MOODY'S

Assessing Hurricane Models: Testing Enhanced Hurricane Forecast Verification Methods

James Dunphy and Dr. Jeremiah Otero

FSU UNDERGRADUATE RESEARCH OPPORTUNITY PROGRAM
CENTER FOR UNDERGRADUATE RESEARCH & ACADEMIC ENGAGEMENT
FSU | FLORIDA STATE UNIVERSITY

Introduction

- Hurricane forecast models guide emergency response and evacuations, especially along the Gulf and Atlantic coasts
- Wind speed and track errors create uncertainty in public safety and disaster management
- Model verification compares forecasts to observed data to identify weaknesses and biases
- This study analyzes wind speed and track errors in the forecast models HAFS-A, HWRF, AVNO (GFS), and OFCL (NHC)
- Enhanced qualitative and quantitative verification methods are used: signed wind error distribution and storm direction relative track error
- Evaluating these errors helps refine modeling techniques and improve hurricane forecast reliability.

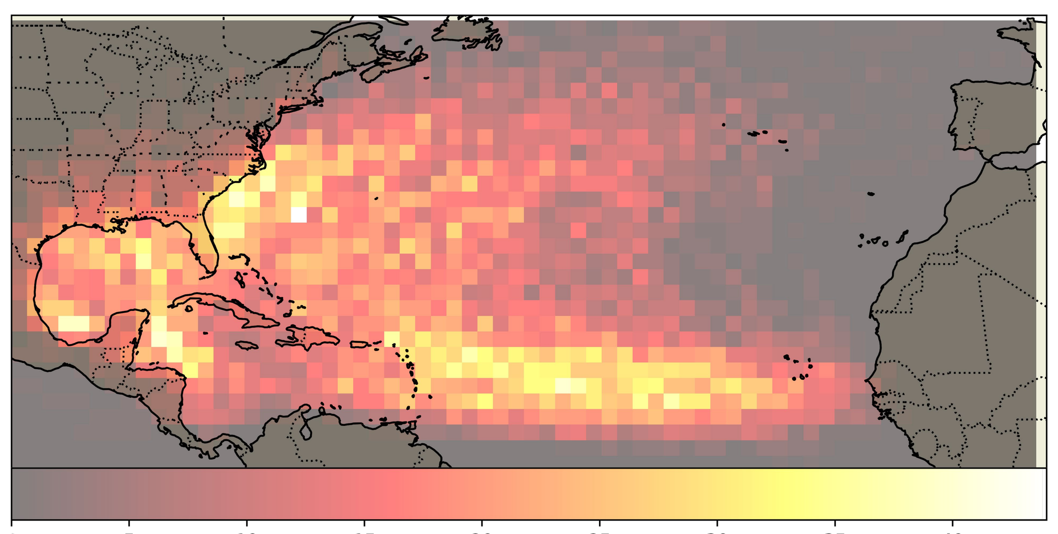

Fig. 1: A density heat map of all tropical cyclones in the Atlantic Basin from 1970-2023. The scale is based on how many distinct tropical cyclones passed through a given bin. The biggest hotspots are the MDR (Main Development Region) in the Atlantic, the Gulf of Mexico (especially near the Yucatán Peninsula), and the United States east coast near the Carolinas.

Methods

- The performance of HAFS-A, HWRF, AVNO, and OFCL models was verified against the NHC Best Track (BEST/B-Deck) dataset.
- Dataset includes all tropical cyclones in the Atlantic Basin from 2017–2024 (HAFS-A model only has data from 2023-2024)
- These models represent a range of hurricane forecasting approaches, including new-gen, legacy, global, and human adjusted models.
- Verified wind speed error and bias (signed and unsigned)
- Verified track error and bias relative to storm direction (signed and unsigned) with great circle distance

$$d_{\text{storm}} = \text{hav}(\phi_{\text{init}}, \lambda_{\text{init}}, \phi_{\text{act}}, \lambda_{\text{act}}) \quad \theta_{\text{storm}} = \text{azm}(\phi_{\text{init}}, \lambda_{\text{init}}, \phi_{\text{act}}, \lambda_{\text{act}})$$

$$d_{\text{error}} = \text{hav}(\phi_{\text{act}}, \lambda_{\text{act}}, \phi_{\text{pred}}, \lambda_{\text{pred}}) \quad \theta_{\text{error}} = \text{azm}(\phi_{\text{act}}, \lambda_{\text{act}}, \phi_{\text{pred}}, \lambda_{\text{pred}})$$

$$\Delta\theta = \theta_{\text{error}} - \theta_{\text{storm}} \qquad \begin{aligned} e_\perp &= d_{\text{error}} \sin(\Delta\theta) \\ e_\parallel &= d_{\text{error}} \cos(\Delta\theta) \end{aligned} \qquad \mathbf{E} = \begin{bmatrix} e_\perp \\ e_\parallel \end{bmatrix}$$

Fig. 2: Formulas used to get storm direction relative error vector. "ϕ" is latitude, "λ" is longitude, "$\text{hav}()$" is the haversine formula, "$\text{azm}()$" is the azimuth formula, "d" is the distance, "$e/\mathbf{E}$" is the error vector, "init" is the initial storm position, "act" is the actual position of the storm for a given forecast interval, "pred" is the predicted position of the storm for a given forecast interval.

- Forecast errors were evaluated at 6-hour intervals, with primary analysis from 12h, 24h, 48h, 72h, and 120h forecasts

Results: Wind Speed Error

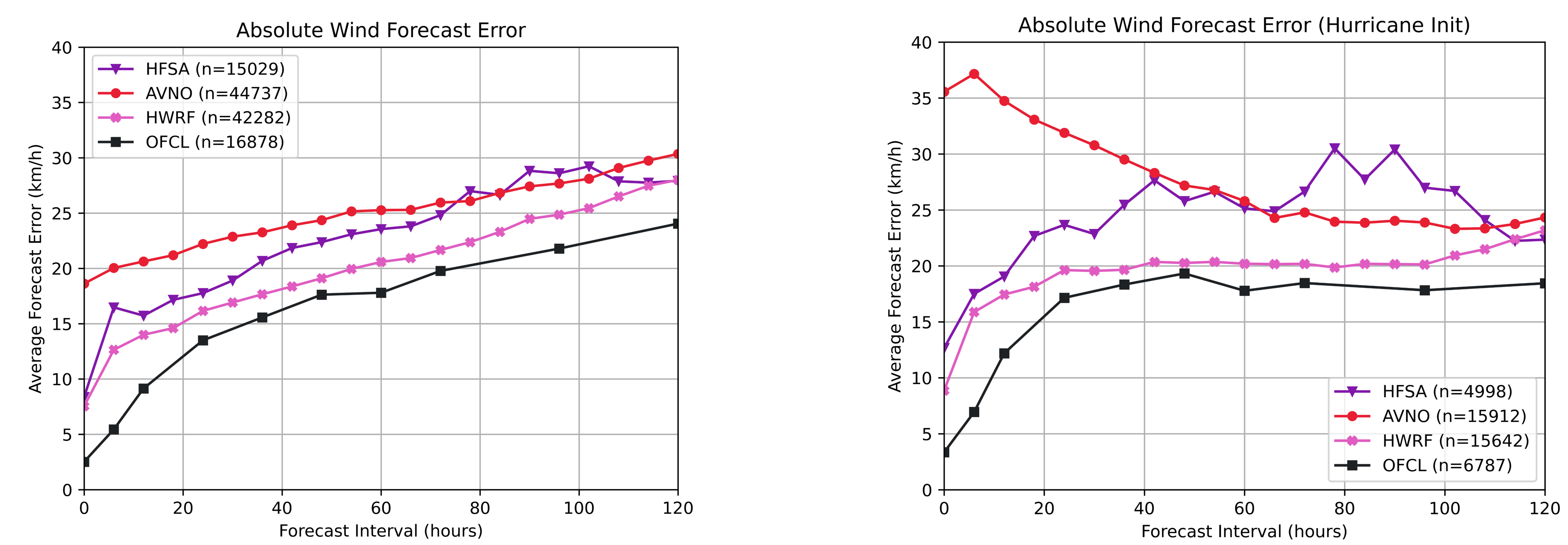

Fig. 3/4: Time series graphs of mean unsinged wind speed forecast error throughout forecast intervals for various models from 2017-2024. The left graph includes all storms, while the right graph is only storms that are at hurricane intensity (according to the Saffir-Simpson scale) at the time of the forecast initialization. Vertical scale is in km/h (1 km/h ≈ 0.62 mph ≈ 0.54 kt ≈ 0.28 m/s) and horizontal scale is in hr. Scales are constant between graphs.

- Wind error increases steeply within the first 24 hours of forecasts intervals, especially HWRF and HAFS-A, then the increase in error tapers off, especially for storms at hurricane strength
- After 120 hours, the error of all models converges, except for OFCL, which maintains lower error
- AVNO has much more severe error at initialization, which leads to much higher error across the board, and even more increased error for storms at hurricane strength
- OFCL was overall most accurate, followed by HWRF, then HAFS-A, then AVNO

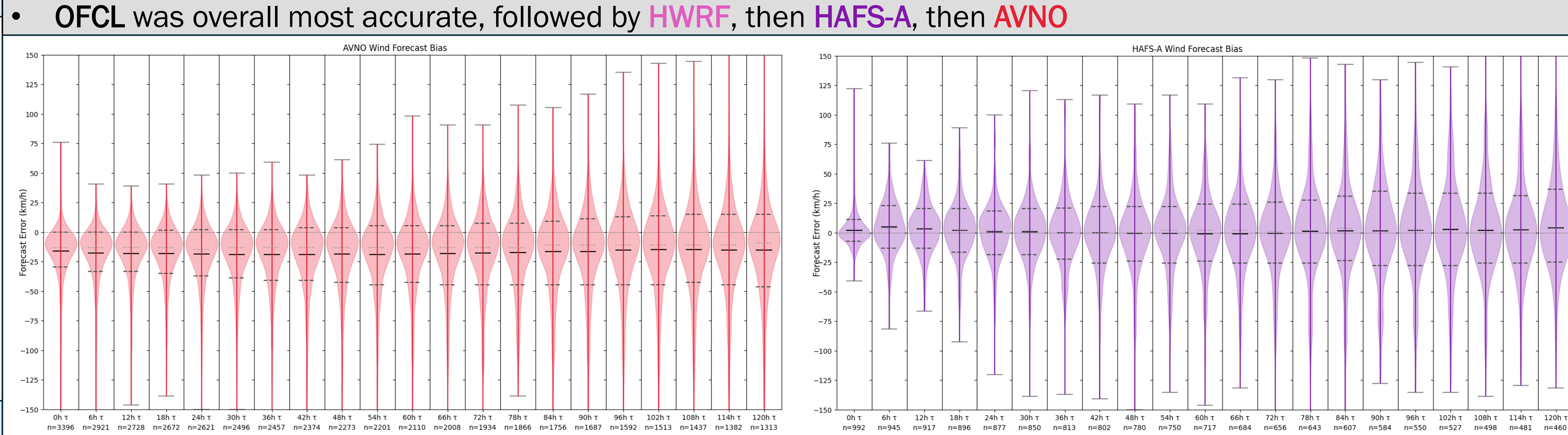

Fig. 5/6: Stacked violin plots for individual forecast intervals plotting wind speed bias, showing distribution of errors. Vertical scale is in km/h (1 km/h ≈ 0.62 mph ≈ 0.54 kt ≈ 0.28 m/s) and horizontal scale is forecast hour intervals. Scales are constant between graphs. The solid line is the mean, the dark dashed line is 1 StDev, the light dashed line is the median. The left graph is AVNO, and the right graph is HAFS-A.

- AVNO consistently underpredicts intensity, with the distribution of wind errors being skewed negatively
- HAFS-A overpredicts very early on, causing a greater magnitude of errors for later forecast intervals
- When corrected for bias, AVNO's wind speed error is comparable to HAFS-A throughout forecast intervals

Results: Track Error

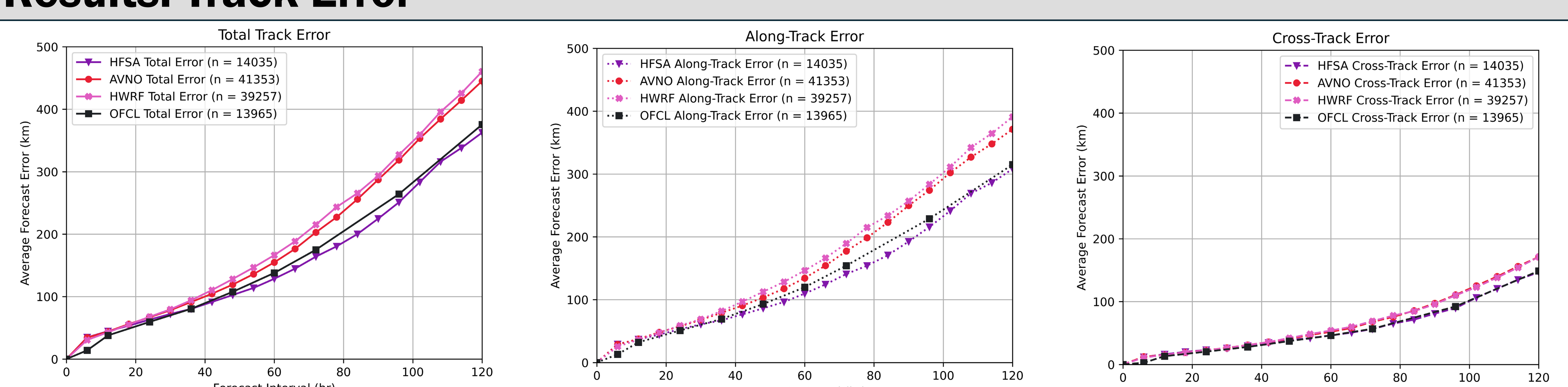

Fig. 7/8/9: Time series plots for different unsigned components of track error. Total error (left graph) is the magnitude of $\mathbf{E}$ (in Fig. 2) and measures general displacement, along-track error (center graph) is the value of $e_\parallel$ and measures over/undershooting distance, cross-track error (right graph) is the value of e and measures the accuracy of the steering. Vertical scale is in km (1 km ≈ 0.62 mi ≈ 0.54 n mi) and horizontal is in hr. Scales constant between all graphs.

- HAFS-A overall performs the best in track error, especially along-track error, slightly outperforming OFCL and distinctly outperforming AVNO and HWRF in all components of track error
- Unlike wind error, track error in all components grew at an exponential rate with increasing forecast intervals
- All models were much more accurate at cross-track forecasts than along-track forecasts, with all models being fairly close together in cross-track error
- HAFS-A tended to be less accurate than other models in the first 24 hours, but more accurate afterwards
- Despite accuracies in wind error, HWRF was the least accurate in track error, especially along-track error

Results: Track Error (cont.)

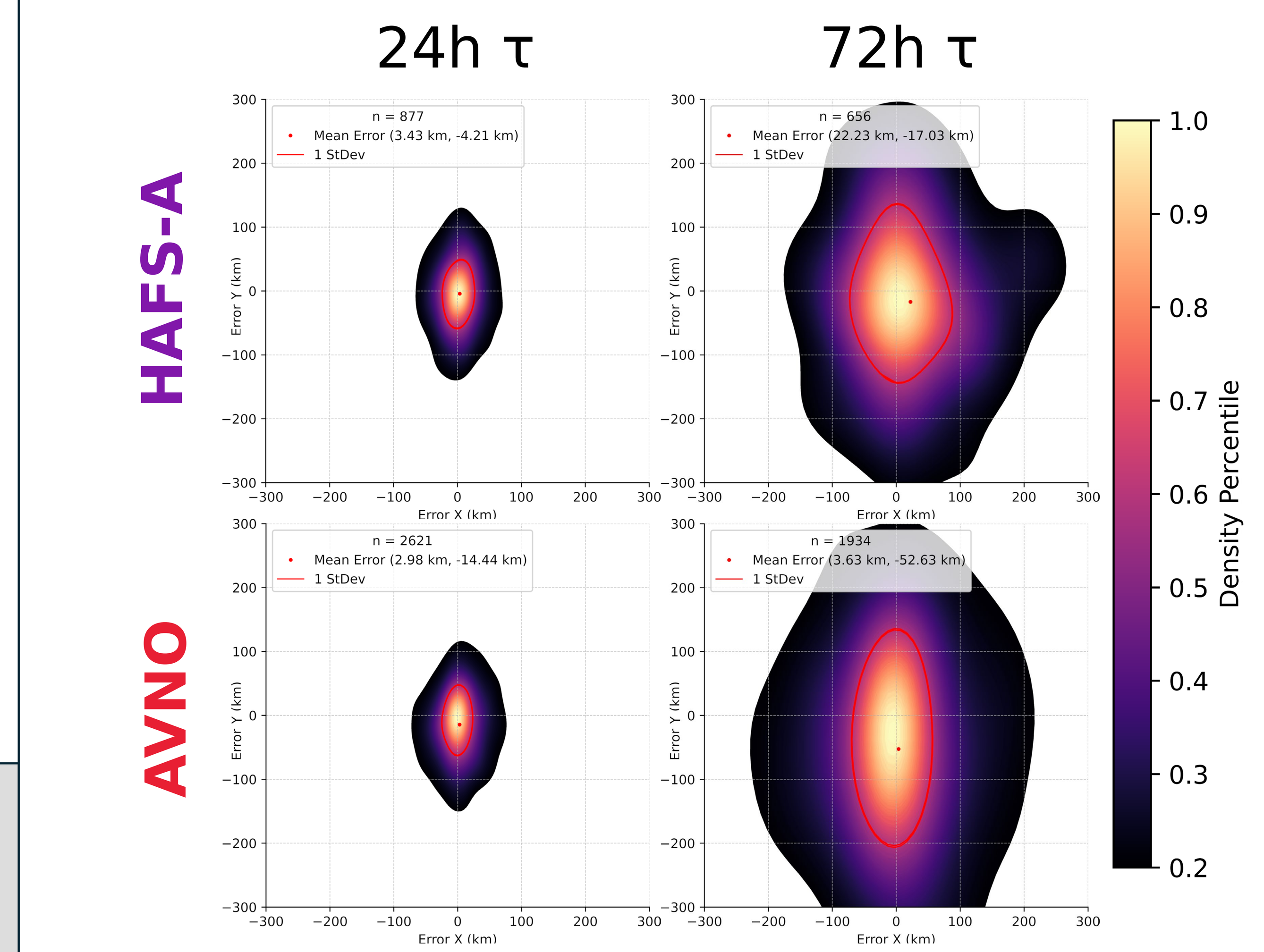

Fig. 10/11/12/13 (top left to bottom right): Series of 2-D KDE plots representing signed track error vectors "$\mathbf{E}$" (in Fig. 2). Top row is HAFS-A, bottom row is AVNO, left column is the 24-hour forecast interval, right column is the 72-hour forecast interval. The red dot is the mean of all the vectors, the red circle is ~1 StDev (68% of data) of the vectors. The vertical and horizontal scales are in km (1 km ≈ 0.62 mi ≈ 0.54 n mi). Scales constant between graphs.

- AVNO has distinctly more along track error than HAFS-A
- AVNO has a substantial "slow" or negative along-track bias, but as minimal "left/right" or cross-track bias, as seen in Fig. 13
- HAFS-A has a "right" or positive cross-track bias and a slight "slow" or negative along track bias, as seen in Fig. 11
- The distribution of HAFS-A is more compact and has less outliers in comparison to AVNO

Conclusions

- OFCL is the overall most accurate model, with the lowest wind speed errors and close to lowest track errors
- Both HAFS-A and HWRF seem to make trade-offs in their forecasts, with each of them being very accurate in one aspect and very inaccurate in the other
- The lower resolution of AVNO in comparison to specialized hurricane models had a big impact on wind speed error and bias, but not as much on track error
- Future research assessing model performance in more specific initial atmospheric conditions, such as wind shear or sea surface temperature, could lead to more insights into model error and bias

Acknowledgements

I would like to first and foremost express my gratitude for my research mentor, Dr. Jeremiah Otero, for his invaluable guidance, feedback, and support throughout the whole project. I would also like to thank Moody's HWind and all its employees for their resources, time, and assessments, which have been of tremendous support for my research. Finally, I would like to thank UROP and my UROP leaders for their advising and feedback and the opportunity to conduct this research.

References