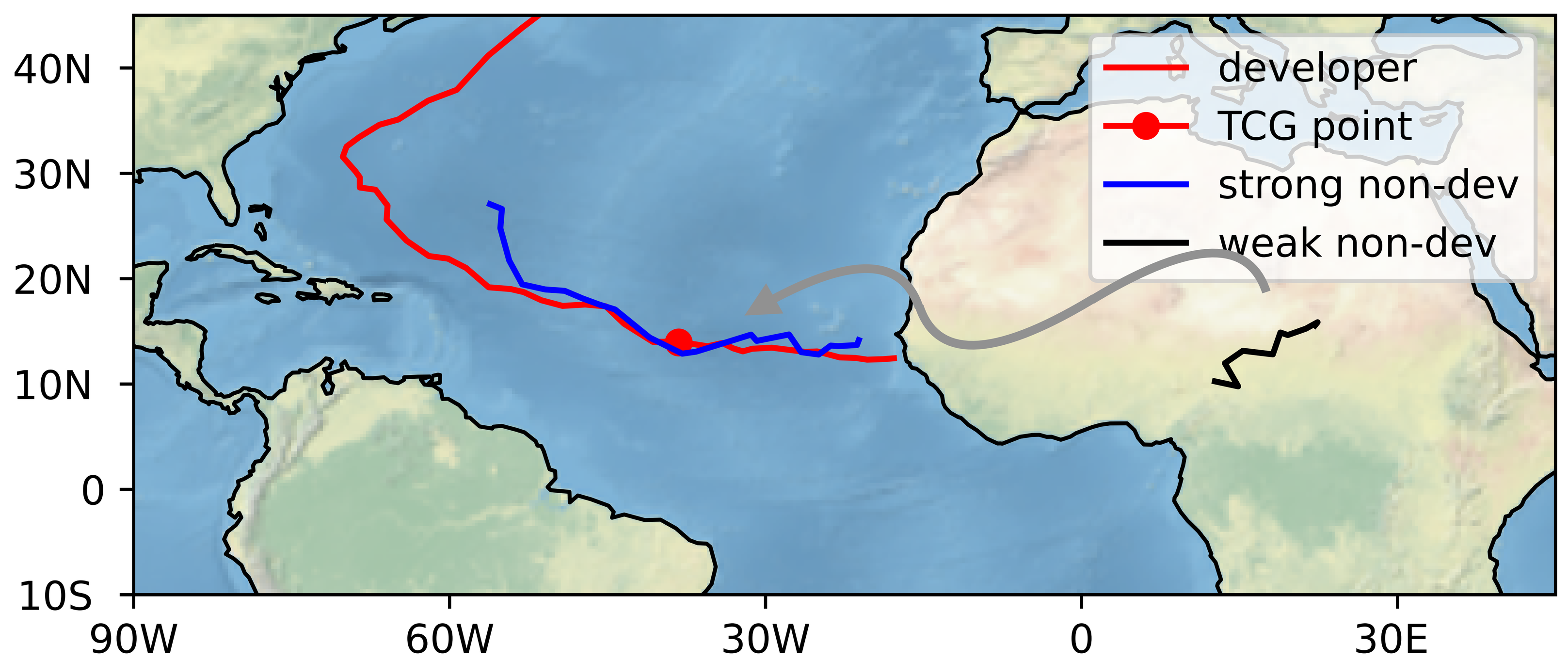
Machine Learning Approaches on Identifying Tropical Waves That Develop into Hurricanes

Haochang Luo[1], Spencer A. Hill[1,2] Mentors from CCAI mentorship program: Manmeet Singh[3], Naveen Sudharsan[3]

[1]Department of Earth and Atmospheric Sciences, City College of New York, New York, NY ; [2]Lamont-Doherty Earth Observatory, Columbia University, New York, NY; [3]Department of Earth and Planetary Sciences, Jackson School of Geosciences, University of Texas, Austin, TX. email: hluo@ccny.cuny.edu

CCNY CUNY

1. Introduction



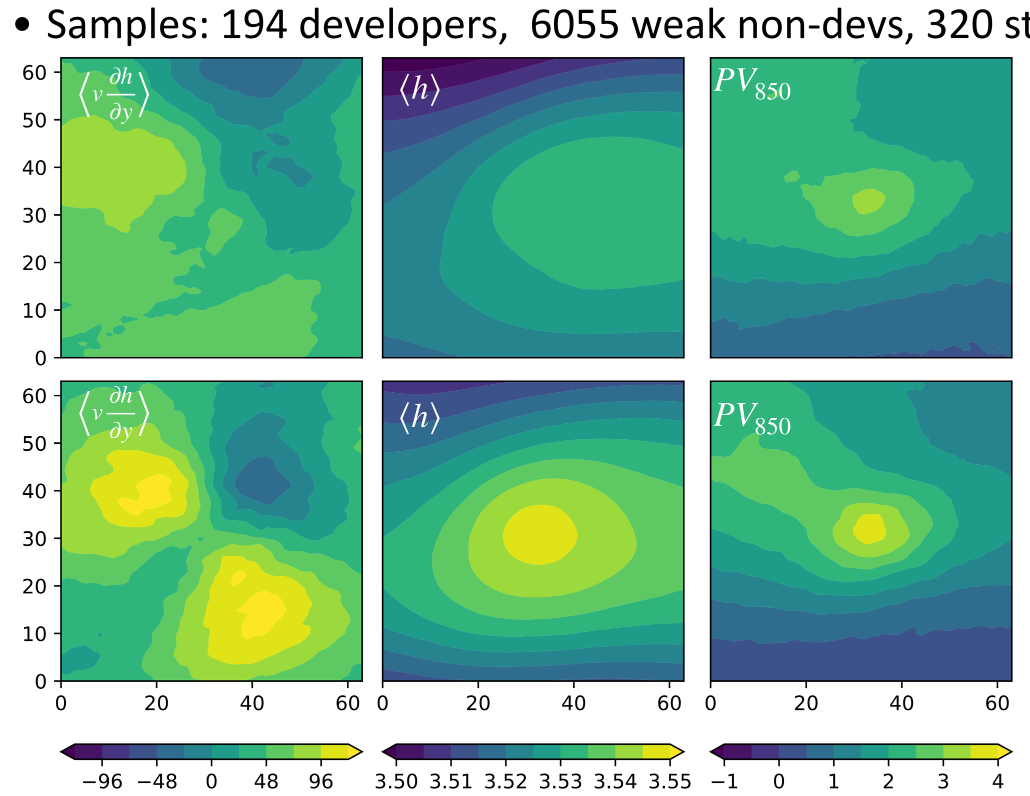
Example tracks of developing AEW (developer, red line); strong non-developing AEW(non-dev, blue) track; weak non-dev track (black); Schematic of African Easterly Jet (AEJ, grey); the point when the AEW turns into a hurricane (red dot).

- The tropical cyclones at the Atlantic basin caused immense socioeconomic losses in the American each year.
- African Easterly Waves (AEWs) are westward-propagating storms that grow from the instabilities in the African Easterly Jet (grey curve). They are weaker than hurricanes.
- ~85% of major hurricanes (Categories 3-5) start out as AEWs and then develop into hurricanes. But only 3% of AEWs develop into hurricanes. The potential vorticity (PV) at 850 hPa of ~95% of non-devs is smaller than $4.1 \times 10^{-5}\ s^{-1}$ throughout their lifetime (weak cases).
- Whether a given AEW will develop into a hurricane or not remains hard to predict.

So can machine learning (ML) approaches be used to differentiate developing and non-developing AEWs?

4. Wave-centered composites and preprocessing

- Wave-centered composites: At each 6-hourly time step, $16^\circ \times 16^\circ$ snapshots of each variable, tracking along the wave center. Temporal composites are the average of snapshots from the beginning of each wave's lifecycle till the end for non-devs and till becoming a hurricane for developers.
- Strong non-devs: lifetime maximum $PV_{850} \geq 4.1 \times 10^{-5}\ s^{-1}$
- Samples: 194 developers, 6055 weak non-devs, 320 strong non-devs



Top: Composites of all strong non-devs throughout their lifetimes.
Bottom: Composites of all developers before they become hurricanes.

80% of the data is used as training set, 10% as validation set, 10% as testing set.
97% of the storms are non-devs, extremely unbalanced. The best f1-score of developers obtained using the full non-devs is 0.7 but not robust in 5-folding cross validation.
So it is crucial to split the non-devs into strong and weak cases and omit the weak ones. Strong non-devs has comparable quantities with developers, the results are more robust.
The training set is balanced using Synthetic Minority Oversampling Technique (SMOTE) and standardized.

2. Theories for AEW development

The prevailing theoretical frameworks for understanding AEW vortex growth (which is required for eventual development into a hurricane) identify at least three key terms:

1. Column-integrated moist static energy (MSE): MSE is conserved when integrated throughout the whole column during moist adiabatic processes. It is closely related to precipitation. Latent energy ($L_v q$) released by water vapor during phase changing is a leading term in MSE and an important factor to the growth of AEWs:

pressure at the top of the atmosphere; column MSE; air heat capacity at constant pressure; specific latent heat through phase changing

$$\langle h \rangle = \frac{1}{g}\int_{p_t}^{p_s} h\, dp = \langle C_p T \rangle + \langle L_v q \rangle + \langle \Phi \rangle \simeq \langle L_v q \rangle$$

temperature; geopotential; surface pressure; specific humidity; gravitational acceleration

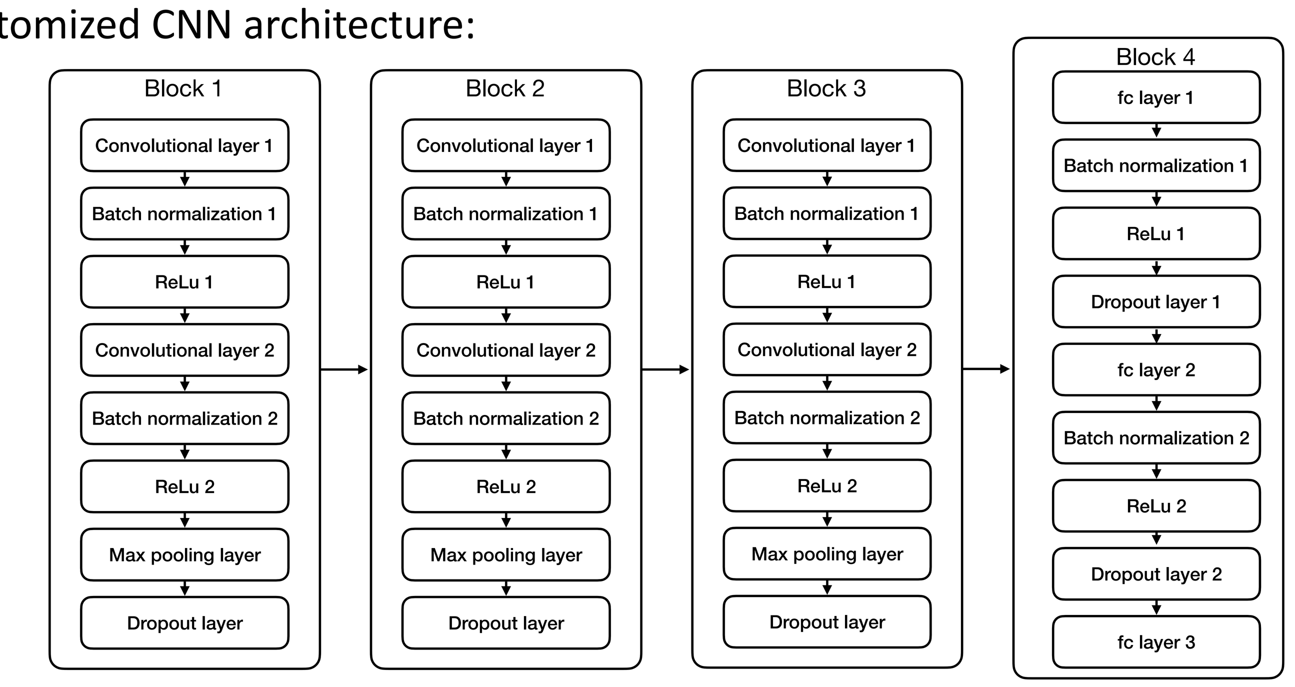
2. Column-integrated meridional advection of MSE: $\left\langle v\frac{\partial h}{\partial y}\right\rangle$

3. Potential vorticity at 850 hPa: PV combines rotation and stratification effects and is conserved following fluid motion in the absence of diabatic heating and friction. The 850 hPa level captures the low-level circulation associated with AEWs while avoiding boundary layer complexities.

Are the variables that are significant in the physical theories also important to the ML models?

5. ML model architectures

- Customized CNN architecture:



- Pre-existing architectures:

Model name	Function
ResNet	resnet18
Wide ResNet	wide_resnet50_2
ResNeXt	resnext50_32x4d
EfficientNet	efficientnet_b0
VGG	vgg16
ViT	vit_b_16

- Loss functions:

name	Function
Binary Cross Entropy (BCE)	BCEWithLogitsLoss
Weighted BCE	pos_weights= Strong non-dev/ developer

- Optimizer:

name	Function
Adaptive Moment Estimation (Adam)	Adam
Adam with weight decay	AdamW
Stochastic Gradient Descent	SGD

- Other hyperparameters:

name	range/value
Random seed	42
Weight decay rate	0.0001
Scheduler	ReduceLROnPlateau
Learning rate	$[10^{-5}, 10^{-2}]$
Batch size	16, 32, 64

Hyperparameters are optimized using Optuna, a Bayesian optimization framework that automatically searches better hyperparameters based on a Tree-structured Parzen Estimator (TPE). The validation f_β-score ($\beta = 0.75$) is used as optimization parameter. 100 trials for each model.

3. Datasets

- Column-integrated MSE, column-integrated meridional MSE advection, and PV at 850 hPa are calculated using the fifth reanalysis product from the European Centre for Medium-Range Weather Forecasts reanalysis data (ERA5)
- ERA5: $0.25^\circ \times 0.25^\circ$ horizontal resolution, 6-hourly temporal resolution, 1979-2018
- Storm tracking dataset: latitudes & longitudes of storm center (developers and non-devs), storm types, hurricane names (developers only), center PV at 850 hPa, $0.25^\circ \times 0.25^\circ$ horizontal resolution, 6-hourly temporal resolution, 1979-2018, June-October. (Cheng et al. 2019)
- Storm center tracking method: Averaging mass-weighted curvature vorticity extrema at 500, 600, 700 hPa and relative vorticity at 700, 850 hPa using ERA-interim reanalysis datasets. Smoothing over 300-km radius. Comparing with National Hurricane Center's Atlantic Hurricane Database version 2 (HURDAT2) to label the storm types.

6. ML model results

Model evaluation on testing sets show that the precision of developers tends to be the lowest statistic, so developer f_β-score is used here as an metric with a inclination to precision ($\beta = 0.75$).

Model	Developer fbeta	Precision	Recall
Customized CNN	0.69	0.64	0.80
ResNet	0.70	0.68	0.72
Wide ResNet	0.77	0.82	0.69
ResNeXt	0.75	0.81	0.67
EfficientNet	0.77	0.92	0.59
VGG	0.75	0.83	0.64
ViT	0.71	0.71	0.64

7. Variable control experiments

Using the customized CNN as a baseline, different combinations of physical variables are tested.

Combination	Developer fbeta	Precision	Recall
MSE_adv+MSE+PV	0.69	0.64	0.80
MSE_adv	0.58	0.54	0.65
MSE	0.66	0.71	0.60
PV	0.65	0.65	0.65
MSE_adv+MSE	0.64	0.67	0.60
MSE_adv+PV	0.64	0.67	0.60
MSE+PV	0.66	0.63	0.75

8. Conclusions and future work

1. The performance of all models are close, with f_β-score (developer) ranging from 0.69 to 0.77. The precision has a higher deviations, ranging from 0.64 to 0.92. And the recall ranges from 0.59 to 0.8.
2. Wide ResNet achieves the highest f_β-score, 0.77, EfficientNet has a comparable f_β-score with ResNet but a higher precision, 0.92. Our customized CNN achieves the highest recall, 0.8. In such an extremely unbalanced classification problem, filtering out weaker ones is important for a robust result.
3. Adding more features improves the results but they should be weighted according to their physical importance. A deep understanding of the physics is needed.

Future work:

- Create more samples using NWP models or using data augmentation methods, e.g. adding noise to existing samples.
- Study the time series instead of using composites.
- Better models and adding more physical variables such as barotropical conversion.