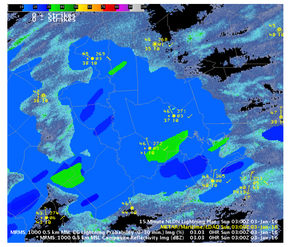
The MRMS Cloud-to-Ground Probability in the next 30 minutes product is currently available in National Weather Service operations, however needs a lot of improvement. Currently the algorithm uses a neural net to perform the probability calculations, with the inputs specified below. The **left image below** shows the MRMS CG Probability product, station observations and NLDN CG Lightning plotted in the AWIPS-2 software. The **right image below** has MRMS Composite Reflectivity with the same observation and lightning overlays. Note, there are no CG lightning flashes, minimal reflectivity (<30dBZ), so very little chance of CG lightning in the next 30 minutes. However, the CG probabilities still range from 20%-35%.

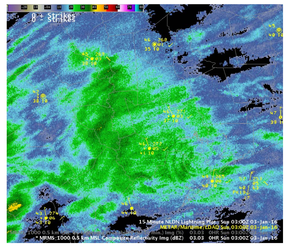
Current Inputs into Neural Net:

Storm Attributes
- Age(s)
- Size(km2)
- Speed(MetersPerSecond)

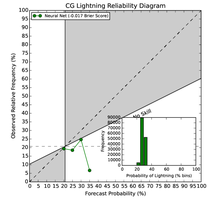
Multi-Radar/ Multi-Sensor
- LayerAverageRef(dBZ)
- LifetimeMaxVIL(kg/m^2)
- MaxRef(dBZ)
- MaxVIL(kg/m^2)
- Reflectivity_-10C(dBZ)
- Reflectivity_-10CIncr(dBZ)
- VIL(kg/m^2)
- VILincr(kg/m^2)

CG Lightning
- Lightning Density (flash/km^2/sec)

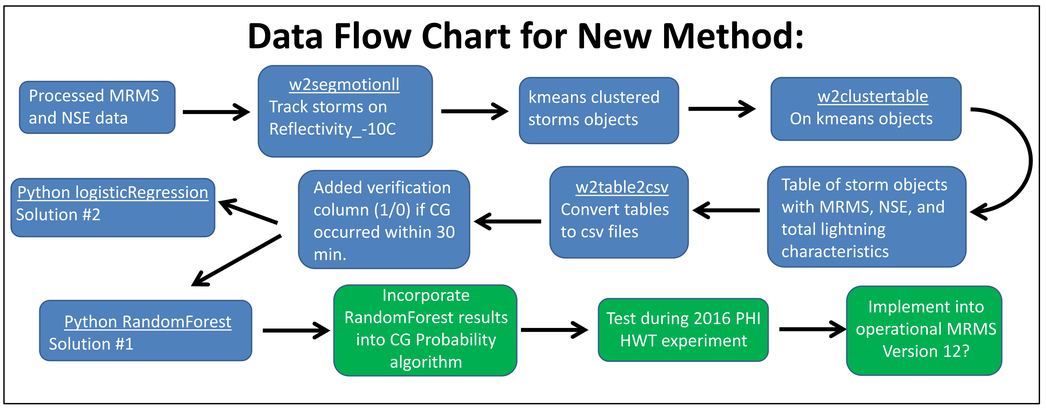




(Right) Distribution of probabilities for the current algorithm. This shows an over-forecast for all probabilities between 20% and 35% with most relative frequencies are under "no skill" level. Note, no probabilities below 20% or above 35% - not helpful!



Data Flow Chart for New Method:



RandomForestClassifier Parameters:
- n_estimators: 300 (Number of trees in the forest)
- max_features: sqrt (Number of features to consider when looking for the best split)
- max_depth: 6 (maximum depth of the tree)
- n_jobs: 5 (number of jobs to run in parallel for both fit and predict)

NEW MRMS Cloud-to-Ground Probability in next 30 min

New Inputs into Random Forest:

Storm Attributes
- Age(s)
- Latitude(Degrees)
- LatRadius(km)
- Longitude(Degrees)
- LonRadius(km)
- MotionEast(MetersPerSecond)
- MotionSouth(MetersPerSecond)
- Orientation(degrees)
- Size(km2)
- Speed(MetersPerSecond)

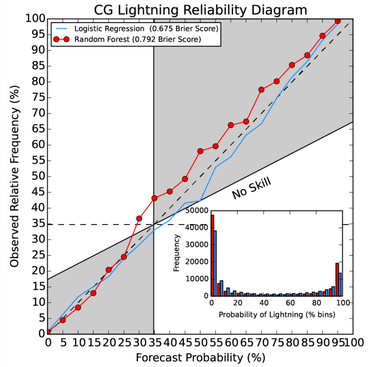
NSE Data
- BRNShear(SquareMetersPerSquareSecond)
- LapseRate_700to500mb(DegreeCPerKilometer)
- LapseRate_850to500mb(DegreeCPerKilometer)
- MeanShear_0-6km(MetersPerSecondPerKilometer)
- MUCAPE(SquareMetersPerSquareSecond)
- SfcCAPE(SquareMetersPerSquareSecond)
- SfcCIN(SquareMetersPerSquareSecond)
- SfcDewPoint(degreeC)
- SfcRH(%)
- SfcTemperature(degreeC)
- SfcThetaE(degreeK)
- SRHelicity0-3km(SquareMetersPerSquareSecond)

Multi-Radar/ Multi-Sensor
- LifetimeMESH(mm)
- LLReflectivity(dBZ)
- LowLvlShear(s^-1)
- MaxRef(dBZ)
- MaxVIL(kg/m^2)
- MeanRef(dBZ)
- MESH(mm)
- MidLvlShear(s^-1)
- Reflectivity_0C(dBZ)
- Reflectivity_-10C(dBZ)
- Reflectivity_-20C(dBZ)
- TotalVIL(kg/m^2)
- VILAreaGT40(km^2)

Total Lightning
- CGCount_15min(flashes)
- CGCount_2min(flashes)
- ENI_ICcount_15min(flashes)
- ENI_ICcount_2min(flashes)
- IC_FlashesPerCellArea
- Lifetime_CGCount(flashes)

New additions to input list:
- NSE data
- MRMS data (more)
- Storm Attributes (more)
- Total Lightning

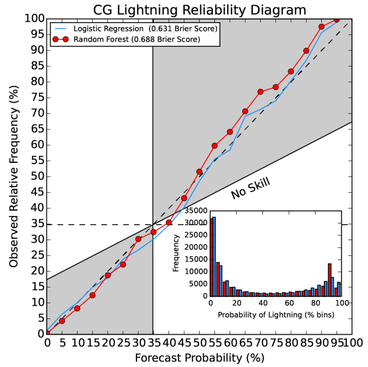
MRMS, NSE, & Total Lightning



Features	Importance
CGCount_15min(flashes)	0.214149
ENI_ICcount_15min(flashes)	0.177857
ENI_ICcount_2min(flashes)	0.142617
IC_FlashesPerCellArea	0.122897
CGCount_2min(flashes)	0.086506

- Uses all inputs for random forest and logistic regression calculations
- Overall best Brier Score for both random forest and logistic regression solutions
- Both solutions fit well with the perfect reliability line
  - Random forest has a slight under-forecasting above 25%
  - Logistic regression has slight over-forecasting in middle probabilities and under-forecasting towards the extremes
- Table shows the most important inputs/features to the random forest solution
  - Lightning was most important-including both CG and IC (in-cloud lightning)

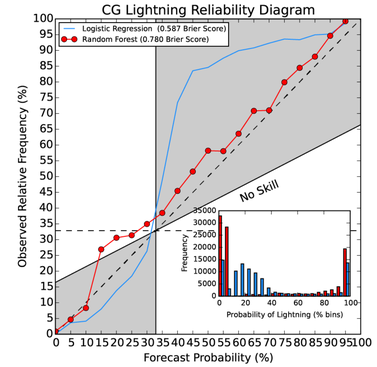
MRMS & NSE



Features	Importance
MaxVIL(kg/m^2)	0.13844
Reflectivity_-10C(dBZ)	0.124486
MESH(mm)	0.084991
LLReflectivity(dBZ)	0.077284
Reflectivity_-20C(dBZ)	0.070433

- Uses only MRMS, NSE, and storm attribute inputs for random forest and logistic regression calculations
  - No total lightning inputs
- Worst Brier Score for random forest
- Both solutions fit well with the perfect reliability line
- Table shows the most important inputs/features to the random forest solution
  - Most important features were all MRMS data-majority being Reflectivity products

Total Lightning

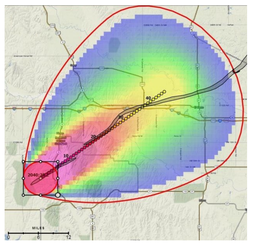


Features	Importance
CGCount_15min(flashes)	0.28843
ENI_ICcount_15min(flashes)	0.226849
IC_FlashesPerCellArea	0.1793
ENI_ICcount_2min(flashes)	0.171245
CGCount_2min(flashes)	0.101159

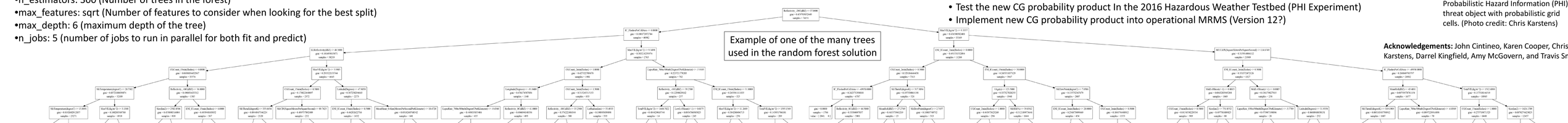
- Uses only total lightning and storm attribute inputs for random forest and logistic regression calculations
  - No MRMS or NSE inputs
- Worst Brier Score for logistic regression
- Random Forest line bounced around a lot
  - Due to small sample size
- Logistic Regression way over-forecasted for events lower than 35% and under-forecasted for events above 35%
- Table shows the most important inputs/features to the random forest solution
  - Most important features were all total lightning products

Discussion/Future Work

- The current/NWS operational solution to CG Probability in the next 30 minutes is:
  - A neural net which forecasters view as a black box-not being able to easily tweak what goes into the calculations
  - limited to probabilities of only 20-35%
  - New total lightning products haven't been added
- The new random forest approach:
  - Uses new total lightning data as well as more MRMS and NSE data as inputs
  - Has realistic probabilities with a good Brier Score
- In the future, need to:
  - Incorporate random forest decision trees into WDSSII CG probability algorithm-including pairing down the inputs
  - Test the new CG probability product In the 2016 Hazardous Weather Testbed (PHI Experiment)
  - Implement new CG probability product into operational MRMS (Version 12?)



Probabilistic Hazard Information (PHI) threat object with probabilistic grid cells. (Photo credit: Chris Karstens)



Example of one of the many trees used in the random forest solution

Acknowledgements: John Cintineo, Karen Cooper, Chris Karstens, Darrel Kingfield, Amy McGovern, and Travis Smith.