An Updated Machine Learning Ensemble for Atlantic Tropical Cyclone Rapid Intensification Forecasting

Andrew Mercer

Alexandria Grimes

Kimberly Wood

Department of Geosciences, Mississippi State University

74th Interdepartmental Hurricane Conference

Tropical Cyclone Operations and Research Forum

February 26, 2020

STATE

MISSISSIPPI STATE UNIVERSITY™

USWRP

Joint Hurricane Testbed

Work funded by NOAA Fund #NA17OAR4590140

Background of Project

- Considerable previous work has investigated predictability of rapid intensification (RI) within Atlantic tropical cyclones (TCs)
  - RI defined as an increase of peak wind speed of 30 kt in 24 hours, current operational definition by NHC
- 2018-2019 Testbed experiments included AI-based ensemble for predicting Atlantic RI using SHIPS-RII predictors and GFS predictors
- Could the AI ensemble be improved using better features from GFS/FNL data in combination with SHIPS-RII predictors?

2019 Testbed Example Output

** MSU experimental Atlantic RI AI ensemble prediction: AL52019 Dorian 08/27/19 0

Global probability for RI (30kt/24h) 13.2%

This AI ensemble run is NOT EXPECTING RI (RI prob <= 21%)

Individual Ensemble Forecasts

Ens Name	RI/no RI	RI prob
SVM1	RI	57.9%
SVM2	no RI	16.4%
SVM3	no RI	7.3%
SVM4	no RI	3.8%
SVM5	no RI	3.1%
SVM6	no RI	15.5%
SVM7	no RI	7.0%
SVM8	no RI	4.2%
SVM9	no RI	3.2%

(RI AI ensemble predictor table for 30 KT OR MORE MAXIMUM WIND INCREASE IN NEXT

Predictor	Value	RI Predictor Range	Scaled Value
TLAT	13.3	8.9 to 34.3	-1.31
SHRG	10.3	4.8 to 31.9	-1.37
BTMAX	-56.2	-84.9 to 4.1	-0.68
KPC9_VVEL200_7	-4.25e-09	-1.84E-08 to 1.84E-08	0.08
KPC8_T600_7	3.77e-05	-2.29E-04 to 1.35E-04	1.25
VMFX	7.9	0.99 to 12.7	0.80
RHMD	45	38.0 to 84.0	-1.00
RI_COUNT	0	0.0 to 8.0	-0.60
V500	3.6	0.6 to 11.1	-0.43
RI_FLAG	0	0.0 to 1.0	-0.70
KPC9_VVEL200_9	-1.16e-04	-1.29E-04 to 1.13E-04	-1.17
KPC4_VVEL200_9	1.90e-06	-1.10E-04 to 4.09E-05	0.90
KPC1_VVEL200_9	3.63e-05	-3.92E-05 to 5.20E-05	0.40
KPC4_T600_7	1.49e-06	-3.87E-05 to 9.58E-05	0.34

Preliminary Verification Skill, 2017 – 2019 seasons (old ensemble)

- Verification skills from old model did not align with the training phase BSS values (which were roughly 0.4)
  - 2017: 0.042
  - 2018: 0.034
  - 2019: -0.08

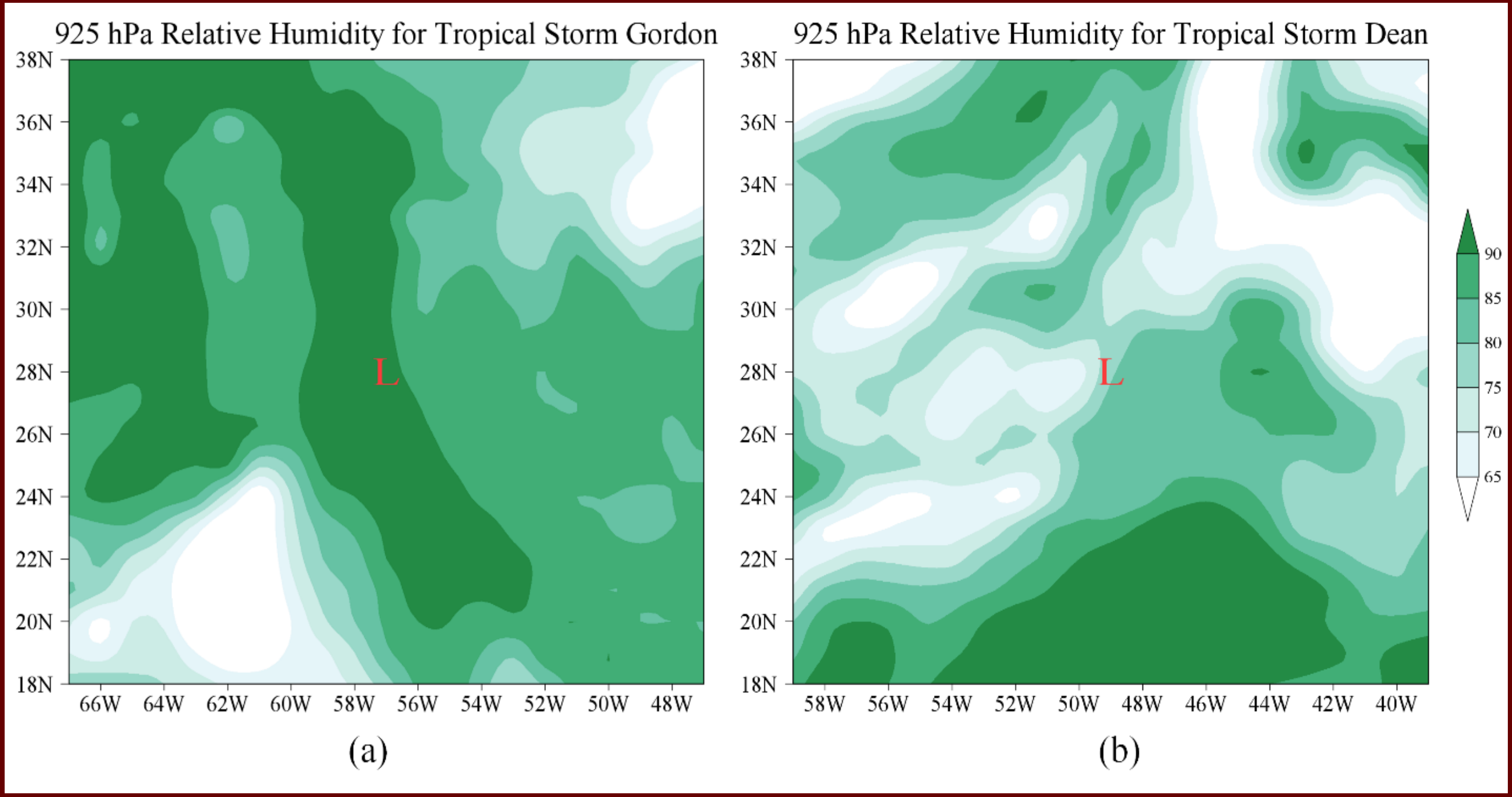
    (preliminary based on b-decks for 11 storms – through Lorenzo)
- Poor performance attributed to overfitting in the cross-validation phase when building the ensemble
- New ensemble generation has been finalized and is now being verified, as described below

Data and Methods – Updates

- RI dataset – HURDAT2
  - All Atlantic Basin TCs spanning 2004-2016 were included, a total of 3605 timesteps (observations taken every 6 hours)
  - Used HURDAT2 only to categorize RI (no real-time best-track data used for forecasts)
- Statistical Hurricane Intensification Prediction System (SHIPS) Rapid Intensification Index (RII) parameters
  - Initial set of 109 predictors from SHIPS used prior to feature selection
    - 71 SHIPS predictors
    - 17 GOES IR imagery predictors
    - 21 Precipitable water predictors
- Also employed 5 persistence predictors (6, 12, and 24-hour previous intensity change, previous RI flag, previous RI count)

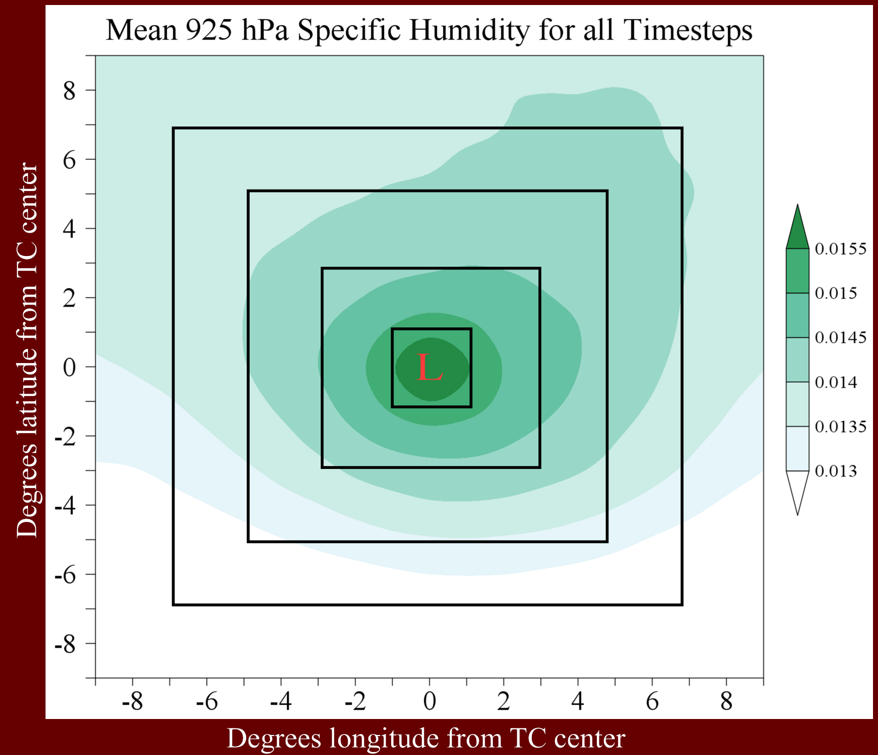
Data and Methods

- SHIPS predictors do not offer spatial insight into TC structure (e.g. 850 hPa specific humidity for two TCs with same low-level mean RH predictor value of 74)



Data and Methods

- Multiple GFS analysis grids retained for each TC timestep
  - 98 total grids
    - 6 three-dimensional variables on 11 isobaric surfaces, including temperature, *u* and *v* wind, vertical velocity, absolute vorticity, equivalent potential temperature, specific humidity
    - Static stability on 9 isobaric levels
    - 12 single-level grids, including MSLP, skin temperature, tropopause *u, v, T,* elevation, and pressure, surface-based CAPE and CIN, 850-200 hPa shear, and 200 hPa divergence



Data and Methods

- PCA and KPCA formulated on GFSA grids individually to quantify maximum separation
  - Tested 17 different kernel functions in KPCA, as well as RPCA
    - RBF kernels with separation parameter of 5, 10, 25, 50, 75, 100, 200, 500, 1000
    - Polynomial kernels with degrees from 2 to 10
  - Separation quantified by *k*-means cluster analysis with 2 clusters, quantifying RI/non-RI separation between clusters
  - Separability metric the sum of the RI percentage in RI cluster with the non-RI percentage in the non-RI cluster
  - All PC scores whose clustering yielded a separation metric exceeding the 99th percentile were retained
  - Analyses repeated for each RI category in Kaplan et al. (2015), generated unique predictors for each RI definition

Data and Methods

- PCA/KPCA separability results

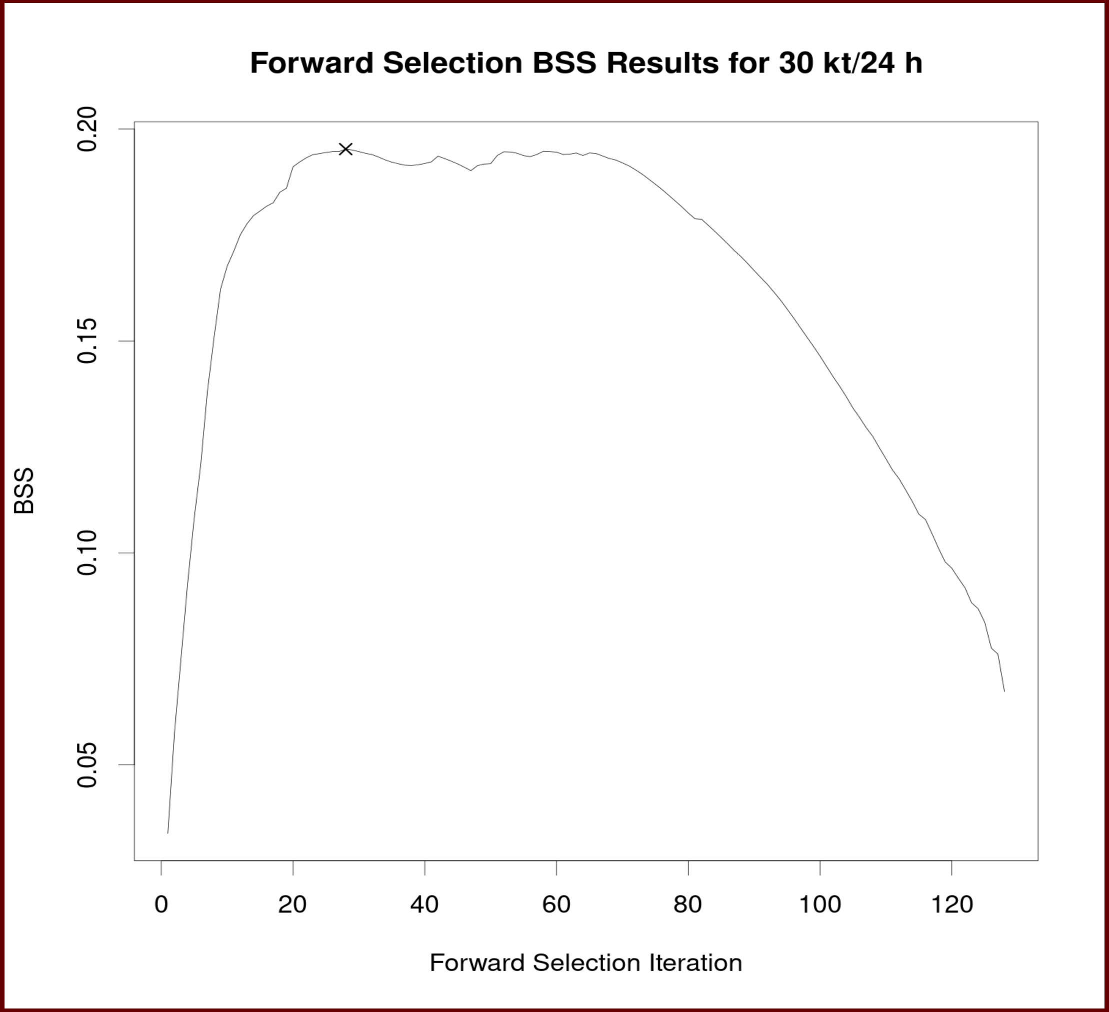
GFSA Field	Grid Size	PCA Method	PCs Kept	RI Frequency	Non-RI Frequency
30 kt/24 h					
400 hPa absolute vorticity	15° x 15°	RPCA	5	0.87	0.43
500 hPa absolute vorticity	11° x 11°	KPCA (σ=50)	6	0.74	0.55
850-200 hPa shear	11° x 11°	RPCA	2	0.87	0.42
600 hPa absolute vorticity	7° x 7°	RPCA	3	0.79	0.51

Feature Selection

- After including PCA-derived features and SHIPS predictors, used feature selection to reduce the 120+ feature sets for each RI category
- Feature selection done using forward selection on all 128 possible predictors (8256 combinations)
  - Forward selection used leave-one-season out approach to minimize overfitting issues from before
- Identified global BSS maximum on testing set, this value was used as a measure of best performance for the given predictor set
- Very computationally expensive but should find global maximum in performance for the given predictor set

Feature Selection

- Example – 30 kt/24 h



New Ensemble Members

- After establishing the feature selection procedure, building AI ensemble was next
- Tested multiple support vector machine (SVM) configurations, as well as a logistic regression model
  - Logistic regression deemed proxy for current SHIPS-RII since it gave similar performance and utilized similar predictors
- SVM configuration parameters tested; all possible permutations from set below (40 possible permutations)
  - Cost function – 0.1, 1, 10, 100
  - Values for γ in radial basis function kernel – 0.01, 0.05, 0.1, 0.25, 0.5, 0.75, 1, 2, 5, 10
- 41 total ensemble members considered

New Ensemble Members

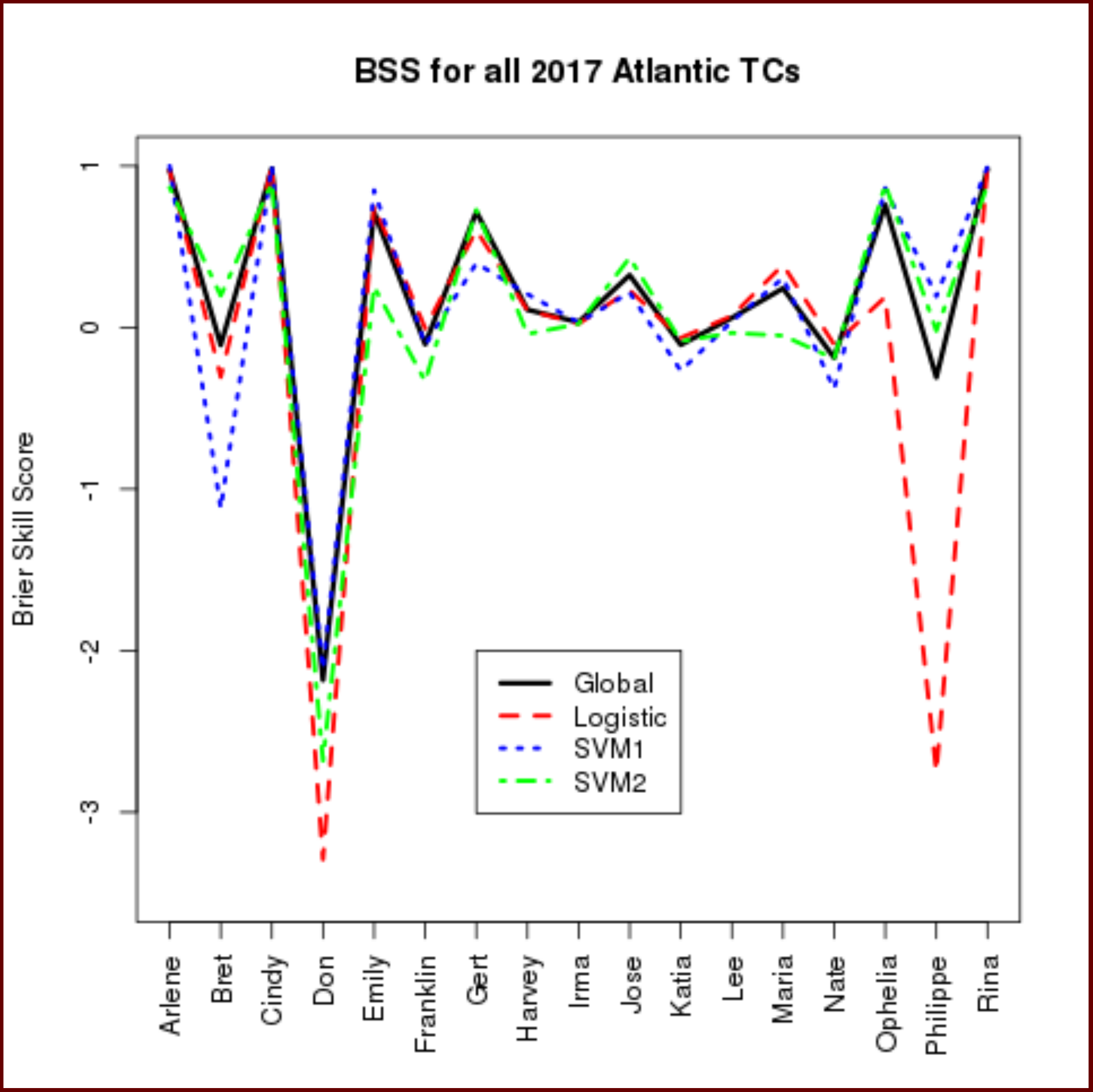
- Logistic regression performance was baseline from which ensemble members were selected
- Logistic BSS from optimal predictor (tuned for logistic model) – 0.155
- When comparing 40 SVM configurations against this value, only two members outperformed the logistic model
  - SVM – γ = 0.01, cost = 10 (called SVM1) – BSS = 0.169
  - SVM - γ = 0.1, cost = 10 (called SVM2) – BSS = 0.164
- Boost of roughly 10% over logistic model

New Predictors

- Each ensemble member retained its own unique set of predictors optimized to that member
- Predictor set sizes were still large:
  - Logistic regression – 21 predictors
  - SVM1 – 45 predictors
  - SVM2 – 36 predictors
- Common variables in all ensemble members included
  - KPC5 for 400 mb Absolute Vorticity (15 x 15 grid)
  - Generalized Shear Predictor (SHRG)
  - Percent area with brightness temperature < -10C
  - 6-hour intensity change
  - Number of previous RIs
  - Previous RI flag (1=RI, 0=no RI)

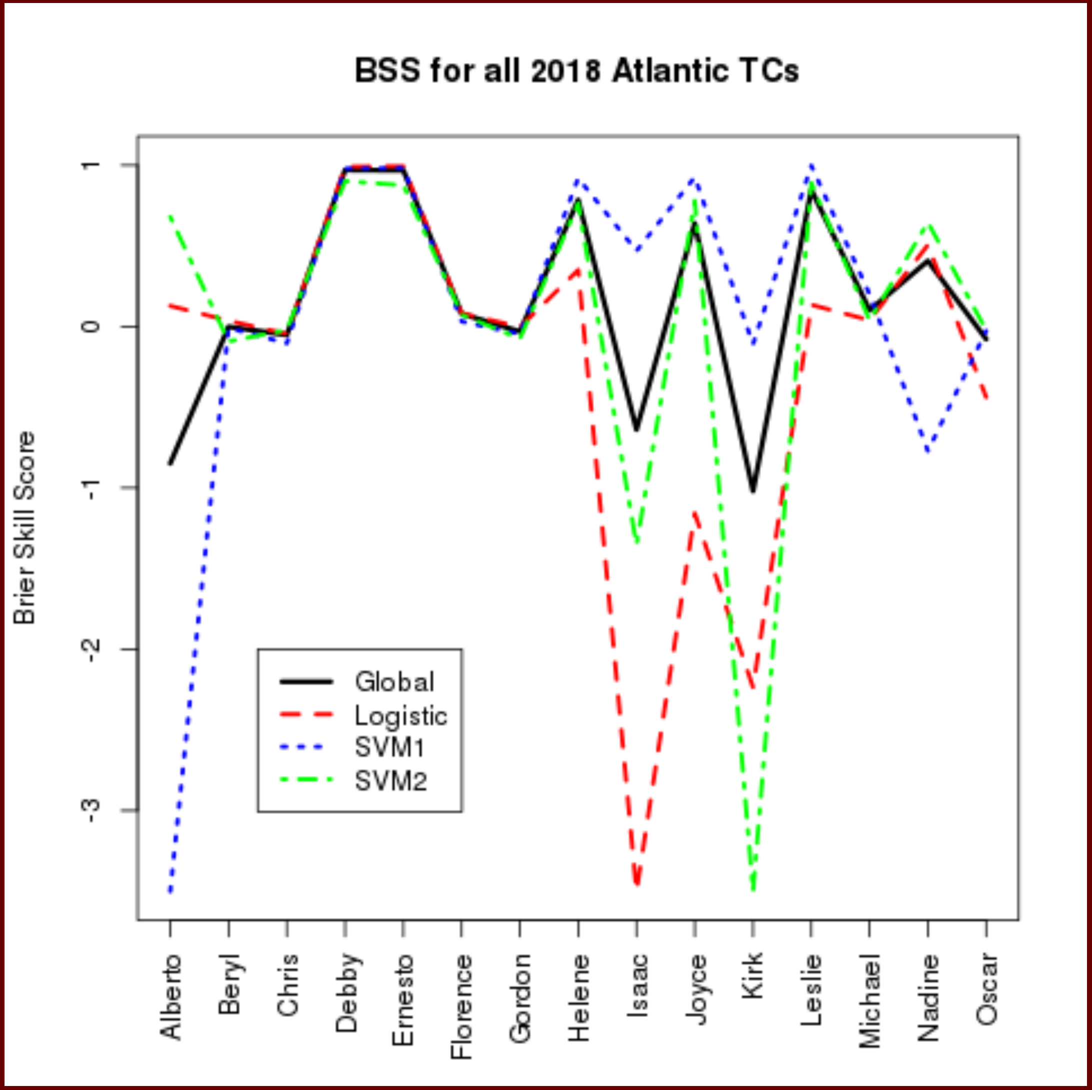
Results, New Ensemble - 2017

- All timesteps
  - Global BSS: 0.111
  - Logistic: 0.125
  - SVM1: 0.089
  - SVM2: 0.038
- Only RI timesteps
  - Global BSS: 0.123
  - Logistic: 0.152
  - SVM1: 0.141
  - SVM2: 0.041



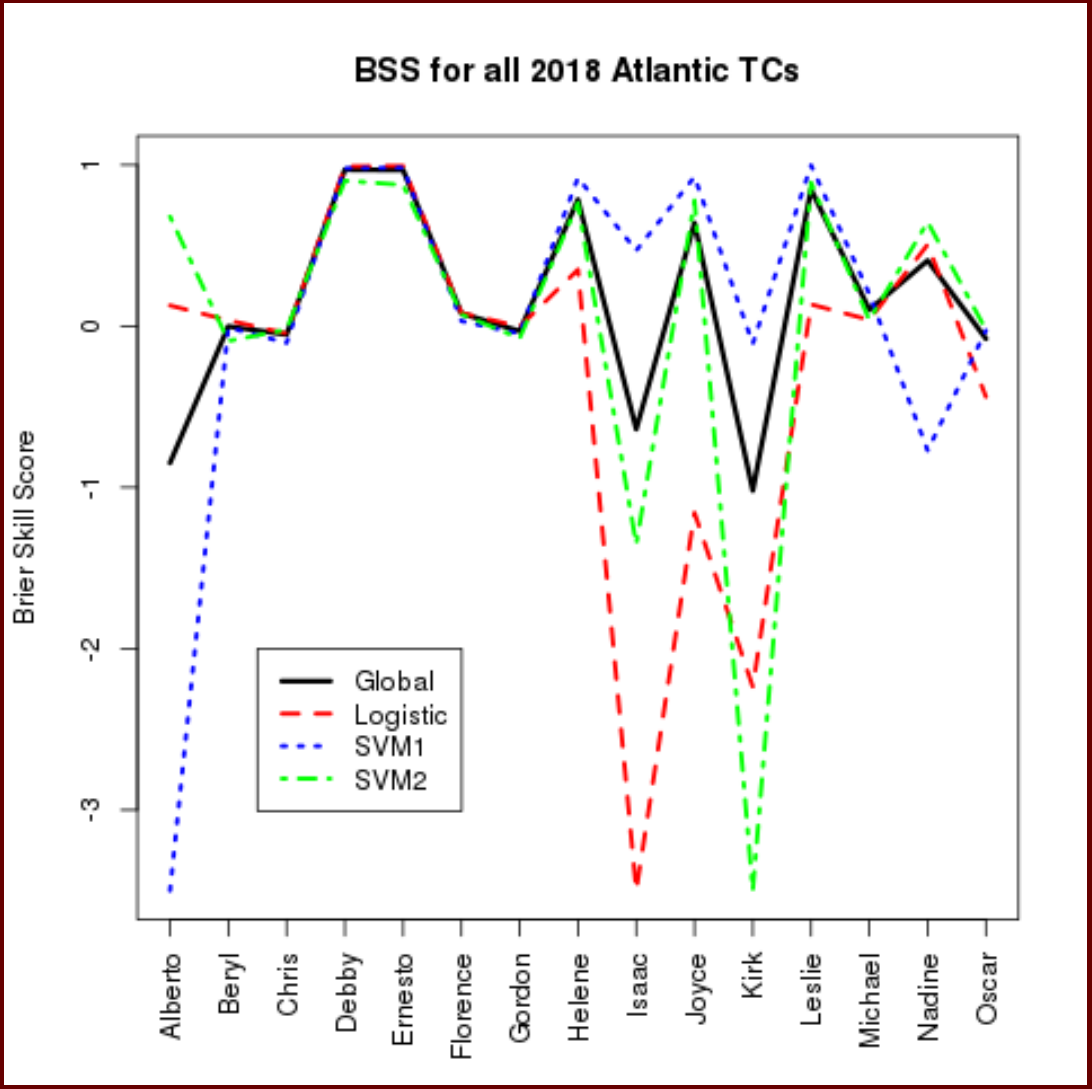
Results, New Ensemble - 2018

- Global BSS: 0.067
  - Logistic: -0.018
  - SVM1: 0.075
  - SVM2: 0.023
- Only RI timesteps
  - Global BSS: 0.037
  - Logistic: 0.020
  - SVM1: 0.053
  - SVM2: 0.027



Results, New Ensemble - 2018

- Global BSS: 0.067
  - Logistic: -0.018
  - SVM1: 0.075
  - SVM2: 0.023
- Only RI timesteps
  - Global BSS: 0.037
  - Logistic: 0.020
  - SVM1: 0.053
  - SVM2: 0.027



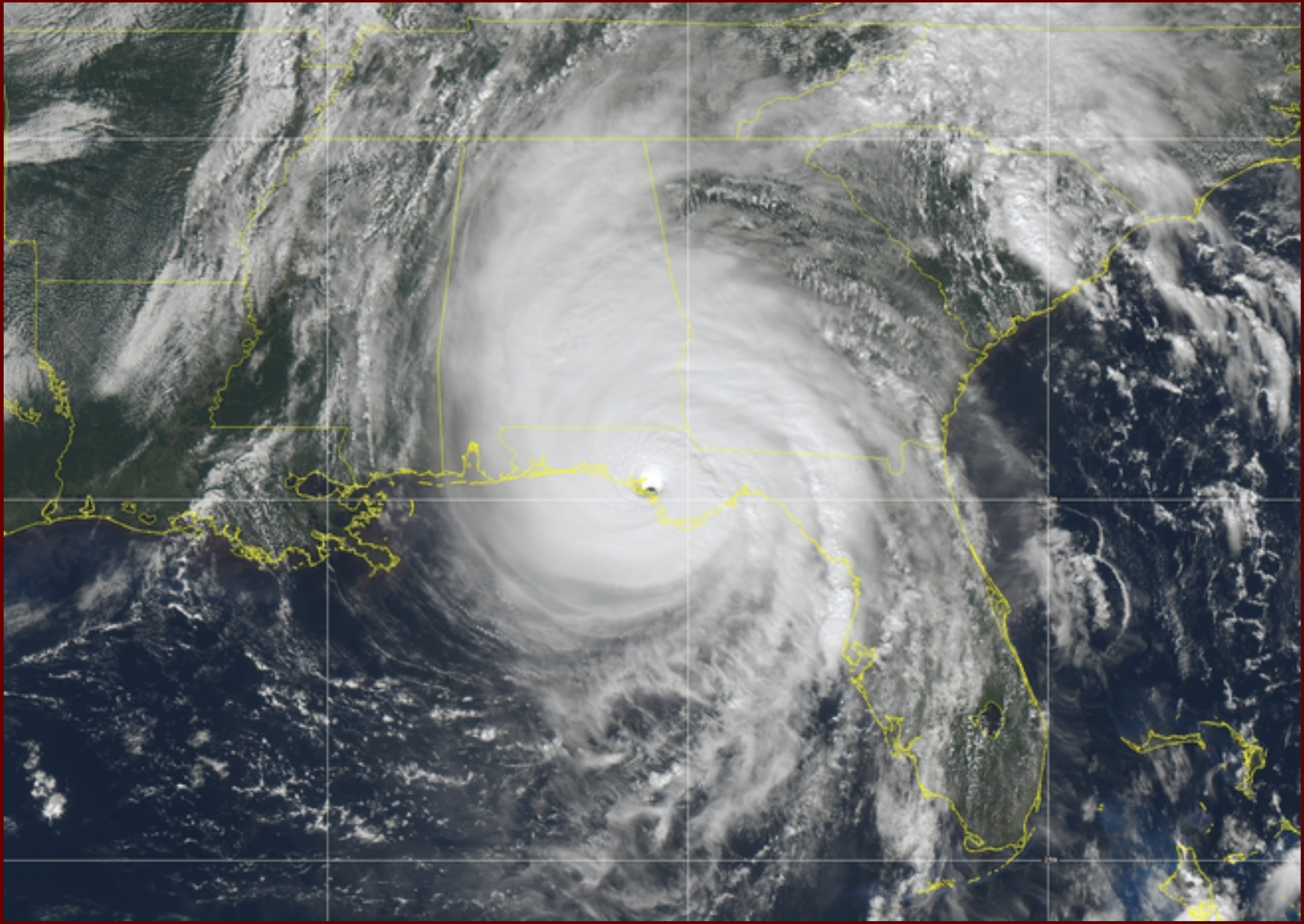
Results

- Individual member performance poorer than global results for 2017-2018. Importantly, results are less variable in global model, which is good for forecast applications
- All years results
  - Global BSS: 0.097
  - Logistic: 0.078
  - SVM1: 0.085
  - SVM2: 0.033

Conclusions

- Updated cross-validation routine providing realistic BSS values during model training phase, boosting results over logistic regression by about 10%
- GFS predictors included in all three retained ensemble members
  - Mid-level vorticity clearly has importance in identifying RI environments
- 2017-2018 results were improved on many storms that were difficult to forecast, though some poor performing storms remain
  - Should improve with additional years of data and more training of ensemble
- Working now on identifying relationships between predictors and forecasts for interpretation

Questions?



Results, New Ensemble - 2019

- Global BSS: -0.059
  - Logistic: -0.051
  - SVM1: -0.132
  - SVM2: -0.125
- Only RI timesteps
  - Global BSS: 0.003
  - Logistic: 0.078
  - SVM1: -0.088
  - SVM2: -0.013