Image2image Tropical Cyclone wind field diagnosis with Pix2Pix generative adversarial networks (GANs)

S. J. Ollier[1], **E. Siman**[1], **D. S. Nolan**[1], **B. D. McNoldy**[2]
[1]Worldsphere.ai, Coral Gables, FL 33124
[2]University of Miami, Coral Gables, FL 33124
sjo@worldsphere.ai, es@worldsphere.ai

Abstract

Recent AI advances enable tropical cyclone (TC) diagnosis from satellite imagery, extracting wind intensity and storm size for operational analyses and forecasts. While complete and sufficiently detailed horizontal wind fields are rarely observed directly, this study presents pix2pix generative adversarial networks (GANs) that predict surface wind fields from real and simulated infrared brightness temperatures using thousands of realistic TC image pairs. Applied to modified real infrared images, the algorithm produces realistic wind fields comparable to observations (6.7 knots root mean square error). Insights from this project will inform future development for live satellite imagery diagnosis to enable higher spatio-temporal resolution TC wind information earlier and faster for meteorologists and emergency managers.

1 Introduction

Most tropical cyclone (TC) maximum wind (intensity) estimates are derived from subjective [1] or objective [2] analyses of Infrared (IR) brightness temperature satellite images. Wind field size and asymmetry are also estimated through subjective and objective methods, such as the operational Multi-platform Tropical Cyclone Surface Wind Analysis (MTCSWA) [3]. Current operational methods measure storm size from IR and calculate radius of maximum wind speed (RMW) and wind radii at 34-, 50-, and 64-knot in quadrants around the storm center using a parametric model [4, 5]. These are calculated using IR observations at 6-hourly intervals and forecast up to 120 hours [4, 5].

Building upon current operational methods [4, 5] and recent advances in AI/ML approaches for tropical cyclone intensity diagnosis [6], size [7] and inner core wind field estimation [8], this study presents a novel pix2pix [9] generative adversarial network (GAN) algorithm designed to simultaneously estimate both the intensity and complete structural characteristics of tropical cyclone surface wind fields directly from infrared satellite imagery. Our approach leverages a comprehensive training dataset comprising thousands of images from two complementary sources:

1. High resolution simulations of TCs produced using the Weather Research and Forecasting (WRF) model.
2. IR observations of real TCs paired with corresponding parametric wind fields based on analyses from operational forecasting centers.

Using a pix2pix GAN [9] framework, we demonstrate the ability to transform single-channel IR satellite observations into detailed, spatially resolved surface wind field reconstructions, offering significant potential to improve both operational forecasting capabilities and our fundamental understanding of tropical cyclone structure and evolution. This methodology represents a substantial advancement over traditional diagnostic approaches by providing complete wind field characterization

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2025.

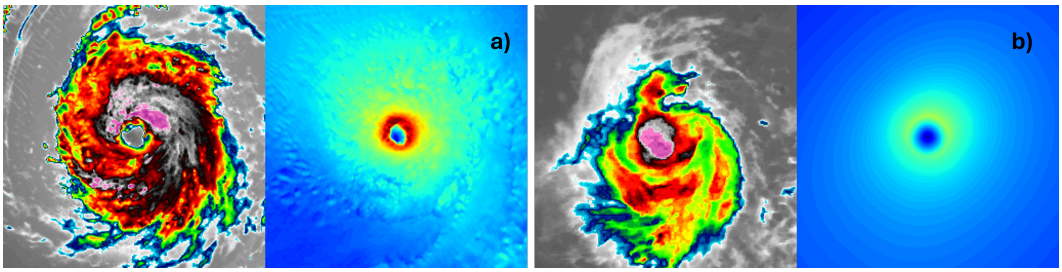

Figure 1: Example input images a) WRF IR and wind field, b) MergeIR and parametric wind field

rather than point estimates, with the potential to enhance both intensity forecasting accuracy and storm structure analysis for improved hazard assessment and disaster preparedness.

2 Methodology

2.1 Datasets

2.1.1 WRF infrared satellite and wind fields

We use synthetic satellite images generated from a collection of fourteen realistic WRF simulations of hurricanes. All simulations were produced with WRF versions 3.1 through 4.3.3, using the same grid spacings, vertical levels, and physical parameterizations of clouds and turbulence [10,11]. Simulated satellite images and surface wind fields were taken every 60 minutes over many days of simulations to produce a set of 1844 corresponding images.

2.1.2 Parametric TC wind fields with concurrent infrared satellite images

Parametric TC wind fields were generated using a parametric model developed to approximate TC surface wind fields. The size, strength, and asymmetries of the wind field are derived from the IBTraCS (International Best Track Archive for Climate Stewardship) dataset. The storm parameters from 6-hourly synoptic data were used to create a smooth wind field on a polar coordinate grid (radial spacing 4 km, angular spacing 2°) with translation asymmetry following [12] and size asymmetry controlled by tropical storm wind radii in four quadrants. Wind fields at each 6-hourly time step were then interpolated onto a Cartesian coordinate grid (2 km resolution). IR brightness temperature data [13] for concurrent time stamps to the parametric wind fields were obtained from NASA Earthdata [14], hereafter referred to as *MergeIR*. These were combined to produce 1984 paired images.

2.2 Data pre-processing

All images were plot as $6^\circ \times 6^\circ$ images with 256×256 pixels centered around the storm eye before satellite and wind field images were combined into 512×256 pixels images. To emphasize key storm wind speeds, WRF and Parametric wind field images were plot with jet colormap, color range between 0-100 knots with 100 contour levels. WRF and MergeIR IR images were plot using a custom colormap, color range between -80, 20 tb with 100 contour levels. Example input data for the models are shown in Figure 1.

Both datasets were categorized by Saffir-Simpson Hurricane wind scale per storm and per image before train, test and validation set creation. To prevent data leakage between sets, the images were split by storm, aiming for representative spread of storm categories and individual image maximum wind speeds (V_{max}) across all 3 sets.

2.3 Models and evaluation

Both models, hereafter named *wrf2wrf* and *mir2para*, utilize Pix2pix GAN [9] architecture (Figure 2, Appendix A.1). Both models ran on NVIDIA L4 GPUs for computational and energy efficiency.

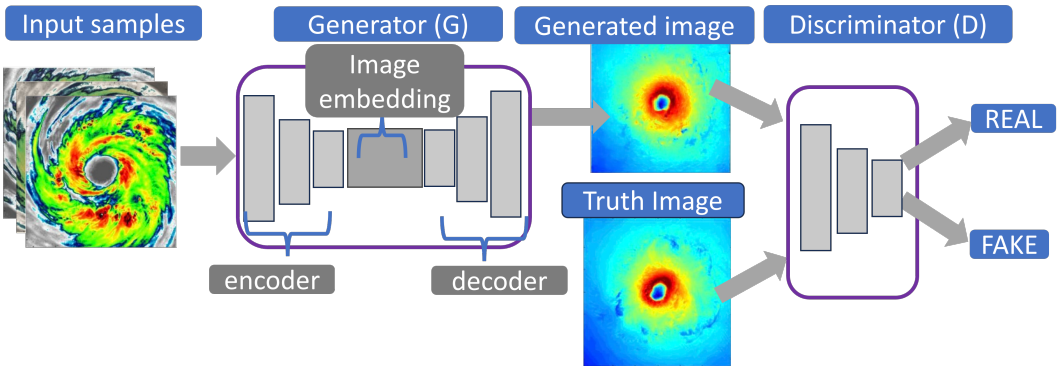

Figure 2: pix2pix GAN [9] for wrf2wrf. The generator (G) aims to create convincing copies of the truth image and the discriminator (D) tries to determine if these are real or fake, each training iteration this improves the skill of both the G and the D.

Model performance was quantitatively assessed using pixel-wise wind speed Root Mean Square Error (RMSE). This was computed by first converting both model-generated outputs and ground truth targets from their colormap representation back to original wind speed values, then calculating the RMSE on a per-pixel basis across the entire image domain. For comparison with operational IBTrACS data [15,16], model results were post-processed to calculate key operational parameters: radius of maximum wind (RMW) and radius of 34, 50 and 64 knot winds (R34, R50, R64).

3 Results and Discussion

Example test-set validation results for wrf2wrf and mir2para are shown in Figure 3. During validation, both models produced realistic storm wind fields with mean pixel RMSE of 7.3 knots for 298 wrf2wrf images and 6.7 knots for 276 mir2para images. Minimum RMSE was 3.6 knots and 1.2 knots for wrf2wrf and mir2para, respectively. Performance for both models varies based on storm V_{max} (maximum sustained wind speed) with higher V_{max} storms producing lower RMSE. These higher V_{max} storms are more organized, providing structured patterns for the models to learn and infer.

With 73 unseen MergeIR test images both models produced realistic wind fields with mean RMSE between MergeIR truth and model prediction of 10.4 knots for wrf2wrf and 7.5 knots for mir2para (Appendix A.1). mir2para was trained on images similar to this test set, so it is unsurprising that this model performs better. However, wrf2wrf which was trained on WRF images, translates well to this new data. This provides encouraging evidence that both models will perform well on real-time operational data.

When mir2para 64-knot quadrant radii (R64) were compared with IBTrACS operational analyses [15,16], Mean Absolute Error (MAE) between the prediction and operational data was as low as 8.9 NM for NW quadrants (8.9 - 12.0 NM for all quadrants) with a -3.3 NM bias (Figure 4). For RMW, the MAE between predictions and operational data was 12.8 NM with a bias of 0.9 NM and RMSE of 18.1 NM ($\sim$34 km), less than the 53 km RMSE reported between predictions and operational flight measurements in [8]. See Appendix A.2, A.3 for more detailed quadrant radii error information.

Our models construct the whole wind field rather than point estimates and fixed radii in current operational methods [3, 4, 5], offering improvements in spatial resolution. Additionally, our $6^\circ \times 6^\circ$ wind fields captures the storm inner-core and outer rainbands, compared to the $\sim 3.6^\circ \times 3.6^\circ$ inner-core focus, with a recommended usable radius of 1.25° due to edge artifacts in [8]. Further, as brightness temperatures images are available as frequently as every 10-15 minutes, our models could increase the frequency of operational wind field data. Similarly, we provide increased temporal resolution than AI-generated wind fields in [8] which rely on microwave satellite image inputs to supplement IR brightness temperature data, reducing temporal resolution to a few times daily.

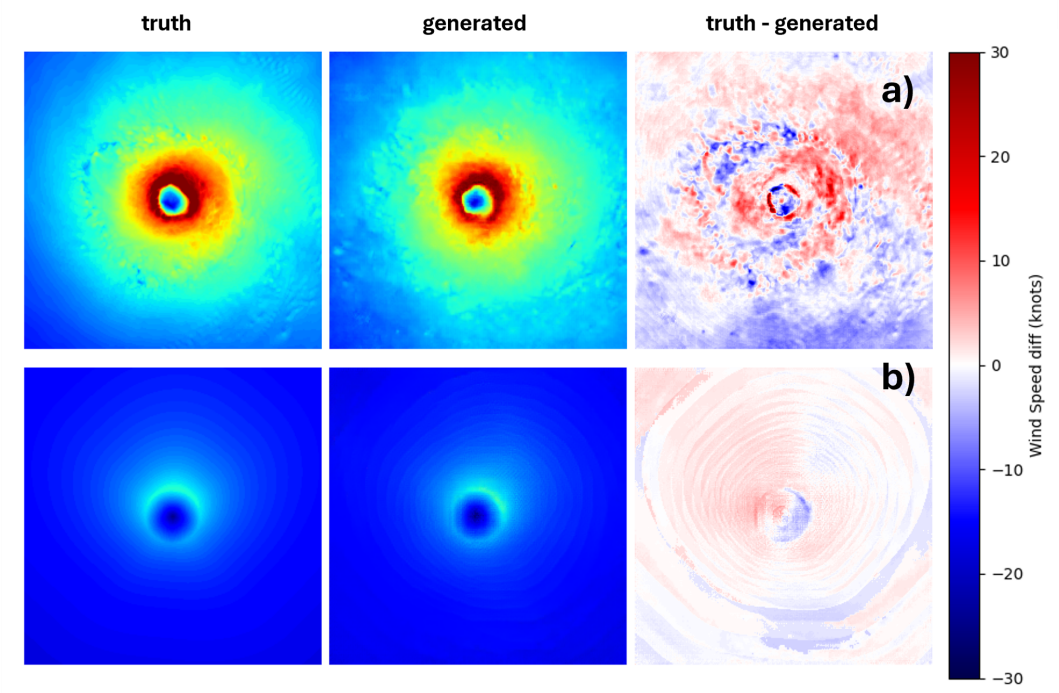

Figure 3: Test-set validation images for a) wrf2wrf Hurricane Florence 2018. b) mir2para Hurricane Larry 2021. From left to right truth image, generated image, difference truth - generated. Legend shows wind speed difference in knots.

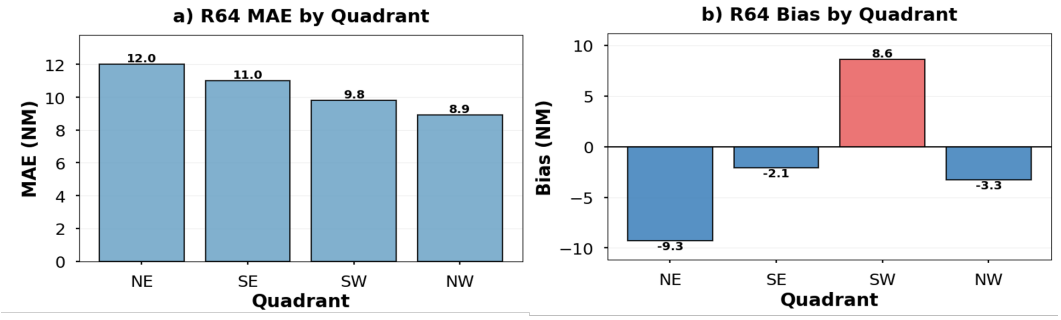

Figure 4: mir2para test set inference result comparison with operational IBTrACS data [15,16] a) Mean Absolute Error (MAE) per quadrant in nautical miles (NM) b) Bias per quadrant (NM).

4 Conclusions and future work

Our pix2pix GAN models, trained with low computational cost, generate realistic TC wind fields from real and synthetic satellite imagery. Our model trained on real satellite imagery compares well to operational best track data. With higher spatial and temporal resolution than existing operational and AI methods, these findings have high potential for development towards highly efficient operational models. Future work will focus on:

1. Tuning models to storm category to better resolve poorer performing categories. Images would be categorized by a classification model before category specific model inference.
2. Physics conditioned learning and inference based on storm characteristics that influence wind field structure.

References

[1] Velden, C., and Coauthors, 2006: The Dvorak tropical cyclone intensity estimation technique: A satellite-based method that has endured for over 30 years. *Bull. Amer. Meteor. Soc.*, **87**, 1195–1210, https://doi.org/10.1175/BAMS-87-9-1195.

[2] Olander, T. L., and C. S. Velden, 2019: The advanced Dvorak technique (ADT) for estimating tropical cyclone intensity: Update and new capabilities. *Wea. Forecasting*, **34**, 905–922, https://doi.org/10.1175/WAF-D-19-0007.1.

[3] Knaff, J. A., and Coauthors, 2011: An automated, objective, multiple-satellite platform tropical cyclone surface wind analysis. *J. Appl. Meteor. Climatol.*, **50**, 2149–2166, https://doi.org/10.1175/JAMC-D-11-09.1.

[4] Knaff, J. A., C. R. Sampson, and G. Chirokova, 2017: A global statistical–dynamical tropical cyclone wind radii forecast scheme. Wea. Forecasting, 32, 629–644, https://doi.org/10.1175/WAF-D-16-0168.1.

[5] DeMaria, M., J. L. Franklin, R. Zelinsky, D. A. Zelinsky, M. J. Onderlinde, J. A. Knaff, S. N. Stevenson, J. Kaplan, K. D. Musgrave, G. Chirokova, and C. R. Sampson, 2022: The National Hurricane Center Tropical Cyclone Model Guidance Suite. Wea. Forecasting, 37, 2141–2159, https://doi.org/10.1175/WAF-D-22-0039.1.

[6] Chen, B. F., B. Chen, H.-T. Lin, and R. L. Elsberry, 2019: Estimating tropical cyclone intensity by satellite imagery utilizing convolutional neural networks. *Wea. Forecasting*, **34**, 447–465, https://doi.org/10.1175/WAF-D-18-0136.1.

[7] Wang, C., and X. Li, 2023: A deep learning model for estimating tropical cyclone wind radius from geostationary satellite imagery. *Mon. Wea. Rev.*, **151**, 403–417, https://doi.org/10.1175/MWR-D-22-0106.1.

[8] Wimmers, A. J., S. Griffin, and C. Velden, 2024: A U-Net retrieval of tropical cyclone inner-core wind fields from microwave and infrared satellite imagery. Artif. Intell. Earth Syst., 3, e230084, https://doi.org/10.1175/AIES-D-23-0084.1.

[9] Isola, P., J.-Y. Zhu, T. Zhou, and A. A. Efros, 2017: Image-to-image translation with conditional adversarial networks. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 1125–1134, https://doi.org/10.1109/CVPR.2017.632.

[10] Nolan, D. S., R. Atlas, K. T. Bhatia, and L. R. Bucci, 2013: Development and validation of a hurricane nature run using the joint OSSE nature run and the WRF model. *J. Adv. Model. Earth Syst.*, **5**, 382–405, https://doi.org/10.1002/jame.20031.

[11] Klotz, B. W., and D. S. Nolan, 2019: SFMR surface wind undersampling over the tropical cyclone lifecycle. *Mon. Wea. Rev.*, **147**, 247–268, https://doi.org/10.1175/MWR-D-18-0296.1.

[12] Knaff, J. A., C. R. Sampson, M. DeMaria, T. P. Marchok, J. M. Gross, and C. J. McAdie, 2007: Statistical tropical cyclone wind radii prediction using climatology and persistence. *Wea. Forecasting*, **22**, 781–791, https://doi.org/10.1175/WAF1026.1.

[13] Janowiak, J., B. Joyce, and P. Xie, 2017: NCEP/CPC L3 half hourly 4km global (60S–60N) merged IR V1. Goddard Earth Sciences Data and Information Services Center (GES DISC), accessed 1 January 2025, https://doi.org/10.5067/P4HZB9N27EKU.

[14] NASA Earthdata, accessed 2024: MERGIR dataset search. [Available online at https://search.earthdata.nasa.gov/search?q=MERGIR.]

[15] Knapp, K. R., M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neumann, 2010: The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying tropical cyclone best track data. Bull. Amer. Meteor. Soc., 91, 363–376, https://doi.org/10.1175/2009BAMS2755.1.

[16] Gahtan, J., K. R. Knapp, C. J. Schreck, H. J. Diamond, J. P. Kossin, and M. C. Kruk, 2024: International Best Track Archive for Climate Stewardship (IBTrACS) Project, Version 4r01. NOAA National Centers for Environmental Information, https://doi.org/10.25921/82ty-9e16.

A Appendix

A.1 Model Architecture and Hyperparameters

Details of the model hyperparameters are shown in Table A.1.

Table A.1: Pix2pix GAN parameters for wrf2wrf and mir2para Models

parameters	Wrf2wrf model	mIR2para model
Batch_size	1	1
Train_dataset_size	1249	1285
Train dataset storms	8	35
Validation_dataset_size	437	283
Validation dataset storms	4	7
Test_dataset_size	298	276
Test dataset storms	2	7
Buffer_size	Train_dataset_size	Train_dataset_size
Lr_disc	2e-4	2e-4
Lr_gen	2e-4	2e-4
Optimizer	Adam	Adam
L1 λ	100	100
Total epochs	170	150

A.2 Model comparison with MergeIR data

Figure A.1 shows a comparison of both models when tested with 73 unseen MergeIR images. For both models RMSE is calculated between the model prediction and the MergeIR truth image.

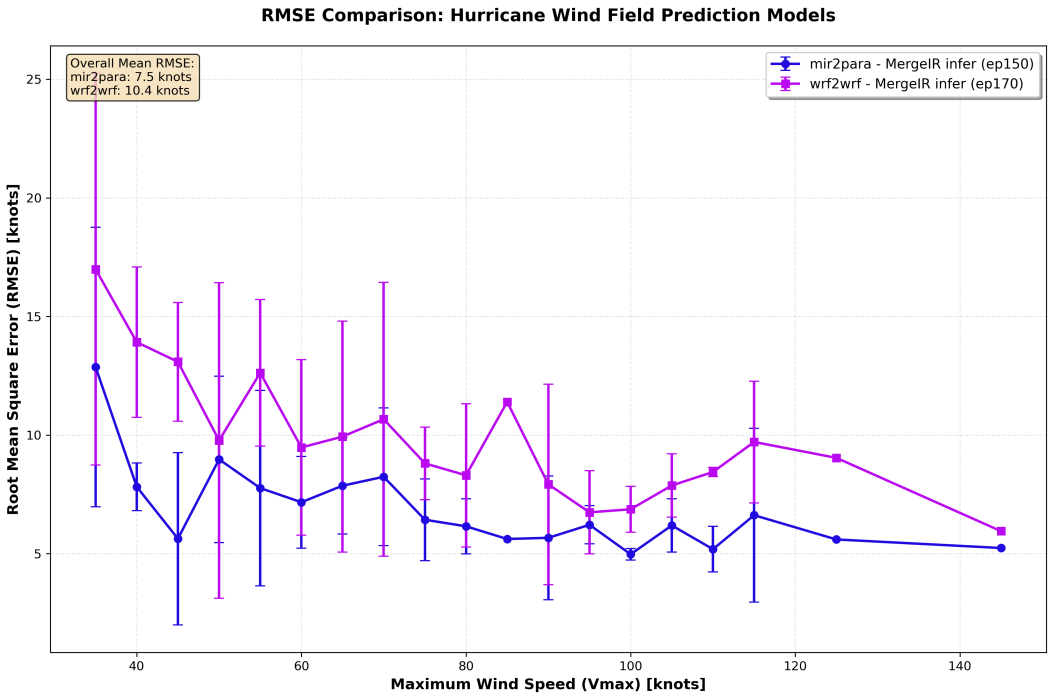

Figure A.1: wrf2wrf (pink) and mir2para (purple) inference pixel-wise wind speed RMSE by V_{max} with MergeIR test dataset

A.3 Model comparison with operational IBTrACS data

Figure A.3 shows MAE and Bias between mir2para test set quadrant radii and IBTrACS operational analyses.

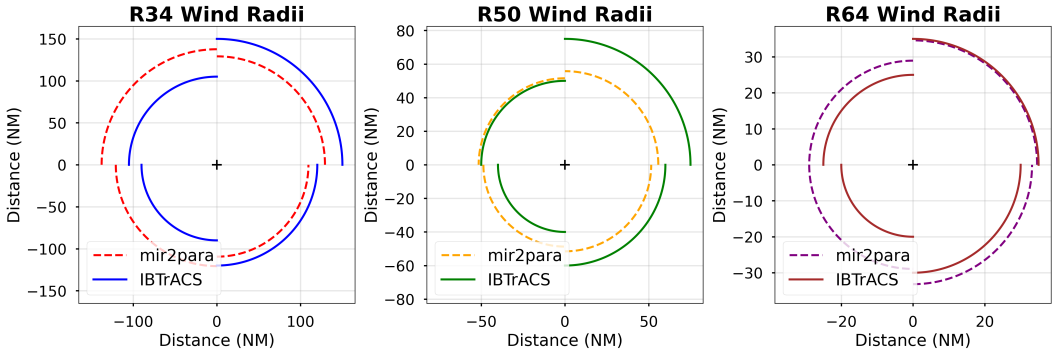

Figure A.2: mir2para test set generated image radii and IBTrACS operational radii comparison for Hurricane Danielle 2010. From left to right truth R34, R50, R64. Dashed lines show mir2para, IBTrACS in solid line.

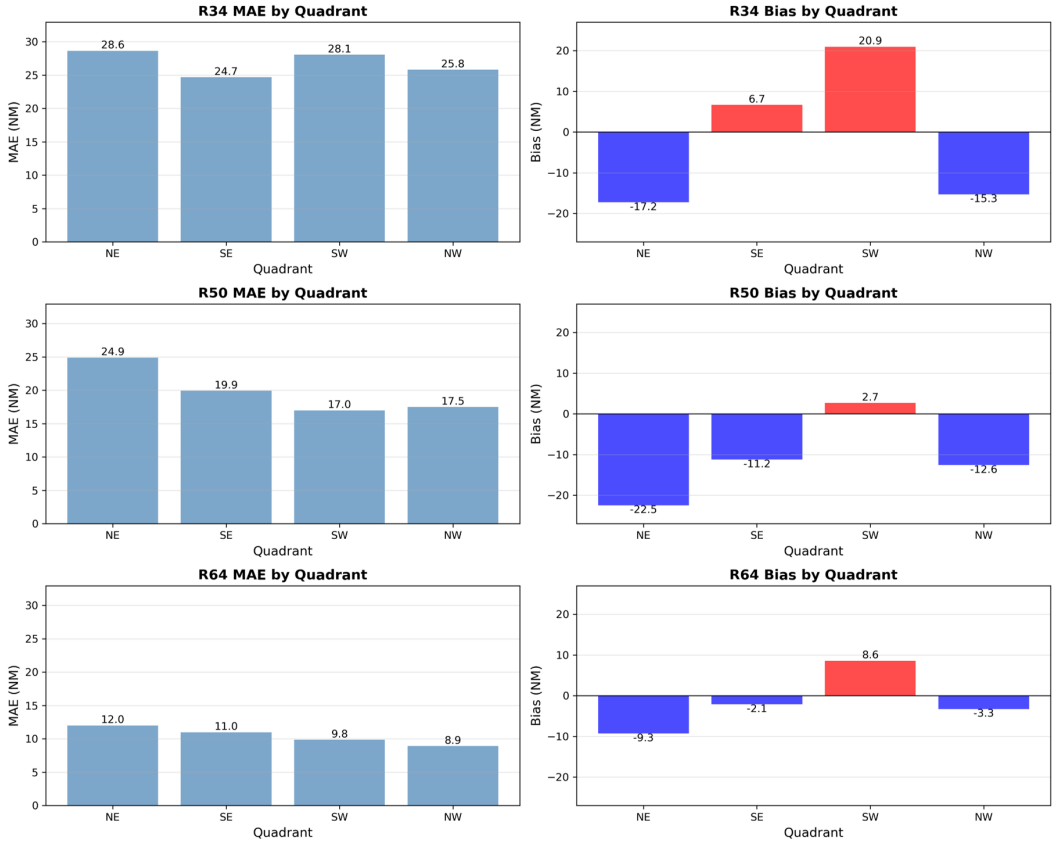

Figure A.3: mir2para test set quadrant radii comparison with operational IBTrACS radii based on 276 images. Left column shows Mean Absolute Error (MAE) per quadrant in nautical miles (NM) for R34, R50 and R64. Right column shows Bias per quadrant (NM) for R34, R50 and R64.