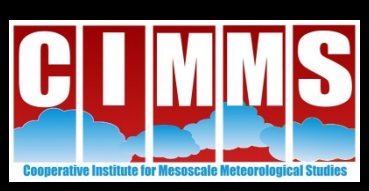
Using Random Forest to Generate Cloud-to-Ground Lightning Probabilities

Tiffany Meyer

Kristin Calhoun

Darrel Kingfield

Chris Karstens



Lightning is a BIG hazard!

Currently in the National Weather Service:

- No lightning alerts/warnings issued

Lightning kills more people than tornadoes

- Forecasters are consistently asked by Emergency Managers to provide Decision Support Services (DSS) for lightning
  - Outdoor concerts
  - Fairs
  - Sporting events

Lightning is a BIG hazard!

GOAL:
Create an automated probability for a storm, that it would produce cloud-to-ground lightning in the next 30 minutes

Random Forest Dataset

GOAL: Create an automated probability for a storm, that it would produce cloud-to-ground lightning in the next 30 minutes

June 2014 – June 2015

- Randomly selected 3-set days
- 12Z on day1 – 12Z on day 3

Training	Validation
3-4-5	8-9-10
18-19-20	
26-27-28	

CONUS Wide

- Vaisala Cloud-to-Ground Lightning
- Earth Networks In-Cloud Lightning
- RUC/RAP Near Storm Environment Data
- Multi-Radar/Multi-Sensor Severe Product Suite

Dataset

RUC Model Data (Near Storm Environment)

- BRN Shear (m^2/s^2)
- Lapse Rate 700-500mb ($^oC/km$)
- Lapse Rate 850-500mb ($^oC/km$)
- Mean Shear 0-6km (m/s/km)
- MUCAPE (m^2/s^2)
- Surface CAPE (m^2/s^2)
- Surface CIN (m^2/s^2)
- Surface Dewpoint (^oC)
- Surface Temperature (^oC)
- Surface ThetaE (K)
- Surface Relative Humidity (%)
- SR Helicity (m^2/s^2)

Multi-Radar Multi-Sensor (MRMS)

- Maximum MESH (mm)
- Maximum Low Level Reflectivity (dBZ)
- Maximum Reflectivity (dBZ)
- Mean Reflectivity (dBZ)
- Max Reflectivity at 0^oC (dBZ)
- Max Reflectivity at -10^oC (dBZ)
- Max Reflectivity at -20^oC (dBZ)
- Low Level Shear (s^{-1})
- Mid Level Shear (s^{-1})
- Maximum VIL (km/m^2)
- Total VIL (km/m^2)

Lightning Data

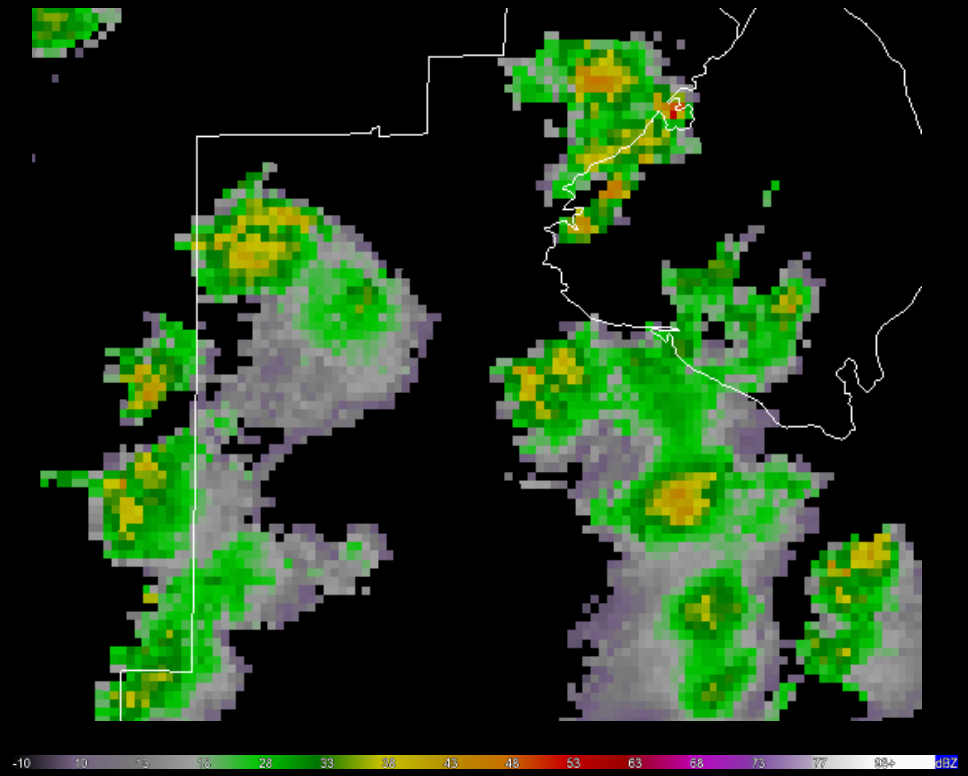
- Vaisala Cloud-to-Ground Lightning (15min)
- Vaisala Cloud-to-Ground Lightning (2min)
- Earth Networks In-Cloud Lightning (15min count)
- Earth Networks In-Cloud Lightning (2min count)
- Earth Networks In-Cloud Lightning (Flashes per storm)

Random Forest Dataset

Used enhanced k-means watershed technique

- Tracked storms over time using MRMS Reflectivity at -10^{o}C
  - Minimum reflectivity of 20 dBZ
  - 3 scales (200km^2, 600km^2, 1000km^2)

Temporal dataset for each storm based on MRMS/NSE/Lightning attributes



Dataset

	Lightning					MRMS				NSE
StormID	CG_15min (flashes)	CG_2min (flashes)	IC_15min (flashes)	IC_2min (flashes)	IC_Flashes (PerCellArea)	MESH (mm)	MaxRef (dBZ)	Refl_-10C (dBZ)	Refl_-20C (dBZ)	MUCAPE (m^2/s^2)
230	7	0	16	1	0.0032	0.87	48.86	40.31	35.39	1291.46
310	13	3	8	0	0.0000	3.54	53.17	44.12	46.96	2627.76
331	101	8	16	2	0.0030	1.06	52.34	40.30	34.77	1484.21
405	11	0	17	2	0.0034	2.49	50.66	43.93	40.54	1410.59
412	1	0	7	2	0.0065	8.36	54.80	54.80	50.77	3706.28
415	0	0	0	0	0.0000	6.51	52.03	50.40	34.14	2423.28
427	0	0	4	2	0.0068	2.38	50.10	45.33	38.54	2382.55
428	22	12	72	25	0.0768	6.01	50.82	45.03	43.57	2845.49
431	0	0	0	0	0.0000	-99900.00	35.30	34.40	34.40	2256.99
440	0	0	0	0	0.0000	2.17	49.73	42.50	30.44	1785.61

Trained Dataset - “Find the Answer”

- Mark if object had CG lightning in the next 30 minutes
  - Added it to the table

	Lightning					MRMS				NSE	Answer
StormID	CG_15min (flashes)	CG_2min (flashes)	IC_15min (flashes)	IC_2min (flashes)	IC_Flashes (PerCellArea)	MESH (mm)	MaxRef (dBZ)	Refl_-10C (dBZ)	Refl_-20C (dBZ)	MUCAPE (m^2/s^2)	CG? (0-30min)
230	7	0	16	1	0.0032	0.87	48.86	40.31	35.39	1291.46	1
310	13	3	8	0	0.0000	3.54	53.17	44.12	46.96	2627.76	1
331	101	8	16	2	0.0030	1.06	52.34	40.30	34.77	1484.21	1
405	11	0	17	2	0.0034	2.49	50.66	43.93	40.54	1410.59	1
412	1	0	7	2	0.0065	8.36	54.80	54.80	50.77	3706.28	0
415	0	0	0	0	0.0000	6.51	52.03	50.40	34.14	2423.28	0
427	0	0	4	2	0.0068	2.38	50.10	45.33	38.54	2382.55	0
428	22	12	72	25	0.0768	6.01	50.82	45.03	43.57	2845.49	1
431	0	0	0	0	0.0000	-99900.00	35.30	34.40	34.40	2256.99	0
440	0	0	0	0	0.0000	2.17	49.73	42.50	30.44	1785.61	0

What is a Random Forest?

- Start with trained dataset
  - Features – every storm attribute tracked (ex. MESH, MaxRefl, CG_2min, IC_2min)
  - Answer – whether or not CG flash happened for that storm

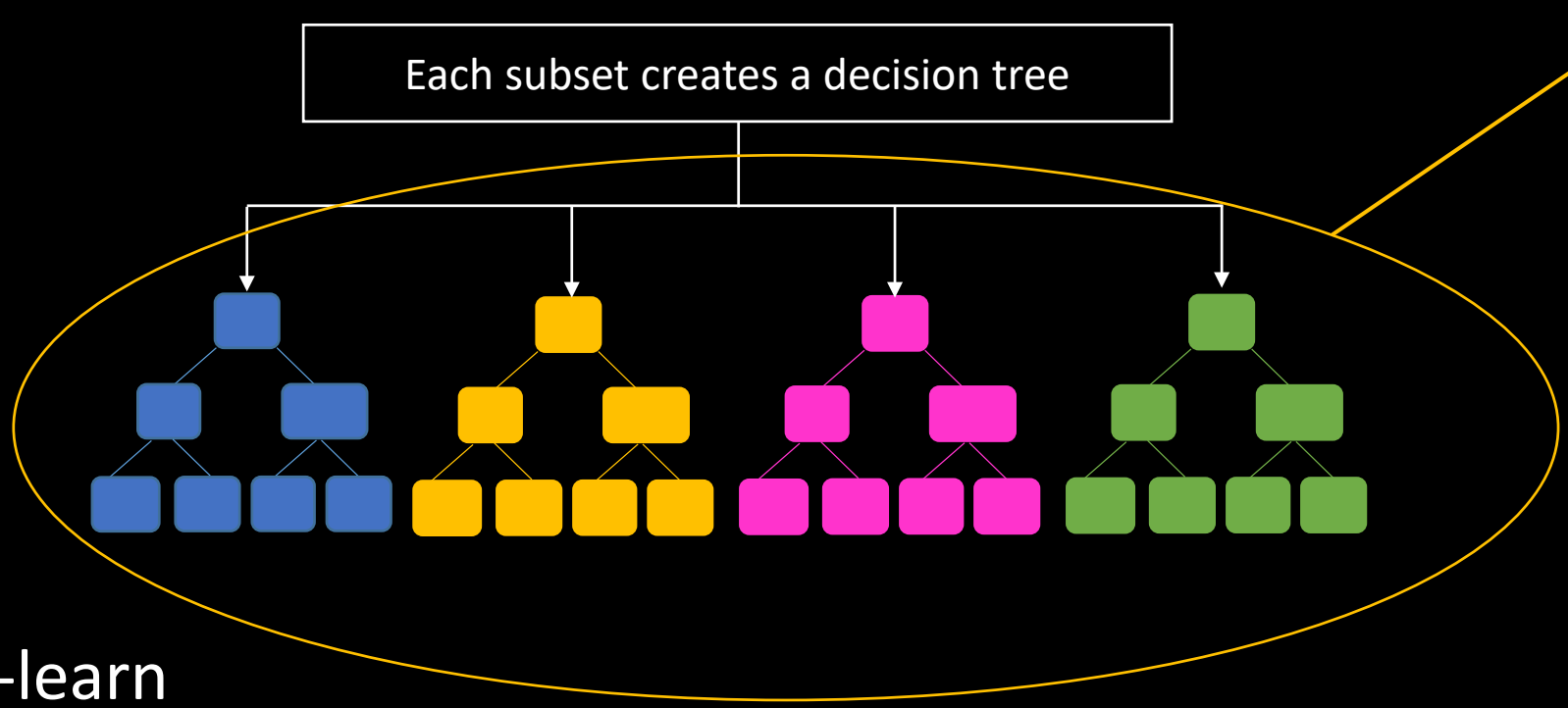
	Features										Answer
StormID	CG_15min (flashes)	CG_2min (flashes)	IC_15min (flashes)	IC_2min (flashes)	IC_Flashes (PerCellArea)	MESH (mm)	MaxRef (dBZ)	Refl_-10C (dBZ)	Refl_-20C (dBZ)	MUCAPE (m^2/s^2)	CG? (0-30min)
230	7	0	16	1	0.0032	0.87	48.86	40.31	35.39	1291.46	1
310	13	3	8	0	0.0000	3.54	53.17	44.12	46.96	2627.76	1
331	101	8	16	2	0.0030	1.06	52.34	40.30	34.77	1484.21	1
405	11	0	17	2	0.0034	2.49	50.66	43.93	40.54	1410.59	1
412	1	0	7	2	0.0065	8.36	54.80	54.80	50.77	3706.28	0
415	0	0	0	0	0.0000	6.51	52.03	50.40	34.14	2423.28	0
427	0	0	4	2	0.0068	2.38	50.10	45.33	38.54	2382.55	0
428	22	12	72	25	0.0768	6.01	50.82	45.03	43.57	2845.49	1
431	0	0	0	0	0.0000	-99900.00	35.30	34.40	34.40	2256.99	0
440	0	0	0	0	0.0000	2.17	49.73	42.50	30.44	1785.61	0

What is a Random Forest?

- Start with trained dataset
  - Features – every storm attribute tracked (ex. MESH, MaxRefl, CG_2min, IC_2min)
  - Answer – whether or not CG flash happened for that storm
- Random subsets are created
  - Random rows and columns
  - Answer

StormID	CG_15min (flashes)	CG_2min (flashes)	IC_15min (flashes)	IC_2min (flashes)	IC_Flashes (PerCellArea)	MESH (mm)	MaxRef (dBZ)	Refl_-10C (dBZ)	Refl_-20C (dBZ)	MUCAPE (m^2/s^2)	Answer: CG? (0-30min)
230	7	0	16	1	0.0032	0.87	48.86	40.31	35.39	1291.46	1
310	13	3	8	0	0.0000	3.54	53.17	44.12	46.96	2627.76	1
331	101	8	16	2	0.0030	1.06	52.34	40.30	34.77	1484.21	1
405	11	0	17	2	0.0034	2.49	50.66	43.93	40.54	1410.59	1
412	1	0	7	2	0.0065	8.36	54.80	54.80	50.77	3706.28	0
415	0	0	0	0	0.0000	6.51	52.03	50.40	34.14	2423.28	0
427	0	0	4	2	0.0068	2.38	50.10	45.33	38.54	2382.55	0
428	22	12	72	25	0.0768	6.01	50.82	45.03	43.57	2845.49	1
431	0	0	0	0	0.0000	-99900.00	35.30	34.40	34.40	2256.99	0
440	0	0	0	0	0.0000	2.17	49.73	42.50	30.44	1785.61	0

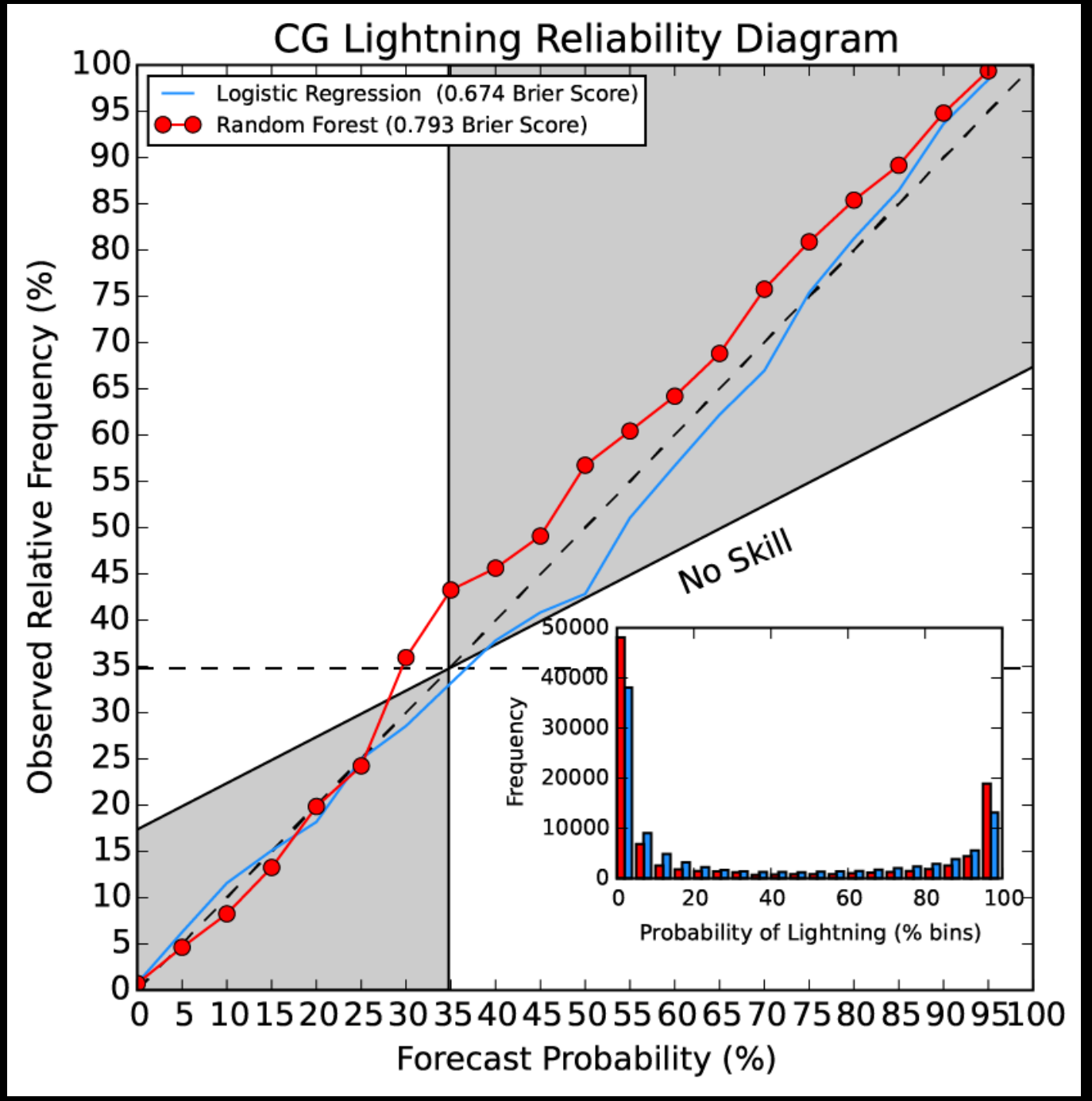
Random Forest



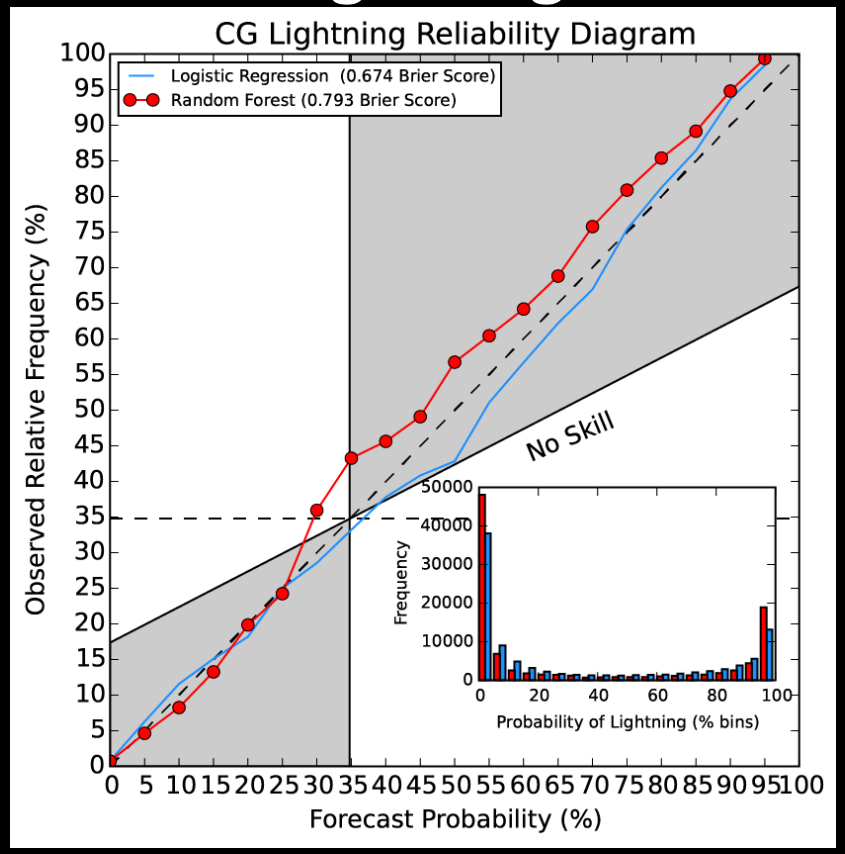
- Python – scikit-learn
  - Python modules for machine learning
  - RandomForestClassifier:
    - *n_estimators*: 300 (number of trees in forest)
    - *max_depth*: 6 (max depth of the tree)
    - *n_jobs*: 5 (number of jobs to run in parallel)

$$\%\ Probability = \frac{\#\ \boldsymbol{YES's}}{\#\ \boldsymbol{TREES}}$$

MRMS, NSE, & Total Lightning

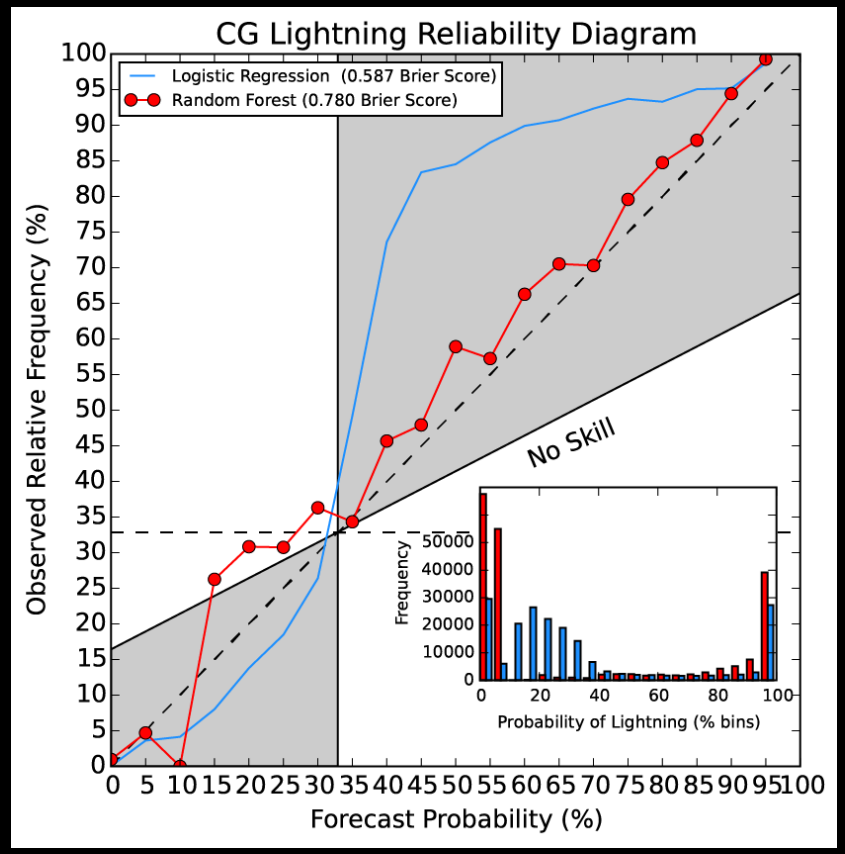


MRMS, NSE, & Total Lightning



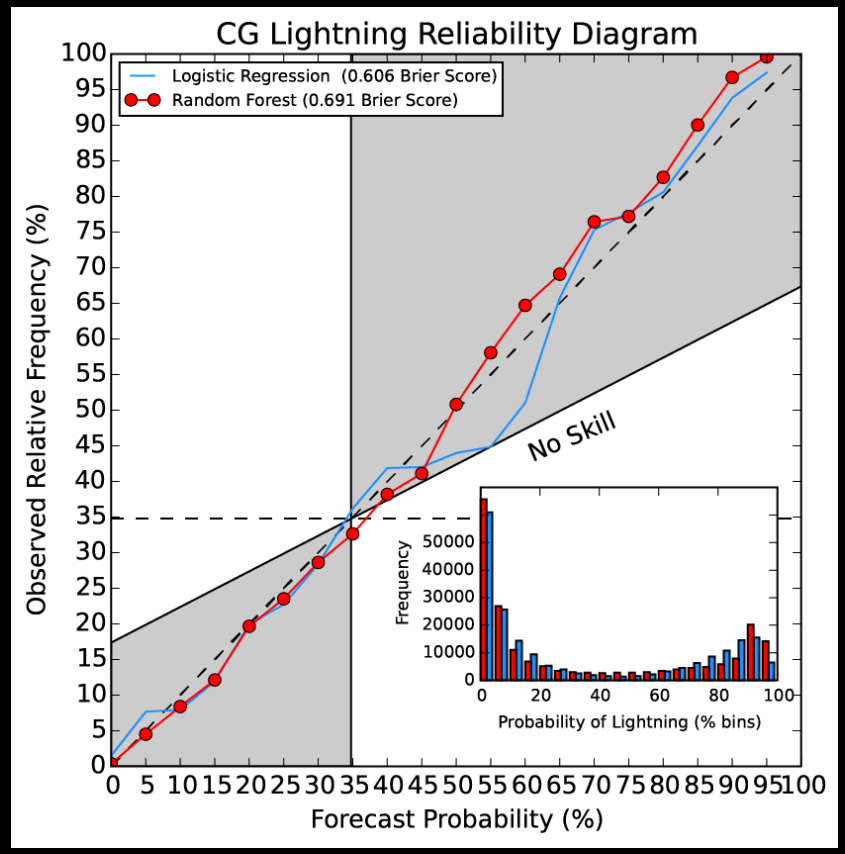
Features	Importance
CG Count 15min(flashes)	0.214149
IC Count 15min(flashes)	0.177857
IC Count 2min(flashes)	0.142617
IC FlashesPerCellArea	0.122897
CG Count 2min(flashes)	0.086506

Total Lightning



Features	Importance
CG Count 15min(flashes)	0.28843
IC Count 15min(flashes)	0.226849
IC FlashesPerCellArea	0.1793
IC Count 2min(flashes)	0.171245
CG Count 2min(flashes)	0.101159

MRMS & NSE



Features	Importance
Max VIL (kg/m^2)	0.13844
Reflectivity @-10C(dBZ)	0.124486
MESH (mm)	0.084991
Low Level Reflectivity (dBZ)	0.077284
Reflectivity @-20C(dBZ)	0.070433

Storm Attributes

28

RUC Model Data (Near Storm Environment)

- BRN Shear (m^2/s^2)
- Lapse Rate 700-500mb (^oC/km)
- Lapse Rate 850-500mb (^oC/km)
- Mean Shear 0-6km (m/s/km)
- MUCAPE (m^2/s^2)
- Surface CAPE (m^2/s^2)
- Surface CIN (m^2/s^2)
- Surface Dewpoint (^oC)
- Surface Temperature (^oC)
- Surface ThetaE (K)
- Surface Relative Humidity (%)
- SR Helicity (m^2/s^2)

Multi-Radar Multi-Sensor (MRMS)

- Maximum MESH (mm)
- Low Level Reflectivity (dBZ)
- Maximum Reflectivity (dBZ)
- Mean Reflectivity (dBZ)
- Reflectivity at 0^{o}C (dBZ)
- Reflectivity at -10^{o}C (dBZ)
- Reflectivity at -20^{o}C (dBZ)
- Low Level Shear (s^{-1})
- Mid Level Shear (s^{-1})
- Maximum VIL (km/m^2)
- Total VIL (km/m^2)

Lightning Data

- Vaisala Cloud-to-Ground Lightning (15min)
- Vaisala Cloud-to-Ground Lightning (2min)
- Earth Networks In-Cloud Lightning (15min count)
- Earth Networks In-Cloud Lightning (2min count)
- Earth Networks In-Cloud Lightning (Flashes per storm)

Storm Attributes

18

Multi-Radar Multi-Sensor (MRMS)

- Max MESH (mm)
- Max Composite Reflectivity (dBZ)
- Max Low Level Reflectivity (dBZ)
- Reflectivity at 0°C (dBZ)
- Reflectivity at -10°C (dBZ)
- Reflectivity at -20°C (dBZ)
- Low Level Shear (s^{-1})
- Mid Level Shear (s^{-1})
- Maximum VIL (km/m^2)

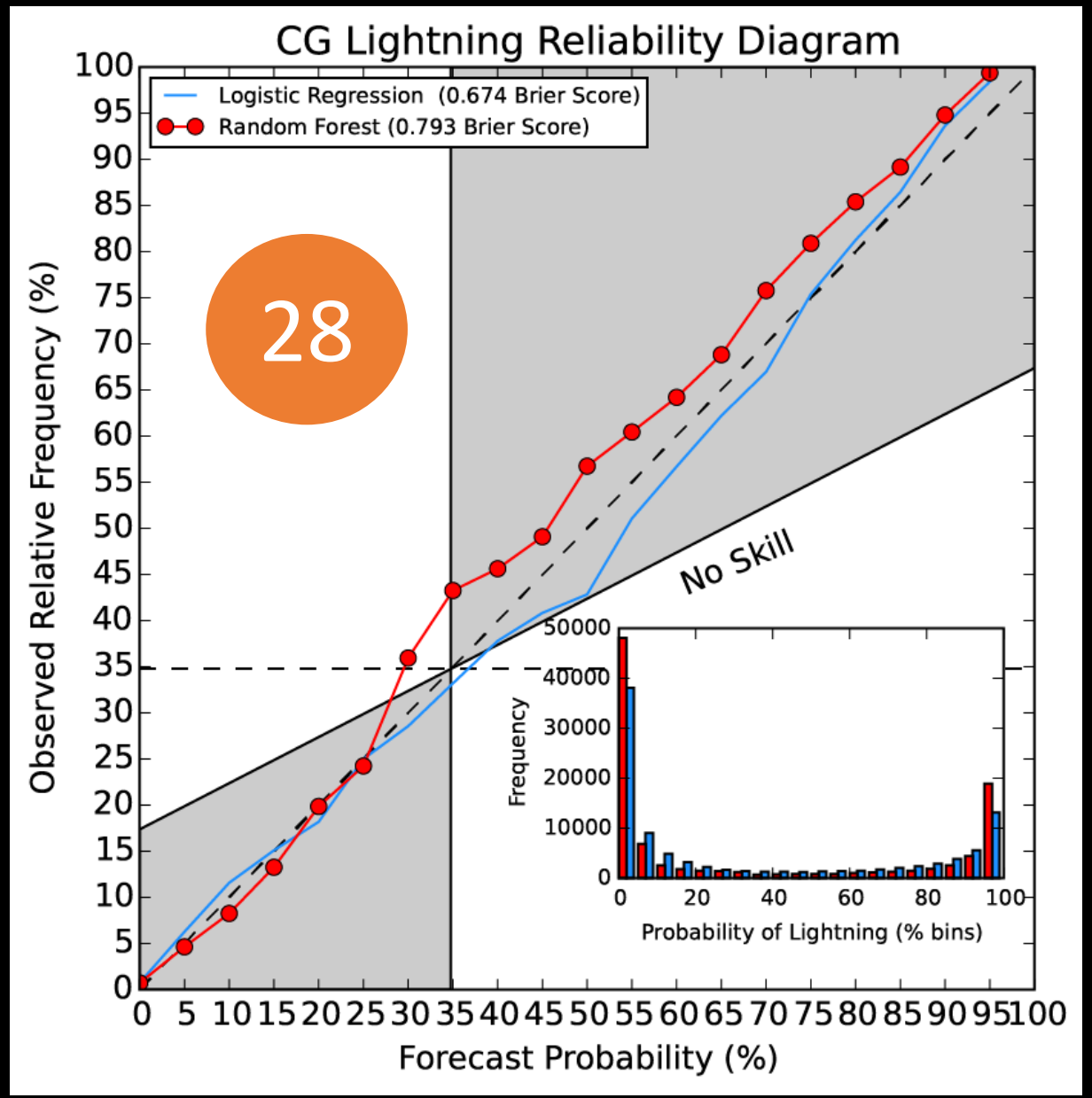
RUC Model Data (Near Storm Environment)

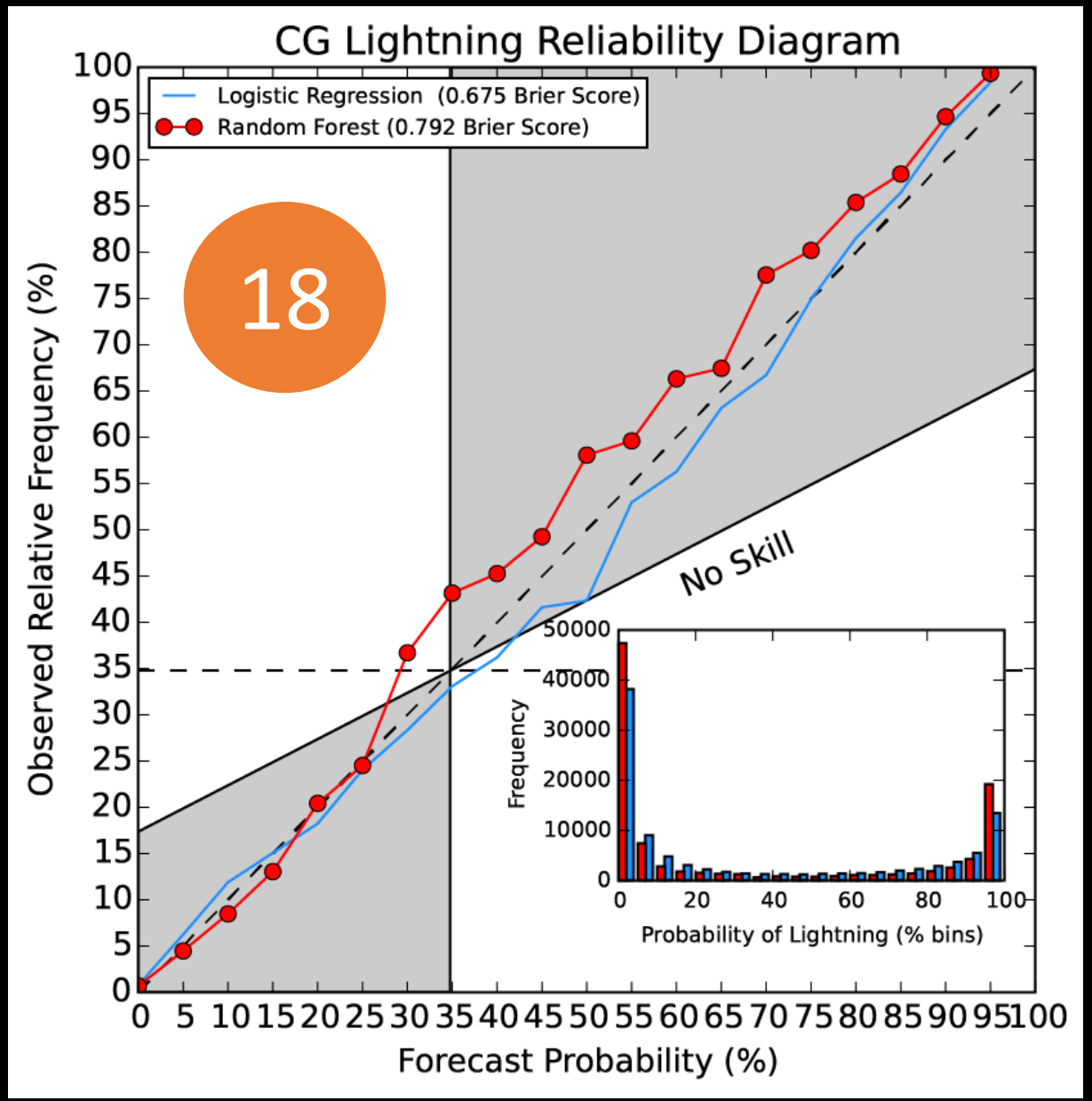
- MUCAPE (m^2/s^2)
- Surface CAPE (m^2/s^2)
- Surface Temperature (°C)
- Surface ThetaE (K)

Lightning Data

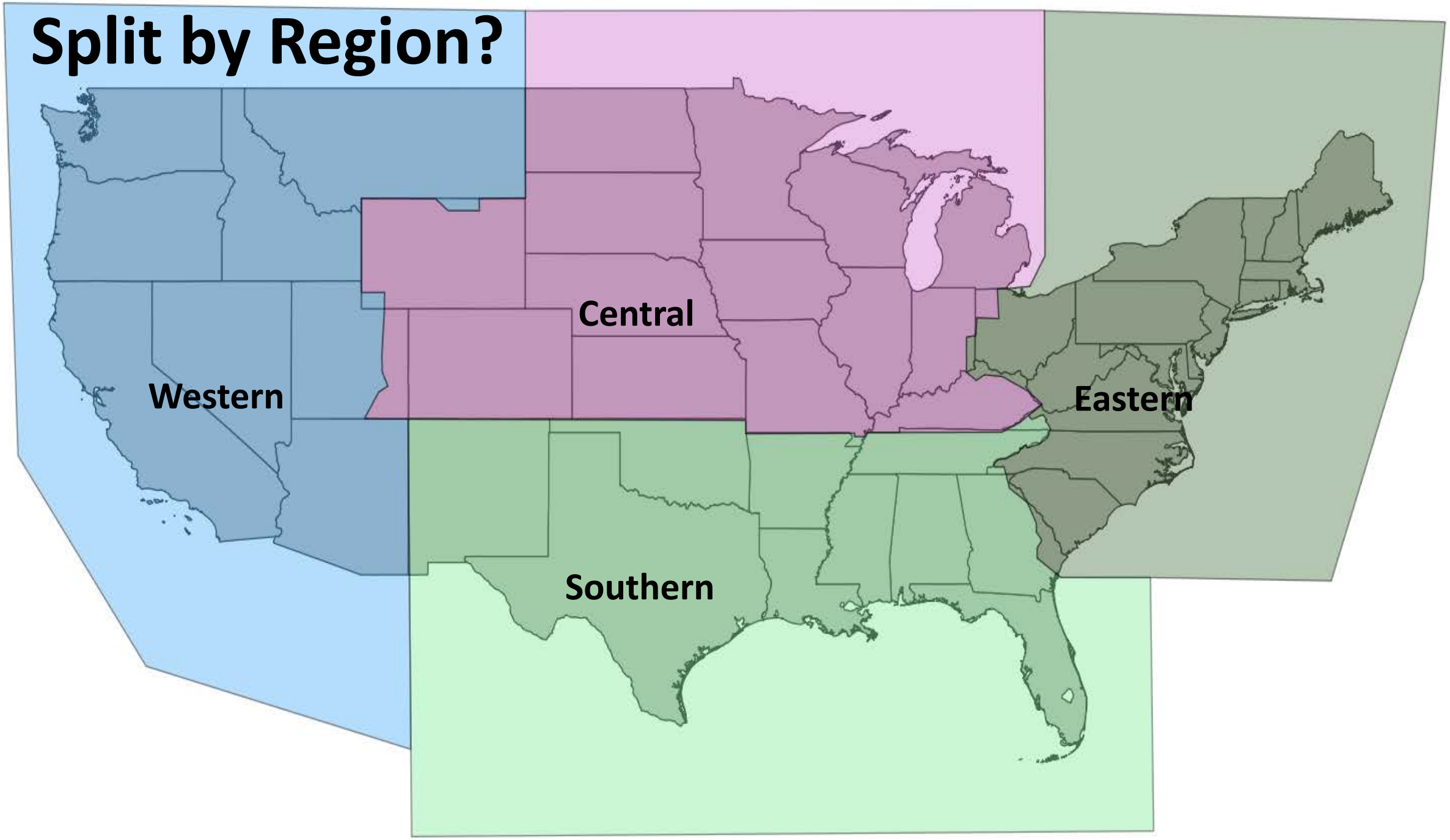
- Vaisala Cloud-to-Ground Lightning (15min)
- Vaisala Cloud-to-Ground Lightning (2min)
- Earth Networks In-Cloud Lightning (15min count)
- Earth Networks In-Cloud Lightning (2min count)
- Earth Networks In-Cloud Lightning (Flashes per storm)

Verification of CG Probabilities





Split by Region?



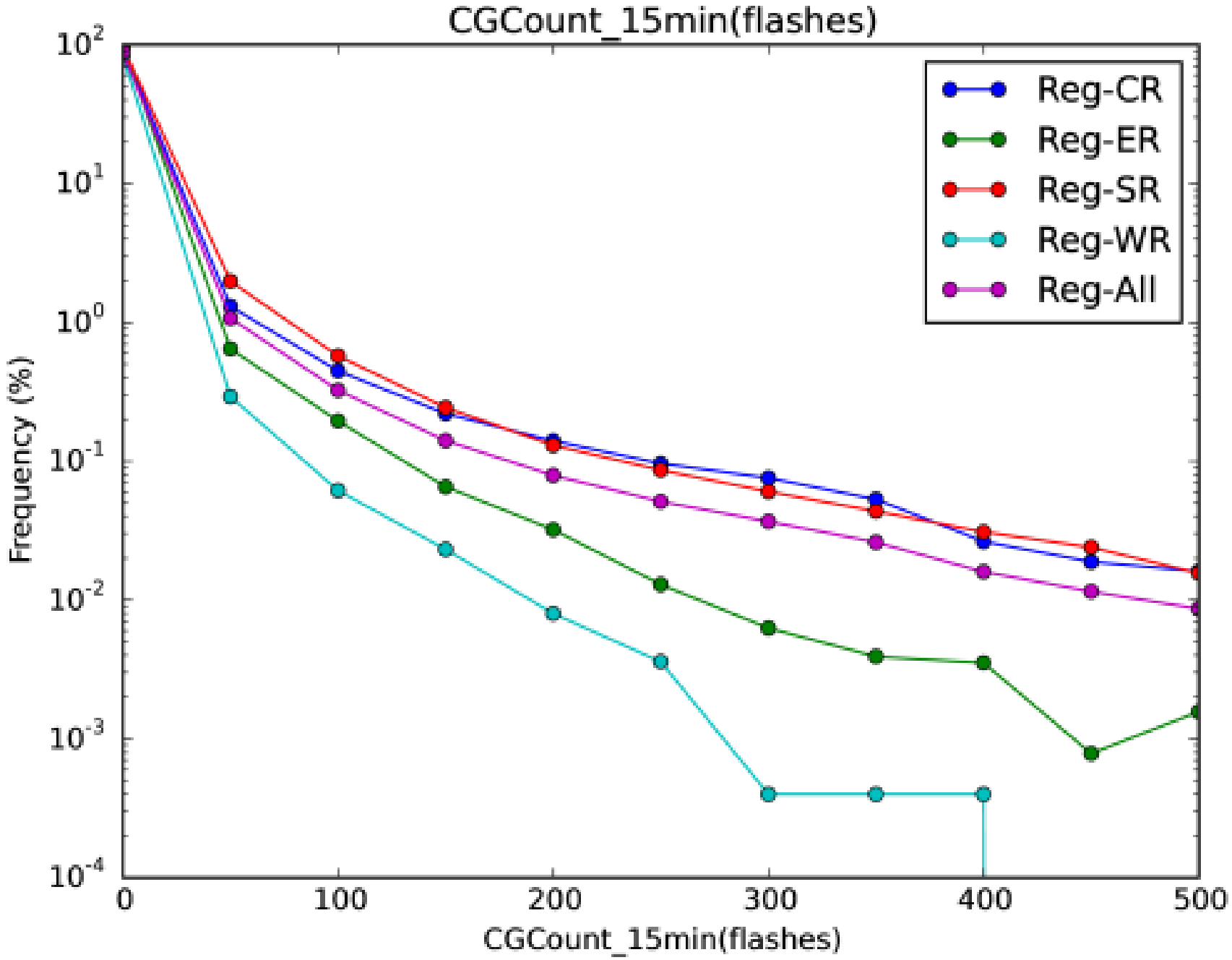
Feature Importance

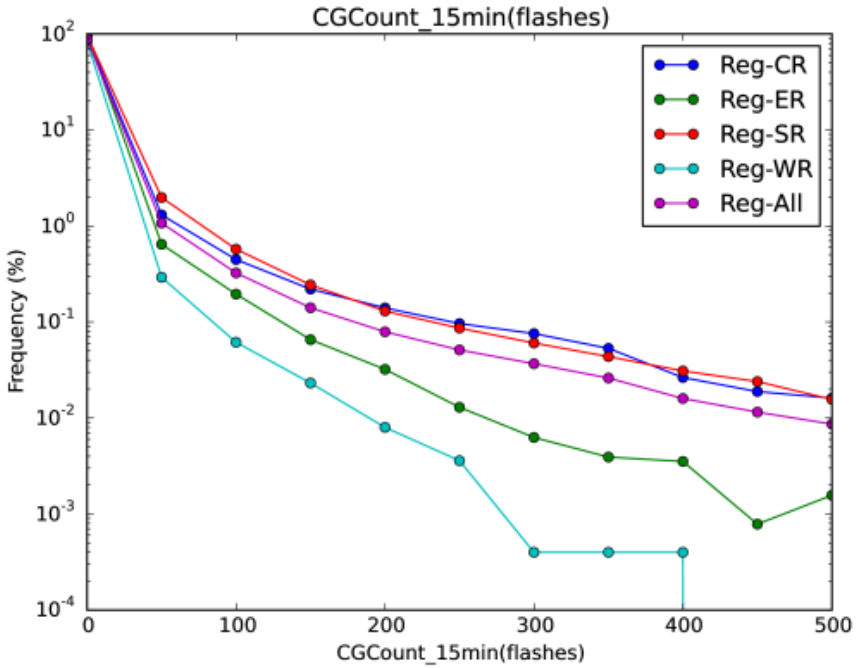
Central & Southern Region	
CG Count 15min (flashes)	.22659
IC Count 15min (flashes)	.16599
IC Flashes PerCellArea	.12469
IC Count 2min (flashes)	.12263

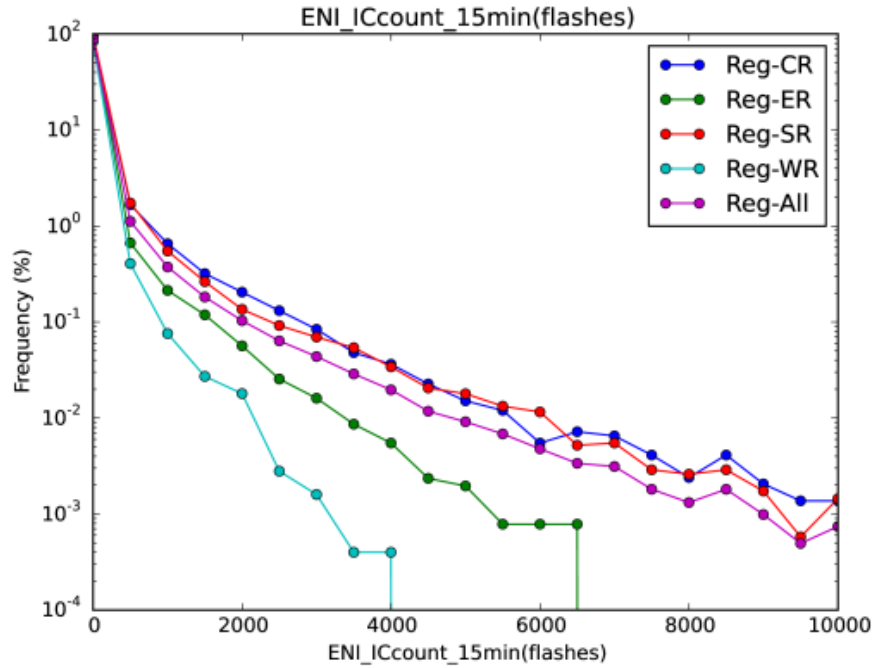
Region	# of Storms
Central	381,308
Eastern	322,684
Southern	481,024
Western	321,234

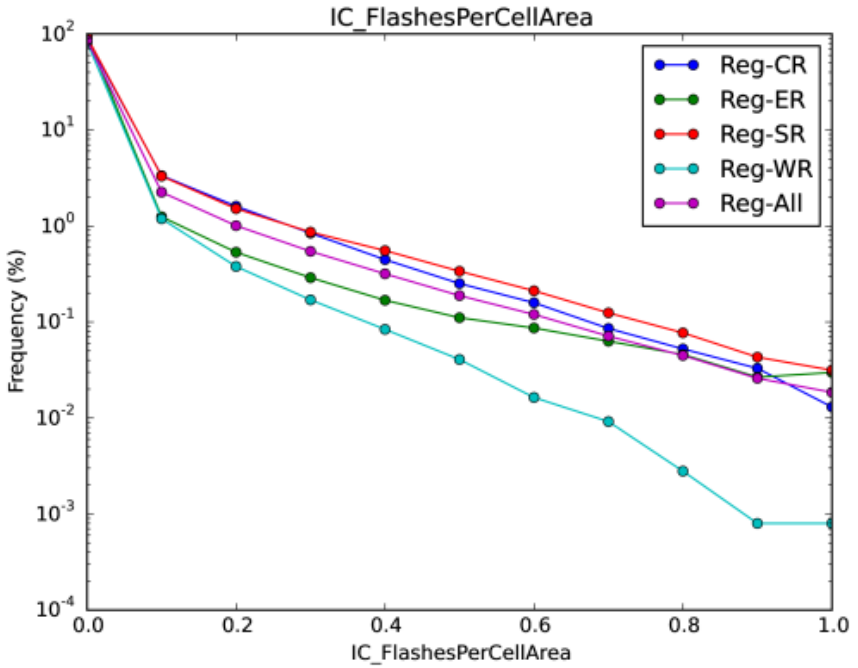
Eastern Region	
CG Count 15min (flashes)	.24269
IC Count 15min (flashes)	.1862
IC FlashesPerCellArea	.13513
IC Count 2min (flashes)	.1221

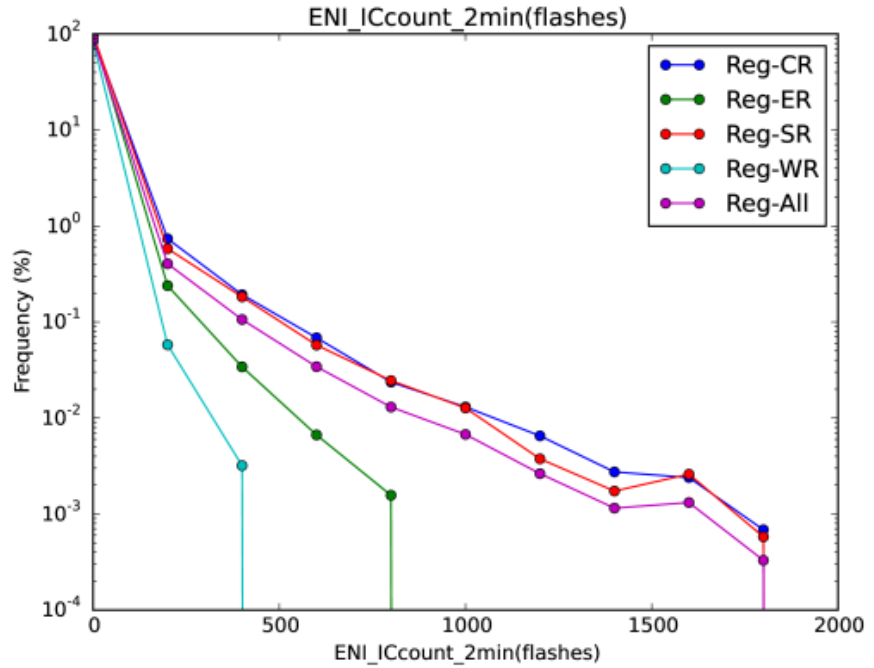
Western Region	
CG Count 15min (flashes)	.24966
IC Count 15min (flashes)	.17732
IC FlashesPerCellArea	.13014
IC Count 2min (flashes)	.11943

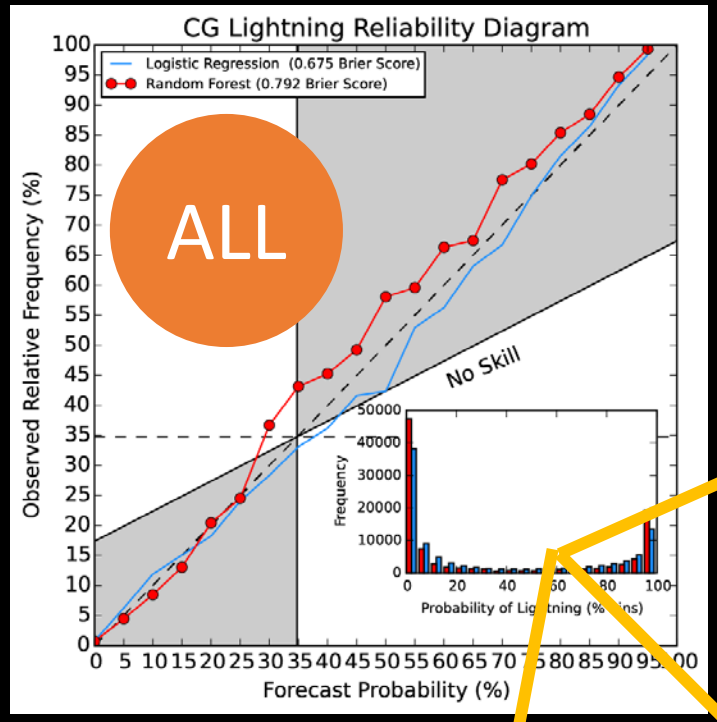

CGCount_15min(flashes)
Reg-CR
Reg-ER
Reg-SR
Reg-WR
Reg-All
Frequency (%)
10^{2}
10^{1}
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
0
100
200
300
400
500
CGCount_15min(flashes)

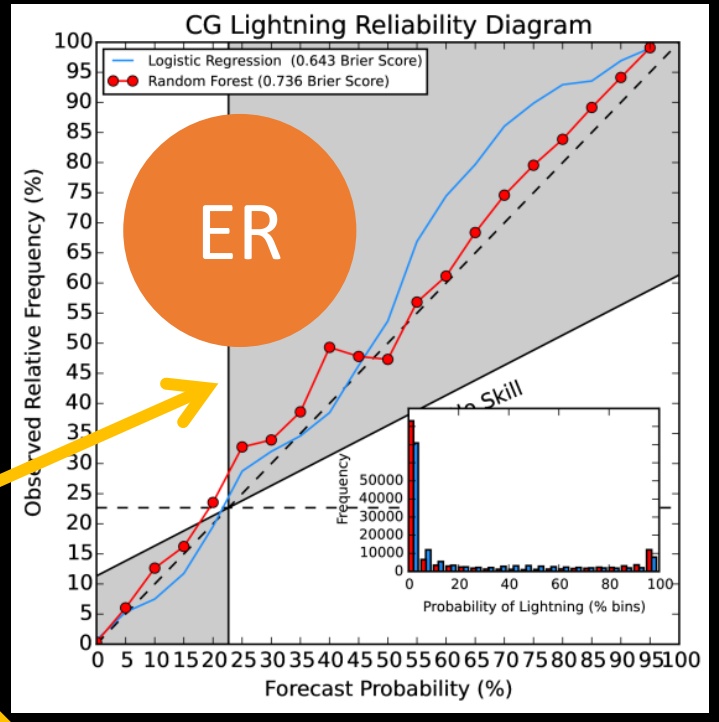

CGCount_15min(flashes)
Reg-CR
Reg-ER
Reg-SR
Reg-WR
Reg-All
Frequency (%)
CGCount_15min(flashes)

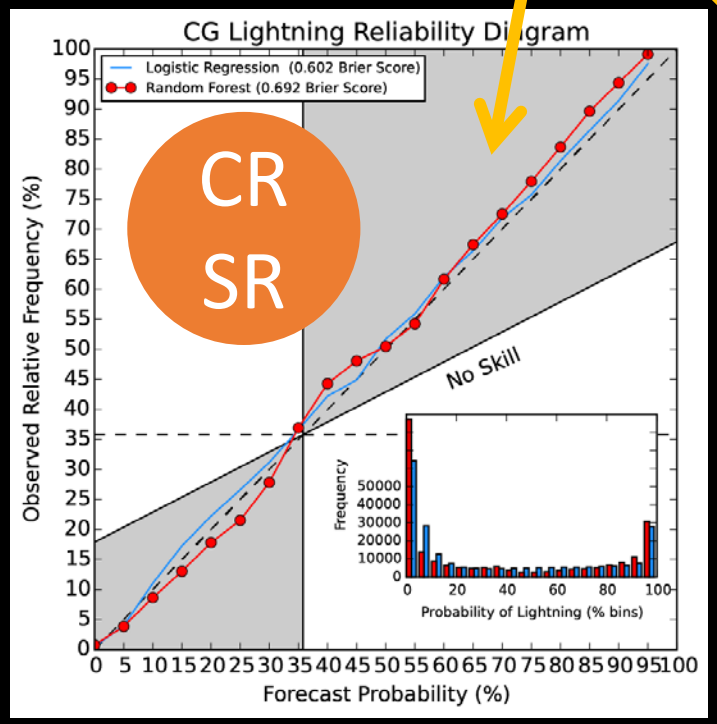

ENI_ICcount_15min(flashes)
Reg-CR
Reg-ER
Reg-SR
Reg-WR
Reg-All
Frequency (%)
ENI_ICcount_15min(flashes)

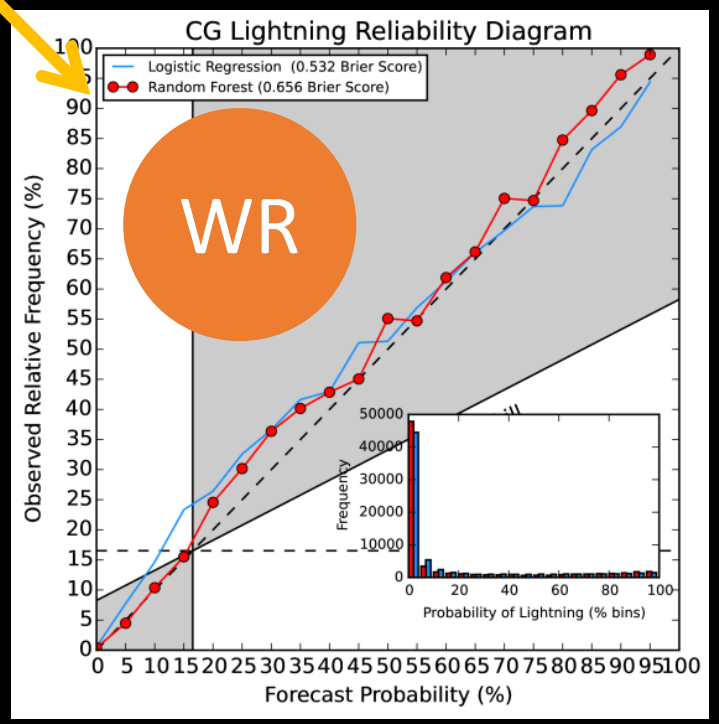

IC_FlashesPerCellArea
Reg-CR
Reg-ER
Reg-SR
Reg-WR
Reg-All
Frequency (%)
IC_FlashesPerCellArea


ENI_ICcount_2min(flashes)
Reg-CR
Reg-ER
Reg-SR
Reg-WR
Reg-All
Frequency (%)
ENI_ICcount_2min(flashes)







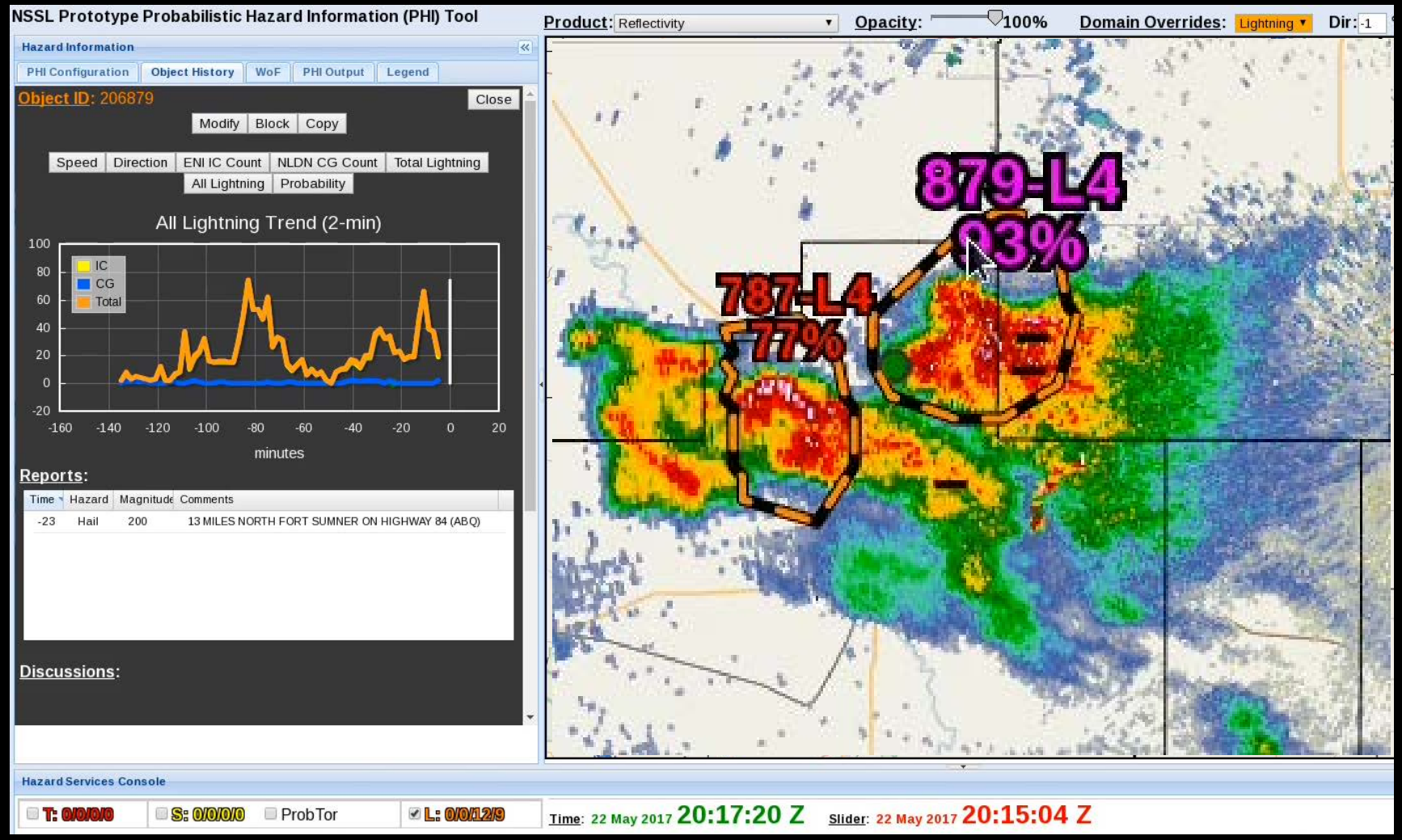


Region	BSS
Central & Southern	.692
Eastern	.736
Western	.656
ALL	*.792*

Real-time use of Probability of CG Lightning

- Created new WDSSII algorithm (*rfprob*)
  - used Python decision trees to create probabilities in realtime
- Forecasters used this product the Probabilistic Hazard Information Experiment in the Hazardous Weather Testbed to issue probabilistic lightning hazards
- Lightning Hazards went to Emergency Managers

- Both Emergency Managers and Forecasters responded positively to this product
  - EM wanted this immediately



Summary

- More data isn't always better
  - Can slow down processing
  - Not all of it is important
- Random Forests can easily take large dataset and create decision trees
  - Shows which features are most important
- Probability Lightning
  - very high Brier Skill Scores
  - Broken out via NWS region

- Future work:
  - Test on smaller scales (20km^2)
  - Incorporate Geostationary Lightning Mapper (on GOES-16)
  - Push this into NWS Operations (DSS)

Tiffany Meyer
Tiffany.Meyer@noaa.gov