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

Andrew Mercer

Alexandria Grimes

Kimberly Wood

Department of Geosciences, Mississippi State University

73rd Interdepartmental Hurricane Conference

Tropical Cyclone Operations and Research Forum

March 13, 2019

Worked 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
- Forecast skill for Atlantic TC RI only slightly positive
- 2018 Testbed experiments included AI-based ensemble for predicting Atlantic RI using SHIPS-RII predictors
- Could the AI ensemble be improved using better features from GFS/FNL data in combination with SHIPS-RII predictors?

2018 Testbed Results, Old Model

TC Event	SHIPS-RII	LOGISTIC	BAYESIAN	CONSENSUS	AI Ensemble
Alberto	--	--	--	--	0.77
Beryl	-0.26	0.066	0.07	0.052	-0.075
Chris	0.035	-0.023	-0.085	0.011	0.01
Debby	-1.42	0.98	1	0.68	--
Ernesto	-1.62	0.97	1	0.66	0.95
Florence	0.14	0.14	-0.015	0.086	-0.03
Gordon	-2.18	-3.47	0.92	-0.75	0.02
Helene	-0.19	-0.15	0	-0.071	0.05
Isaac	-4.07	-2.28	-2.02	-2.18	0.66
Joyce	0.65	1	1	0.95	1
Kirk	-2.04	-1.87	0.91	-0.42	0.11
Leslie	-1.78	0.92	1	0.59	0.75
Michael	0.26	0.1	-0.066	0.11	0.08
Nadine	-0.77	0.89	1	0.69	0.59
Oscar	-0.48	0.91	1.00	0.75	0.78
2018 Season	0.0063	0.042	0.0022	0.073	0.043

Updates to AI Ensemble

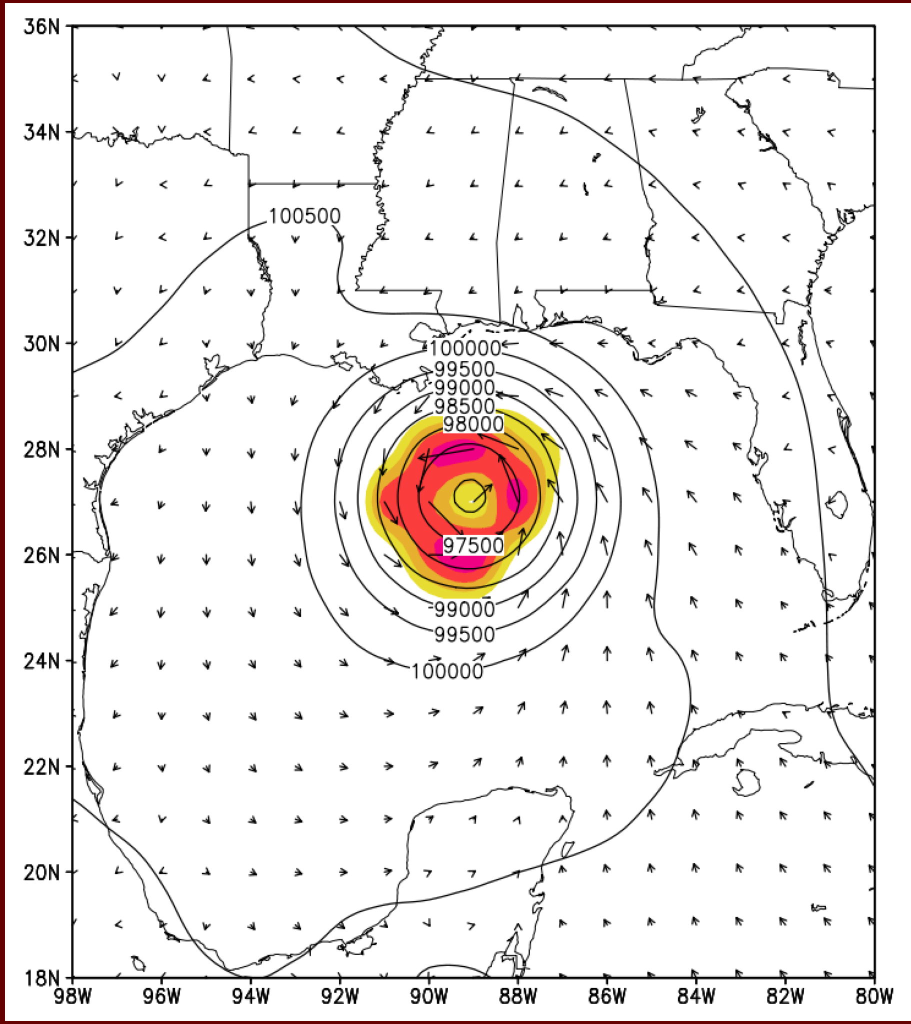
- RI dataset – HURDAT2
  - RI defined as increase of 30 kt in wind speed in 24 hours (5.8% of the timesteps considered)
  - All Atlantic Basin TCs spanning 1999-2016 were included, a total of 5409 timesteps (observations taken every 6 hours)
  - Also obtained storm characteristic predictors (8 total), including max wind speed, TC center latitude/longitude, 24-hour, 12-hour, and 6-hour intensity change, and flags indicating previous RI for a storm
- SHIPS-RII predictors retained (16 total)
  - Predictors included RSST, U200, VMPI, RHLO, Z850, D200, REFC, T200, SHRD, PENV, POT, MTPW, CFLX, RHCN, SDBT, BT30

Updates to AI Ensemble

- FNL data (1° global data every 6 hours)
  - GFS final analysis fields, updated GFS analysis fields with additional observations included, typically produced an hour after GFS operational forecast
  - Used to provide robust training database from which AI ensemble may be improved
- Fields selected from FNL (98 total TC-centric domains)
  - Isobaric fields – Temperature, relative humidity, absolute vorticity, *u* and *v* wind components, and vertical velocity
    - Retained at 1000, 925, 850, 700, 600, 500, 400, 300, 250, 200, 100 mb
  - Single-level fields – CAPE, CIN, sea level pressure, precipitable water, surface/skin temperature
  - Calculated fields from FNL – static stability, equivalent potential temperature, vertical shear, divergence

Data and Methods

- TC-center centric domain
  - Based on FNL-analyzed minimum SLP (using NHC as first guess) to minimize bias
  - Extends 9° latitude longitude in all directions, total of 19 x 19 grid (361 points)
  - 98 layers with 361 points per layer gives 37358 total features
  - Dire need for feature reduction method



Data and Methods

- Kernel PCA – feature reduction technique
  - Project datasets into higher dimensional Hilbert space, as in SVMs, via kernel matrix (similar to correlation matrix in PCA)
  - Compute KPC loadings by scaling the eigenvectors of the kernel matrix by the square root of the eigenvalues
- Kernel PCA cannot yield unique patterns as it is not possible to back-transform data from Hilbert space back to original space
- KPC loadings a square matrix with dimensionality as the number of timesteps (5409)
- Difficult to choose proper number of KPCs to retain as they do not contribute linear variance explained as in RPCA. Tested between 2 and 20 KPCs (trial and error)

Data and Methods

- Kernel functions tested
  - Radial basis function (RBF) kernel - $k(x, x') = \exp[-\frac{\|x-x'\|}{2\sigma^2}]$
  - Polynomial kernel - $k(x, x') = (xx' + 1)^d$
  - For a kernel matrix, $x' = x^T$
- Kernel functions have tunable parameters. For RBFs, the σ value is tunable (represents spread in the Gaussian function). In polynomial kernels, the degree *d* is tunable
- Tested σ = 1, 5, 10, 25, 50, 75, 100, 200, 500, 1000 for RBF, *d* = 1 to 10 for polynomial kernel
- KPC loading matrix with given configuration used to identify which loadings split most effectively into RI and non-RI clusters
  - Used K-means clustering with 2 clusters
  - Calculated percentage of RI in each cluster, maximum difference between the two was the KPC loading that separated the data most effectively

Data and Methods

- Best KPCA configurations by layer (top 9 kept, 55 KPCA predictors)

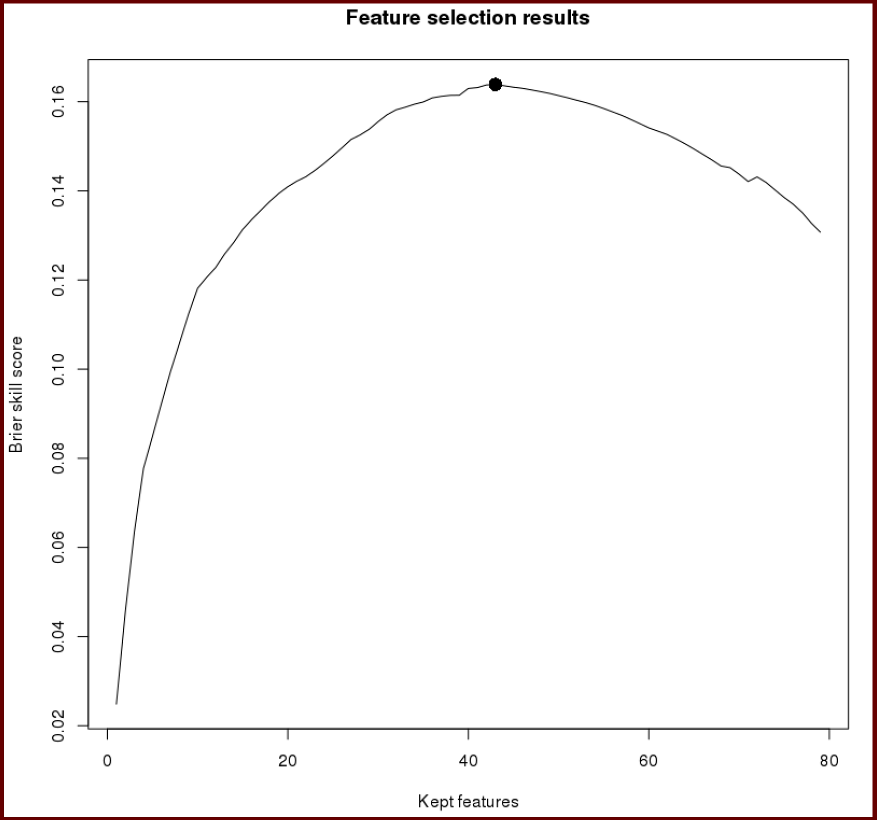
Layer Name	# KPCs Kept	Kernel	Separation Percentage
200 mb Temperature	2	Poly (deg=7)	0.203
200 mb v-wind	19	RBF (σ=1)	0.241
850 mb Specific Humidity	5	RBF (σ=1)	0.207
500 mb Vertical Velocity	11	Poly (deg=5)	0.21
850 mb Theta-E	2	RBF (σ=1)	0.21
600 mb Theta-E	6	Poly (deg=7)	0.215
500 mb Theta-E	2	Poly (deg=8)	0.205
200 mb Theta-E	6	Poly (deg=7)	0.244
600 mb Static Stability	2	Poly (deg=2)	0.231

Feature Selection

- 55 total KPC predictors, 8 NHC-derived predictors and 16 SHIPS predictors (a total of 79)
- Required feature selection method to reduce number of predictors and optimize skill
- Jackknife (leave-one-out) cross-validation for 5409 timesteps (5408 training, 1 testing, repeated for all timesteps)
- Forward selection logistic regression fitting using cross-validation to identify optimal feature combination
  - Optimization done using Brier skill score

Feature Selection

- 43 predictors kept
  - FNL data (31 predictors)
    - 200-mb *v* 13 KPCs
    - 850-mb *q* 3 KPCs
    - 200-mb *T* 2 KPCs
    - 200-mb θ_e 2 KPCs
    - 500-mb θ_e 1 KPC
    - 600-mb θ_e 2 KPCs
    - 500-mb *ω* 7 KPCs
    - 600-mb *σ* 1 KPC
  - NHC data (4 predictors)
    - Maximum wind speed, RI occurrence flag, RI count, 24-hour intensity change
  - SHIPS data (8 predictors)
    - -30C brightness temperature, low-level relative humidity, mean 200 mb temperature, ocean heat content, average surface pressure, 850 mb-250 mb shear magnitude, brightness temperature standard deviation, Reynolds SST



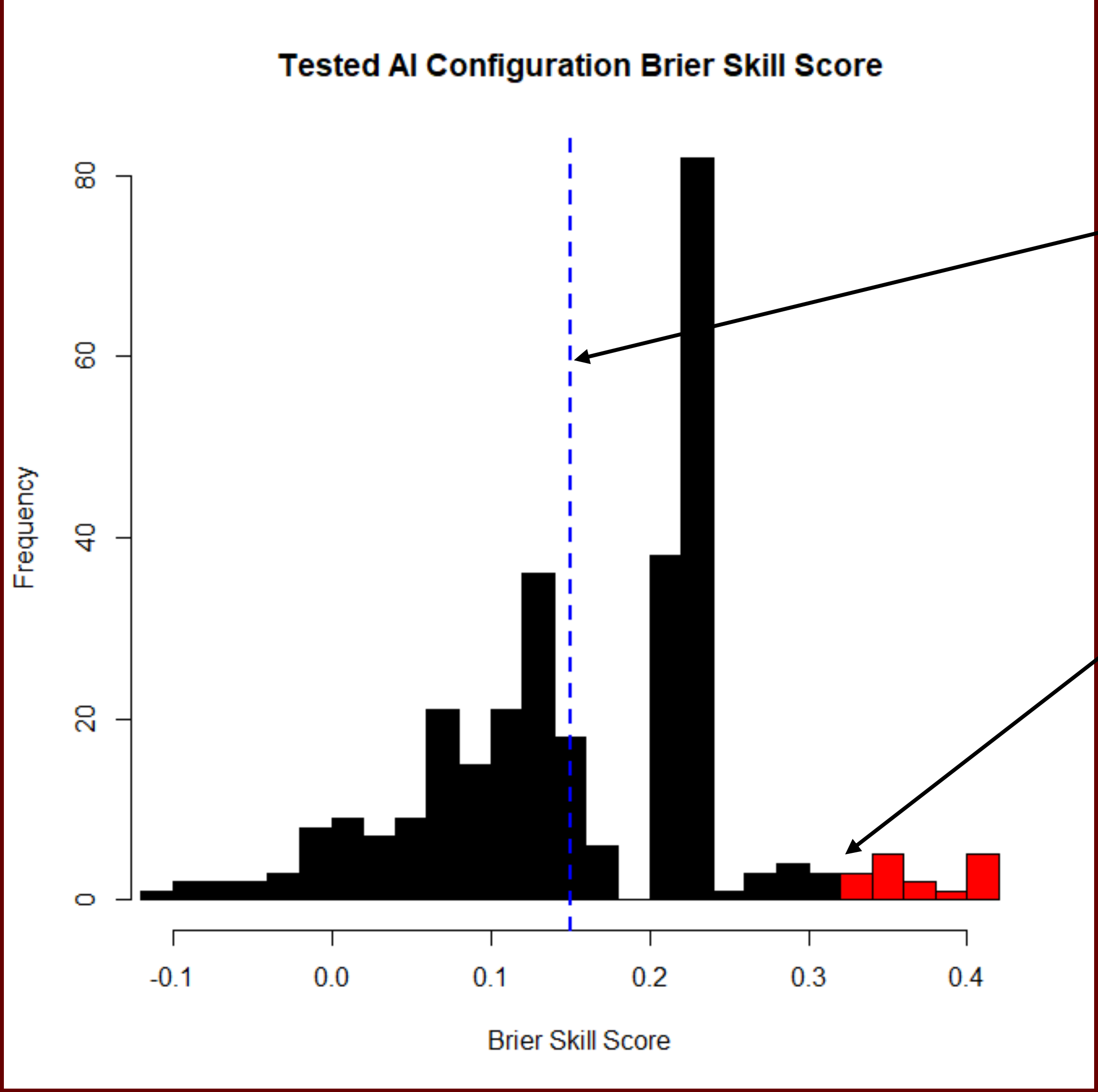
Developing AI Ensemble

- After identifying the optimal feature combination, used numerous AI configurations to train RI prediction ensemble (307 total configurations)
  - Support vector machines (27 configurations)
    - RBF kernels (γ = 0.01, 0.025, 0.05, 0.075, 0.1, 0.125, 0.15, 0.175, 0.2)
    - Costs (1, 10, 100)
  - Simple artificial neural networks (160 configurations)
    - 1 hidden layer for all networks, hidden nodes in the layer between 2 and 5
    - Weight decay rate (0.0025 to 0.025 by 0.0025)
    - Maximum iterations/epochs (50000, 100000, 200000, 500000)
  - Random Forests (120 configurations, all possible permutations)
    - Number of variables used for each tree, 2 to 5
    - Number of grown trees (1000 to 30000 by 1000)

Developing AI ensemble

- Use same jackknife cross-validation methodology to identify optimal ensemble members
- Used probabilistic output from each AI tested, identifying an RI/non-RI cutoff probability that optimized Heidke skill score
- Top AI ensemble performers retained as part of the AI ensemble (top 10%)
  - No neural networks or random forests were retained
- Global AI ensemble probability and individual AI ensemble member probabilities used to generate Brier skill scores, compared against logistic baseline and current SHIPS-RII results

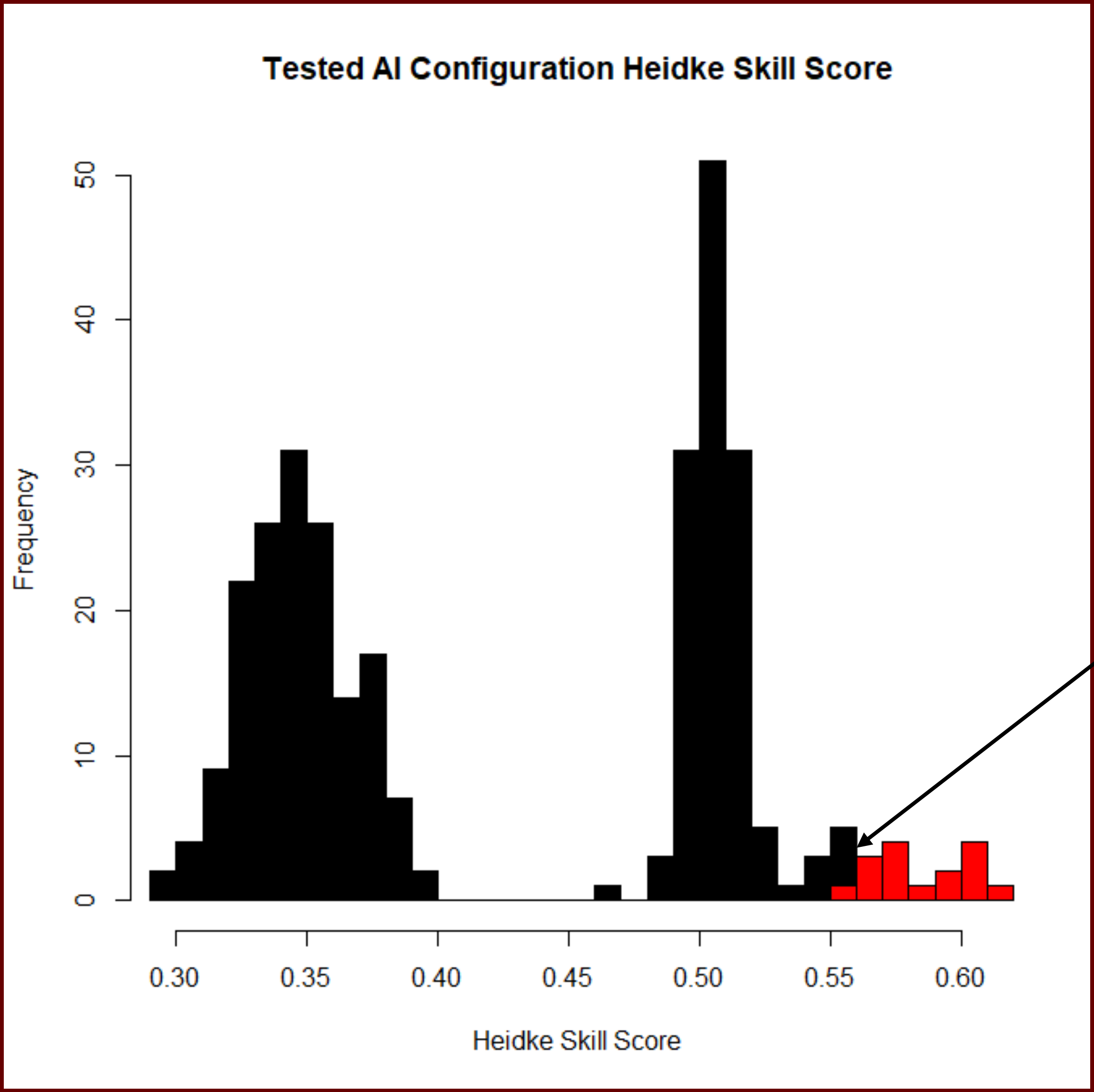
Ensemble Member BSS



SHIPS-RII Consensus

95th percentile

AI ensemble results (contingency)



95th percentile

Note: All of the top performers were SVM configurations

Global and Member Contingency and BSS Statistics

Member	RI Cutoff	POD	FAR	Bias	HSS	BSS
SVM (cost=1,γ=0.05)	0.21	0.57	0.39	0.94	0.57	0.34
SVM (cost=1,γ=0.075)	0.15	0.64	0.43	1.11	0.58	0.36
SVM (cost=1,γ=0.1)	0.16	0.63	0.42	1.10	0.58	0.35
SVM (cost=10,γ=0.025)	0.22	0.67	0.40	1.12	0.61	0.41
SVM (cost=10,γ=0.05)	0.20	0.69	0.42	1.18	0.60	0.42
SVM (cost=10,γ=0.075)	0.19	0.67	0.41	1.14	0.61	0.40
SVM (cost=10,γ=0.1)	0.17	0.65	0.42	1.13	0.59	0.37
SVM (cost=10,γ=0.125)	0.26	0.57	0.38	0.92	0.57	0.35
SVM (cost=10,γ=0.15)	0.21	0.57	0.40	0.96	0.56	0.32
SVM (cost=10,γ=0.175)	0.14	0.60	0.43	1.06	0.56	0.29
SVM (cost=100,γ=0.025)	0.24	0.65	0.42	1.11	0.59	0.39
SVM (cost=100,γ=0.05)	0.19	0.70	0.43	1.22	0.60	0.42
SVM (cost=100,γ=0.075)	0.20	0.67	0.40	1.12	0.61	0.40
SVM (cost=100,γ=0.1)	0.17	0.66	0.42	1.14	0.60	0.37
SVM (cost=100,γ=0.125)	0.13	0.67	0.46	1.24	0.57	0.35
SVM (cost=100,γ=0.15)	0.16	0.61	0.42	1.06	0.57	0.32

Global Ensemble Simple Mean: BSS = 0.398 Weighted Mean: BSS = 0.399

Notes and Limitations

- At this point, testing on 2017 and 2018 hurricane seasons is still ongoing
- Preliminary improvements to this point for 2017 and 2018 are minimal, possibly due to differences in GFS and FNL (operational forecasts will use GFS analysis fields) and the difficult forecasts those years held
- New AI ensemble will be ready for operations prior to the Atlantic 2019 Hurricane Season

Conclusions

- Results from 2018 Testbed in line with current SHIPS-RII consensus members, though not outperforming
- Integration of FNL data into AI ensemble is showing promise for improving RI predictions, testing still ongoing
- Updated ensemble will be participating in 2019 Joint Hurricane Testbed experiment

Questions?