

Article

TC-RadiiFormer: Current-Time Quadrant Wind-Radius Estimation from Multispectral Satellite Sequences

Yiming Li ¹ , Xiang Wang ¹, Hai Hu ^{2,*}, Junxing Zhu ¹ , Wenhao Qian ³ and Liang Guo ⁴

¹ College of Meteorology and Oceanography, National University of Defense Technology, Changsha 410073, China; liyiming23@nudt.edu.cn (Y.L.); xiangwangcn@nudt.edu.cn (X.W.); zhujunxing@nudt.edu.cn (J.Z.)

² Municipal Geomatics Center of Guangzhou, Guangzhou 510180, China

³ Hunan Institute of Advanced Technology, Changsha 410073, China; qianwenhao@hunanxianjin.cn

⁴ Guangzhou Urban Planning & Design Survey Research Institute, Guangzhou 510060, China; solenduo@126.com

* Correspondence: huhai@gz.gov.cn

Highlights

What are the main findings?

- TC-RadiiFormer reduced MAE averaged over R34, R50 and R64 by 27.1%, 25.4% and 13.0% relative to ResNet-50, Swin V2-T and Swin V2-T with a temporal Transformer.
- Multispectral and historical observations improved accuracy; nested decoding maintained ordered quadrant wind radii.

What are the implications of the main findings?

- Geostationary sequences provide complementary information on the extent and asymmetry of tropical-cyclone winds.
- The joint probabilistic output supports assessment of wind-footprint extent and uncertainty in each geographic quadrant.

Abstract

The extent of tropical-cyclone winds is essential for assessing marine and coastal hazards, but surface observations remain sparse over the open ocean. Geostationary satellites continuously observe cloud systems, offering complementary spectral and temporal information for estimating wind radii. We present TC-RadiiFormer to jointly estimate the current radii of 34, 50 and 64 kt winds (R34, R50 and R64) in four geographic quadrants from multispectral satellite sequences. The model combines interactions among spectral channels with historical observations at their actual acquisition times, then decodes geographically aligned, ordered radius distributions. Training used 2017–2022 observations, model selection used 2023, and evaluation used 2024 final best-track wind radii across three basins. Three-seed positive-radius mean absolute errors (MAEs) were 13.164, 8.384 and 5.731 nautical miles for R34, R50 and R64. MAE averaged over the three thresholds was 27.1%, 25.4% and 13.0% lower than ResNet-50, Swin V2-T and Swin V2-T with a temporal Transformer, respectively. Matched comparisons supported contributions from multispectral observations, historical information and actual acquisition timing. Spatial and temporal case analyses connected these aggregate improvements to estimates of wind extent and its evolution. These results support joint use of spectral and temporal satellite information for geographically resolved wind-radius estimation.



Academic Editor: Mark Bourassa

Received: 18 August 2026

Revised: 3 October 2026

Accepted: 6 October 2026

Published: 7 October 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: tropical cyclone; wind-radius estimation; geostationary satellite; multispectral imagery; irregular time series; structured prediction; ordinal regression

1. Introduction

Maximum wind speed describes tropical-cyclone (TC) intensity at one location; wind radii describe the area reached by damaging surface winds. R34, R50, and R64 mark the maximum extent of the 34, 50, and 64 kt thresholds in the northeast (NE), southeast (SE), southwest (SW), and northwest (NW) geographic quadrants [1]. The twelve radii capture both overall TC size and azimuthal asymmetry, including the contraction of stronger-wind regions toward the inner circulation. These properties shape marine exposure, coastal warning areas, surge response, and hazard assessment beyond what a single intensity value can express [2,3]. Their physical ordering, $R34 \geq R50 \geq R64$, further makes wind-radius estimation a joint geographic and cross-threshold problem. Together, intensity and quadrant radii determine how hazards extend around a TC and which coastlines, vessels, and offshore assets face exposure.

Operational analyses assemble wind radii from heterogeneous observations and expert interpretation. Aircraft, buoys, scatterometers, passive-microwave radiometers, and synthetic aperture radar provide surface-sensitive evidence, but each source has limitations in revisit time, swath, rain sensitivity, saturation, or basin availability [4–7]. These measurements therefore offer valuable but intermittent views of the open-ocean circulation. Geostationary imagers provide a different advantage: frequent, broad-coverage observations of cloud-top temperature, phase, moisture, and vertical structure from a nearly fixed viewing geometry [8–11]. Although those radiances do not measure surface wind directly, their spatial organization and evolution contain information related to circulation size and asymmetry. The complementarity between surface-sensitive observations and continuous geostationary coverage has consequently made satellite data central to both operational analysis and objective wind-structure estimation.

Satellite-based methods have progressed from empirical cloud-pattern relationships and radial brightness-temperature profiles to learned multispectral estimators [12–17]. Recent methods also retain geographic asymmetry: Lu et al. estimated quadrant R34/R50/R64, TC-WREM estimated four-quadrant gale-force radii, and DL-TCR used two infrared image times in threshold-by-quadrant models [18–20]. Together, these studies show the value of satellite information while leaving a question for joint estimation: How should complementary spectral measurements and an irregular observation history be combined to describe twelve related wind radii?

Three gaps remain for joint wind-radius estimation across the North Atlantic (AL), Eastern North Pacific (EP), and Western North Pacific (WP). First, cloud-top radiance is indirect evidence of near-surface wind extent, and corresponding Advanced Baseline Imager and Advanced Himawari Imager channels retain sensor-specific spectral responses and radiometric differences [21]. Second, basin-spanning multispectral observation histories are irregular: scans become available at unequal offsets, cadence varies, and missing history is sample-dependent; explicit handling of unavailable inputs is motivated by missing-modality studies [22]. Prescribing one or two image times cannot fully represent that temporal support. Third, fitting thresholds or quadrants separately does not encode their joint structure. Independent outputs can lose fixed geographic identity under transformed views and can violate $R34 \geq R50 \geq R64$. The central methodological problem is thus to integrate physically identified multispectral observations and their actual timing while producing one geographically registered, cross-threshold radius distribution.

We address this problem with TC-RadiiFormer, which connects the available observations to the joint structure of quadrant wind radii. Spectral graph-set encoding combines physically identified channel measurements and their differences; current-relative polar memory incorporates irregular historical observations using their actual acquisition times; and inverse geographic alignment followed by nested decoding produces ordered radius distributions in fixed quadrants. These choices permit spectral and temporal information to inform all twelve outputs together. Complete-system benchmarks test the resulting estimator, matched spectral and history controls quantify information contributions, and separate structural and probability diagnostics assess output consistency and distribution quality. The channel-role organization supports joint ABI/AHI modeling; source-only transfer and limited output-head adaptation assess its empirical limits.

2. Related Work

Satellite wind-radius estimation uses different views of TC structure. Infrared cloud patterns and radial brightness-temperature profiles provide indirect indicators of circulation size, supporting azimuthal wind-profile and R34 estimates [12–14]. Operational formulations combine routinely available TC information with infrared- or model-derived size to estimate quadrant R34, R50 and R64 [23]. Surface-sensitive observations offer a complementary view: scatterometers support quadrant R34 estimation [6], while synthetic aperture radar (SAR) can resolve maximum wind, radius of maximum wind (RMW) and threshold radii [24].

These observation types also define the interpretation of an evaluation. SAR and best-track structure can differ with the parameter and analysis convention [25], and final best-track radii synthesize heterogeneous, spatially incomplete evidence [23,26]. Satellite estimators trained against best-track radii therefore learn agreement with that retrospective reference. Independent surface-wind observations provide a separate route to assessing the underlying wind structure.

The output object is equally important when comparing learned estimators. TC-SEM, DeepTCNet and later transfer-learning studies estimate scalar size or azimuthally averaged structure [15,16,27]. TC-ResNet combines infrared, water-vapor and microwave information for quadrant-mean R34 [17]. Physics-incorporated, multimodal and multitask approaches extend the estimated size quantities and input information [28–32]. These studies establish useful links between satellite structure and TC size, but their aggregate radii do not retain all twelve threshold-by-quadrant values considered here.

The closest comparisons retain geographic quadrants. Lu et al. use radial infrared profiles, TC center and intensity information, with estimated R34 contributing to the R50 and R64 models [18]. TC-WREM combines TC attributes with a cloud image for quadrant gale-force radii [19]. DL-TCR uses current and 3 h earlier infrared imagery in separate threshold-by-quadrant submodels [20]. Sequential and missing-modality studies further show the value of temporal information and explicit treatment of unavailable observations [22,33]. Table S4c, Panel A compares these tasks, inputs and outputs; the observation-matched DL-TCR and Lu adaptations in Section 5.2 supply the closest task-specific tests.

Combining observations across channels and times requires preserving their measurement meaning. Corresponding ABI and AHI bands retain distinct spectral responses, calibration and viewing geometry [21]; descriptor-aware set processing offers a way to combine identifiable channel roles [34]. Studies of sensor adaptation and geographic or temporal shifts show why a common input format alone cannot establish transfer [35,36]. Historical scans likewise arrive at unequal intervals. Irregular-series methods [37,38],

missing-observation handling [22] and high-cadence TC-center estimation [39] motivate using actual acquisition times rather than only image order.

The joint output adds geometric and cross-threshold requirements. TC-centered polar coordinates represent radial and azimuthal structure [40], quarter-turn group processing supports shared spatial transformations [41], and ordinal regression provides ordered categorical outputs [42]. Continuous wind-field reconstruction and downscaling [43–46], aircraft-constrained quadrant analysis [47], future-size prediction [48] and SAR profile analysis [49] address related aspects of wind structure. Here the target is the current extent of three wind thresholds in fixed geographic quadrants. The following method combines that target structure with multispectral observations and their recent history.

3. Data

3.1. Geostationary Satellite Observations

The predictor source comprised observations from the Advanced Baseline Imager (ABI) on the GOES-R series and the Advanced Himawari Imager (AHI) on Himawari-8/9. The multiplatform ABI/AHI acquisition record was harmonized through ten physical channel roles before sample construction. ABI Level-1b full-disk radiances (ABI-L1b-RadF) and AHI observations from the AHI-L1b-FLDK collection, decoded as JMA Himawari Standard Data v1.3, were calibrated and navigated according to provider documentation [9–11]. Interband acquisition times within a physical observation had to agree within 120 s, and the complete multichannel scan span was limited to 900 s. Scans with unsupported navigation, metadata, or source geometry were excluded.

Predictors comprised calibrated radiometry and spectral, geometric, temporal, quality and validity metadata. Basin, intensity, pressure, radius of maximum wind, TC motion, historical wind radii, model fields and platform identity were excluded from the numerical predictor tensor. Platform identity supported provenance and quality control. External TC centers from the retrospective best-track context supported platform screening, scan selection and north-up, TC-referenced azimuthal-equidistant (AEQD) geometry.

The observations span three basins and a wide range of TC locations (Figure 1). Table 1 links this geographic coverage to the year-separated data roles and threshold-specific evaluation populations; Table S1a gives the yearly basin composition.

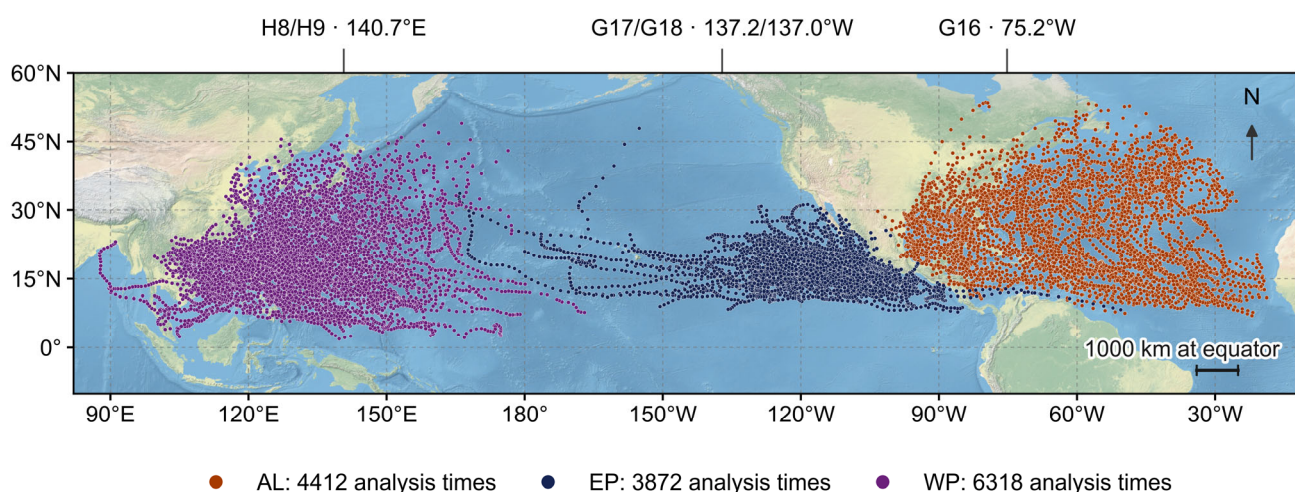


Figure 1. Study region and all 14,602 TC analysis locations during 2017–2024: 4412 in AL, 3872 in EP and 6318 in WP. Points identify analysis times, not independent storms. Upper labels give the geostationary platforms and orbital longitudes. The Plate Carrée map uses Natural Earth I land cover, shaded relief and ocean bathymetry as geographic background; these layers are not model inputs. Dashed meridians and parallels are spaced by 30° and 15°. The scale refers to the equator.

Table 1. Data partitions and positive-reference test populations. (a) Year-separated analysis-time samples; (b) Positive-reference populations in the 2024 test set.

(a)				
Role	Years	Times	TCs	Basins
Training	2017–2022	11,161	428	AL/EP/WP
Validation	2023	1906	60	AL/EP/WP
Test	2024	1535	62	AL/EP/WP
(b)				
Radius	TCs	Analysis times	Quadrant labels	
R34	56	1012	3443	
R50	39	577	2029	
R64	31	378	1306	

Each radius has its own positive-reference population. Valid zeros, missing labels and overflow are summarized in Table S2a,b. The 428 distinct training TCs contribute 430 annual occurrences because two storms span calendar years (Table S6e). The separate 2022 evaluation used 2017–2020 for training and 2021 for validation.

3.2. Retrospective Reference Wind Radii

Wind radii came from retrospective final ATCF-format BEST analyses: National Hurricane Center/Central Pacific Hurricane Center records for the North Atlantic and eastern/central North Pacific, and Joint Typhoon Warning Center records for the Western North Pacific. The dataset used zero-hour BEST records at 00, 06, 12, and 18 UTC, with Central North Pacific records included in the Eastern North Pacific population. The native target unit was nautical miles (nmi), and the output quadrant order was NE, SE, SW, and NW.

Each R34, R50, and R64 quadrant label had one of four meanings: missing, valid zero, positive radius, or overflow beyond 800 nmi. Explicit masked zeros were retained as valid zeros, while negative ATCF filler values remained missing. When the analysis maximum wind was below a radius threshold, an otherwise unreported cell without a negative filler was assigned valid zero; maximum wind was used only for this target-construction rule. Positive labels through 800 nmi entered finite ordinal shells, and larger labels entered the overflow category. Explicit zeros and Vmax-inferred zeros were recorded separately. Missing cells were marginalized in training and excluded from scored endpoints (S2).

Final BEST radii combine observations, retrievals, analysis rules and forecaster judgment into a retrospective expert-synthesis reference [23,26]. Their coverage and uncertainty vary among thresholds, basins and observing periods. This study evaluates agreement with that reference; independent surface-wind validation is discussed in Section 6.

Visible reflectance resolves illuminated cloud texture, while water-vapor and thermal infrared brightness temperatures describe complementary moisture and cloud-top features (Figure 2). These channels provide indirect information about the cloud system associated with the circulation. Figure S1 and Table S1b give the full ten-band example and ABI/AHI mapping.

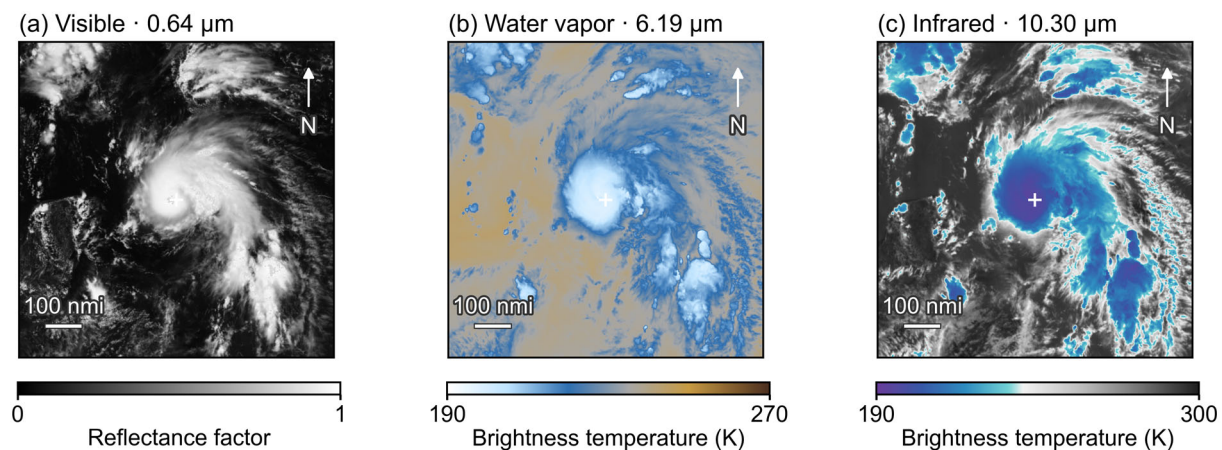


Figure 2. Complementary ABI observations of Ida on 27 August 2021: (a) visible reflectance, (b) water-vapor brightness temperature and (c) thermal infrared brightness temperature. The three panels share the same center and geographic extent, with separate physical color scales and 100-nmi scale bars. The scan starts at 17:50:20 UTC and ends at 17:59:52 UTC, before the 18:00 analysis time. These calibrated images illustrate the channel information; bilinear display resampling retains the original masks and does not increase the observation resolution. Section 3.3 defines input eligibility.

3.3. Analysis-Time Sample Construction

Each sample was defined at analysis time t_0 . The anchor observation was the latest complete physical scan whose acquisition ended no later than t_0 and no more than 15 min before it. Candidate history covered the preceding 6 h and was eligible only after acquisition had finished. Scan completion governed causal availability; acquisition midpoint time t defined the temporal offset used by the model. A scan could therefore be available at t_0 while retaining a nonzero midpoint offset.

Physical scans were stably ordered by actual midpoint offset, and duplicate source identities were removed before the two history-window caps were applied independently. Missing history remained missing instead of being filled by a duplicate observation. Every AEQD grid contained 448×448 samples. Grid spacing was 2, 4, and 8 km for the anchor observation, 2 and 4 km for observations within the preceding hour, and 8 km for observations from 1 to 6 h; the three grid spacings correspond to nominal full widths of 896, 1792, and 3584 km. Radiometric and brightness-temperature-difference fields used channel-wise median and median absolute deviation (MAD) normalization and clipping to $[-8, 8]$.

A 15 min completion window was specified before evaluation to select the latest completed multichannel observation; tighter-window results and their different sample populations are reported in Table S6b.

The fixed archive contains 11,161 training, 1906 validation and 1535 test analysis times. The 2024 test set contains 62 TCs before positive-radius filtering (Table 1).

4. Methodology

4.1. Task Formulation

For analysis time t_0 , let $I(t_0)$ index the anchor observation and the preceding 6 h observations retained after stable sorting, source-identity deduplication, and the two window-specific caps. For observation i , x_i denotes its calibrated multichannel spatial tensor, t_i its acquisition midpoint, d_i its observation-associated descriptors, and m_i its

validity and quality masks. The complete input set X_{t_0} and the model mapping f_θ to the predictive-median radius matrix are

$$X_{t_0} = \{(x_i, t_i, d_i, m_i) : i \in I(t_0)\},$$

$$\hat{R}_{t_0}^{(0.5)} = f_\theta(X_{t_0}) \in \mathbb{R}^{3 \times 4}. \quad (1)$$

Equation (1) defines current-time wind-radius estimation of R34, R50, and R64 in NE, SE, SW, and NW from observations completed by t_0 . The final BEST reference supplies the same-time matrix.

4.2. TC-RadiiFormer Architecture

The architecture combines complementary spectral measurements, an irregular observation history and ordered quadrant outputs (Figure 3). TC-RadiiFormer has 27.45 million parameters. Shared encoding and historical processing precede inverse geographic alignment, averaging and a single decoding step; Table S3a, Panel A and Figure S2 provide implementation details.

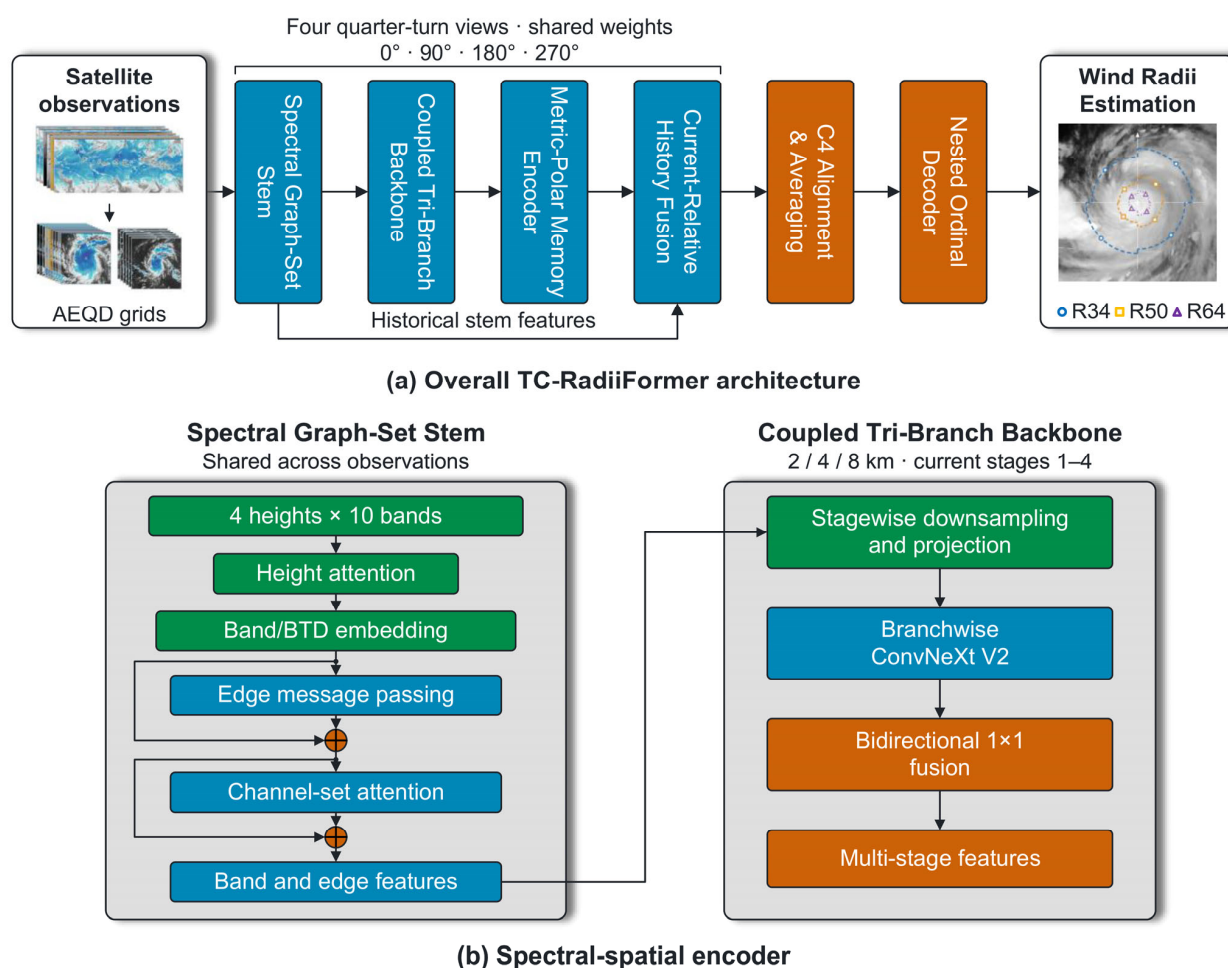


Figure 3. TC-RadiiFormer architecture. **(a)** Four shared-weight quarter-turn views combine current features and stem-derived history before inverse geographic alignment, averaging and one nested ordinal decoder. **(b)** Spectral graph-set encoding and the coupled tri-branch backbone provide multiscale features. Current observations use all four backbone stages; historical memory reads stem features. Arrows indicate the direction of information flow. Box colors serve as visual grouping cues, while labels identify the operations; blue circles, gold squares and purple triangles denote R34, R50 and R64, respectively. Satellite images and radius overlays are illustrative. Figure S2, Table S3a and Algorithm S1 provide implementation details.

4.2.1. Spectral Graph-Set Encoder

Channels respond to different cloud and moisture properties, so the encoder retains their physical identities while learning their interactions. Ten roles span visible, near-infrared, shortwave-infrared, water-vapor and thermal measurements. Provider calibration, channel descriptors and training-only median/MAD standardization accommodate heterogeneous observations [21]. Reflective-band reliability decreases smoothly over solar zenith angles of 82–90°; S1 gives mappings and masks.

Wavelength, spectral-response, resolution and validity descriptors condition mixing across projection heights of 0, 6, 12 and 18 km. These hypotheses differ from horizontal sampling scales. Weights are normalized over locally valid supports.

The spectral graph-set stem uses seven signed brightness-temperature differences (Figure 3b). Two rounds of edge-conditioned message passing and permutation-equivariant set attention [34] feed coupled band, edge and spatial streams through ConvNeXt V2 blocks [50]. The current observation passes through all four stages at three scales, yielding twelve feature representations. Historical encoding shares the stem and first-stage weights and stops after stage 1; temporal memory reads the stem features. Section 5.3 evaluates channel diversity and these interactions; S4 gives the edges and dimensions.

4.2.2. Metric-Polar Temporal Memory

Actual acquisition-time offsets and validity masks preserve irregular timing and missing-history information within the fixed-capacity input representation. Threshold-conditioned cross-attention forms current-observation memory from the full multiscale features. Recent observations at 2/4 km spacing and earlier observations at 8 km spacing use stem-derived historical features. Section S3 defines the time windows, scan caps and missing-history rules.

The recent pathway aligns historical features using bounded radial and angular offsets, then combines their differences and interactions with recency weighting. The longer pathway uses actual interscan intervals in an exponential state update and a support-weighted mean of filtered anomalies. Figure S2 relates these operations to the input history; Algorithm S1 specifies the recurrence.

A masked gate combines the historical updates as a bounded residual of the current features. The current-observation path has coefficient one and remains unchanged when no valid history exists. Equation (S3) supplies the 25% root mean square (RMS) bound and numerical safeguards. The history and time-encoding controls test these information sources separately.

4.2.3. C4-Aligned Nested Decoder

The decoder must preserve north-referenced quadrant identity while restricting the three wind thresholds to their physical order. Polar structure, discrete group-aware learning, and ordinal regression provide the geometric and ordered precedents [40–42]. Let X denote the complete spatial predictor family, $g \in \{0, 1, 2, 3\}$ the quarter-turn index, ρ_g and ρ_g^{-1} the forward and inverse transforms, Φ the complete encoder-plus-history path, and Ψ the decoder. The aligned memory and output distribution are

$$\bar{h} = \frac{1}{4} \sum_{g=0}^3 \rho_g^{-1} \Phi(\rho_g X), P = \Psi(\bar{h}). \quad (2)$$

Equation (2) applies synchronized C4 quarter turns to the spatial inputs. Shared encoder and history outputs are inverse-aligned, averaged, and decoded once, keeping estimates tied to geographic quadrants. The rotation controls evaluate this processing separately from historical information.

Twelve threshold-by-quadrant queries combine the aligned memory, shared TC state, threshold states and quadrant-local summaries. Cyclic interactions connect neighboring quadrants. A fixed reordering converts counterclockwise polar sectors to NE, SE, SW and NW.

Four nested presence states per quadrant encode absence and the three admissible threshold combinations. A cyclic conditional random field couples quadrant states. Conditional positive-radius transports restrict finite support to ordered thresholds and treat overflow as absorbing. The marginals contain valid zero, 80 finite positive shells and overflow; Section S2 defines all four branches, their normalization, missing-label marginalization and the non-nested-label fallback; Table S3a, Panel A gives dimensions.

For the all-present state, let $k, j, i \in \{0, \dots, 80\}$ index conditional positive R64, R50, and R34 states; let $p_{64,k}$, $p_{50,j}$, and $p_{34,i}$ be their distributions; and let T_{kj}^{50} and T_{ji}^{34} be the learned R64-to-R50 and R50-to-R34 transports. For batch row n , threshold r , quadrant q , and output category $v \in \{0, \dots, 81\}$, P_{nrqv} is the marginal probability and c_v its radius representative. A separate cumulative-category index a defines the predictive-median readout. The nested transports and deterministic median are

$$\begin{aligned} p_{50,j} &= \sum_k p_{64,k} T_{kj}^{50}, T_{kj}^{50} = 0 (j < k), \\ p_{34,i} &= \sum_j p_{50,j} T_{ji}^{34}, T_{ji}^{34} = 0 (i < j), \\ \hat{R}_{nrq}^{(0.5)} &= c_{\min\{a: \sum_{v=0}^a P_{nrqv} \geq 0.5\}}. \end{aligned} \quad (3)$$

Equation (3) orders finite-shell radii as $R34 \geq R50 \geq R64$. Shell spacing is 5 nmi to 200 nmi, 10 nmi to 400 nmi, and 20 nmi to 800 nmi. The predictive median minimizes expected absolute error under the model distribution, with overflow represented at 800 nmi. Section 5.4 evaluates point accuracy separately from output ordering.

4.3. Training and Inference

The joint negative log-likelihood marginalizes missing labels and includes a marginal fallback for observed non-nested labels. Together, these terms define L_{joint} . A dimensionless finite-grid cumulative distribution function (CDF) loss regularizes the objective. This upper-edge auxiliary differs from the exact discrete continuous ranked probability score (CRPS) used for evaluation [51]; both are defined in S2.

Let L_{sup} be the supervised loss. It combines the joint likelihood and finite-grid CDF loss in the default configuration; the loss controls replace the likelihood term:

$$L_{\text{sup}} = L_{\text{joint}} + 0.25L_{\text{grid}}. \quad (4)$$

Evaluation CRPS uses the energy expression for discrete radius representatives, with targets and overflow mass capped at 800 nmi. It is reported in nmi over valid labels; positive-radius comparisons use the median in Equation (3). All neural models start without pretraining.

5. Experiments and Results

5.1. Experimental Setup and Evaluation

Observations from 2017 to 2022 formed the training set, 2023 supplied validation, and 2024 supplied evaluation. Learning rates and checkpoints were selected using 2023 data. Domain-transfer experiments restricted training, validation and normalization to the source domain (S7).

The 2024 positive-radius populations contained 56/39/31 eligible TCs, 1012/577/378 analysis times and 3443/2029/1306 quadrant cells for R34/R50/R64 (Table 1). Each thresh-

old has its own positive-label population. AdamW used a learning rate of 3×10^{-4} , weight decay of 0.05, an effective batch size of 32, 10 warmup epochs, cosine decay, a 120-epoch ceiling, and patience of 20. Checkpoints minimized 2023 positive-radius mean MAE, with the earliest best epoch breaking ties. Seeds were 3401, 3402 and 3403.

The primary metric was MAE of predictive-median radii conditional on a positive reference. Absolute errors were averaged over eligible quadrants, then analysis times within each TC, then eligible TCs. Seeds 3401, 3402 and 3403 contributed equally. Let $\mathcal{B}$ denote the set of basins AL, EP and WP; $M_{m,b,r}$ denotes model m 's TC-aggregated MAE in basin b at radius r , and $N_{b,r}$ denotes its eligible TC count. The all-basin value is

$$M_{m,\text{all},r} = \frac{\sum_{b \in \mathcal{B}} N_{b,r} M_{m,b,r}}{\sum_{b \in \mathcal{B}} N_{b,r}}. \quad (5)$$

Equation (5) applies eligible-TC weighting separately within each radius. Mean MAE is the arithmetic mean of the three all-basin MAEs for R34, R50 and R64; each wind-speed threshold therefore receives equal weight. Valid zeros and missing labels are distinct from the positive-radius endpoint. Probability and presence scores use their stated label populations. Distances are in nmi (1 nmi = 1.852 km), and lower MAE indicates smaller error.

ResNet-50, Swin V2-T and Plain CNN represent residual convolution, hierarchical window attention and a simple convolutional system [52,53]. They receive the same radiometric information, descriptors and observation slots through a common adapter. A fourth system adds a two-layer temporal Transformer to Swin V2-T, replacing concatenation fusion while retaining the independent distributional objective. Table S4a and Section S5 specify the implementations. Learning-rate selection uses 2023 positive-radius mean MAE from 10^{-4} , 3×10^{-4} and 10^{-3} ; all neural systems start without pretraining. DL-TCR and Lu adaptations provide task-specific comparisons on matched WP observations (S5).

On an NVIDIA A100 80 GB GPU (NVIDIA Corporation, Santa Clara, CA, USA), the 27.45 M-parameter model required 48.973 ms per batch of 16 samples (3.061 ms per sample, amortized), using BF16, 20 warmups and 100 synchronized repetitions. This computation is small relative to the six-hour analysis cycle; acquisition, preprocessing and delivery add to operational latency. Table S4e gives baseline timings and training costs.

For each contrast, $\Delta M = M_{\text{control}} - M_{\text{reference}}$; positive values favor the reference. Paired confidence intervals use 50,000 whole-TC bootstrap draws within each basin, shared across compared outputs. Bonferroni correction controls the family-wise error rate at 5% within 20 system/component, nine validation, three domain-transfer, seven probabilistic, eight adapted-method, five label-state and six input-perturbation comparisons. Four additional mean-MAE comparisons of temporal fusion and point ordering form a separate family. Each interval has confidence $1 - 0.05/K$ for family size K (Table S5i,h). Intervals condition on the fitted seeds and exclude retraining uncertainty; crossing zero does not establish equivalence. Basin-radius summaries with fewer than ten eligible TCs are descriptive. Readout, head-adaptation and radius-calibration comparisons use additional families of four, two and two endpoints, respectively (Tables S6f,g and S8e).

Training-distribution diagnostics use predictions from the fitted models on the 2017–2022 training set. Alternative point estimates use the same predicted probabilities. Output-head adaptation and calibration use earlier-year/2023 target data, with 2024 reserved for evaluation (Tables S6e–g and S8e).

5.2. Wind-Radius Estimation Performance

Combining spectral and historical information yielded lower positive-radius MAE at all three thresholds than the four generic estimation systems (Table 2). Relative to

ResNet-50 and Swin V2-T, mean MAE fell by 27.1% and 25.4%; comparator-minus-TC-RadiiFormer differences were 3.386 nmi (CI 1.917 to 4.911) and 3.092 nmi (CI 1.349 to 4.737). The advantage remained when Swin V2-T included a temporal Transformer: 13.0%, or 1.364 nmi (CI 0.588 to 2.337). Table S5d provides extended and quadrant metrics.

Table 2. Positive-radius accuracy of complete estimation systems on the 2024 test set.

Model	R34	R50	R64	Mean MAE	Difference	Confidence Interval
TC-RadiiFormer	13.164	8.384	5.731	9.093	0.000	Reference
ResNet-50	17.202	11.278	8.956	12.479	3.386	[1.917, 4.911]
Swin V2-T	16.684	10.615	9.257	12.185	3.092	[1.349, 4.737]
Plain CNN	28.938	16.759	12.888	19.528	10.435	[8.676, 12.395]
Swin V2-T + temporal Transformer	15.139	9.641	6.590	10.457	1.364	[0.588, 2.337]

Errors and differences are in nmi. Mean MAE gives equal weight to R34, R50 and R64; the difference is comparator minus TC-RadiiFormer. All rows use uncalibrated predictions and the same 56/39/31 eligible TCs. Confidence intervals for ResNet-50, Swin V2-T and Plain CNN are 99.75%; the temporal-Transformer comparison uses 98.75%, following the respective comparison families in Section 5.1.

Swin with temporal attention had a lower mean-MAE point estimate than concatenation fusion: 10.457 versus 12.185 nmi. The paired difference was -1.728 nmi (CI -3.484 to 0.253), so an improvement between these fusion variants was not established. TC-RadiiFormer had lower error than both, with adjusted paired intervals supporting each comparison.

The R34/R50/R64 MAEs were 13.164 ± 0.157 , 8.384 ± 0.068 and 5.731 ± 0.169 nmi, respectively, where dispersion is the sample standard deviation across three fitted models. Their arithmetic mean was 9.093 nmi. Figure 4 shows basin-specific errors; Table S5c gives the estimates and evaluation populations.

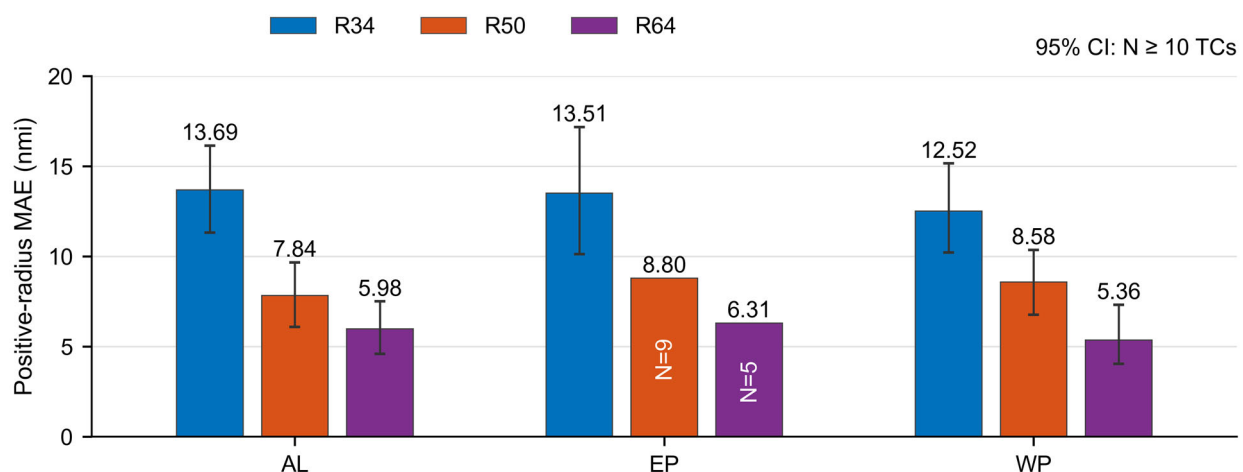


Figure 4. Positive-radius MAE by basin. Blue, orange and purple denote R34, R50 and R64. All bars are solid. Error bars show descriptive 95% basin-specific whole-TC bootstrap confidence intervals. EP R50 and R64 each have fewer than ten eligible TCs and are displayed without error bars. Radius-specific TC counts appear in Table S5c; the vertical axis begins at zero.

Signed errors complement the accuracy comparison. TC-RadiiFormer biases were -6.250 , -2.924 and 1.482 nmi for R34/R50/R64, indicating underestimation of the outer thresholds and a smaller positive R64 bias. Figure 5 relates predictions to the reference-radius distribution, and Figure S6 shows the corresponding residuals. Section 5.5 examines their variation across TC conditions and individual sequences.

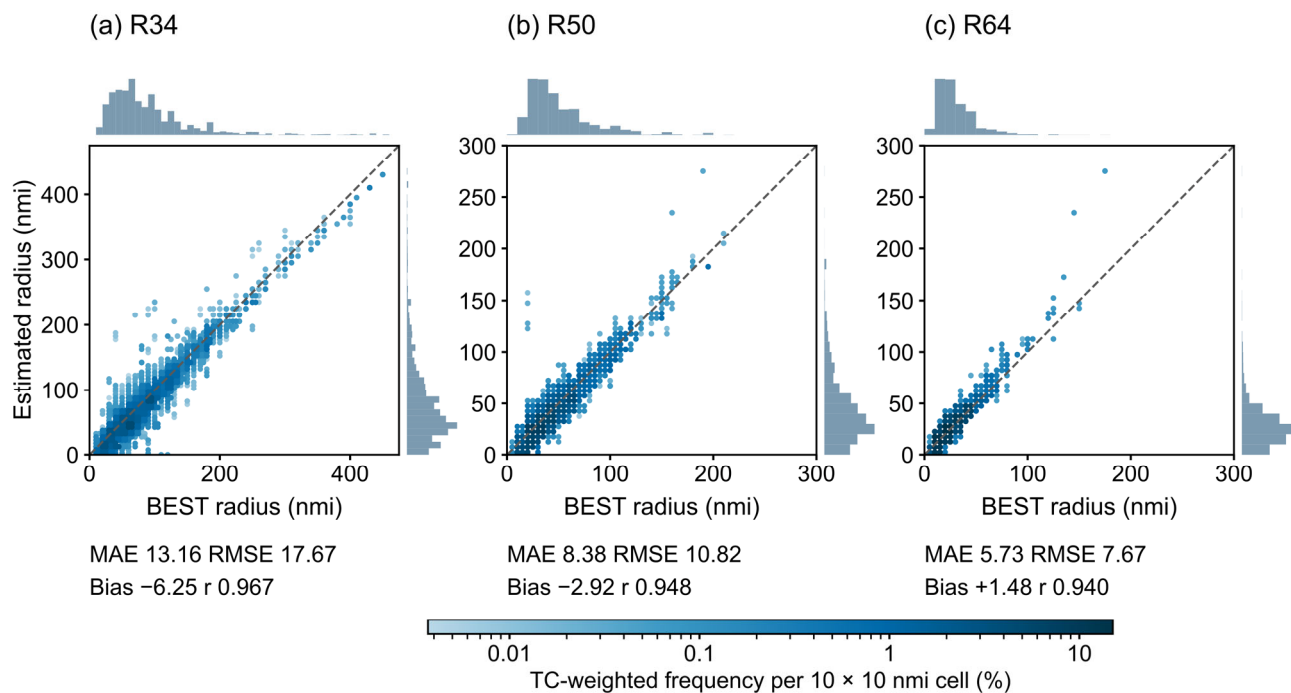


Figure 5. Estimated and BEST positive radii for (a) R34, (b) R50 and (c) R64. Circular points include every prediction from the three seeds, without displacement or trimming. Color is TC-weighted frequency in each 10×10-nmi cell, normalized to 100% per radius, with equal seed weights and a shared logarithmic scale. Upper and right histograms show the corresponding weighted BEST and estimated-radius distributions; their height scales are independent. Dashed lines mark equality. MAE, RMSE, bias (nmi) and weighted correlation r are scored separately per seed, then averaged. Repeated predictions do not add independent TCs.

DL-TCR and Lu et al. [18,20] provide image-based and profile-based wind-radius comparisons (Table 3). Each adaptation and its TC-RadiiFormer reference use the same eligible observations and spatial support while retaining their native feature representations; Lu uses radial profiles whereas TC-RadiiFormer retains two-dimensional IR structure. Mean differences were 3.952 nmi (CI 2.626 to 5.212) and 9.707 nmi (CI 7.798 to 11.318), respectively, on 341 WP times and 22/17/15 positive TCs. Table S5e reports the extended metrics.

Table 3. Observation-matched comparisons with literature-guided wind-radius models.

Model	R34	R50	R64	Mean MAE	Difference	Confidence Interval
TC-RadiiFormer (DL inputs matched)	12.428	6.907	7.093	8.809	0.000	Reference
DL-TCR adaptation	17.416	11.480	9.386	12.761	3.952	[2.626, 5.212]
TC-RadiiFormer (Lu inputs matched)	14.074	8.938	7.194	10.068	0.000	Reference
Lu adaptation	29.933	17.145	12.248	19.775	9.707	[7.798, 11.318]

All errors are in nmi. Mean MAE is averaged over the three thresholds; the difference is adaptation minus its matched TC-RadiiFormer reference. Each paired block shares observations, spatial coverage and stated auxiliary inputs, with 341 WP times and 22/17/15 positive TCs. Native feature representations differ. Confidence intervals are 99.375%, adjusted for eight adapted-method comparisons.

5.3. Contributions of Spectral and Temporal Information

Multispectral observations improved the estimates beyond a single infrared view (Figure 6a). Ten channels reduced mean MAE by 2.987 nmi (CI 2.106 to 3.881), while removing graph messages increased it by 3.136 nmi (CI 2.203 to 4.184; Figure 6b). The TC distributions show how these gains extend across storms: R34 error decreased for 50 of 56 TCs with ten-channel input and for 53 of 56 with graph messages (Figure 6c,d). Individual gains varied, but the positive median differences at all three thresholds support combining complementary channels and their learned interactions.

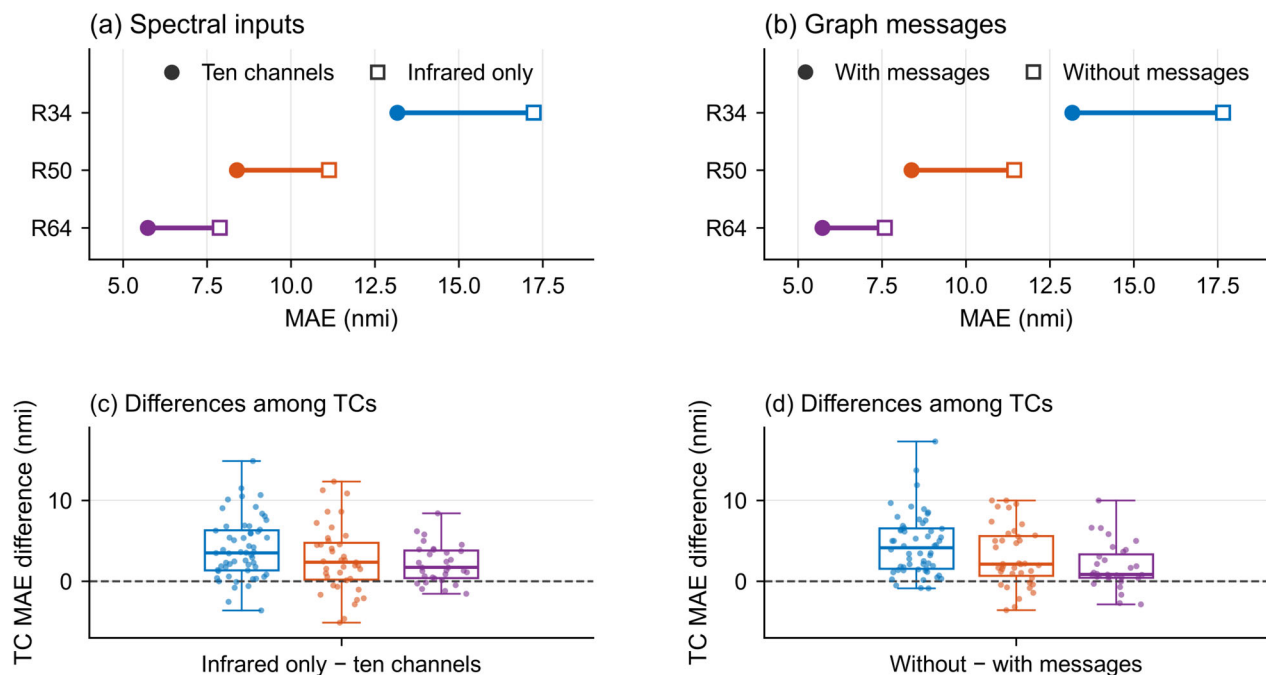


Figure 6. Spectral contributions to radius estimation. Upper panels show MAEs scored per seed and averaged equally for (a) ten-channel versus infrared-only inputs and (b) with versus without graph messages; filled circles denote the complete model and open squares its matched control. Lower panels show seed-averaged TC MAE differences for (c) infrared-only minus ten-channel inputs and (d) without minus with graph messages; positive values favor the complete model. Blue/orange/purple identify R34/R50/R64 (56/39/31 TCs). Boxes span quartiles with medians inside; capped lines span all points. These distributions describe TC variability; adjusted confidence intervals for mean differences are in the text and Table S5i.

Historical observations improved estimates of the current wind extent (Figure 7a). Mean MAE decreased by 2.839 nmi relative to current-only input (CI 1.945 to 3.742) and by 2.477 nmi relative to two observations (CI 1.716 to 3.241). R34 error decreased in 51 and 49 of the 56 TCs, respectively (Figure 7d); the median TC improvement over current-only input was 3.189 nmi. Replacing actual acquisition offsets with uniform intervals increased mean MAE by 2.952 nmi (CI 2.085 to 3.863), with the images and ordering unchanged (Figure 7b,e). Thus, observation timing contributed alongside image content. Using up to 33 historical observations, in addition to the anchor observation, six-hour histories had lower mean MAE than nine- and twelve-hour ones by 1.348 nmi (CI 0.581 to 2.165) and 2.213 nmi (CI 1.480 to 2.937; Figure 7c,f). This identifies the best of the tested windows; changing the window also changes selected observations and their spacing.

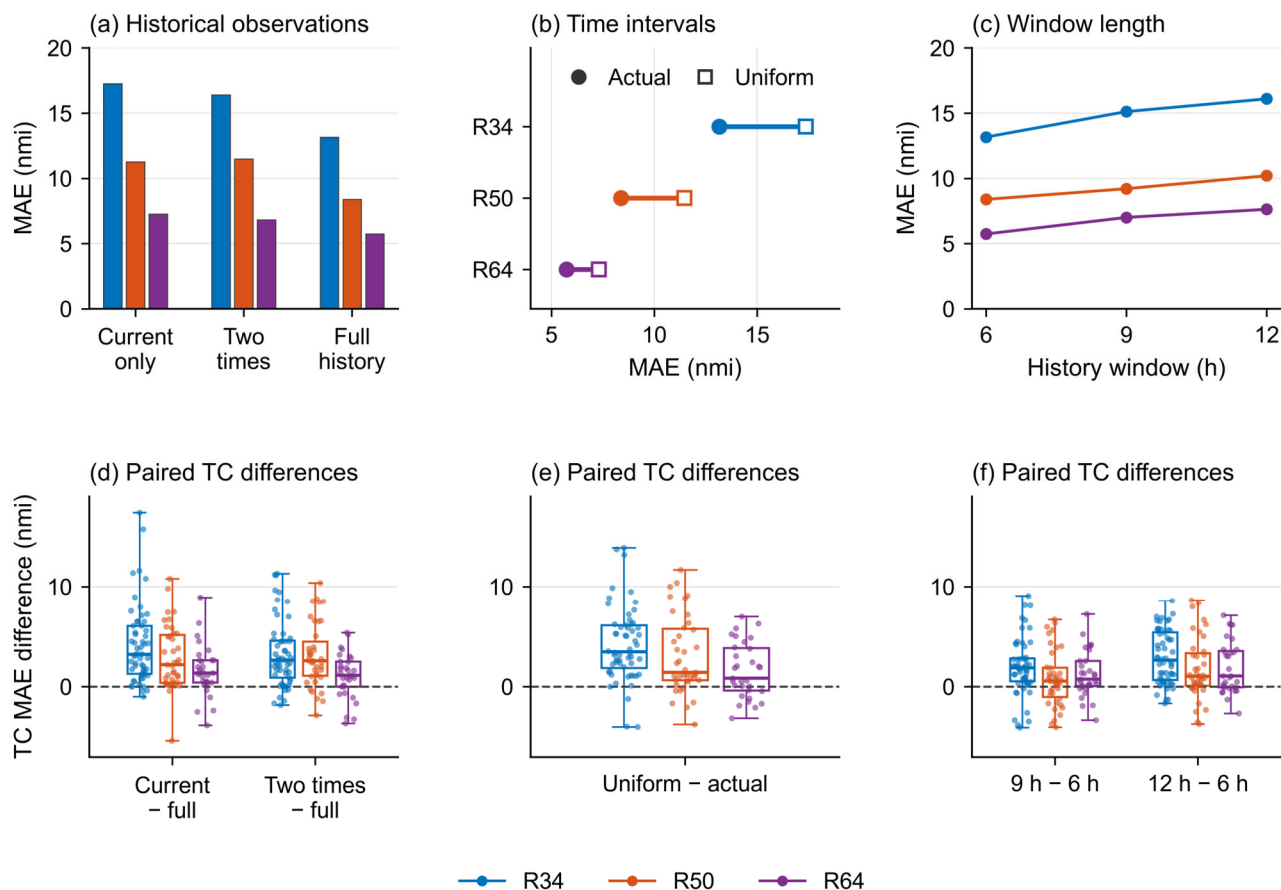


Figure 7. Contributions of historical observations and acquisition timing. Upper panels show absolute MAE for (a) current-only, two-time and full-history inputs; (b) actual versus uniform time intervals; and (c) 6, 9 and 12 h windows with the same cap of 33 historical observations. Lower panels show each corresponding control-minus-reference TC difference: full history is the reference in (d), actual intervals in (e), and 6 h in (f). Blue/orange/purple identify R34/R50/R64. Points, boxes and capped lines follow Figure 6, retaining all 56/39/31 eligible TCs and all five temporal contrasts. These distributions describe between-TC variation; adjusted mean-difference intervals are given in the text and Table S5i, Panel A.

Additional controls support the spatial processing within the integrated system (Table S5a,i). Removing the broader C4 configuration increased mean MAE by 2.983 nmi (CI 2.010 to 4.046), while disabling only outer rotation averaging increased it by 0.959 nmi (CI 0.178 to 1.721). Replacing the spatial backbone with capacity-matched Swin increased error by 1.962 nmi (CI 1.190 to 2.852). These component changes are distinct from the complete-system comparisons in Table 2.

5.4. Effects of Output Constraints

Nested decoding supplies a joint distribution restricted to ordered radii in each geographic quadrant. Independent outputs produced 21.996% point-order violations on the common R64-positive population specified in S4. Under the same marginal likelihood loss, the independent-minus-nested mean-MAE difference was 1.257 nmi (CI −0.744 to 3.372; Figure S3). The nested head therefore guarantees admissible ordering, while its independent point-accuracy advantage remains unresolved.

A parameter-free isotonic projection also removed the independent head's point-order violations and reduced mean MAE from 12.030 to 11.417 nmi (Table S5a). The projected-minus-unprojected difference was −0.613 nmi (adjusted CI −1.144 to −0.183); the projected-minus-nested difference under the common marginal loss was 0.643 nmi

(CI -0.856 to 2.147 ; Table S5i, Panel G). Point projection enforces ordering on median estimates; nested decoding embeds the threshold relationships in the joint predictive distribution used for training and uncertainty estimation.

5.5. Error Characteristics and Case Studies

R34 underestimation occurred across basins: signed biases were $-4.689/-7.751/-6.625$ nmi in AL/EP/WP (Table S5g). Below and at or above 96 kt, MAEs were 13.182/15.375 nmi and biases were $-6.201/-5.178$ nmi. Before, within ± 12 h of, and after peak intensity, biases were $-7.031/-7.010/-3.981$ nmi. These descriptive patterns show a broadly negative offset, with a smaller offset after peak intensity, alongside substantial variation among TCs (Figure 8). Geographic-quadrant MAEs were 14.076, 12.904, 11.611 and 13.077 nmi for NE, SE, SW and NW.

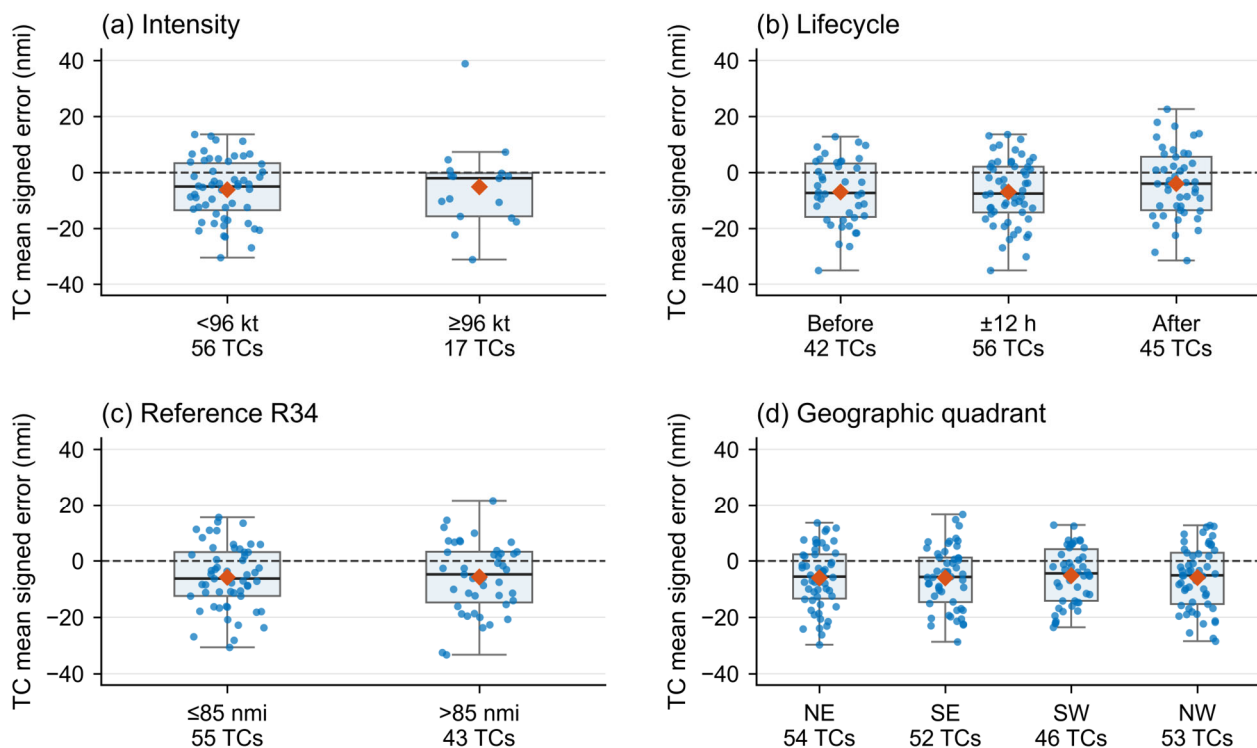


Figure 8. R34 signed errors by (a) reference intensity, (b) time relative to the first lifetime intensity maximum, (c) reference quadrant radius and (d) geographic quadrant. Each blue point is one TC's mean signed error over eligible times and quadrants, averaged equally across seeds; counts give contributing TCs. Orange diamonds mark the arithmetic mean across contributing TCs within each stratum. Boxes span the interquartile range (IQR), with the median inside; whiskers reach the most extreme observations within 1.5 IQR of the quartiles. A TC may contribute to multiple strata. Before/after peak excludes the central ± 12 h window.

The R34 offset was also present across training-defined radius strata: biases below q90, from q90 to q95, and at or above q95 were $-6.199/-5.601/-6.354$ nmi. Thus, it was not confined to the largest reference radii. Figure S7 and Table S6c extend the descriptive analysis to R50/R64; all positive-label residuals are included in Figure S6.

R34 prediction-minus-reference quantile differences at q10/q50/q90 are $-7.500/-0.833/+2.500$ nmi in training and $-12.500/-7.500/+2.500$ nmi in 2024 (Table S6f). The training median is close to BEST, whereas the 2024 median is lower. These distributional differences describe a nonuniform shift across quantiles; they are distinct from paired residual quantiles.

Among 23,537 positive training R34 labels, 1681 (7.142%; TC-weighted 4.043%) exceed 200 nmi and 51 (0.217%; 0.111%) exceed 400 nmi, spanning 90 and 11 TCs (Table S6e). Given the correct finite shell, midpoint errors are bounded by half its width (2.5/5/10 nmi). The nearest-representative oracle MAE is 2.584 nmi, so shell spacing alone does not impose the observed R34 MAE.

With predicted probabilities fixed, using the upper edge of the median category changes R34 bias from -6.250 to -3.631 nmi but increases R64 MAE. The probability mean reduces mean MAE to 8.702 nmi (difference -0.391 ; adjusted CI -0.551 to -0.267), while R34 bias remains -5.952 nmi (Table S6f). Such empirical gains are compatible with imperfect predicted distributions; the median remains the primary readout.

The three preselected cases connect the error distributions to spatial wind extent (Figure 9). The AL and WP estimates capture the broad reference footprint at different scales. In EP032024, R34 errors range from -22.5 to -30.8 nmi across the four quadrants, illustrating the negative offset in Section 5.5. The quadrant error matrices and Table S7a quantify this spatial underestimation.

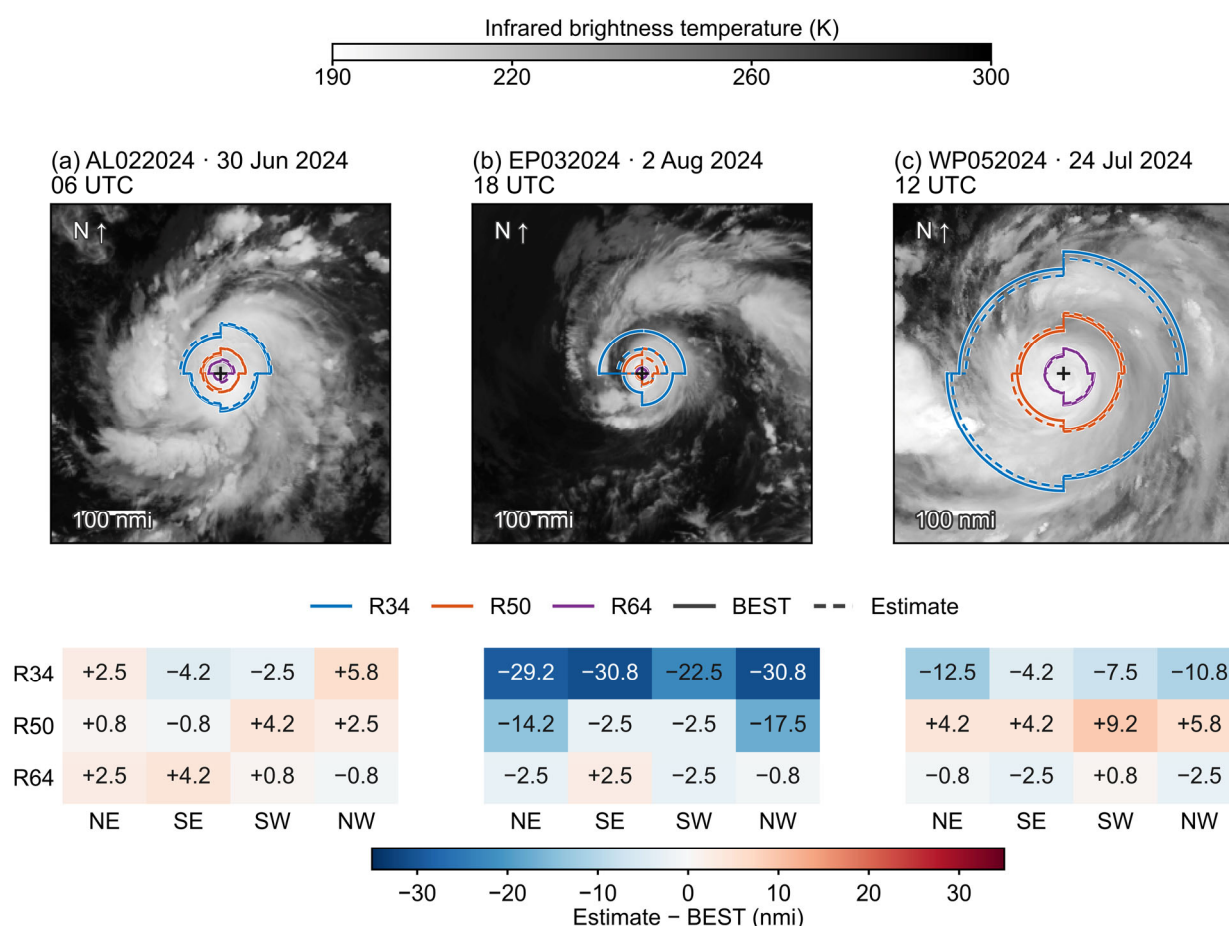


Figure 9. Cloud context and quadrant-radius errors for (a) AL022024, 30 June 2024 at 06 UTC; (b) EP032024, 2 August 2024 at 18 UTC; and (c) WP052024, 24 July 2024 at 12 UTC. Solid BEST contours and dashed three-seed-mean predictive medians overlay north-up ± 512 km infrared scenes on a shared 190–300 K scale. A 100-nmi scale appears at the lower left inside each scene. Blue/orange/purple identify R34/R50/R64; lower panels show estimate-minus-BEST differences in NE/SE/SW/NW. Infrared scenes provide retrospective cloud context; scan times are given in S8. Contours depict quadrant maxima; study-wide scores are computed separately for each seed before equal averaging.

The same cases show how estimates follow the changing wind extent (Figure 10a–c). The AL sequence captures expansion and contraction, while WP includes growth to a much broader R34 footprint. Separating the signed errors makes the history comparison visible (Figure 10d–f): positive-quadrant R34 MAE decreased from 8.521 to 6.875 nmi in AL, 23.322 to 18.880 nmi in EP and 12.665 to 11.068 nmi in WP (Table S7b). EP remained underestimated over much of the evaluated sequence, extending the spatial pattern in Figure 9b through time. These trajectories describe current-time estimation over the complete evaluated case sequences.

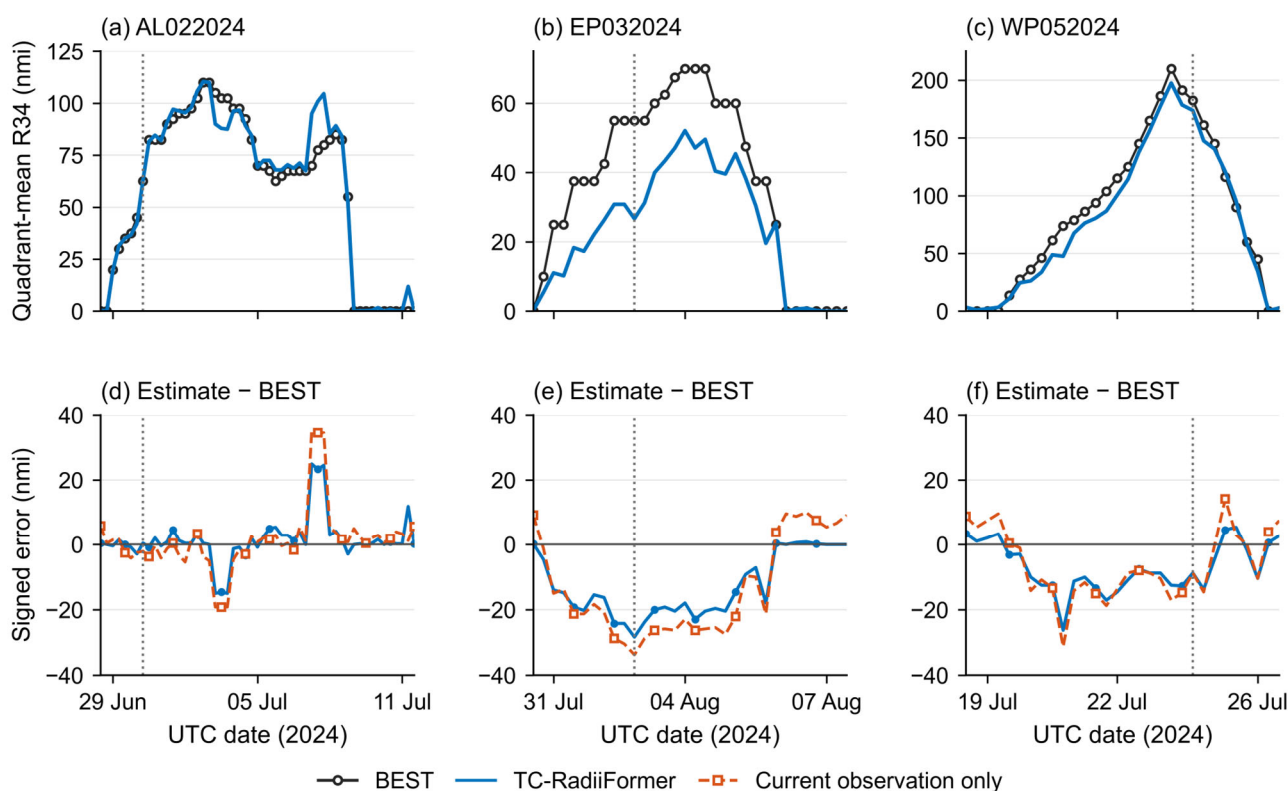


Figure 10. R34 evolution and signed error for the three cases in Figure 9, using all 53, 32 and 30 evaluated times. Panels (a,d), (b,e) and (c,f) correspond to AL022024, EP032024 and WP052024, respectively. Upper panels show four-quadrant-mean BEST radii (black, open circles) and TC-RadiiFormer (blue). Lower panels show estimate-minus-BEST errors for TC-RadiiFormer (blue, solid) and the current-only control (orange, dashed, open squares), on a common error scale. Valid zeros are included. Prediction curves average quadrants within each seed, then seeds equally; markers in the lower row appear every fourth time for readability, while lines retain every evaluated time. Dotted vertical lines mark Figure 9 snapshots. Upper axes start at zero with case-specific limits. Figure S9 and Table S7b provide other radii and positive-quadrant scores.

5.6. Generalization Across Basins and Sensors

Atlantic-withheld training and ABI↔AHI source-only transfer increase mean MAE on matched target populations (Table 4a). Model selection and normalization use source-domain data. A separate retrospective evaluation for 2022 gives a mean MAE of 13.077 nmi using earlier training and validation years (Table S6a).

Table 4. Cross-domain performance on matched 2024 target populations. (a) Source-only transfer; (b) target-assisted output-head adaptation.

(a)				
Training Comparison	TCs R34/50/64	Full-Domain MAE	Transfer MAE	Difference [CI]
Atlantic withheld	19/13/11	9.167	15.334	6.167 [1.889, 11.233]
ABI to AHI	24/17/15	8.874	12.610	3.736 [0.895, 6.880]
AHI to ABI	33/22/16	9.278	15.400	6.122 [3.075, 8.994]
(b)				
Transfer	Source-Only MAE	Adapted MAE	Adapted – Source [CI]	Gap Recovered (%)
ABI→AHI	12.610	12.185	−0.424 [−0.598, −0.265]	11.36
AHI→ABI	15.400	14.279	−1.121 [−1.508, −0.767]	18.30

(a) Difference = transfer minus full-domain MAE; $K = 3$; 98.33% intervals. AHI/ABI target sets contain 779/756 times. Sensor membership may change within a TC, so sensor-specific TC counts need not sum to the global count; (b) Differences are adapted minus source-only MAE; $K = 2$, 97.5% intervals. Recovery (%) = $100 \times (\text{source-only} - \text{adapted}) / (\text{source-only} - \text{full-domain})$, using the matched reference in (a). Differences are calculated before rounding.

Adapting only the nested output head to target 2017–2022 labels, with 2023 checkpoint selection and fixed feature/query modules and source normalization, improves MAE in all six runs (Table S6g). Within the 1000-update budget, adaptation recovers 11.36% and 18.30% of the ABI→AHI and AHI→ABI gaps; residual gaps are 3.311 and 5.001 nmi (Table 4b). These fractions quantify transfer improvement within the learned representations.

5.7. Probabilistic Performance and Calibration

The uncalibrated TC-RadiiFormer also improved probabilistic scores relative to the spatial baselines. Its all-valid mean CRPS was 3.519 nmi, compared with 5.852 for ResNet-50 and 5.488 for Swin V2-T. Paired improvements were 2.332 nmi (CI 1.821 to 2.842) and 1.968 nmi (CI 1.449 to 2.475; Tables S8a and S5i, Panel C). Figure 11 complements CRPS with central-interval coverage and width. Interval endpoints use category shell edges, whereas medians and CRPS use discrete representatives (S2). At the 90% level, R34/R50/R64 coverage was 95.9/99.1/99.8% over all valid labels, with widths of 37.6/23.8/19.0 nmi; positive-label coverage was 91.8/96.5/98.6%, with widths of 49.0/33.9/27.5 nmi.

Shell boundaries and zero probability mass affect coverage. A shared CDF map selected in 2023 reduces all-valid mean CRPS from 3.519 to 3.117 nmi, but increases positive-radius mean MAE from 9.093 to 10.447 nmi. With $a = 1.5$ and $b = 1$, the transformed median corresponds to raw quantile 0.339, shifting estimates toward smaller categories and changing zero/positive boundary crossings. CRPS includes valid zeros and evaluates the full distribution; positive-radius MAE evaluates its median on a different population. Positive CRPS worsens for all radii (Table S8d).

Radius-specific CDF maps preserve marginal ordering through a common positive slope and ordered intercepts (Table S8e). CRPS-only selection yields mean CRPS 3.019 nmi and positive-radius mean MAE 10.567 nmi. Selecting with a per-radius validation-MAE constraint gives corresponding scores of 3.232 and 9.080 nmi, but positive-radius 90% coverage falls from 91.8/96.5/98.6% to 84.6/88.7/97.8%. This selection constraint does not guarantee test MAE or positive-reference coverage. The uncalibrated model remains the primary result.

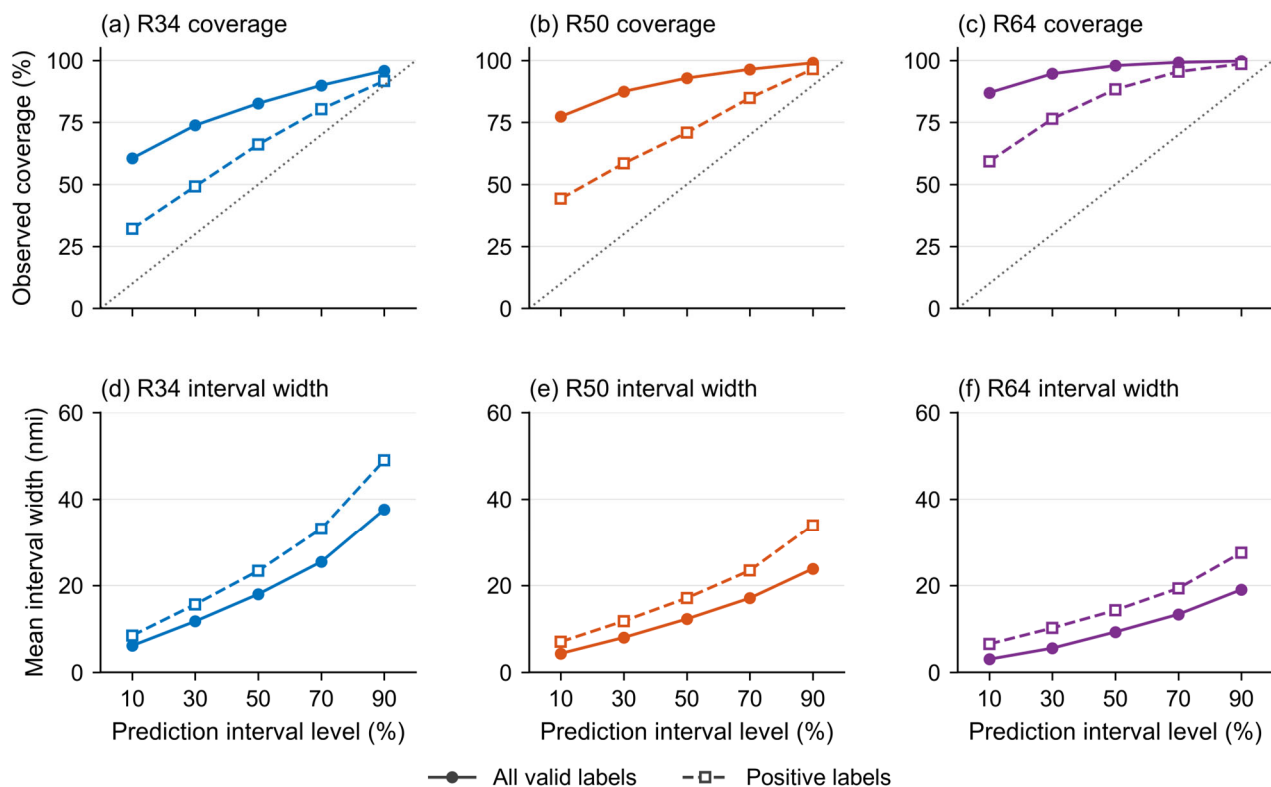


Figure 11. Coverage and width of the uncalibrated TC-RadiiFormer's central prediction intervals. Panels (a,d), (b,e) and (c,f) show R34, R50 and R64, respectively; panels (a–c) show coverage and panels (d–f) show mean width. Solid circles use all valid labels; open squares use positive labels. Dotted diagonals indicate equality between interval probability level and observed coverage. Each population uses its own TC weights and equal seed weights. Endpoints use shell edges (S2). Positive-label curves are stratified diagnostics; prediction intervals describe the output distribution, whereas bootstrap confidence intervals quantify uncertainty in model comparisons.

Presence and zero-radius scores complete the evaluation across observed label states (Table S8b). At the 0.5 presence threshold, R34/R50/R64 sensitivity was 0.989/0.985/0.968 and specificity 0.619/0.653/0.741, indicating stronger detection of present wind radii than identification of absent radii. Valid-zero MAEs were 3.028/2.421/1.572 nmi. Recorded and intensity-inferred zeros are reported separately in S9, together with the available overflow diagnostics.

6. Discussion

TC-RadiiFormer provides a geographically resolved estimate of wind extent from complementary spectral measurements and their observation history. The system comparisons show lower errors than convolutional, spatial-Transformer, and temporal-Transformer baselines on the same retrospective population. Matched controls then identify which information supports that result: full history improved mean MAE by 2.839 nmi over current-only input and 2.477 nmi over fixed two-time input, while multispectral input improved it by 2.987 nmi over infrared alone. Actual acquisition times also contributed when image content and ordering were held fixed. These tests connect the model design to the information gained from frequent satellite observations.

The spatial controls further support outer quarter-turn averaging and the chosen backbone within this estimator. Ordered decoding serves the complementary purpose of maintaining a valid joint output support; its accuracy is assessed separately in the common-loss and projection controls. The lower CRPS of the uncalibrated TC-RadiiFormer

shows that its distributional output provides useful information beyond the median. The calibration experiment illustrates why point accuracy, coverage and distribution scores must be assessed together: optimizing an all-valid probability score can shift medians and change performance within positive-label subsets.

Cloud and moisture structure provide established indicators of circulation size [12–18]. Performance nevertheless varies among storms, as illustrated by EP’s persistent R34 offset and WP R50 favoring current-only input. Training quantiles, large-radius frequencies and alternative point estimates identify distributional and representation effects relevant to R34 underestimation. The probability mean improves aggregate MAE while leaving most of the R34 bias, suggesting that changing the point estimate alone is insufficient. Separating cloud-to-surface correspondence from reference uncertainty requires independent wind observations.

BEST supplies both training and evaluation targets, so the scores measure agreement with retrospective operational analyses. Residuals reflect uncertainty in both model estimates and reference radii. Similarly, cross-domain results combine sensor, geographic and storm-population differences. Output-head adaptation improves accuracy with limited target supervision, while the remaining gap indicates incomplete transfer under this budget. These experiments quantify transfer performance without separating the contributions of the co-varying domain factors.

A proposed independent validation would use TCs excluded from training and selection, collocated ASCAT Level-2 and suitable SAR wind products, and a ± 30 min window with $\pm 15/\pm 60$ min sensitivity checks. Overpass-time GEO inputs, center uncertainty, wind height and averaging convention, quality flags, rain/coastal effects, quadrant coverage and retrieval ranges would be recorded. Only observed threshold boundaries would be scored; incomplete swaths would be censored or unscorable, with uncertainty summarized by TC. Sentinel-1 Ocean Wind Field processing uses atmospheric-model wind information [54], which must inform the assessment of product independence. Prospective assessment would also evaluate operational centers and acquisition-to-delivery latency.

7. Conclusions

TC-RadiiFormer jointly estimates current R34, R50 and R64 in four geographic quadrants from multispectral geostationary observations and their actual acquisition times. Against 2024 best-track radii, positive-radius MAEs were 13.164, 8.384 and 5.731 nmi. Mean MAE was 27.1%, 25.4% and 13.0% lower than ResNet-50, Swin V2-T and Swin V2-T with a temporal Transformer. Matched comparisons supported gains from complementary spectra, historical observations and actual timing, while the uncalibrated TC-RadiiFormer also improved CRPS relative to the spatial baselines. By evaluating accuracy, output ordering and probability quality separately, the study supports joint wind-radius estimation from frequent satellite observations, with further assessment needed against independent surface-wind measurements.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/rs18193428/s1>: Sections S1–S5, implementation details; Sections S6–S9, extended evaluations; Figure S1, Ten-channel ABI observations for the Ida example; Figure S2, Actual-time historical memory update and fusion; Figure S3, Nested, independent and projected output heads; Figure S4, TC-level error differences between current-only and full-history inputs; Figure S5, TC-level R34 errors across residual strata; Figure S6, Signed residual distributions for R34, R50 and R64; Figure S7, Upper-tail MAE and bias; Figure S8, Prediction and error responses to input perturbations; Figure S9, R50 and R64 evolution for three case studies; Figure S10, Calibration of 90% interval coverage by label state; Figure S11, Randomized within-category probability integral transform diagnostics; Figure S12, Nominal versus empirical prediction-interval coverage; Table S1a,

Labeled analysis times by year and basin; Table S1b, Physical channel roles and ABI/AHI band mapping; Table S1c, Sensor and basin counts for 2024 observations; Table S2a, Validation and test label populations; Table S2b, Origins of valid zeros in 2024; Table S3a, Model dimensions, outputs and configuration controls; Table S3b, Quarter-turn consistency of point predictions; Table S4a, Generic baseline implementations; Table S4b, Generic adapter field order; Table S4c, Related radius estimators and their geographic and cross-threshold outputs; Table S4d, Published MAE on study-specific populations; Table S4e, Training and inference resources for complete systems, matched adaptations and structural controls; Table S4f, Capacity selection using 2023 validation data; Table S5a, Positive-radius MAE for complete and component systems; Table S5b, Absolute scores for three training seeds; Table S5c, Basin-by-radius MAE on the 2024 test set; Table S5d, Complete-system RMSE, bias, correlation and quadrant MAE; Table S5e, Matched-adaptation errors and quadrant MAE; Table S5f, Paired TC error differences for spectral and historical controls; Table S5g, R34 residual strata; Table S5h, Multiplicity adjustment for paired comparisons; Table S5i, Adjusted comparisons of systems, components, transfer, probabilistic endpoints, adaptations, perturbations, temporal fusion and point projection; Table S6a, Retrospective evaluation on 2022 data; Table S6b, MAE on tighter anchor-completion subsets; Table S6c, Training-defined upper-tail error strata; Table S6d, Prediction response to input perturbations; Table S6e, Training positive-radius populations and tail support by basin; Table S6f, R34 distributions, shell representation, alternative point estimates and paired readout differences; Table S6g, Matched-target transfer and output-head adaptation; Table S7a, Wind radii for the three case-study snapshots; Table S7b, Positive-reference errors for the case studies; Table S8a, CRPS for all valid labels; Table S8b, Zero and presence metrics and zero-origin diagnostics; Table S8c, Prediction-interval coverage and width at five probability levels; Table S8d, Probability and point diagnostics by reference-label state; Table S8e, Radius-specific calibration, accuracy trade-offs, complexity and interval diagnostics; Algorithm S1, Actual-time history update. Full captions and legends are provided in the supplementary file. References [9–11,15–21,28,31,32,34,50,51] are cited in the Supplementary Materials.

Author Contributions: Conceptualization, X.W., H.H. and L.G.; Methodology, Y.L. and W.Q.; Software, Y.L.; Validation, Y.L.; Formal Analysis, Y.L.; Investigation, Y.L., W.Q. and J.Z.; Resources, X.W., H.H. and L.G.; Data Curation, Y.L.; Writing—Original Draft Preparation, Y.L.; Writing—Review and Editing, X.W., W.Q. and J.Z.; Visualization, Y.L.; Supervision, X.W. All authors have read and agreed to the published version of the manuscript and agree to be accountable for all aspects of the work.

Funding: This research was partially supported by the National Natural Science Foundation of China (Grant No. 62372460), the Hunan Provincial Natural Science Foundation of China (Grant No. 2024JJ4042), the Science and Technology Innovation Program of Hunan Province (Grant No. 2024RC3134), and the Guangzhou Integrated Marine Disaster Prevention System Development Project—Information System and Supporting Infrastructure Construction (Guangzhou Government Procurement No. CZ2024-1224).

Data Availability Statement: GOES-R Advanced Baseline Imager (ABI) Level-1b full-disk radiances (ABI-L1b-RadF) are publicly distributed through NOAA/NCEI under dataset DOI <https://doi.org/10.7289/V5BV7DSR> and the NOAA Open Data catalog <https://registry.opendata.aws/noaa-goes/> (accessed on 30 September 2026). Himawari-8/9 Advanced Himawari Imager (AHI) observations are publicly distributed through the NOAA Open Data AHI-L1b-FLDK collection <https://registry.opendata.aws/noaa-himawari/> (accessed on 30 September 2026) and were decoded as JMA Himawari Standard Data v1.3. Final AL/EP/CP ATCF BEST b-decks are available from the NOAA/NHC/CPHC archive <https://ftp.nhc.noaa.gov/atcf/archive/> (accessed on 30 September 2026), and final WP ATCF BEST b-decks are available from the JTWC best-track archive <https://www.metoc.navy.mil/jtwc/jtwc.html?best-tracks> (accessed on 30 September 2026). The manuscript reports the fixed split definitions, population summaries, and evaluation tables used in the analysis. Processed TC-referenced samples and the reproducibility code are available from the corresponding author, Hai Hu, upon reasonable request. Source satellite files and source ATCF decks remain available from the providers above and are not redistributed with the manuscript.

Conflicts of Interest: Xiang Wang serves as a Guest Editor of the Special Issue to which this manuscript is submitted. This relationship is disclosed, and independent editorial handling is requested in accordance with MDPI policy. The authors declare no other conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. National Hurricane Center. Tropical Cyclone Surface Wind Field—Methodology. Available online: <https://www.nhc.noaa.gov/aboutwindfield3.shtml> (accessed on 13 August 2026).
2. Irish, J.L.; Resio, D.T.; Ratcliff, J.J. The influence of storm size on hurricane surge. *J. Phys. Oceanogr.* **2008**, *38*, 2003–2013. [CrossRef]
3. Knaff, J.A.; Longmore, S.P.; Molenaar, D.A. An objective satellite-based tropical cyclone size climatology. *J. Clim.* **2014**, *27*, 455–476. [CrossRef]
4. Zhang, B.; Zhu, Z.; Perrie, W.; Tang, J.; Zhang, J.A. Estimating tropical cyclone wind structure and intensity from spaceborne radiometer and synthetic aperture radar. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 4043–4050. [CrossRef]
5. Fang, H.; Perrie, W.; Zhang, G.; Xu, C.; Chan, P.; Su, H.; Zheng, G.; Yang, J. Radiometer-only tropical cyclone center location, intensity and wind structure estimation in eight years of SMAP data. *Atmos. Res.* **2025**, *316*, 107945. [CrossRef]
6. Qian, Z.; Yu, W.; Guo, W.; Bai, L.; Lu, X. Intercomparison, fusion and application of FY-3E/WindRAD and HY-2B/SCA ocean surface wind products for tropical cyclone monitoring. *Remote Sens.* **2025**, *17*, 3809. [CrossRef]
7. Fang, H.; Fan, G.; Perrie, W.; Li, Z.; Zhang, C.; Zhang, G.; Yang, J. Tropical cyclone ocean winds and structure parameters retrieved from cross-polarized SAR measurements. *J. Meteorol. Res.* **2024**, *38*, 937–953. [CrossRef]
8. GOES-R Program. ABI Bands Quick Information. Available online: <https://www.goes-r.gov/mission/ABI-bands-quick-info.html> (accessed on 13 August 2026).
9. GOES-R Program. Product Definition and Users' Guide, Volume 3: Level 1B Products. Available online: <https://www.goes-r.gov/users/docs/PUG-L1b-vol3.pdf> (accessed on 13 August 2026).
10. Japan Meteorological Agency. Himawari-8/9 Himawari Standard Data User's Guide, Version 1.3. Available online: https://www.data.jma.go.jp/mscweb/en/himawari89/space_segment/hsd_sample/HS_D_users_guide_en_v13.pdf (accessed on 13 August 2026).
11. Japan Meteorological Agency. Himawari-8/9 Calibration Portal. Available online: https://www.data.jma.go.jp/mscweb/en/oper/calibration/calibration_portal.html (accessed on 13 August 2026).
12. Mueller, K.J.; DeMaria, M.; Knaff, J.A.; Kossin, J.P.; Vonder Haar, T.H. Objective estimation of tropical cyclone wind structure from infrared satellite data. *Weather Forecast.* **2006**, *21*, 990–1005. [CrossRef]
13. Dolling, K.; Ritchie, E.A.; Tyo, J.S. The use of the deviation angle variance technique on geostationary satellite imagery to estimate tropical cyclone size parameters. *Weather Forecast.* **2016**, *31*, 1625–1642. [CrossRef]
14. Lu, X.; Yu, H.; Yang, X.; Li, X. Estimating tropical cyclone size in the Northwestern Pacific from geostationary satellite infrared images. *Remote Sens.* **2017**, *9*, 728. [CrossRef]
15. Baek, Y.-H.; Moon, I.-J.; Im, J.; Lee, J. A novel tropical cyclone size estimation model based on a convolutional neural network using geostationary satellite imagery. *Remote Sens.* **2022**, *14*, 426. [CrossRef]
16. Zhuo, J.-Y.; Tan, Z.-M. Physics-augmented deep learning to improve tropical cyclone intensity and size estimation from satellite imagery. *Mon. Weather Rev.* **2021**, *149*, 2097–2113. [CrossRef]
17. Xu, J.; Wang, X.; Wang, H.; Zhao, C.; Wang, H.; Zhu, J. Tropical cyclone size estimation based on deep learning using infrared and microwave satellite data. *Front. Mar. Sci.* **2023**, *9*, 1077901. [CrossRef]
18. Lu, X.; Wong, W.-K.; Yu, H.; Yang, X. Tropical cyclone size identification over the western North Pacific using support vector machine and general regression neural network. *J. Meteorol. Soc. Jpn.* **2022**, *100*, 927–941. [CrossRef]
19. Zhou, B.; Lu, X.; Wu, X.; Qiu, X.; Xie, H.; Zhu, Z.; Zheng, J. Research on the estimation of tropical cyclone gale radius based on a fusion TC-WREM model. *J. Trop. Meteorol.* **2024**, *40*, 736–744. [CrossRef]
20. Wang, C.; Li, X. A deep learning model for estimating tropical cyclone wind radius from geostationary satellite infrared imagery. *Mon. Weather Rev.* **2023**, *151*, 403–417. [CrossRef]
21. Chang, T.; Xiong, X. GOES-16/ABI thermal emissive band assessments using GEO-LEO-GEO double difference. *Earth Space Sci.* **2019**, *6*, 2303–2316. [CrossRef] [PubMed]
22. Jin, Y.; Liu, J.; Ren, K.; Wang, X.; Deng, K.; Fan, Z.; Deng, C.; Yue, Y. Toward robust tropical cyclone wind radii estimation with multimodality fusion and missing-modality distillation. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–18. [CrossRef]
23. Knaff, J.A.; Slocum, C.J.; Musgrave, K.D.; Sampson, C.R.; Strahl, B.R. Using routinely available information to estimate tropical cyclone wind structure. *Mon. Weather Rev.* **2016**, *144*, 1233–1247. [CrossRef]

24. Combot, C.; Mouche, A.; Knaff, J.; Zhao, Y.; Zhao, Y.; Vinour, L.; Quilfen, Y.; Chapron, B. Extensive high-resolution synthetic aperture radar (SAR) data analysis of tropical cyclones: Comparisons with SFMR flights and best track. *Mon. Weather Rev.* **2020**, *148*, 4545–4563. [\[CrossRef\]](#)
25. Shimada, U.; Hayashi, M.; Mouche, A. A comparison between SAR wind speeds and western North Pacific tropical cyclone best track estimates. *J. Meteorol. Soc. Jpn.* **2024**, *102*, 575–593. [\[CrossRef\]](#)
26. Landsea, C.W.; Franklin, J.L. Atlantic hurricane database uncertainty and presentation of a new database format. *Mon. Weather Rev.* **2013**, *141*, 3576–3592. [\[CrossRef\]](#)
27. Zhuo, J.-Y.; Tan, Z.-M. A deep-learning reconstruction of tropical cyclone size metrics 1981–2017: Examining trends. *J. Clim.* **2023**, *36*, 5103–5123. [\[CrossRef\]](#)
28. Yuan, S.; You, Q.; Mu, B.; Qin, B.; Xu, J. Estimating the tropical cyclone wind structure using physics-incorporated networks. *Front. Earth Sci.* **2023**, *10*, 1024979. [\[CrossRef\]](#)
29. Bai, C.; Yan, H.; Mu, P.; Huang, C.; Zhang, J.; Shu, S. Estimating tropical cyclone maximum wind speed and radius using a multimodal hybrid guided network. *J. Geophys. Res. Atmos.* **2026**, *131*, e2025JD044721. [\[CrossRef\]](#)
30. Chen, T.; Zhuge, X.; Yu, H.; Liu, D. Retrieving tropical cyclone radius of maximum wind from geostationary satellite infrared imagery with SAR observations as ground truth. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2026**, *19*, 11411–11426. [\[CrossRef\]](#)
31. Liu, J.; Yue, Y.; Jin, Y.; Ren, K.; Wang, X.; Deng, K.; Deng, C.; Deng, K. Collaborative estimation of tropical cyclone wind radii with multitask learning and multiteacher distillation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, *18*, 12545–12558. [\[CrossRef\]](#)
32. Wang, C.; Li, X. Physically consistent tropical cyclone monitoring from infrared satellite imagery via knowledge-guided deep learning. *J. Geophys. Res. Mach. Learn. Comput.* **2026**, *3*, e2025JH000875. [\[CrossRef\]](#)
33. Zhang, C.-J.; Geng, B.-F.; Ma, L.-M.; Lu, X.-Q. Estimation of tropical cyclone size by combining sequential infrared satellite images with multitask deep learning. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 1–15. [\[CrossRef\]](#)
34. Lee, J.; Lee, Y.; Kim, J.; Kosiorek, A.; Choi, S.; Teh, Y.W. Set Transformer: A framework for attention-based permutation-invariant neural networks. *Proc. Mach. Learn. Res.* **2019**, *97*, 3744–3753.
35. Choo, M.; Kim, Y.; Lee, J.; Im, J.; Moon, I.-J. Bridging satellite missions: Deep transfer learning for enhanced tropical cyclone intensity estimation. *GISci. Remote Sens.* **2024**, *61*, 2325720. [\[CrossRef\]](#)
36. Yan, H.; Mu, P.; Zhang, S.; Zhu, Y.; Zhang, J.; Bai, C. IDOL: Meeting diverse distribution shifts with prior physics for tropical cyclone multi-task estimation. *Adv. Neural Inf. Process. Syst.* **2025**, *38*, 109612–109642. [\[CrossRef\]](#)
37. Kidger, P.; Morrill, J.; Foster, J.; Lyons, T. Neural controlled differential equations for irregular time series. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 6696–6707.
38. Dumeur, I.; Valero, S.; Inglada, J. Paving the way toward foundation models for irregular and unaligned satellite image time series. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 1–21. [\[CrossRef\]](#)
39. Lagerquist, R.; Chirokova, G.; DeMaria, R.; DeMaria, M.; Ebert-Uphoff, I. Center fixing of tropical cyclones using uncertainty-aware deep learning applied to high-temporal-resolution geostationary satellite imagery. *Weather Forecast.* **2026**, *41*, 265–294. [\[CrossRef\]](#)
40. Chen, B.; Chen, B.-F.; Hsiao, C.M. CNN Profiler on polar coordinate images for tropical cyclone structure analysis. *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 991–998. [\[CrossRef\]](#)
41. Cohen, T.; Welling, M. Group equivariant convolutional networks. *Proc. Mach. Learn. Res.* **2016**, *48*, 2990–2999.
42. Fu, H.; Gong, M.; Wang, C.; Batmanghelich, K.; Tao, D. Deep ordinal regression network for monocular depth estimation. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2018; pp. 2002–2011. Available online: https://openaccess.thecvf.com/content_cvpr_2018/html/Fu_Deep_Ordinal_Regression_CVPR_2018_paper.html (accessed on 13 August 2026).
43. Wimmers, A.J.; Griffin, S.; Velden, C. A U-Net retrieval of tropical cyclone inner-core wind fields from microwave and infrared satellite imagery. *Artif. Intell. Earth Syst.* **2024**, *3*, e230084. [\[CrossRef\]](#)
44. Hu, Y.; Deng, K.; Su, Q.; Zhang, D.; Shi, X.; Ren, K. CycloneWind: A dynamics-constrained deep learning model for tropical cyclone wind field downscaling using satellite observations. *Remote Sens.* **2025**, *17*, 3134. [\[CrossRef\]](#)
45. Cheng, Y.-Y.; Lo, H.-H.; Chen, B.-F.; Kuo, H.-C. Analyzing tropical cyclone surface wind asymmetry with a polar-coordinate periodic-domain deep learning model. *Mon. Weather Rev.* **2026**, *154*, 1109–1132. [\[CrossRef\]](#)
46. Zhang, E.; Su, H.; Chan, P.W.; Zhai, C.; Zhou, W.; Xu, W.; Hu, M.; Li, X.; Jia, Y.; Song, Y. Deep learning-based multisource satellite data fusion and downscaling for tropical cyclone wind fields. *J. Geophys. Res. Mach. Learn. Comput.* **2025**, *2*, e2025JH000792. [\[CrossRef\]](#)
47. Uhlhorn, E.W.; Tolwinski-Ward, S.; Lorsolo, S. An asymmetric model of the tropical cyclone surface wind field and probabilistic climatology of structural parameters. *J. Appl. Meteorol. Climatol.* **2024**, *63*, 1569–1587. [\[CrossRef\]](#)
48. Guo, R.; Xu, J.; Yang, C.; Zhou, M. Physics-informed deep-learning prediction of tropical cyclone wind size in the western North Pacific. *J. Meteorol. Res.* **2026**, *40*, 636–652. [\[CrossRef\]](#)

49. Wang, P.; Bernier, N.B.; Xie, S. Synthetic aperture radar (SAR)-based evaluation of tropical cyclone wind profiles and a theory-based radius of maximum wind estimation. *Geophys. Res. Lett.* **2026**, *53*, e2026GL122677. [[CrossRef](#)]
50. Woo, S.; Debnath, S.; Hu, R.; Chen, X.; Liu, Z.; Kweon, I.S.; Xie, S. ConvNeXt V2: Co-designing and scaling ConvNets with masked autoencoders. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2023; pp. 16133–16142. [[CrossRef](#)]
51. Gneiting, T.; Raftery, A.E. Strictly proper scoring rules, prediction, and estimation. *J. Am. Stat. Assoc.* **2007**, *102*, 359–378. [[CrossRef](#)]
52. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2016; pp. 770–778. [[CrossRef](#)]
53. Liu, Z.; Hu, H.; Lin, Y.; Yao, Z.; Xie, Z.; Wei, Y.; Ning, J.; Cao, Y.; Zhang, Z.; Dong, L.; et al. Swin Transformer V2: Scaling up capacity and resolution. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2022; pp. 11999–12009. [[CrossRef](#)]
54. Copernicus. Sentinel-1 Processing: Ocean Wind Field Processing. Available online: <https://sentiwiki.copernicus.eu/web/s1-processing> (accessed on 2 October 2026).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.