Data Fusion: A Machine Learning Tool for Forecasting Winter Mixed Precipitation Events

Brian Filipiak, Kristen Corbosiero, Andrea Lang, Ross Lazear, and Nick Bassill

University at Albany, SUNY Albany, NY

102nd AMS Annual Meeting 01/26/2022

NWS Focal Points: Christina Speciale (ALB), Neil Stuart (ALB)





Background

- Numerous sources of data for forecasters
- Develop algorithm to breakdown conditions surrounding an event
- Identify important environmental variables for each type of precipitation
- Synthesize data and conditions to improve forecasting ability



Random Forest

- 'Forest' of decision trees
- Identify patterns and nonlinear interactions in data
- Train the trees to make a prediction from its previous knowledge
- Generate a probabilistic outcome and relative feature importance
- Promotes explainable AI





Data Sources

- CoCoRaHS Reports
 - Trained, consistent observations
- New York State Mesonet (NYSM)
 - Hourly Statistics
 - 5-minute observations
- Upper Air Soundings
 - BUF, ALB, OKX offices and WMW (Canada)
- North American Mesoscale (NAM) Forecast Model
 - 4km resolution from BUFKIT



10 m Winds (sonic)
10 m Winds (prop)
9 m Temperature
2 m Temperature
Relative Humidity
Solar Insolation
Precipitation
Snow Depth
Camera
Pressure
5 cm Soil
25 cm Soil
50 cm Soil



Random Forest Methods & Evaluation

- Configuration: 650 trees, 75/25 training and testing split, stratified
- Accuracy, Precision, Recall, F1 Scores





Random Forest Framework

State University of New York

Testing Data Sets



Random Forest Results (NYSM)



Original Soundings-NWS Buffalo, Albany, and Upton HAVG RCO: NYSM Hourly Averaged Raw and Calculated Original HAVG RO: NYSM Hourly Averaged Raw Original HAVG CO: NYSM Hourly Averaged **Calculated Original** OBS5 RCO: NYSM 5minute observations Raw and Calculated Original OBS5 RO: NYSM 5minute observations Raw Original OBS5 CO: NYSM 5minute observations **Calculated Original** ALL RCO: ALL NYSM Raw and Calculated Original ALL RO: ALL NYSM Raw Original ALL CO: ALL NYSM **Calculated Original**



Random Forest Results (NYSM)





Random Forest Results (NAM)

- NAM 4km sounding profile dataset
- Uses forecast hour timing to match with events



NAM_RCO: NAM Raw and Calculated Original NAM_RO: NAM Raw Original NAM_CO: NAM Calculated Original



Random Forest Results (NAM)



NAM_RCO: NAM Raw and Calculated Original NAM_RO: NAM Raw Original NAM_CO: NAM Calculated Original

NAM_RCN: NAM Raw and Calculated New NAM_RN: NAM Raw New NAM_CN: NAM Calculated New



Random Forest Results (NAM)





Random Forest Results (Updated NYSM)



Original Soundings- NWS Buffalo, Albany, and Upton Updated Soundings- NWS Buffalo, Albany, Upton and Maniwaki, Quebec or NAM ALL_RCO: ALL NYSM Raw and Calculated Original ALL_RO: ALL NYSM Raw Original ALL_CO: ALL NYSM Calculated Original

ALL_RCN: ALL NYSM Raw and Calculated New ALL_RN: ALL NYSM Raw New ALL_CN: ALL NYSM Calculated New

ALL_NAM_RCN: ALL NYSM and NAM Raw and Calculated New ALL_NAM_RN: ALL NYSM and NAM Raw New

ALL_NAM_CN: ALL NYSM and NAM Calculated New



Operational Product

- Develop web-based product for operational use
- Live updating map of probabilities with radar/reflectivity
- Incorporate information about most important variables





Operational Product

Data Fusion Operational Website: http://www.atmos.albany.edu/student/ filipiak/op/



SCAN ME



NYSM and Upper Air Raw and Calculated Data

Home NYSM & Upper Air NAMNEST Data and Methods Training

****EXPERIMENTAL****

Please click any button to view the map of your choice

|--|

Top 10 Most Important Variables

temp_2m_min	temp_2m_avg	temp_2m_max	temp_2m	Geopotential	Temperature at	Temperature at	Temperature at	700hpa Wet Bulb	850hpa Wet Bulb
[degC]	[degC]	[degC]	[degC]	Height at 500hpa	850hPa	700hPa	925hPa	Temperature	Temperature
0.07293	0.068204	0.06617	0.065254	0.032656	0.031072	0.02952	0.02824	0.024154	0.023014

Dec 7, 2021

01:31:11 UTC

Contact: Brian Filipiak bfilipiak@albany.edu







NAMNEST Raw and Calculated Data						
Home	NYSM & Upper Air	NAMNEST	Data and Methods	Training		
****EX	KPERIMENTAI	Dec 7, 2021 01:31:24 UTC				
Please click any button to view the map of your choice						

Sleet

All Mixed

Freezing Rain

State University of New York

Dominant

Feature Table for Each Forecast Hour

Contact: Brian Filipiak bfilipiak@albany.edu

Snow

Rain



Dominant Precipitation Probabilities with NAM Composite Reflectivity Valid at 01_19_2022_1800Z



г 75

- 65

55

- 45 25 Refectivity dBZ

25

- 15

>



Recent/Future Work

- Studied individual variable impact more in depth
- Add in additional data sources and combinations
- Develop metrics to identify how our product compares to other methods

Albany

 Evaluate and complete verification of website products over winter months

State University of New York



	Phenomenon obse			
Forecast	Yes	No	Metrics	
Yes	True positives (TP)	False positives (FP)	Precision	
	Hits	False alarms	ТР	
			$\overline{TP + FP}$	
No	False negatives (FN)	True negatives (TN)	Negative predictive value	
	Misses	Correct negatives	TN	
			FN + TN	
Metrics	Sensitivity	Specificity	Accuracy	
	TP	TN	TP + TN	
	$\overline{TP + FN}$	$\overline{FP + TN}$	TP + FP + FN + TN	

Conclusions

- The random forest can accurately identify different winter precipitation types
- Specific data types and combinations of may need to be treated differently
- There will always be room for human interpretation of ML generated guidance





Questions?

Contact: Brian Filipiak, bfilipiak@albany.edu

http://www.atmos.albany.edu/student/filipiak/ op/



